DYNAMIC EMOTIONAL COMMUNICATION

EDITED BY: Wataru Sato, Eva G. Krumhuber, Tjeerd Jellema and
Justin H. G. Williams

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

DYNAMIC EMOTIONAL COMMUNICATION

Topic Editors:
Wataru Sato, Kyoto University, Japan
Eva G. Krumhuber, University College London, United Kingdom
Tjeerd Jellema, University of Hull, United Kingdom
Justin H. G. Williams, University of Aberdeen, United Kingdom

This eBook aims to deepen our understanding of emotional communication by introducing "dynamic" perspectives.

Facial and bodily expressions of emotion functions as indispensable communicative signals for human beings. People decode the emotional information conveyed by facial/bodily expressions and use this to coordinate cooperative or competitive social relationships. Experimental psychological research has long investigated these important means of emotional communication. However, this was typically done by using static stimuli of facial/bodily expressions to assess the detection and interpretation of emotions. This paradigm was also adopted in neuropsychological, neurophysiological, and neuroimaging studies. Although researchers accumulated valuable information regarding the psychological and neural mechanisms underlying these processes, the static nature of the stimuli may have resulted in important phenomena remaining unexamined.

Recently, scientists have begun to explore dynamic emotional communication, in particular by using dynamic facial/bodily expressions of emotion, instead of static photographs, as stimuli. This is having important consequences for emotion research. As dynamic emotional expressions have increased ecological validity and as there are differences in the visual processing of dynamic and static information, a host of novel aspects of the psychological and neural processing of emotional expressions have been elucidated. For example, it has been shown that motor resonance and the recruitment of motor areas are fundamental to dynamic emotional communication. Researchers have also started to investigate the encoding of dynamic emotional interactions and have clarified the messages embedded in the temporal aspects and the patterns of reciprocal inter-individual coordination. Moreover, investigations of dynamic emotional communication have identified heretofore unrecognized impairments in the social functioning of individuals with psychiatric disorders, such as autism spectrum disorder and schizophrenia.

Table of Contents

Editorial: Dynamic Emotional Communication

Wataru Sato [1]*, Eva G. Krumhuber [2], Tjeerd Jellema [3] and Justin H. G. Williams [4]

[1] *Kokoro Research Center, Kyoto University, Kyoto, Japan,* [2] *Department of Experimental Psychology, University College London, London, United Kingdom,* [3] *Department of Psychology, University of Hull, Hull, United Kingdom,* [4] *Translational Neuroscience Group, Institute of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom*

Editorial on the Research Topic

Dynamic Emotional Communication

INTRODUCTION

Psychological research has a long history of investigating facial and bodily expressions associated with emotion. This is partly due to the fact that non-verbal behaviors are indispensable communicative signals during the creation and maintenance of social relationships. A number of neuroscientific studies have also investigated the neural mechanisms underlying the processing of these emotional signals.

However, most previous research assessing emotional communication has been conducted using static stimuli. Although researchers have accumulated valuable information about the psychological and neural mechanisms underlying the processing of emotional signals using such stimuli, their static nature may have left important phenomena unexamined.

To address this issue, recent studies have explored emotional communication using dynamic facial and bodily expressions of emotion, which has had important consequences for emotion research. Because dynamic emotional expressions are associated with increased ecological validity, resulting in a number of important differences in the psychological/neural processing between dynamic and static information, a host of novel aspects of emotional communication have been elucidated. Furthermore, the dynamic perspective can be applied to broader methodological and conceptual areas.

The present Research Topic brings together a collection of new articles that have investigated dynamic emotional communication and demonstrates recent advances in this field of research. Here, we introduce these articles and discuss them in the context of related studies by grouping them into the following four areas: (a) decoding of dynamic emotional signals, (b) moderators of dynamic emotional signal decoding, (c) encoding of dynamic emotional signals, and (d) other dynamic aspects of emotional communication. The term "decoding" was used to refer to various types of processing (e.g., perceptual and motor) in addition to the recognition of emotions. The term "encoding" was used to refer to the production of emotional signals.

DECODING OF DYNAMIC EMOTIONAL SIGNALS

Seminal research has demonstrated that emotional recognition based on dynamic facial expressions is more efficient than that based on static expressions (Bassili, 1978), with several subsequent studies investigating this issue (for a review, see Krumhuber et al., 2013; Krumhuber and Skora, 2016). In this Research Topic, Dobs et al. reviewed the literature and reported that there are evident dynamic advantages for subtle expressions or for full-blown expressions under suboptimal conditions. Additionally, these authors provided an overview of the methods used to present dynamic facial expressions (e.g., videos and point lights) as well as their advantages and disadvantages.

Several studies have reported that the genuineness of an emotional message is decoded more effectively from dynamic, compared with static, facial expressions. For example, Zloteanu et al. investigated the discrimination performance of genuine expressions vs. deliberate expressions of surprise that were presented in both dynamic and static formats. These authors found that dynamic genuine expressions are perceived as more genuine-looking than static ones and that the presentation format modulated the genuineness ratings of deliberate expressions. In a similar vein, Namba et al. investigated whether decoders could distinguish between genuine and deliberate facial expressions of some emotions when they are presented in dynamic and static formats. The discriminability of the genuineness of an expression was enhanced for dynamic displays, in comparison to static displays. Busin et al. assessed the judgements of genuine vs. masked emotions in dynamic facial expressions rotated to the left or right side. Eye movement patterns revealed preferential attention to the left hemi-face, which has been previously reported during the processing of static expressions. Other studies have revealed that the dynamic nature (e.g., speed) of facial expressions provides information about the naturalness (Sato and Yoshikawa, 2004), genuineness (Krumhuber and Kappas, 2005), and trustworthiness (Krumhuber et al., 2007) of the portrayed emotion.

Various types of other information can be decoded from dynamic emotional signals. Orlowska et al. evaluated the recognition of reward, affiliative, and dominance smiles during dynamic and static presentations and found that the recognition of affiliative smiles is more accurate for dynamic expressions than static expressions. The authors also assessed the effects of facial muscle restriction and suggested that facial mimicry is unlikely to be critical to this process. Other studies have shown that, compared with static expressions, dynamic facial expressions facilitate the detection of an expression (Ceccarini and Caudeka, 2013), the experience of emotional arousal (Sato and Yoshikawa, 2007a), and facial mimicry (Weyers et al., 2006; Sato and Yoshikawa, 2007b). Different visual styles between dynamic and static facial expressions have been suggested in the context of eye fixation patterns (e.g., more fixation on the center for dynamic expressions; Blais et al., 2017).

Some studies have investigated multimodal dynamic emotional signals, which are more natural than those from a single modality. Garrido-Vásquez et al. recorded event-recorded potentials (ERPs) to investigate the priming effects of dynamic facial expressions (angry, happy, and neutral) on the processing of emotionally intoned sentences (angry and happy). The amplitudes of auditory-related components at ∼100 ms are higher in response to incongruently primed sentences than other conditions, suggesting the occurrence of rapid cross-modal emotional interactions. Mortillaro and Dukes reviewed studies investigating the decoding and encoding of facial and bodily expressions of positive emotions. They proposed that the inclusion of dynamic information and facial as well as bodily signals is important when distinguishing between expressions of positive emotions (e.g., joy and pride).

Valid stimulus sets are needed to investigate the decoding of emotional signals. For this purpose, Calvo et al. developed a database of dynamic emotional facial expressions by creating morphing animations. They validated these novel stimuli via human observer judgments as well as automated assessment of facial expressions. Several other studies have developed stimulus databases (for a review, see Krumhuber et al., 2016), allowing for the selection of an appropriate database based on the researcher's needs.

A number of neuroimaging studies have investigated the neural mechanisms underlying the processing of dynamic emotional signals (e.g., Sato et al., 2004). Zinchenko et al. conducted meta-analysis of functional magnetic resonance imaging (fMRI) studies including dynamic facial expressions. They found that some brain regions (e.g., the fusiform and middle temporal gyri, amygdala, and inferior frontal gyrus) are robustly activated during the observation of dynamic facial expressions. The involvement of action observation network (AON; e.g., the middle temporal gyrus/superior temporal sulcus and inferior frontal gyrus), which can match the observation and execution of actions (cf. Rizzolatti et al., 2001), appears to be one of the most distinctive features associated with the neural processing of dynamic, compared with static, facial expressions. To further investigate this issue, Rymarczyk et al. simultaneously recorded facial electromyography (EMG) and fMRI data during the observation of dynamic and static facial expressions of fear and disgust. They reported that facial EMG patterns of facial mimicry are correlated with specific activation in several brain regions, including the AON, under dynamic presentation conditions. There are several other unique aspects of the neural processing of dynamic facial expressions compared with that of static expressions. For example, the observation of dynamic facial expressions evidently induces modulatory influences from the amygdala to the neocortex (Sato et al., 2017) and clearly reveals hemispheric functional asymmetry (right cortical and left cerebellar; Sato et al., 2019). Differences in the decoding of dynamic and static facial expressions have also been suggested by lesion studies (e.g., Humphreys et al., 1993).

Several neurophysiological studies in animals have provided information about the cellular-level neural substrates involved in dynamic emotional signal decoding. For example, Jellema and Perrett (2003) found that some neurons in the superior temporal

sulcus of monkeys fire in response to dynamic bodily actions but not to static postures.

MODERATORS OF DYNAMIC EMOTIONAL SIGNAL DECODING

Several stimulus properties of dynamic emotional signals moderate the decoding processes. For example, Plouffe-Demers et al. compared spatial frequency tuning during the recognition of dynamic and static facial expressions. The results showed that the recognition of dynamic facial expressions relies more strongly on lower spatial frequencies. Rooney and Bálint tested the effects of shot scale (i.e., the apparent distance of characters from the camera) on the tendency to recognize the mental states of others in fictional films. Close-up, compared with long, shots of a character are associated with higher tendencies to attribute emotional and mental states to a character.

Perceiver factors also moderate the decoding process of dynamic emotional signals. Wingenbach et al. investigated the effects of manipulating facial muscles on the recognition of emotion from dynamic facial expressions. Compared to passive viewing, holding a pen in the mouth reduces recognition accuracy of facial expressions based on salient features in the lower face region (e.g., happy expressions), indicating that bodily actions shape the processing of dynamic facial expressions. In a similar vein, Kato et al. explored the role of manual movements in the perception of valence from emotional scenes. Downward manual movements (temporally proximate and after the observation of images) made the scenes appear more emotional negative. Other studies have shown that the processing of dynamic emotional signals could be moderated by stable perceiver characteristics, such as empathic personality traits (e.g., Mailhot et al., 2012).

Psychiatric conditions are considered as moderators of dynamic emotional signal decoding. Okruszek reviewed evidence regarding the decoding performance of patients with various psychiatric conditions, such as schizophrenia, in the context of point-light bodily displays. They found that these patients have unique problems, though the magnitude is weaker than impairments in facial or vocal signal processing. Palumbo et al. compared individuals with autism spectrum disorder (ASD) to matched-controls in terms of the ability to evaluate expressions depicted in the last frames of dynamic facial expression videos. The results, together with their previous finding (Palumbo et al., 2015), suggested that ASD impairs the ability to anticipate immediate future emotional state of others' minds. Other studies have reported that individuals with ASD experience other types of impairments in the processing of dynamic facial expressions such as reduced facial mimicry (Rozga et al., 2013).

The modulatory effects of psychiatric conditions and the underlying neural mechanisms in the decoding of dynamic emotional signals are another topic of scientific interest. Sato et al.'s fMRI study investigated brain activity during the observation of dynamic facial expressions in individuals with ASD and typically developing controls. Atypical modulatory influences were found from the amygdala to the neocortical network, including the AON, during the processing of dynamic facial expressions in the ASD group. This corroborates previous findings showing decreased activity and connectivity within the AON during dynamic facial expression processing in individuals with ASD (Sato et al., 2012), which has been proposed to be a core issue associated with ASD (Williams et al., 2001). Other research has reported patterns of brain activity in response to dynamic emotional signals to differ among various psychiatric conditions, including schizophrenia (e.g., Russell et al., 2007).

ENCODING OF DYNAMIC EMOTIONAL SIGNALS

Studies have begun to explore the encoding of dynamic facial expressions of emotion, which is generally more difficult to assess than the decoding processes. Scherer et al. analyzed the encoding of emotional facial expressions by actors and found that spatial and temporal patterns of facial action units (AUs; Ekman et al., 2002) are largely consistent with dynamic processes as hypothesized by the component process model (Scherer, 2001). Furthermore, the AU patterns are systematically related to the recognition of emotions in decoders. Hyniewska et al. analyzed the AUs of emotional facial expressions, unobtrusively filmed in a real-life emotional situation, and obtained decoder ratings of emotions and appraisals for these expressions. Associations between specific emotions/appraisals and sets of AUs were found, which suggests that the decoding of emotions/appraisals is achieved via the perception of a set of AUs. Grossard et al. investigated the encoding of emotional facial expressions using different tasks (e.g., imitation of a model) and in different regions using a large sample of children. The results suggested that the encoding of emotional facial expressions is a complex developmental process influenced by several factors (e.g., age).

A few previous studies have investigated the neural mechanisms underlying the encoding of dynamic emotion signals. Heller et al. (2014) simultaneously measured fMRI and facial EMG data during the observation of emotional images and found amygdala activity associated with brow muscle activity in response to negative pictures. In the case of some neural lesions affecting higher level motor control, it is possible to retain capacity for emotional expression in the presence of voluntary facial paresis (e.g., Hopf et al., 1992).

OTHER DYNAMIC ASPECTS OF EMOTIONAL COMMUNICATION

The investigation of dynamic, dyadic interactions remains an understudied and interesting field of research. To demonstrate the dynamic nature of emotional communication, Hareli et al. investigated how an observer's perception of power could be influenced by an emotional exchange between members of a dyad. The results revealed that the perception of power changes depending on the emotional response of one's partner. A previous fMRI study has measured the brain activity of two

individuals during face-to-face interactions and observed inter-individual synchronized activity in the lateral occipitotemporal cortex (Koike et al., 2019).

The dynamic perspective can also be applied to the analysis of emotion communication data. Guérin-Dugué et al. jointly recorded ERPs and eye movements during the observation of static emotional facial expressions and applied general linear models to depict the temporal dynamics of neural facial expression processing. Their analyses revealed the emotion-dependent modulation of early components (starting at 20 ms) related to eye fixation in response to facial expressions.

CONCLUSIONS

Together, these findings indicate that a dynamic perspective on emotional communication can provide valuable information. Specifically, the psychological and neural decoding of dynamic facial and bodily signals implies a number of features that differ from those of static displays. Several unique moderators are related to the processing of dynamic emotional messages. Investigation of dynamic facial and bodily expressions are necessary to reveal how emotional messages are encoded. The dynamic perspective can be applied to a broader range of research. Further research should investigate dynamic emotional communication to deepen our understanding of real-life emotional communication.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 373–379. doi: 10.1037/0096-1523.4.3.373

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., and Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion* 17, 1107–1119. doi: 10.1037/emo0000283

Ceccarini, F., and Caudeka, C. (2013). Anger superiority effect: the importance of dynamic emotional facial expressions. *Vis. Cogn.* 21, 498–540. doi: 10.1080/13506285.2013.807901

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *The Facial Action Coding System, 2nd Edn.* Salt Lake City, UT: Research Nexus eBook.

Heller, A. S., Lapate, R. C., Mayer, K. E., and Davidson, R. J. (2014). The face of negative affect: trial-by-trial corrugator responses to negative pictures are positively associated with amygdala and negatively associated with ventromedial prefrontal cortex activity. *J. Cogn. Neurosci.* 26, 2102–2110. doi: 10.1162/jocn_a_00622

Hopf, H. C., Müller -Forell, W., and Hopf, N. J. (1992). Localization of emotional and volitional facial paresis. *Neurology* 42, 1918–1923. doi: 10.1212/WNL.42.10.1918

Humphreys, G. W., Donnelly, N., and Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia* 31, 173–181. doi: 10.1016/0028-3932(93)90045-2

Jellema, T., and Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia* 41, 1728–1737. doi: 10.1016/S0028-3932(03)00175-1

Koike, T., Sumiya, M., Nakagawa, E., Okazaki, S., and Sadato, N. (2019). What makes eye contact special? Neural substrates of on-line mutual eye-gaze: a hyperscanning fMRI study. *eNeuro* 6:e0284–18. doi: 10.1523/ENEURO.0284-18.2019

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonverb. Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., and Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion* 7, 730–735. doi: 10.1037/1528-3542.7.4.730

Krumhuber, E., and Skora, P. (2016). "Perceptual study on facial expressions," in *Handbook of Human Motion,* eds B. Müller and S. Wolf (Heidelberg: Springer-Verlag), 1–15. doi: 10.1007/978-3-319-30808-1_18-1

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2016). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 1–13. doi: 10.1177/1754073916670022

Mailhot, J. P., Vachon-Presseau, E., Jackson, P. L., and Rainville, P. (2012). Dispositional empathy modulates vicarious effects of dynamic pain expressions on spinal nociception, facial responses and acute pain. *Eur. J. Neurosci.* 35, 271–278. doi: 10.1111/j.1460-9568.2011.07953.x

Palumbo, L., Burnett, H. G., and Jellema, T. (2015). Atypical emotional anticipation in high-functioning autism. *Mol. Autism* 6:47. doi: 10.1186/s13229-015-0039-7

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Rozga, A., King, T. Z., Vuduc, R. W., and Robins, D. L. (2013). Undifferentiated facial electromyography responses to dynamic, audio-visual emotion displays in individuals with autism spectrum disorders. *Dev. Sci.* 16, 499–514. doi: 10.1111/desc.12062

Russell, T. A., Reynaud, E., Kucharska-Pietura, K., Ecker, C., Benson, P. J., Zelaya, F., et al. (2007). Neural responses to dynamic expressions of fear in schizophrenia. *Neuropsychologia* 45, 107–123. doi: 10.1016/j.neuropsychologia.2006.04.026

Sato, W., Kochiyama, T., Uono, S., Sawada, R., Kubota, Y., Yoshimura, S., et al. (2019). Widespread and lateralized social brain activity for processing dynamic facial expressions. *Hum. Brain Mapp.* 40, 3753–3768. doi: 10.1002/hbm.24629

Sato, W., Kochiyama, T., Uono, S., Yoshikawa, S., and Toichi, M. (2017). Direction of amygdala-neocortex interaction during dynamic facial expression processing. *Cereb. Cortex* 27, 1878–1890. doi: 10.1093/cercor/bhw036

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., and Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Brain Res. Cogn. Brain Res.* 20, 81–91. doi: 10.1016/j.cogbrainres.2004.01.008

Sato, W., Toichi, M., Uono, S., and Kochiyama, T. (2012). Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. *BMC Neurosci.* 13:99. doi: 10.1186/1471-2202-13-99

Sato, W., and Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Sato, W., and Yoshikawa, S. (2007a). Enhanced experience of emotional arousal in response to dynamic facial expressions. *J. Nonverb. Behav.* 31, 119–135. doi: 10.1007/s10919-007-0025-7

Sato, W., and Yoshikawa, S. (2007b). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1–18. doi: 10.1016/j.cognition.2006.05.001

Scherer, K. R. (2001). "Appraisal considered as a process of multi-level sequential checking," in *Appraisal Processes in Emotion: Theory, Methods, Research,* eds K. R. Scherer, A. Schorr, and T. Johnstone (New York, NY: Oxford University Press), 92–120.

Weyers, P., Mühlberger, A., Hefele, C., and Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology* 43, 450–453. doi: 10.1111/j.1469-8986.2006. 00451.x

Williams, J. H. G., Whiten, A., Suddendorf, T., and Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neurosci. Biobehav. Rev.* 25, 287–295. doi: 10.1016/S0149-7634(01)00014-8

Check for
updates

Hemiface Differences in Visual Exploration Patterns When Judging the Authenticity of Facial Expressions

Yuri Busin[1], Katerina Lukasova[1,2]*, Manish K. Asthana[3] and Elizeu C. Macedo[1]

[1] Social and Cognitive Neuroscience Laboratory and Developmental Disorders Program, Center for Health and Biological Sciences, Mackenzie Presbyterian University, São Paulo, Brazil, [2] Center of Mathematics, Computation and Cognition, Federal University of ABC (UFABC), São Bernardo, Brazil, [3] Department of Humanities and Social Sciences, Indian Institute of Technology Kanpur, Kanpur, India

Past studies have found asymmetry biases in human emotion recognition. The left side bias refers to preferential looking at the left-hemiface when actively exploring face images. However, these studies have been mainly conducted with static and frontally oriented stimuli, whereas real-life emotion recognition takes place on dynamic faces viewed from different angles. The aim of this study was to assess the judgment of genuine vs. masked expressions in dynamic movie clips of faces rotated to the right or left side. Forty-eight participants judged the expressions on faces displaying genuine or masked happy, sad, and fearful emotions. The head of the actor was either rotated to the left by a 45° angle, thus showing the left side of the face (standard orientation), or inverted, with the same face shown from the right side perspective. The eye movements were registered by the eye tracker and the data were analyzed for the inverse efficiency score (IES), the number of fixations, gaze time on the whole face and in the regions of interest. Results showed shorter IESs and gaze times for happy compared to sad and fearful emotions, but no difference was found for these variables between sad and fearful emotions. The left side preference was evident from comparisons of the number of fixations. Standard stimuli received a higher number of fixations than inverted ones. However, gaze time was long on inverted compared to standard faces. Number of fixations on exposed hemiface interacted with the emotions decreasing from happy to sad and fearful. An opposite pattern was found for the occluded hemiface. These results suggest a change in fixation patterns in the rotated faces that may be beneficial for the judgments of expressions. Furthermore, this study replicated the effects of the judgment of genuine and masked emotions using dynamic faces.

Keywords: emotion judgment, dynamic emotions, eye movements, left side preference, genuine emotions, event-elicited masked emotions, gaze pattern

INTRODUCTION

Facial expressions allow the exchange of information about affective states and play a crucial role in social cognition of humans. It has been suggested that human face processing is enhanced by a left gaze bias defined by preferential and longer viewing of the left hemiface (the right side of the viewed face; Gilbert and Bakan, 1973; Sackeim et al., 1978; Heller and Levy, 1981; Hisao and Cottrel, 2008). The left side bias was found in children over 5 years of age, but was reduced in 11-year-olds with autism (Chiang et al., 2000; Taylor et al., 2012), which may indicate links with the development of social recognition and interaction. In addition, preferential left side gaze, particularly when unrelated to faces was found also in human 6-month old babies and rhesus monkeys, which may suggest even broader adaptive significance (Guo et al., 2009).

Assessment of the hemifacial asymmetries in emotional expressions showed that the left side is more emotionally expressive and the left-sided facial movements are more pronounced for negative than positive emotions (Borod et al., 1988; Nicholls et al., 2004). Indeed measuring facial muscle movement during emotional expression demonstrated increased movement of the left in comparison with the right hemiface (Dimberg and Petterson, 2000). These findings are in line with studies using composite photographs, created by mirror-reversed images of left–left and/or right–right hemiface, showing that the left composite of faces are judged as more emotionally expressive than the right one (Moreno et al., 1990). Also for posed smiles, produced by actors in the absence of the real emotion stimuli, the left–left composite photographs were judged as more trustworthy than the right ones (Okubo et al., 2013).

To determine which facial features are selected in visual search for more detailed examination, gaze fixation has been examined during judgment of different emotions. In facial expressions of 2D images people fixate their eyes mainly on the eyes and nose region, followed by the mouth and cheeks (Kret et al., 2013; Miellet et al., 2013). However, these regions seem to contribute differently to the recognition depending on the type of emotion being processed. Happy expressions can be recognized after exposure as brief as 20–40 ms, and the most fixated facial region is the mouth, while other regions make little contribution to this recognition (Nusseck et al., 2008; Calvo and Nummenmaa, 2009; Du and Martinez, 2013). Longer exposure times of approximately 100–250 ms are needed for recognition of sad and fearful expressions (Eisenbarth and Alpers, 2011; Du and Martinez, 2013). For recognition of sadness, mainly the eyes, eyebrows, and mouth are looked at Nusseck et al. (2008), Eisenbarth and Alpers (2011). For fear recognition, people mainly fixate the eyes, and the nose region can provide additional information (Schurgin et al., 2014). Interestingly, visual processing of facial regions correlated with the total number of left hemiface fixations and when the eye movements were reduced by short stimuli presentation time, the left side bias was evident (Butler et al., 2005; Butler and Harvey, 2006).

Much of this research has used static faces, which do not closely reflect a natural social interaction. Therefore, a dynamic presentation should provide a more similar representation of the natural environment, as well as more visual cues for local and global feature processing when compared to the use of static presentations (Atkinson et al., 2004; Krumhuber and Manstead, 2009; McLellan et al., 2010; Harris et al., 2014). In the case of basic expressions, there is a consensus over a stereotypical pattern of facial activation that can be adequately perceived and recognized as one emotion (Nusseck et al., 2008; Cristinzio et al., 2010). This pattern strongly depends on deformation of distinct morphological facial areas [action units (AUs); Ekman and Friesen, 1978]. For example, happy emotions can be produced by AU such as crow's feet wrinkles around the eyes together with pulling up of the lip corners, known as the Duchenne marker (D) (Ekman and Rosenberg, 1997). This marker is produced by the contraction of the orbiculares oculi and zygomaticus major muscles and is thought to be a sign of a genuine smile in static emotional faces (Peron and Roy-Charland, 2013). A study that examined the importance of the D marker in discrimination between spontaneous and deliberate smiles in static and dynamic displays by healthy adults showed that the marker was not the most stable cue for rating smiles and the selection of preferable visual features follows a different pattern (Krumhuber and Manstead, 2009). The importance of dynamic expressions, such as movie clips, lies in the possibility of seeing the onset, apex, and offset phases of the expressed emotion, thus increasing perceptual sensitivity (Krumhuber and Kappas, 2005). Furthermore, it seems that both the features and the event's timing play an important role in facial perception and emotional recognition. The observer may ignore the AU markers of negative emotion in the eye regions when there is a smiling mouth. This effect tended to be bigger if the mouth motion came only after a change in the eyes (Iwasaki and Noguchi, 2016).

Thus the evidence shows that the perception of timing in facial movement enhances the facial expression recognition (Atkinson et al., 2004; Harris et al., 2014; Weyers et al., 2016; Yan et al., 2017). However, not many studies investigated how the left side bias is affected in these dynamic presentations, and the influence of timing. In one study that investigated this question, a stronger left hemiface bias was found in dynamic displays compared to static faces or face-like objects. The preference to explore the right side of the face was most evident in the eye region and it was present even in the mirrored face stimuli (Everdell et al., 2007).

The current study aimed to investigate: (i) the pattern of gaze on rotated dynamic human faces showing three basic human emotions (happy, sad, and fearful), and (ii) the effect of left side bias, showing the same clip from the left (standard) and right (inverted) side in a 45° angle. We hypothesized that recognizing happy emotion in movie clips requires less visual processing, an effect previously reported only in static images (Nusseck et al., 2008; Korb et al., 2014). On the other hand, inverted images pose higher demands on visual processing since they offer a non-preferential side of the human face; thus, we expect to find the left side preference for visual perception (Chelnokova and Laeng, 2011). Additional difficulty is expected when discriminating between genuine and masked expressions due to temporal incongruence and asymmetry of AU markers, since studies indicate that in dynamic faces, the typical AU

marker's deformation may be overridden by other temporal cues (Krumhuber and Kappas, 2005).

MATERIALS AND METHODS

Participants

A total of 47 undergraduate students of the Mackenzie Presbyterian University volunteered for the experiment. This sample size is consistent with many other studies on this subject (Chelnokova and Laeng, 2011; Du and Martinez, 2013). All volunteers had normal or corrected-to-normal vision. Participants with a history of head surgery, head trauma or seizures, drug addiction, psychosis, or dementia were excluded. One participant was later excluded from the experiment due to insufficient eye-tracking data. Thus, 46 participants ($M = 22.65$ years old, $SD = 3.22$) were included in the analyses. Female ($N = 30$) and male participants did not differ with respect to age and handedness ($p > 0.05$). This study was carried out in accordance with the recommendations of Mackenzie Presbiterian University Ethics committee, that reviewed and approved the project. The study was approved by the local ethics committee (CAAE No. 50307815.8.0000.0084) and each participant provided written informed consent prior to the experiment.

Stimuli

Movie clips were selected from the Computerized Test of Primary Emotion Perception (Miguel and Primi, 2014). The test shows genuine and event-elicited masked facial expressions for a variety of human emotions. Each clip depicted the head and the upper part of the shoulders of a person expressing an emotion, with the head rotated horizontally 45° to the left side. Each clip was of 4 s duration.

Miguel and Primi (2014) recorded videos of individuals viewing pictures of different emotional content from the International Affective Pictures System (IAPS) in order to produce genuine emotional expressions. The incongruent emotion videos were produced when individuals had to mask the genuine expressions elicited by the picture with one out of eight primary emotions. For example, when viewing a happy picture, the individual in the video could produce either a sad or another facial expression. These emotions were labeled as event-elicited masked emotions. The videos were administered to 310 naïve participants who judged the videos for the type and veracity of the expressed emotion (Primi, 2014).

For the purpose of this study, only three basic emotions were chosen: happy, sad, and fearful expressions. The emotions were presented by 12 different actors (three men and nine women) and there were four actors per emotion. Each actor performed both genuine and masked expressions. The clips were matched on other physical properties of the image such as the background color, luminosity, and the size and position of the face in the background. Each clip was recorded showing the left side of the face from a 45° angle (labeled as standard) and was mirrored to show the actors from a right-hand 45°

angle (labeled as inverted). Each clip was presented four times in pseudo-randomized sequences in two runs separated by a 5 min rest period. In total, the participants judged 96 clips (48 in each run): 24 standard movie clips for genuine emotions (i.e., happy, sad, and fearful), 24 standard movie clips for masked emotions, 24 inverted movie clips for genuine emotions, and 24 inverted movie clips for masked emotions. In each group with 24 clips, the same number of movies showed happy, sad, and fearful emotions, eight of each. Busin (2011, Unpublished) validated all the clips with healthy participants in a pilot study. In this study ($N = 13$) genuine displays were correctly rated as genuine ($M = 70\%$, $SD = 4.6$) and masked ($M = 43\%$, $SD = 4.9$). Also, all emotional expressions were recognized accordingly, including happy ($M = 80\%$, $SD = 4$), sad ($M = 80\%$, $SD = 3.9$), and fearful ($M = 10\%$, $SD = 3$).

Eye Tracking and Measures

Using the Eye Gaze Edge 1750 eye tracker (LC Technologies, Inc., United States) the current study collected position information related to both eyes. The eye tracking data analysis program NYAN was used for off-line data processing. The default settings for fixation detection considered parameters of gaze deviation from a threshold of 25 pixels for the minimum of six samples, with a recording frequency of 120 Hz. The movie clips were presented on a 19-inch flat screen color monitor (1490 × 900 pixels) at a viewing distance of 60 cm. In addition, the eye position was monitored in real-time by the experimenters on a second monitor used both for instruction and quality check.

Procedure

At the beginning of the experiment, all participants were given detailed instructions and a brief training. The participants were instructed to watch the movie clips and decide whether the presented emotion was genuine or masked. After each movie clip, a black screen with a fixation cross appeared, during which the participant was instructed to respond to the clip by pressing one of two keys on the keyboard: "v" for genuine, "m" for masked. Once the response was given, a new movie clip was presented. All tests were conducted in the same room with the lights off, without sounds, and in the presence of an experimenter.

Data Analysis

All statistical data analyses were performed using the IBM SPSS 20.0 program. For eye-tracking data, we performed conventional analysis of variance (ANOVA) with emotion (happy, sad, fearful), veracity (genuine, masked) and side (standard, inverted) as within-subject factors. Based on previous research findings, three basic dependent measures were considered: (1) *inverse efficiency score (IES)*: computed for each participant's average response time divided by the total of correct responses in order to account for any possibilities of speed-accuracy trade-offs (Townsend and Ashby, 1983); (2) *number of fixations*: the average number of eye fixations in the whole movie clip; and (3) *gaze duration*: the average duration of all fixations in the whole movie clip.

RESULTS

Inverse Efficiency Score (IES)

Using *IES* scores as the dependent variable, a three-way ANOVA was conducted. Results revealed a significant main effect for veracity ($F_{(1,45)} = 6.96$, $p = 0.01$, $n_G^2 = 0.023$) and emotion ($F_{(2,90)} = 4.75$, $p = 0.01$, $n_G^2 = 0.021$). The *post hoc* Bonferroni comparison indicated lower *IES* for happy ($M = 194$ ms) than sad ($M = 239$ ms) but there was not difference for sad and fearful ($M = 243$ ms) emotions. A lower *IES* was found for genuine ($M = 202$ ms) compared to masked ($M = 249$ ms) emotions.

Number of Fixations

Results of a three-way ANOVA examining the number of fixations revealed statistically significant main effects for veracity ($F_{(1,45)} = 4.62$, $p = 0.04$, $n_G^2 = 0.002$) and side ($F_{(1,45)} = 16.48$, $p < 0.001$, $n_G^2 = 0.007$), but not for emotion. More fixations were made on the genuine ($M = 8.87$) compared to masked ($M = 8.61$) expressions and on standard ($M = 9.01$) compared to inverted ($M = 8.47$) faces.

Gaze Duration

A three-way ANOVA showed significant main effects for side ($F_{(1,45)} = 4.18$, $p < 0.05$, $n_G^2 = 0.004$) and emotion ($F_{(2,90)} = 5.36$, $p < 0.01$, $n_G^2 = 0.005$). The *post hoc* Bonferroni comparison indicated shorter gaze duration on happy ($M = 403$ ms) than fearful ($M = 422$ ms) emotions, but no difference was found

for fearful and sad ($M = 428$ ms). Gaze was longer on the inverted ($M = 429$ ms) compared to standard ($M = 406$ ms) faces (Supplementary Material 1).

Analyses of ROI

To better characterize the visual exploration pattern, the number of fixations and gaze time on regions of interest (ROI) was computed (**Figure 1**). ROIs were selected as follows: exposed half-face and occluded half-face (**Figures 1A,B**; ROI a, b). The aim was to show the pattern of visual exploration of the face as a function of veracity and side. The three-way ANOVA was performed for each emotion with veracity (genuine, masked) and side (standard, inverted) as within-subject factors.

For the number of fixations on the exposed half-faces, the main effects were found for side ($F_{(1,45)} = 12.85$, $p < 0.001$, $n_G^2 = 0.053$) and emotion ($F_{(2,90)} = 9.79$, $p < 0.001$, $n_G^2 = 0.007$). Furthermore, there were interactions between emotion and veracity ($F_{(2,90)} = 25.75$, $p < 0.001$, $n_G^2 = 0.013$) and emotion and side ($F_{(2,90)} = 7.55$, $p < 0.01$, $n_G^2 = 0.005$), but not veracity and side. The standard oriented faces received more fixations than inverted faces in all the emotions and the interaction is depicted in **Figure 2**.

For the number of fixations on occluded half-face ROI, the main effects were found for veracity ($F_{(1,45)} = 32.74$, $p < 0.001$, $n_G^2 = 0.04$) and emotion ($F_{(2,90)} = 38.31$, $p < 0.001$, $n_G^2 = 0.116$). There were interactions between emotion and veracity ($F_{(2,90)} = 7.37$, $p < 0.01$, $n_G^2 = 0.023$), emotion and side ($F_{(2,90)} = 39.85$, $p < 0.001$, $n_G^2 = 0.121$) and veracity and side

FIGURE 1 | The regions of interest (ROIs). Faces in standard orientation **(A,C)** and inverted **(B,D)** with ROIs defined as occluded hemi-face (a), exposed hemi-face (b), eyes region (c), nose region (d), mouth region (e). The variables of interest were extracted from ROIs and the three-way ANOVA was used with emotion as a between-groups factor.
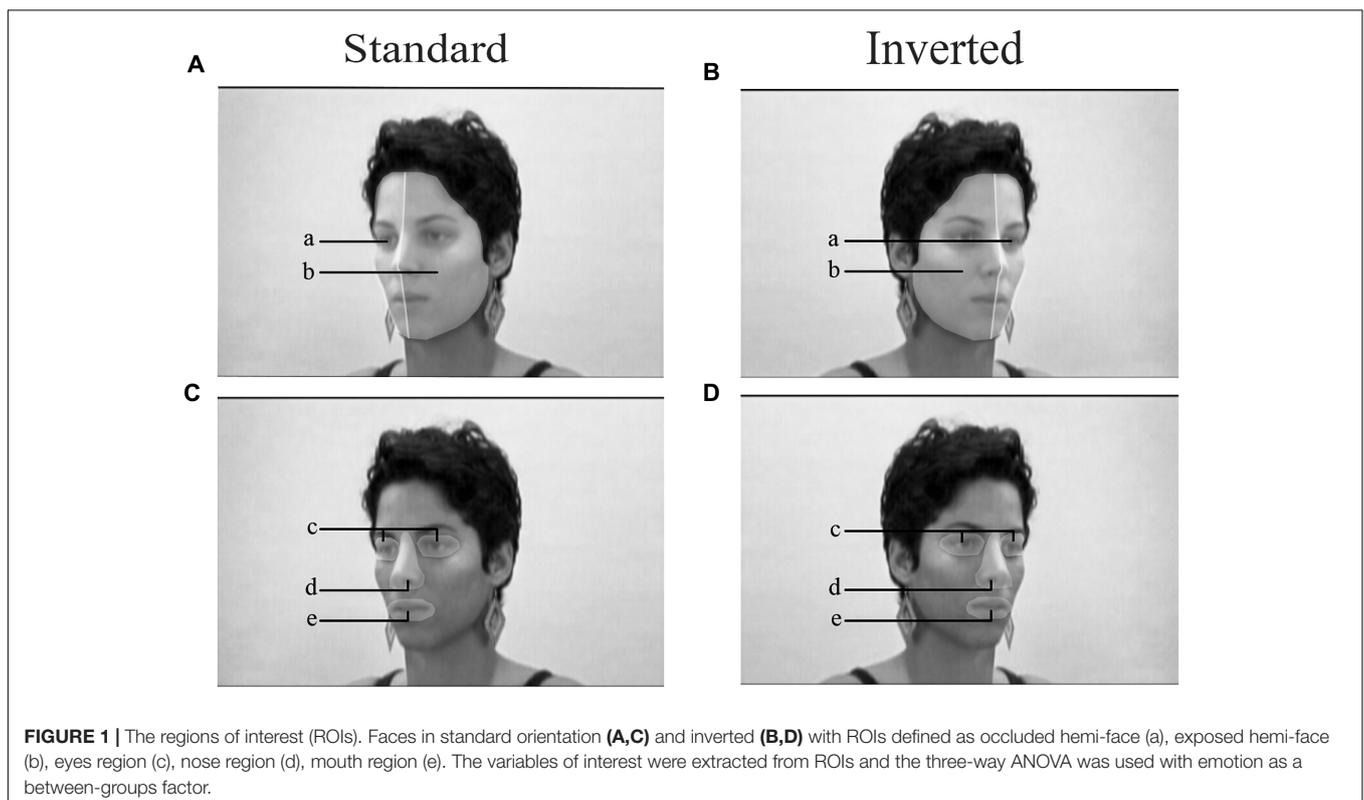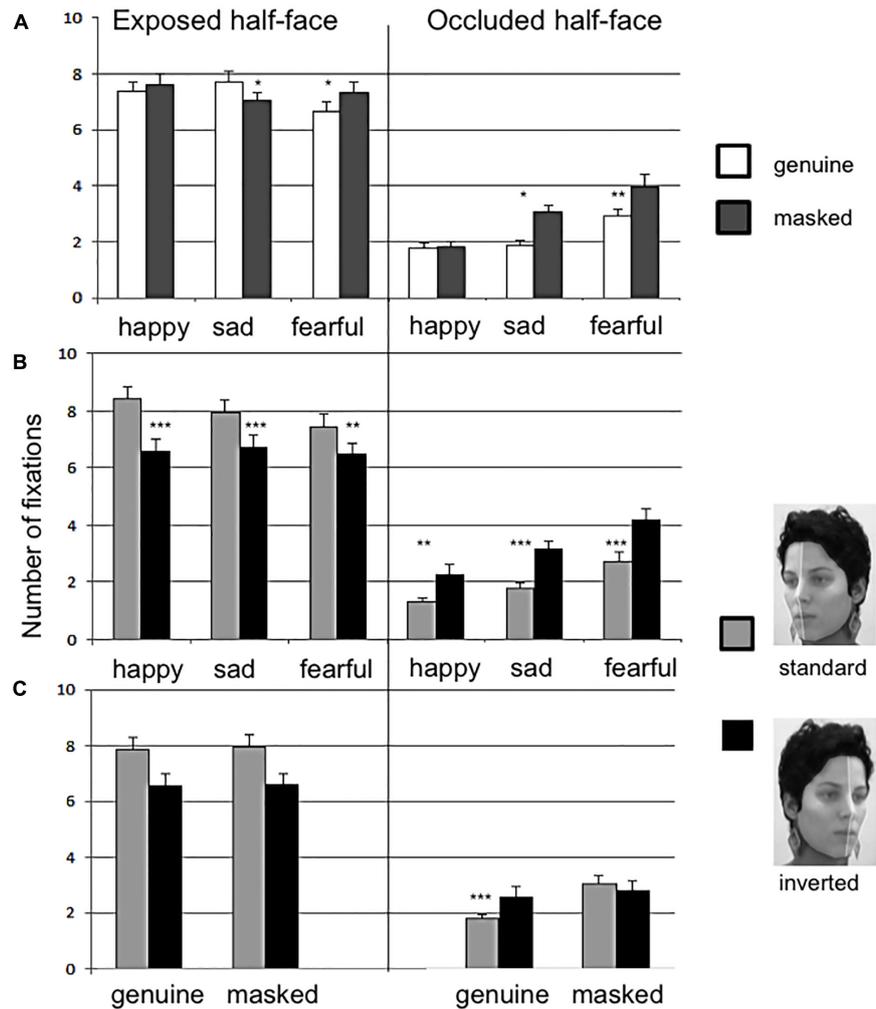
FIGURE 2 | The number of fixations on the ROIs. Exposed half-face (left side) and occluded half-face (right side) graphs show distribution of fixations in interaction with emotion and veracity **(A)**; emotion and side **(B)**; and veracity and side **(C)**. The error bars show standard error and statistical significance is marked by *$p < 0.5$; **$p < 0.01$; ***$p < 0.001$.

$(F_{(2,90)} = 12.32, p < 0.001, n_G^2 = 0.024)$. The direction of the interactions is depicted in **Figure 2**.

The gaze duration on the exposed half-faces showed a significant main effect for emotion $(F_{(2,90)} = 6.30, p < 0.01, n_G^2 = 0.016)$. The exposed half-faces of happy emotions $(M = 388$ ms) received significantly shorter gaze than sad $(M = 423$ ms, $p < 0.05)$, and fearful $(M = 429$ ms) emotions. There was no difference in gaze between sad and fearful (Bonferroni correction). No main effect was found for the gaze duration on occluded half-face ROI.

For the eyes, nose, and mouth ROI, ANOVA (**Figures 1C,D**; ROI c, d, e) was performed on gaze time with emotion (happy, sad, fearful), facial region (eye, nose, mouth), veracity (genuine, masked), and side (standard, inverted) as within-subject factors. The significant two-way interaction were found for emotion and region $(F_{(4,176)} = 9.64, p < 0.001, n_G^2 = 0.022)$ and

region and veracity $(F_{(2,88)} = 11.21, p < 0.001, n_G^2 = 0.025)$. There was a three-way interaction of region, emotion and veracity $(F_{(4,176)} = 6.60, p < 0.001, n_G^2 = 0.016)$. Pairwise comparison indicated longer gaze time on nose and eyes region in genuine happy emotions; longer gaze on eyes in genuine sad; and on nose in genuine fearful emotions. Longer gaze time was found for the mouth region in all masked emotions. The results are depicted in **Figure 3**.

DISCUSSION

The present study revealed that the pattern of gaze on dynamic human faces of three basic human emotions varied according to the side of the rotated face and the type of emotion being judged. Faces exposed from the left side had more fixations
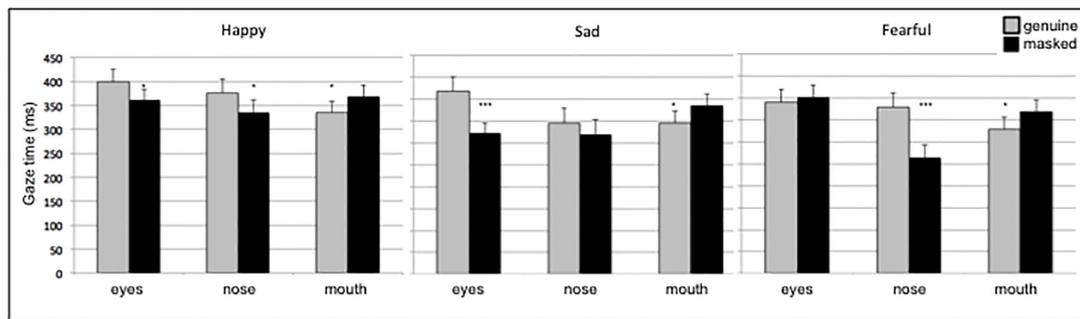
FIGURE 3 | The gaze time on eyes, nose and mouth ROIs. The gaze time duration as a function of region, emotion, and veracity. The error bars show standard error and statistical significance is marked by *p < 0.5; **p < 0.01; ***p < 0.001.

and the number of fixations decreased progressively from happy to sad and then fearful emotions. This pattern was evident mainly in the exposed hemiface, which suggests that subjects directed their gaze toward most salient features of the face. The occluded hemiface was fixated to a smaller extent and a different pattern was found; the number of fixations increased from happy to sad and fearful emotions. Thus, subjects may develop flexible scanning routines in order to gather additional information when facing rotated dynamic human faces. In this case, fixating in occluded facial regions seems to be associated with the increasing difficulty to judge the veracity of the presented emotion. A smaller number of fixations on the exposed right hemiface could evidence more efficient visual processing. However, when we look at the occluded right hemiface, the increase in the number of fixations indicates that there is much more need for additional visual clue than in the left occluded hemiface. These results evidence the presence of an asymmetry bias in dynamic emotions and indicate a specific strategy to extract additional visual clues for correct emotional judgment.

Previous studies showed that the left side of the face is more active than the right side when we express emotions. In addition, the aesthetic feeling is generally better for the left-faced images (Chiang et al., 2000; Adolphs, 2002; Okubo et al., 2013). People more often show the left cheek when they take selfies (Lindell, 2017) and portraits of faces are depicted mainly from the left side (McManus and Humphrey, 1973). Blackburn and Schirillo (2012) investigated preferences for the recognition of emotions according to the face's orientation. They recorded the reaction time and judgment of pleasantness of photos with smiling expressions rotated horizontally by 15°, emphasizing either the left or right side comparisons. The results indicated a left side bias, since it was more pleasurable to look at pictures in which the left side was more apparent, and the recognition time was lower in this condition. The pattern of visual exploration found in the present study is aligned with these findings. However, it is not clear whether this asymmetric bias may be supported by neuro-functional maturation when it comes to the face perception (Chiang et al., 2000; Adolphs, 2002) or is rather a culturally defined viewing preference (Marzoli et al., 2014).

Genuine and masked emotions are characterized by different brain states during their production, since the actor who was asked to produce a happy face was viewing a sad scene. Studies suggest that this incongruence is expected to produce asymmetry in the dynamics of emotion expression, by irregular onset/offset time of the muscle deformation, for example in a fake smile. Iwasaki and Noguchi (2016) showed that the change in mouth movements impaired the emotion perception of micro-expression in the eye regions, but only when showed after and not before the eye change. The diagnostic information for the emotional expression may be concentrated in different regions of the static face (Ekman and Friesen, 1978; Nusseck et al., 2008; Cristinzio et al., 2010). In a dynamic display of rotated faces, length of gaze on preferential facial regions varied as a function of the type of emotion. For genuine sad emotions the eyes were preferred, while for fear the nose was preferentially gazed. The increase in gaze time on mouth region in all masked emotions may be explained by increased difficulty in judging emotion's veracity. Buchan et al. (2007) showed that even modest increases in difficulty alter gaze patterns.

The results of this study showed longer IES in judging sad and fearful expressions compared to the happy expressions, combining that with shorter gaze time on happy faces, it indicates the effect known as happy emotion facilitation. This is in line with other studies of emotions in static faces, which defend that some expressions, such as a smile, are readily recognized due to deformation of muscles in only one or two facial regions (Nusseck et al., 2008; Du and Martinez, 2013). The genuine smile in static emotional faces is judged by the presence of crow's feet wrinkles around the eyes known as the Duchenne marker (Peron and Roy-Charland, 2013). However, as shown by studies with dynamic presentation of emotion, the temporal development of the expressions that change gradually over time produce subtle cues that enhance the perception of embedded information. These additional cues such as mouth deforming and opening reduce the importance of the eye region typically found is static face stimuli (Krumhuber and Manstead, 2009; Krumhuber et al., 2013; Korb et al., 2014). When looking at dynamic emotions, average gaze time was the longest for the nose region of happy faces, the mouth of sad faces, and the eyes of

fearful faces. Considering that this pattern was influenced by the genuine/masked factor, it is plausible that these results indicate a goal-driven viewing strategy.

Some limitations of this study must be acknowledged. First, our sample was limited in diversity (i.e., more than half were psychology undergraduate students). Second, all the movie clips were presented at the center of the screen, and the only manipulation was the mirroring of the faces. Thus, the extrapolation of conclusions on hemifield perception should be careful, since this variable was not controlled in our study. Finally, we also make no claim whether the perception of genuine and masked emotions behaves in a similar fashion for emotions other than happiness, sadness, and fear. Further studies should attend to these questions.

In summary, this study provides insight into the hemiface differences in emotion judgment and evidence of the asymmetry bias in dynamic stimuli contributing to understanding basic processes of social interactions.

AUTHOR CONTRIBUTIONS

YB and EM developed and contributed to the study design. YB and MA performed the data collection. KL performed the data analysis. All authors contributed to data interpretation and writing the manuscript, and approved the final version of the manuscript for submission.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02332/full#supplementary-material

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177. doi: 10.1016/S0959-4388(02)00301-X

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33, 717–746. doi: 10.1068/p5096

Blackburn, K., and Schirillo, J. (2012). Emotive hemispheric differences measured in real-life portraits using pupil diameter and subjective aesthetic preferences. *Exp. Brain Res.* 219, 447–455. doi: 10.1007/s00221-012-3091-y

Borod, J. C., Kent, J., Koff, E., Martin, C., and Alpert, M. (1988). Facial asymmetry while posing positive and negative emotions: support for the right hemisphere hypothesis. *Neuropsicologia* 26, 759–764. doi: 10.1016/0028-3932(88)90013-9

Buchan, J. N., Paré, H., and Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamics face processing. *Soc. Neurosci.* 2, 1–13. doi: 10.1080/17470910601043644

Butler, S., Gilchrist, I. D., Burt, D. M., Perrett, D. I., Jones, E., and Harvey, M. (2005). Are the perceptual biases found in chimeric face processing reflected in eye-movement patterns? *Neuropsychologia* 43, 52–59. doi: 10.1016/j.neuropsychologia.2004.06.005

Butler, S. H., and Harvey, M. (2006). Perceptual biases in chimeric faces processing: Eye-movement pattern biases in chimeric face processing: Eye-movement patterns cannot explain it all. *Brain Res* 1124, 96–99. doi: 10.1016/j.brainres.2006.09.069

Calvo, M. G., and Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: neurophysiological implications. *Cogn. Affect. Behav. Neurosci.* 9, 398–411. doi: 10.3758/CABN.9.4.398

Chelnokova, O., and Laeng, B. (2011). Three-dimensional information in face recognition? an eye-tracking study. *J. Vis.* 11, 1–15. doi: 10.1167/11.13.27

Chiang, C. H., Ballantyne, A. O., and Trauner, D. A. (2000). Development of perceptual asymmetry for free viewing of chimeric stimuli. *Brain Cogn.* 44, 415–424. doi: 10.1006/brcg.1999.1202

Cristinzio, C., N'Diaye, K., Seeck, M., Vuilleumier, P., and Sander, D. (2010). Integration of gaze direction and facial expression in patients with unilateral amygdala damage. *Brain* 133(Pt 1), 248–261. doi: 10.1093/brain/awp255

Dimberg, U., and Petterson, M. (2000). Facial reactions to happy and angry facial expressions: evidence for right hemisphere dominance. *Psychophysiology* 37, 693–696. doi: 10.1111/1469-8986.3750693

Du, S., and Martinez, A. M. (2013). Wait, are you sad or angry? Large exposure time differences required for the categorization of facial expressions of emotion. *J. Vis.* 13, 1–14. doi: 10.1167/13.4.13

Eisenbarth, H., and Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 860–865. doi: 10.1037/a0022758

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Rosenberg, E. L. (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS).* New York, NY: Oxford University Press.

Everdell, I. T., Marsh, H., Yurick, M. D., Munhall, K. G., and Paré, M. (2007). Gaze behaviour in audiovisual speech perception: asymmetrical distribution of face-directed fixations. *Perception* 36, 1535–1545. doi: 10.1068/p5852

Gilbert, C., and Bakan, P. (1973). Visual asymmetry in perception of faces. *Neuropsychologia* 11, 355–362.

Guo, K., Meints, K., Hall, C., Hall, S., and Mills, D. (2009). Left gaze bias in humans, rhesus monkeys and domestic dogs. *Anim. Cogn.* 12, 409–418. doi: 10.1007/s10071-008-0199-3

Harris, R. J., Young, A. W., and Andrews, T. J. (2014). Dynamic stimuli demonstrate a categorical representation of facial expression in the amygdala. *Neuropsychologia* 56, 47–52. doi: 10.1016/j.neuropsychologia.2014.01.005

Heller, W., and Levy, J. (1981). Perception and expression of emotion in right-handers and left-handers. *Neuropsychologia* 19, 263–272.

Hisao, J. H., and Cottrel, G. (2008). Two fixations suffice in face recognition. *Psychol. Sci.* 19, 998–1006. doi: 10.1111/j.1467-9280.2008.02191.x

Iwasaki, M., and Noguchi, Y. (2016). Hiding true emotions: micro-expressions in eyes retrospectively concealed by mouth movements. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep22049

Korb, S., With, S., Niedenthal, P., Kaise, S., and Grandjean, D. (2014). The perception and mimicry of facial movements predict judgments of smile authenticity. *PLOS ONE* 9:e99194. doi: 10.1371/journal.pone.0099194

Kret, M. E., Stekelenburg, J. J., Roelofs, K., and de Gelder, B. (2013). Perception of face and body expressions using electromyography, pupillometry and gaze measures. *Front. Psychol.* 4:28. doi: 10.3389/fpsyg.2013.00028

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonver. Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emotion Review* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., and Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 807–820. doi: 10.1037/a0017844

Lindell, A. (2017). Consistently showing your best side? Intra-individual consistency in selfie pose orientation. *Front. Psychol.* 8:246. doi: 10.3389/fpsyg.2017.00246

Marzoli, D., Prete, G., and Tommasi, L. (2014). Perceptual asymmetry and handedness: a neglected link? *Front. Psychol.* 5:163. doi: 10.3389/fpsyg.2014.00163

McLellan, T., Johnston, L., Dalrymple-Alford, J., and Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cogn. Emot.* 24, 1277–1292. doi: 10.1080/02699930903306181

McManus, I. C., and Humphrey, N. (1973). Turning the left cheek. *Nature* 243, 271–272. doi: 10.1038/243271a0

Miellet, S., Vizioli, L., He, L., Zhou, X., and Caldara, R. (2013). Mapping face recognition information use across cultures. *Front. Psychol.* 4:34. doi: 10.3389/fpsyg.2013.00034

Miguel, F. K., and Primi, R. (2014). Creation of emotional expressions videos using multimedia stimuli. *Teoria Prát.* 16, 155–168.

Moreno, C. R., Borod, J., Welkowitz, J., and Alpert, M. (1990). Lateralization for the perception and expression of facial expression as a function of age. *Neuropsychologia* 28, 199–209. doi: 10.1016/0028-3932(90)90101-S

Nicholls, M. E. R., Ellis, B. E., Clement, J. G., and Yoshino, M. (2004). Detecting hemifacial asymmetries in emotional expression with three-dimensional computerized image analysis. *Proc. R. Soc. Lond.* 271, 663–668. doi: 10.1098/rspb.2003.2660

Nusseck, M., Cunningham, D. W., Wallraven, C., and Bülthoff, H. H. (2008). The contribution of different facial regions to the recognition of conversational expressions. *J. Vis.* 8, 1–23. doi: 10.1167/8.8.1

Okubo, M., Ishikawa, K., and Kobayashi, A. (2013). No trust on the left side: hemifacial asymmetries of trustworthiness an emotional expressions. *Brain Cogn.* 82, 181–186. doi: 10.1016/j.bandc.2013.04.004

Peron, M., and Roy-Charland, A. (2013). Analysis of eye movements in the judgment of enjoyment and non-enjoyment smiles. *Front. Psychol.* 4:659. doi: 10.3389/fpsyg.2013.00659

Primi, R. (2014). Psychometric study of the computerized test of primary emotions perception. *Aval. Psicol.* 13, 1–9.

Sackeim, H. A., Gur, R. C., and Saucy, M. C. (1978). Emotions are expressed more intensely on the left side of the face. *Science* 202, 434–436.

Schurgin, W. M., Nelson, J., Lida, S., Ohira, H., Chiao, J. Y., and Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *J. Vis.* 14, 14. doi: 10.1167/14.13.14

Taylor, S., Workman, L., and Yeomans, H. (2012). Abnormal patterns of cerebral lateralization as revealed by the universal chimeric faces task in individuals with autistic disorder. *Laterality* 17, 428–437. doi: 10.1080/1357650X.2010.521751

Townsend, J. T., and Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.

Weyers, P., Mühleberger, A., Hefele, C., and Pauli, P. (2016). Electromyographic response to static and dynamic avatar emotional facial expressions. *Psychophysiology* 43, 450–453. doi: 10.1111/j.1469-8986.2006.00451.x

Yan, X., Young, A. W., and Andrews, T. J. (2017). Differences in holistic processing do not explain cultural differences in the recognition of facial expression. *Q. J. Exp. Psychol. (Hove)* 70, 2445–2459. doi: 10.1080/17470218.2016.1240816

Watching More Closely: Shot Scale Affects Film Viewers' Theory of Mind Tendency But Not Ability

Brendan Rooney¹* and Katalin E. Bálint²,³*

¹ School of Psychology, University College Dublin, Dublin, Ireland, ² Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands, ³ Institute for Media, Knowledge and Communication, University of Augsburg, Augsburg, Germany

Recent research debates the effects of exposure to narrative fiction on recognition of mental states in others and self, referred to as Theory of Mind. The current study explores the mechanisms by which such effects could occur in fictional film. Using manipulated film scenes, we conducted a between subject experiment ($N = 136$) exploring how film shot-scale affects viewers' Theory of Mind. Specifically, in our methods we distinguish between the trait Theory of Mind abilities (ToM ability), and the state-like tendency to recognize mental states in others and self (ToM tendency). Results showed that close-up shots (compared to long shots) of a character was associated with higher levels of Theory of Mind tendency, when the facial expression was sad but not when it was neutral. And this effect did not transfer to other characters in the film. There was also no observable effect of character depiction on viewers' general Theory of Mind ability. Together the findings suggest that formal and content features of shot scale can elicit Theory of Mind responses by directing attention toward character mental states rather than improving viewers' general Theory of Mind ability.

Keywords: theory of mind, shot scale, close up shot, facial expression, characters, film

INTRODUCTION

Theory of Mind (ToM), the psychological process by which people recognize and understand the mental states of others is arguably the most important process to human social functioning (Premack and Woodruff, 1978; Tomasello, 1999; Frith, 2012). Supporting this idea, marked social difficulties have been associated with deficits in ToM ability (Baron-Cohen et al., 2001; Shamay-Tsoory et al., 2010) and a low level of mind perception is associated with dehumanization or stigmatization of others (Cameron et al., 2015). Researchers distinguish between the representation of thoughts (cognitive ToM), feelings (affective ToM) and motivations (intentional ToM) of the other (e.g., Dziobek et al., 2008; Shamay-Tsoory et al., 2009). A large body of work also links ToM and related social cognition processes with understanding mental states in the self (Gallese, 2003; Decety and Sommerville, 2003; Neal and Chartrand, 2011; Erbas et al., 2016), further demonstrating the importance of ToM skills. Given the high social value of ToM, researchers are particularly interested in ways to elicit ToM and foster interpersonal sensitivity

(Meyer and Lieberman, 2016). Recently, it has been proposed that engagement with narrative fiction is particularly effective in this regard. Drawing on a well-established body of research that identifies the significance of facial-cues in social cognition (e.g., Ekman and Friesen, 1971; Baron-Cohen et al., 2001; van Kleef, 2009), we predict that manipulating film viewers' visual access to such social cues via shot scale in fictional film narrative will affect ToM response toward characters. We use a true experimental design to explore how shot-scale affects viewers' ToM. By embedding our research in the everyday act of natural film-viewing, this study offers a high-level of experimental control and a high-level of ecological validity; two typically conflicting characteristics that have been difficult to resolve in research to date.

Narrative fiction has high potential for evoking ToM responses (Mar et al., 2006). Research has demonstrated that high quality literary fiction (Kidd and Castano, 2013, 2016, 2017; Pino and Mazza, 2016), cinematic fiction (Black and Barnes, 2015) and narrativized video-games (Bormann and Greitemeyer, 2015) can improve ToM performance. These findings, however, seem to be difficult to replicate (Panero et al., 2016; Pino and Mazza, 2016), which may be a symptom of the fact that little is known about the mechanisms (in the viewer or the media) that facilitate the increase in ToM. Researchers draw on the work of Mar and Oatley (2008) and propose that ToM performance was superior because the fictional narratives elicited mental simulation and abstraction of social experience. They attribute the ToM performance effects to the effort involved in constructing a mental model of the characters. If this is true then it is reasonable to predict that features of the media may challenge or guide the construction of mental models and differentially affect ToM.

A growing body of work shows that audio–visual narratives are of special importance in eliciting ToM (see Levin et al., 2013; Tan, 2013). One of the main advantages of film over other media is the central role of faces in telling the story. The visual cues carried within human faces are strongly associated with ToM response (Calder et al., 2002; Mosconi et al., 2005; Itier et al., 2007; Itier and Batty, 2009; Fischer et al., 2012). For example, facial expressions and gaze direction are salient triggers of ToM (Frischen et al., 2007). van Kleef's (2009) Emotions as Social Information (EASI) model explains the link between emotional expression and the observer's response via inferential and affective reactions. Within this framework, numerous studies have demonstrated the effects of emotional expressions on viewers' character judgments (Hareli and Hess, 2010), attributions (van Doorn et al., 2015), and inferences about intentions (van Kleef et al., 2004; de Melo et al., 2014). Specific expressions (such as sadness or fear) include social information that tells a story to the viewer (Parkinson, 1999, 2001; Hareli and Hess, 2010). Testament to the importance of reading facial expressions in narrative, Cutting and Armstrong (2016) demonstrate that filmmakers use longer durations for scenes that present faces at a distance, amongst clutter, and argue that this is because viewers need more time to successfully read character expression in a cluttered

context. This demonstrates the formal features, such as shot-scale, play an important role in mediating social information in a film.

Shot-scale, defined as the apparent distance of characters from the camera, is one of the most effective visual devices in regulating the relative size of characters' faces, the relative proportion of the human figure to the background (Salt, 1992; Bowen and Thompson, 2013), and arranging film content according to its saliency (Carroll and Seeley, 2013). It has an impact on self-reported arousal (Canini et al., 2011), prosocial behavior (Cao, 2013), and character liking (Mutz, 2006). Previously, Bálint et al. (2016) observed a relationship between ToM responding and shot-scale distribution within a film. This study found that films with a higher proportion of closer shots (compared to films with fewer or no close shots) evoked higher levels of ToM responding. While the study statistically controlled for various potentially confounding variables, the test condition stimuli were different films (different stories, with different characters). Thus the study was subject to the typical trade-off between experimental control and ecological validity that has been common in the previous research to date. To overcome this limitation, by working with professional animation designers and filmmakers, the present study manipulates shot-scale (by inserting specifically designed close-up shots) into a film, while holding all other variables constant.

Previous research showing a relationship between shot-scale and ToM have failed to clarify if ToM is specifically targeted toward the character who is shown in close-up. It may be reasonable to predict that showing a close-up of a character would elicit ToM toward that character exclusively, yet previous research seems to claim that engagement with fiction results in a non-specific activation of ToM (e.g., Kidd and Castano, 2013; Black and Barnes, 2015). In that case, we would see a transfer effect of target character close-ups on ToM responses toward non-target characters. Thus the present study distinguishes between references to mental states of the target character (who featured in the close-ups) and a non-target character (a character who is seen only in extreme or very long-shots).

Previous studies exploring the effect of narrative fiction have primarily used tasks that explicitly require participants, in a forced-choice test, to identify emotional states (Reading the Mind in the Eyes test; Baron-Cohen et al., 1985), thoughts (Yoni task; Shamay-Tsoory and Aharon-Peretz, 2007; Kalbe et al., 2010) or beliefs (False-belief task; Wimmer and Perner, 1983; Baron-Cohen et al., 1985) from faces or descriptions of scenarios (Happé, 1993, 1994). While these measures have been widely and reliably used for decades (e.g., see Wellman et al., 2001; Fernández-Abascal et al., 2013; Devine and Hughes, 2016), they prompt ToM by explicitly asking about mental states (Apperly and Butterfill, 2009; Rosenblau et al., 2015). The nature of these tasks allows them to successfully tap into participants' ToM ability (or competence). It has been argued that beyond one's *ability* to understand mental states, people demonstrate individual differences in their *tendency* to do so, resulting in a 'competence–performance gap' (e.g., Meins et al., 2014). Unlike recent distinctions between explicit and

implicit ToM, that concern a person's conscious awareness of their deliberate efforts to mentalize, the distinction between ability and tendency concerns the extent to which a person is prompted or spontaneously models the mental states of another. Apperly (2012) argues that when exploring ToM we must recognize the distinction between the *ability* to conceive of the mind of the other, the mental processes involved in doing so, and the *tendency* to pay attention or care about the mind of the other. Prompting tasks are less sensitive to the absence of mental state references, and are less valid representations of individual differences in adults' spontaneous ToM (Meins and Fernyhough, 1999). This calls for the use of a measure of ToM-tendency, without which we can say little about unprompted social cognition in everyday life.

Addressing these abovementioned issues, this study employed a data collection method that distinguishes TOM-tendency and TOM-ability (Bálint et al., 2014). It also allows us to break ToM down further by coding whether the participant is mentalizing the character's cognition, emotion and intentions. Previous studies demonstrated that emotional and cognitive processes of social cognition are interdependent but separate mechanisms in the brain (Dziobek et al., 2008; Zaki and Ochsner, 2012). Therefore, our coding system differentiated whether the theory of mind response referred to cognitive, emotional or intentional mental states in the character. Our procedure was informed by standardized assessments of ToM processes using story-based stimuli and qualitative data collection (Heavey et al., 2000; Dziobek et al., 2006; Golan et al., 2006; Barnes et al., 2009; Dodell-Feder et al., 2013). We are also interested in exploring the way in which character depiction affects references to one's own mental states (hereafter referred to as ToM-self). This is particularly interesting in light of recent research showing that reading fiction does not elicit a shared emotional state with the characters (Pino and Mazza, 2016).

Our over-arching research question asks how shot-scale affects ToM, that is, the degree to which viewers perceive film-characters as intentional agents with mental states. To partition effects of shot scale from the content of the shot, we also manipulate facial expression of the character in the shot. We refer to these formal and content aspects of shot scale together as "character depiction." The main research question has three parts: we examine the effect of character depiction on ToM-tendency (RQ1), on ToM-ability (RQ2), and on ToM-self (RQ3). In all cases we predict that close ups increase ToM responses compared to long shots, and this effect will be more pronounced when the target is depicted in a close up with a sad facial expression compared to a neutral facial expression. The use of additional facial expressions may lead to interesting results in the context of the current study, but would each require an additional experimental group in the research design (and thus more participants). As an initial exploration, we use a sad facial expression due to its strong congruence with the major themes of separation in the film, the accompanying music and because a sad expression tends to signal affective tendencies in the observer (Knutson, 1996; Hess et al., 2000;

Hareli and Hess, 2010). Aside from testing our main hypotheses, for RQ1 and RQ2, we predict that ToM responses will be higher for the target character than for a non-target character, exploring any possible transfer effects from target to non-target characters.

MATERIALS AND METHODS

Overview

The present study was an online experiment (Qualtrics software) with an incomplete mixed-design. Shot-scale of character (Long-shot vs. Close-up) and Facial expression (Sad vs. Neutral) were levels to the between subject variable collectively referred to as "Character Depiction." The incomplete design was necessary because facial expression can be only manipulated in close-up condition but not in long shot, where character faces are not seen. The study design also included Character (Target vs. Non-target) as a within subject variable. ToM-tendency and ToM-ability were dependent variables.

Participants

Power analysis called for a sample between 117 and 141 so as to achieve sufficient power (0.9; $\alpha = 0.05$) to detect medium effect sizes. Recruiting through a university student participant pool, 170 people started the experiment; 26 of them did not complete the outcome measures and so could not be included in the study. Four participants were excluded due to excessively long duration with the stimulus (>6.5 min) indicating that they did not progress through the study in line with other participants (e.g., rewatching the video or engaging in other tasks). In addition, 2 participants were excluded for reporting to have seen the whole film before and 2 for reporting that they write English at an intermediate level or lower, as this may have affected their ability to express their ToM response (all other participants reported very good, fluent or native-speaker English abilities). Thus the final sample consisted of 136 participants (78 female, 34 male, 24 did not report gender; age: $M = 22.06$, $SD = 8.71$).

Stimulus Material

We used the first two sequences (2 min) of the multi-international-award winning animated film *Father and Daughter* (Dudok de Wit, 2001) with two characters, a man (non-target character) and a girl (target character). This segment included the title screen "*Father and Daughter*" and credits. The film is a two-dimensional hand-drawn animation, created in a simplistic style, characterized by a limited color palette and simple lines [see Bateman (2014) and Suckfüll (2010) for a formal analysis of the film and responses toward moments of narrative impact]. The film is accompanied by instrumental music (Waves of the Danube), but it contains no dialog or lyrics. The first sequence presents a man (non-target character) and a girl (target character) riding bicycles through a landscape. They arrive to a tree at a lake where the man gets off the bicycle. The girl stops and gets off her bicycle too. The man walks down to the water to a boat, then returns to hug the girl. He walks back to the water,

sits into the boat and rows away. The girl stays there standing and watches him rowing away. In the second sequence the girl is again on the same road riding her bicycle. She stops at the same tree, looks at the water, and after a moment she leaves again.

To manipulate the depiction of the target character (shot-scale and facial expression) we developed three different versions of the film excerpt (original version in long-shots, and two manipulated versions in close-ups). The first sequence presents the target character in a point of view shot as she looks at the man rowing away; the second point of view shot at the end of the second sequence presents her again as she looks at the water. In the manipulated versions of the film this long-shot (see **Figure 1**) was replaced by a close-up of the target character with either a sad or neutral facial expression (see **Figure 1**). Animation designers created and edited these close-up shots into the film to be a perfect fit to the style of the original artwork. The length of the films and close-up shots were kept constant. In all conditions the non-target character was depicted in long shots.

In a pilot study we tested the designed close-up shots for emotionality to make sure that the faces were perceived as neutral or sad. Thirty-one participants (15 females; 24 – 38 years old, $M = 31:28$; $SD = 3.96$ years) rated the test faces, after they were given some minor context. The faces were randomly selected by Qualtrics online survey designer, and presented in the order they would appear during the film. For each face, participants had to estimate the age of the depicted character (this is relevant to the narrative), and rate the perceived intensity of discrete emotions (i.e., emotionless, happy, sad, angry, disgusted, fearful, other emotion) on a 9-point scale from "not at all" to "very much." For each face, the average ratings on each emotion were calculated, these were then combined by group to give a group average rating on each emotion. Comparison of the mean ratings for each group showed that neutral faces evoked significantly higher ratings than sad faces on the dimensions of emotionless, $t(21.91) = -5.65$, $p < 0.05$, $CI_{95} = -4.64$, -2.15, and happy, $t(29) = -2.21$, $p < 0.05$, $CI_{95} = -1.42$, -0.056; and significantly lower ratings on dimensions of sad, $t(29) = 5.70$, $p < 0.05$, $CI_{95} = 1.71, 3.62$, angry, $t(29) = 5.63$, $p < 0.05$, $CI_{95} = 1.49, 3.21$, disgust, $t(29) = 3.12$, $p < 0.05$, $CI_{95} = 0.51, 2.42$, and fear $t(29) = 4.98$, $p < 0.05$, $CI_{95} = 1.52$, 3.64.

Procedure

The study was approved by the University College Dublin Research Ethics Committee. Participants were asked to complete the experiment in one sitting in an undistracted environment. First they reported their proficiency in the English language; then they were randomly assigned to one of three conditions (Long-shot, Sad close-up or Neutral close-up). After the film, participants responded to three open-ended questions (see **Table 1**). The first question asked participants to describe the story and was designed to allow for ToM-tendency responses. The second question was designed to capture ToM-ability using a prompt to describe the story from the target character's perspective. Finally, we prompted participants to describe their own experience so as to capture ToM-self. These questions were carefully designed to allow us to explore various ToM effects while minimizing demands on the participants. For example, we decided to use the same question to explore manipulation effects on both target and non-target characters. Once participants responded to these, they completed quantitative control measures of their experience and answered questions about their demographics. At the end of the session, participants were debriefed. Mental state references

FIGURE 1 | Images of character depiction by condition. Each row depicts an experimental condition. Participants watched a sequence containing two scenes. Depending on the condition, the shot presented in the first and third column (Scenes 1 and 2) was followed by a different shot. Condition 1 (top row) = Long shot is followed by another long shot with no visible facial expression; Condition 2 (middle row): Long shot is followed by a close up shot with neutral facial expression; Condition 3 (bottom row): Long shot is followed by a close-up shot with sad facial expression.

TABLE 1 | Description of questions used after viewing and the nature of ToM that they access.

Theory of Mind measures	Question
ToM-tendency	Q.1 Implicit question for unprompted ToM
	Please describe the story of the film scene in as much detail as possible using at least 6–10 sentences.
ToM-ability	Q.2 Explicit question prompting ToM for character
	Try to imagine the story from the perspective of the female character, how would you describe her feelings, thoughts, and intentions? Please write at least 6–10 sentences.
ToM-self	Q.3 Explicit question prompting ToM for self
	Describe your own experience during the movie. What happened to you while watching it? How would you describe your thoughts and feelings? Please write at least 6–10 sentences.

TABLE 2 | Coding frame used to assess frequency of mental state references.

Mental state type*	Reference to...	Example
Affect	Wishes, desires, or feelings	'Anxious,' 'excited,' 'feeling lonely'
Cognition	Memory function	'Forget,' 'remember,' 'was reminded'
	Knowledge	'Realize,' 'understand,' 'assume'
	Other cognition/metacognition	'Imagine,' 'accept,' 'pretend'
Intention	Expressed by an explicit word	'Intend,' 'determined to,' 'attempt'
	Expressed by a preposition	'To,' 'so that,' 'in order to'
	Expressed by a modal verb	'Have to,' 'must,' 'want'

*References to mental states of target (female), non-target (male) and self were coded separately.

were assessed using a quantitative content-analytic method by a trained coder, blind to the experimental conditions, developed in prior work (Bálint et al., 2014, 2016) and detailed in the next section. For each of the coded dependent variables, a randomly selected ten percent of descriptions was coded by another independent rater. Agreement was calculated for each variable using Krippendorf's Alpha; these yielded acceptable levels of agreement ($\alpha = 0.67$ to 1).

Measures
ToM-Tendency
To measure ToM-tendency we coded responses to question 1, identifying where participants made explicit reference to a mental state. These mental state references were also categorized as referring to the target (female) or the non-target (male) character, and by type of mental state (affective, cognitive or intention; see **Table 2**). Once coded, each participant's response was given a score for the frequency of mental state references, where higher scores are indicative of higher levels of ToM-tendency in a category.

ToM-Ability
The ability to use ToM was assessed by coding mental state references occurring in answers to question 2 (which prompted ToM). Again all utterances were coded for explicit references to character mental states and categorized by character (target/non-target) and by type. Higher scores mean more frequent references to mental states, indicating a higher level of ToM-ability.

ToM-Self
References to one's own mental states were coded in responses to question 3 that explicitly prompted reflection on the participant's own experience. Once the mental state reference was coded as a self-reference, it was further classified into one of three ToM types described in **Table 2**.

Controls
Besides gender, and age, we asked participants to indicate the highest level of education they obtained (see **Table 3**).

We also included control variables for familiarity with the film scene (yes or no); perceived quality of the film; self-reported proficiency in the English language (from 0 for basic proficiency in writing to 4 for first language); size of screen used; and word count of response to the open ended questions.

Data Analysis
Open responses were coded and group mean scores were calculated separately for target and non-target characters. Data were cleaned, distributions were explored, and descriptive statistics are reported in **Table 3**. Given the nature of the data (count data) the hypotheses were tested using Poisson regression. The independent (predictor) variables were Character depiction condition (Long-shot vs. Sad close-up vs. Neutral close-up), and Character (Target vs. Non-Target). Frequency of mental state references (categorized as ToM-tendency, ToM-ability, and ToM-self) were offset against the log transformed word count in participants' responses, to account for individual response length in a way that is required for analysis of count data (Agresti, 2003). In addition, to account for the personal relevance of the story, reported gender and age were included as covariates.

RESULTS

Before testing the hypotheses, a series of one-way ANOVA revealed no significant difference between the experimental groups in their level of English, $F(2,133) = 1.373$, $p > 0.05$, education, $F(2,109) = 0.266$, $p > 0.05$, age, $F(2,109) = 1.383$, $p > 0.05$, or the size of the screen that they viewed the film on, $F(2,109) = 0.472$, $p > 0.05$ (see **Table 3**). Importantly, there was no significant difference observed between the groups in perceived quality of the film, $F(2,109) = 1.133$, $p > 0.05$ demonstrating that the manipulation did not significantly detract from the viewing experience.

ToM-Tendency
To answer RQ1, we tested how Character depiction (close-up and facial expression) affected participants' ToM-tendency,

TABLE 3 | Means and standard deviations for output and other variables.

	Long Shot (Original)		Close-up Neutral		Close-up Sad		Scale
	M	**SD**	**M**	**SD**	**M**	**SD**	
ToM-tendency for target	2.29	2.14	1.97	1.52	3.17	2.14	~
ToM-tendency for target (Affective)	1.11	1.60	1.14	1.36	1.86	1.53	~
ToM-tendency for target (Cognitive)	0.74	0.85	0.51	0.90	0.72	0.97	~
ToM-tendency for target (Intentions)	0.43	0.61	0.32	0.53	0.58	0.65	~
ToM-tendency for non-target	1.23	1.68	0.76	0.95	1.28	1.47	~
ToM-ability for target	9.69	4.54	8.81	4.06	10.69	5.13	~
ToM-ability for non-target	0.54	1.01	0.35	0.72	0.36	0.54	~
ToM-self	5.80	3.15	6.21	3.74	6.67	3.84	~
ToM-self (Affective mental states)	3.89	2.35	4.62	3.18	5.08	3.20	~
ToM-self (Cognitive mental states)	1.74	1.75	1.68	1.53	1.39	1.18	~
ToM-self (Intentions)	0.17	0.45	0.05	0.23	0.19	0.47	~
Level of english	1.05	0.22	1.06	0.02	1.00	0.00	(First language) 1 to 5 (basic proficiency)
Highest level of formal education	4.00	1.47	3.76	1.38	3.92	1.42	1 (none), 2 (Secondary education, not completed), 3 (Secondary education, completed), 4 (Trade/Apprenticeship), 5 (Higher cert/Diploma), 6 (Bachelors Degree, 7 (Masters/Ph.D.).
Age	21.83	8.46	23.82	11.09	20.53	5.60	Age in years
Screen size	3.75	0.94	3.87	0.99	3.66	0.91	(Cinema size) 1 to 6 (<than smart phone)
Perceived film quality	4.68	1.60	5.11	1.56	4.63	1.46	(Bad) 1 to 7 (good)

~Average number of mental state references in a response that was 6 – 10 sentences (in the inferential analysis, Poisson Regression, this was offset against the log transformed word count of the response).

and if this differs for the target and non-target character. Analysis revealed a significant interaction between the depiction and the character (target/non-target), $F(5,214) = 17.43$, $p < 0.01$. Results demonstrated that the manipulation affected responses toward the target but not the non-target character (see **Figure 2**). Pairwise contrasts (using least significant difference) demonstrated that participants in the sad close-up condition made significantly more references to target character's mental states than those in the long-shot condition, $b(0.053) = 0.104$, $p = 0.05$, and the neutral close-up condition, $b(0.051) = 0.128$, $p = 0.013$. This pattern of findings is in line with our prediction that participants in the sad close-up condition would demonstrate the highest level of ToM-tendency, and that it was directed toward the target character's mental states (rather than the non-target character).

To explore the effect of character depiction further, we tested its effect on the type of mental states for the target character. Results revealed a significant interaction effect between depiction and type of mental state, $F(8,322) = 7.781$; $p < 0.05$. Pairwise comparisons demonstrated that the sad close-up condition was associated with significantly more references to the target character's affective mental states, than the neutral close-up, $b(0.047) = 0.095$, $p = 0.045$, or long-shot conditions, $b(0.046) = 0.104$, $p = 0.025$. No significant effects of character depiction were evident for the mental state references to cognitions or intentions.

ToM-Ability

RQ2 explored the effect of character depiction on ToM-ability. While mean levels of mental state references where higher for all conditions in question 2 (which explicitly prompted ToM) compared to question 1, using the same analysis, no significant effects of depiction were observed for the target, $F(2,214) = 0.38$, $p > 0.05$ or for the non-target characters, $F(2,214) = 1.27$, $p > 0.05$ (see **Figure 2**). These results do not support our prediction that the inclusion of close-up shots (especially emotional close-up shots) elicits participants' ToM-ability toward the target character, and thus hypothesis 2 was not supported.

ToM-Self

Finally, we tested hypothesis 3 predicting that Character depiction would affect references to one's own mental states (ToM-self). Results showed a marginally significant effect of depiction on the frequency of ToM-self, $\chi^2(4) = 9.16$, $p = 0.057$. Relative to the long-shot condition, participants in the neutral close-up condition referred to their own mental states more frequently, $\chi^2(1) = 3.137$, $Exp(B) = 1.13$; $CI_{95} = [0.987, 1.29]$; $p = 0.077$. This effect was even stronger for the sad close-up condition, $\chi^2(1) = 3.713$, $Exp(B) = 1.139$; $CI_{95} = [0.998, 1.30]$; $p = 0.054$, with no significant effect observed between the neutral and sad close-up conditions, $\chi^2(1) = 0.023$, $Exp(B) = 0.990$; $CI_{95} = [0.870, 1.126]$; $p = 0.879$. Thus it seems shot-scale and facial expression affected ToM-self, in line with hypothesis 3.
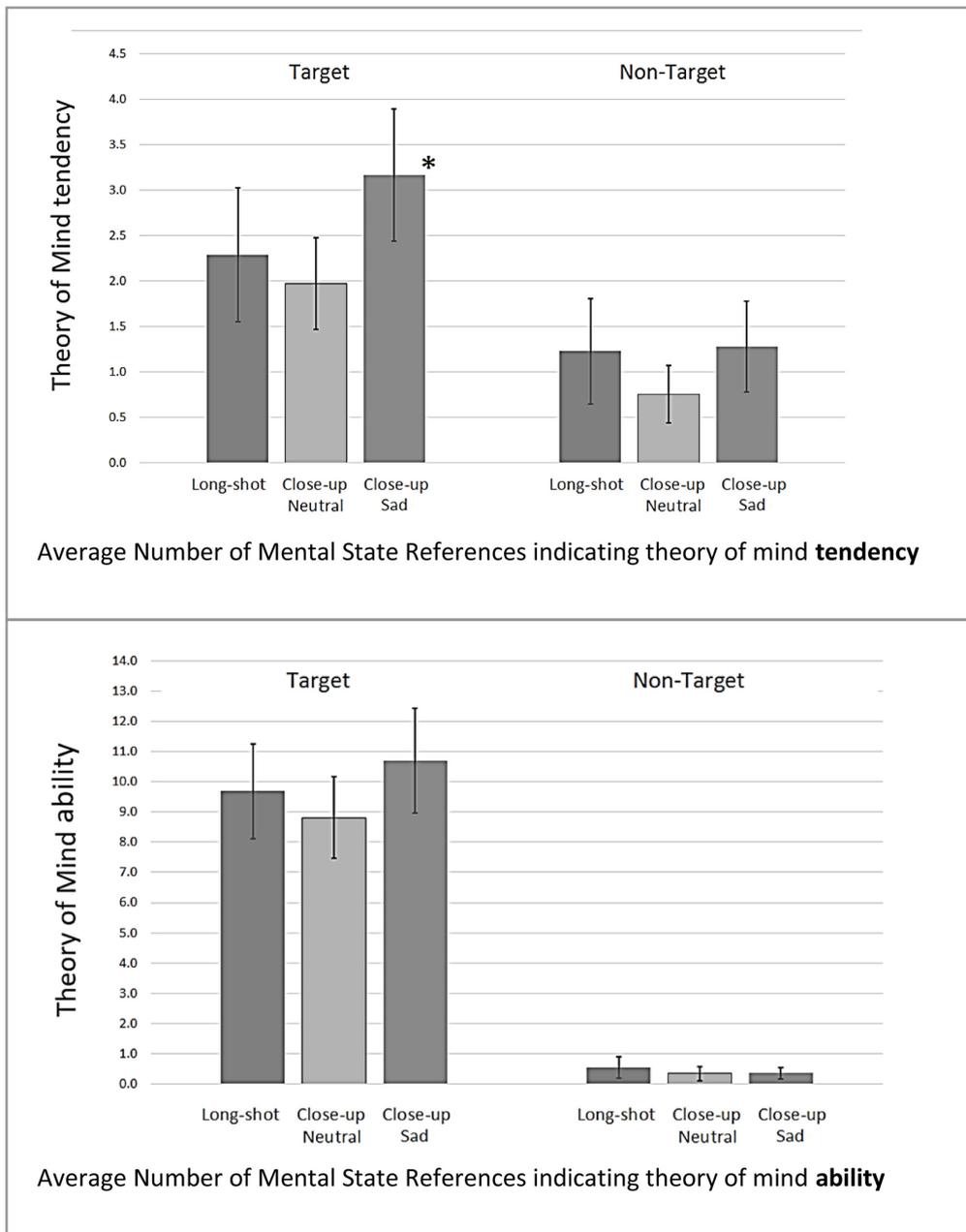
FIGURE 2 | Average number of mental state references indicating theory of mind tendency (top) and ability (bottom) with 95% Confidence Intervals (means shown regardless of word-count). Mental state references are presented categorized by whether they referred to the Target character or Non-target character. The conditions listed on the X-axis indicate the way in which the target character was depicted (non-target character was depicted in long shot for each condition). *ToM-tendency was significantly higher in for the target character when she was presented in close-up with a sad facial expression.

DISCUSSION

Using highly controlled yet ecologically valid film stimuli in a true experimental design, we explored the effect of character depiction on viewers' social cognition. Specifically, we were interested in viewers' tendency to reference character mental states (ToM-tendency) and their ability to do so when prompted (ToM-ability). Our findings demonstrate that shot-scale and facial expression do affect social cognition. Specifically, we observed that the close-ups of sad faces produced significantly higher ToM-tendency than other conditions, and that the use of a neutral close-up produced no more ToM-tendency than the long-shot version. This suggests that the increase in ToM-tendency response is not driven by merely presenting the character's face larger in the frame (i.e., at a smaller spatial distance from the viewer), but rather it is the social and

emotional information carried by the face that drives ToM-tendency responses. Importantly, this work extends the findings of previous research which demonstrated that exposure to fiction films (as opposed to documentary films) can elicit ToM response (Black and Barnes, 2015) by further exploring the way in which formal features of the narrative can effect types of ToM responding.

Supporting hypothesis 1, the current findings demonstrate an effect of character depiction on participants' ToM-tendency. More specifically, the close-up shots of the target character with a sad facial expression were associated with higher tendency to refer to the target character's mental states. Breaking down this finding into the different types of ToM response, we found that the effect was driven by affective ToM. That is, the increase in ToM response primarily consisted of references to the target character's *affective* mental states, rather than her cognition or intention. The manipulation of facial expression was one of emotional valence; the faces presented were either sad or emotionless. Thus this finding is in line with that of previous research showing that sad expressions elicit affective responses in observers (Knutson, 1996; Hess et al., 2000; Hareli and Hess, 2010). In line with this work, we predict that the ratio of references to the target character's feelings, thoughts or intentions may change in the context of a different film or if future researchers use different manipulations of facial expression, e.g., a thoughtful face.

An important aim of the present study was to explore whether the ToM-eliciting effect of seeing characters in close-up transfers to character depicted only in long shot (non-target character). Results of the current study showed no effect of character depiction on ToM responding toward the non-target character. Given that the inserted close-up shots did not feature the male character, this is perhaps not surprising. Indeed the characters in the stimulus of the current study differed not only in shot scale but along other dimensions (e.g., gender, age, appearance) which may have also inhibited a transfer effect. Nevertheless it is important because it demonstrates no effect of character depiction on any general form of ToM responses, where previous researchers have reported such general ToM effects using other media formats (e.g., Kidd and Castano, 2013, 2016, 2017; Black and Barnes, 2015; Pino and Mazza, 2016). In line with this, when prompted to recount the narrative events from the perspective of the target character (question 2), all groups demonstrated a higher frequency of mental state references to the target character, with no difference between conditions. This demonstrates that when called upon to do so, there was no difference between groups in terms of participants' *ability* to mentalize. Thus the use of close-up shots does not increase ToM responding by activating some enhanced mentalizing ability toward all characters, but rather it demonstrates, that close-ups work by directing our attention to the salient aspects of particular characters in the narrative. This is in line with Peskin and Astington's (2004) findings that adding metacognitive language (words expressing character mental states) into stories improved children's vocabulary on mental states, but not their performance in a false belief test. It seems that emotional

words in printed media have similar function to emotional faces in visual media. Furthermore, filmmakers are skilled in their ability to direct attention toward such important social cues (Loschky et al., 2015; Cutting and Armstrong, 2016).

Character depiction also appears to have affected references to one's own mental states (ToM-self). Close-ups of sad faces produced higher levels of ToM-self than other conditions. Results show that the neutral close-up condition produced more references to participants' own emotions than the long-shot condition, and the sad close-up condition produced even more references to participants' own emotions. These findings show a similar pattern as ToM-tendency responses for the target. They suggest that shot-scale and facial expression do not increase ToM-ability in general, but rather it increases one's tendency to mentalize toward the target, and in doing so may facilitate identification of their own mental states. This finding is in line with the large body of research linking the processes of social cognition of others, with self (e.g., Vygotsky, 1978; Neisser, 1988; Gallese, 2003) and the evidence for overlapping neural mechanisms in these processes (Decety and Jackson, 2004; Gallese, 2007; Lieberman, 2007; Rooney et al., 2012). Drawing on this work, we argue that directing attention to others' mental states, aids recognition of one's own mental states.

Synthesis

Taken together, the findings have implications for our understanding of the nature of ToM responses toward characters. They demonstrated that viewers did not differ in their ToM-ability, but rather they differed in their ToM-tendency. Showing the sad facial expression of a fictional character makes viewer mental states more readily available and featured more in their unprompted responses. But when prompted, all groups demonstrated the ability to call on social cognitive faculties to model the characters' mental states. These findings have important implications for the way in which ToM responses are measured in future research studies, and how they have been measured in the past. Here we show the way in which participants are asked about the experience can have a large impact on the findings. Accessing unprompted ToM responses may show differences that are not evident in prompted responses. This is particularly important given that so many ToM measures use direct questions to assess participants' ability to mentalize, rather than observing their uncontaminated responses. The failure to distinguish between these aspects of ToM may explain why previous research has presented conflicting and ambiguous results (e.g., Kidd and Castano, 2013, 2016, 2017; Panero et al., 2016; Pino and Mazza, 2016). In line with researchers such as Apperly (2012), Meins et al. (2014), and Rosenblau et al. (2015), we argue that capturing unprompted ToM responses taps in to participant's ToM-tendency and is representative of how ToM manifests in everyday life. Thus we too, call on researchers to give careful consideration to the operational definitions of social cognition they employ and the claims that can be made from their findings.

Limitations and Implications for Future Work

The strength of our own claims is somewhat limited by our focus on a single emotion manipulation, in a single film stimulus. Indeed the stimulus used was an animation rather than live action. This means that our findings presented in the context of simple highly designed visual information and call on future research to extend the findings with even more ecological validity. Nevertheless, we argue that this is an important strength of our work too. The stimulus used (its design and manipulation) offers a degree of experimental control that is typically difficult to achieve, without contaminating the ecological validity of the study. This major strength of the current study, compliments previous research that explored the relationship between ToM responses and shot-scale distribution in different films (Bálint et al., 2016). Taken together these studies, using various films (Bálint et al., 2016) and in a single experimentally manipulated film (the present study) provide evidence that the distribution of close-up shots may be utilized to increase ToM responding. Importantly, here we do not propose that simply inserting close-up shots into film will automatically generate increased ToM responses in viewers. Indeed, our findings that show an effect for the facial expression demonstrate that the social information presented in the close-up is particularly important in directing attention toward character mental states. In addition, we recognize that other ways in which the close-up is used will drive the ToM responses. Future research needs to explore these subtleties further by, for example, manipulating the number and position of the close-ups used, or how the depiction of the character might interact with viewer identity or personal relevance of the narrative.

We propose that using close-up shots of a sad expression drew participants' attention to the character's mental states, made character mental states more accessible and thus more likely to be integrated into viewers' models of the narrative. To be clear, we make this proposal for the current sample, and those within a population that they represent. The current sample of participants where relatively young adults in university education and our findings demonstrated that when eventually prompted to take the perspective of the target character, all groups regardless of condition, were able to do so. It is clear that the nature of our sample (convenient sample of volunteers) limits the extent to which the findings might generalize. While we stand by the way in which these findings speak to previous research, with similar limitations, we expect future research to build upon this limitation and design novel ways in which data can be collected (ethically) from a more representative and diverse population. For example, it remains to be seen how these findings may be extended to populations with deficits in social cognition such as participants with autism or schizophrenia. These populations may not be able to mentalize when prompted to do so. We might speculate that simply inserting close-ups would not increase ToM responding for an autistic population without some form of guidance or scaffolding, i.e., additional resources to draw attention to relevant social information.

CONCLUSION

Using a true experimental design, with highly controlled visual stimuli in an ecologically valid activity, the present study makes an important contribution to our understanding of theory of mind response. The findings indicate that depiction of the character can direct attentional focus toward their mental states, making them more accessible to the viewer and thus increasing viewers' *tendency* to use those mental states in a representation of the narrative. However, mere exposure to close-up faces of characters does not enhance general theory of mind *ability*, nor does it transfer to mentalizing with other characters depicted in long shots. Finally, the findings demonstrate that directing viewers' attention to the mental states of characters also elicits viewers' modeling of their own mental state, supporting the idea that understanding mental states in others is linked to understanding self. Findings of the present study show that shot scale and facial expression of character depiction is a powerful tool for shaping viewers' recognition of mental states in characters on screen and in self.

ETHICS STATEMENT

The current study was considered by the host institution's ethics review board and deemed exempt from full review due to the low risk involved. All participants read the full information sheet and gave electronic consent to participate. Participation was online, so no individual participant was directly canvassed and all were free to discontinue anonymously.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Agresti, A. (2003). *Categorical Data Analysis*. Hoboken, NJ: Wiley.

Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Q. J. Exp. Psychol.* 65, 825–839. doi: 10.1080/17470218.2012.676055

Apperly, I. A., and Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953–970. doi: 10.1037/a0016923

Bálint, K., Klausch, T., and Pólya, T. (2016). Watching closely shot scale influences theory of mind response in visual narratives. *J. Media Psychol.* doi: 10.1027/1864-1105/a000189

Bálint, K., Nagy, T., and Csabai, M. (2014). The effect of patient-centeredness and gender of professional role models on trainees' mentalization responses: implications for film-aided education. *Patient Educ. Couns.* 97, 52–58. doi: 10.1016/j.pec.2014.06.005

Barnes, J. L., Lombardo, M. V., Wheelwright, S., and Baron-Cohen, S. (2009). Moral dilemmas film task: a study of spontaneous narratives by individuals with autism spectrum conditions. *Autism Res.* 2, 148–156. doi: 10.1002/aur.79

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The reading the mind in the eyes test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatry* 42, 241–251. doi: 10.1111/1469-7610.00715

Bateman, J. A. (2014). "Looking for what counts in film analysis: a programme of empirical research," in *Visual Communication*, ed. D. Machin (Berlin: Mouton de Gruyter), 301–329.

Black, J., and Barnes, J. L. (2015). Fiction and social cognition: the effect of viewing award-winning television dramas on theory of mind. *Psychol. Aesthet. Creat. Arts* 9, 423–429. doi: 10.1037/aca0000031

Bormann, D., and Greitemeyer, T. (2015). Immersed in virtual worlds and minds: effects of in-game storytelling on immersion, need satisfaction, and affective theory of mind. *Soc. Psychol. Pers. Sci.* 6, 646–652. doi: 10.1177/1948550615578177

Bowen, C. J., and Thompson, R. (2013). *Grammar of the Shot*, 2nd Edn. Amsterdam: Focal Press.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., et al. (2002). Reading the mind from eye gaze. *Neuropsychologia* 40, 1129–1138. doi: 10.1016/S0028-3932(02)00008-8

Cameron, C. D., Harris, L. T., and Payne, B. K. (2015). The emotional cost of humanity: anticipated exhaustion motivates dehumanization of stigmatized targets. *Soc. Psychol. Pers. Sci.* 7, 105–112. doi: 10.1177/1948550615604453

Canini, L., Benini, S., and Leonardi, R. (2011). "Affective analysis on patterns of shot types in movies," in *Proceedings of the 7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011)*.

Cao, X. (2013). The effects of facial close-ups and viewers' sex on empathy and intentions to help people in need. *Mass Commun. Soc.* 16, 161–178. doi: 10.1080/15205436.2012.683928

Carroll, N., and Seeley, W. P. (2013). "Cognitivism, psychology, and neuroscience: movies as attentional engines," in *Psychocinematics: Exploring Cognition at the Movies*, ed. A. P. Shimamura (New York, NY: Oxford University Press), 53–75. doi: 10.1093/acprof:oso/9780199862139.003.0003

Cutting, J. E., and Armstrong, K. L. (2016). Facial expression, size, and clutter: inferences from movie structure to emotion judgments and back. *Atten. Percept. Psychophys.* 78, 891–901. doi: 10.3758/s13414-015-1003-5

de Melo, C. M., Carnevale, P. J., Read, S. J., and Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *J. Pers. Soc. Psychol.* 106, 73–88. doi: 10.1037/a0034251

Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71–100. doi: 10.1177/1534582304267187

Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533. doi: 10.1016/j.tics.2003.10.004

Devine, R. T., and Hughes, C. (2016). Measuring theory of mind across middle childhood: reliability and validity of the silent films and strange stories tasks. *J. Exp. Child Psychol.* 149, 23–40. doi: 10.1016/j.jecp.2015.07.011

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., and Hooker, C. I. (2013). Using fiction to assess mental state understanding: a new task for assessing theory of mind in adults. *PLOS ONE* 8:e81279. doi: 10.1371/journal.pone.0081279

Dudok de Wit, M. (2001). *Father and Daughter*. Amsterdam: CinéTé Filmproductie BV.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., et al. (2006). Introducing MASC: a movie for the assessment of social cognition. *J. Autism Dev. Disord.* 36, 623–636. doi: 10.1007/s10803-006-0107-0

Dziobek, I., Rogers, K., Fleck, S., Bahnemann, M., Heekeren, H. R., Wolf, O. T., et al. (2008). Dissociation of cognitive and emotional empathy in adults with Asperger syndrome using the multifaceted empathy test (MET). *J. Autism Dev. Disord.* 38, 464–473. doi: 10.1007/s10803-007-0486-x

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377

Erbas, Y., Sels, L., Ceulemans, E., and Kuppens, P. (2016). Feeling me, feeling you: the relation between emotion differentiation and empathic accuracy. *Soc. Psychol. Pers. Sci.* 7, 240–247. doi: 10.1177/1948550616633504

Fernández-Abascal, E. G., Cabello, R., Fernández-Berrocal, P., and Baron-Cohen, S. (2013). Test-retest reliability of the "Reading the Mind in the Eyes" test: a one-year follow-up study. *Mol. Autism* 4:33. doi: 10.1186/2040-2392-4-33

Fischer, A. H., Gillebaart, M., Rotteveel, M., Becker, D., and Vliek, M. (2012). Veiled emotions: the effect of covered faces on emotion perception and attitudes. *Soc. Psychol. Pers. Sci.* 3, 266–273. doi: 10.1177/1948550611418534

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694–724. doi: 10.1037/0033-2909.133.4.694

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2213–2223. doi: 10.1098/rstb.2012.0123

Gallese, V. (2003). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36, 171–180. doi: 10.1159/000072786

Gallese, V. (2007). Before and below "theory of mind": embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 659–669. doi: 10.1098/rstb.2006.2002

Golan, O., Baron-Cohen, S., Hill, J. J., and Golan, Y. (2006). The "Reading the Mind in Films" Task: complex emotion recognition in adults with and without autism spectrum conditions. *Soc. Neurosci.* 1, 111–123. doi: 10.1080/17470910600980986

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48, 101–119. doi: 10.1016/0010-0277(93)90026-R

Happé, F. G. E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093

Hareli, S., and Hess, U. (2010). What emotional reactions can tell us about the nature of others: an appraisal perspective on person perception. *Cogn. Emot.* 24, 128–140. doi: 10.1080/02699930802613828

Heavey, L., Phillips, W., Baron-Cohen, S., and Rutter, M. (2000). The awkward moments test: a naturalistic measure of social understanding in autism. *J. Autism Dev. Disord.* 30, 225–236. doi: 10.1023/A:1005544518785

Hess, U., Blairy, S., and Kleck, R. E. (2000). The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation. *J. Nonverbal Behav.* 24, 265–283. doi: 10.1023/A:1006623213355

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Itier, R. J., Villate, C., and Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: evidence from eye movement monitoring. *Neuropsychologia* 45, 1019–1028. doi: 10.1016/j.neuropsychologia.2006.09.004

Kalbe, E., Schlegel, M., Sack, A. T., Nowak, D. A., Dafotakis, M., Bangard, C., et al. (2010). Dissociating cognitive from affective theory of mind: a TMS study. *Cortex* 46, 769–780. doi: 10.1016/j.cortex.2009.07.010

Kidd, D., and Castano, E. (2013). Reading literary fiction improves theory of mind. *Science* 342, 377–380. doi: 10.1126/science.1239918

Kidd, D., and Castano, E. (2016). How levels of familiarity with literary and genre fiction relate to mentalizing. *Psychol. Aesthet. Creat. Arts* 11, 474–486. doi: 10.1037/aca0000069

Kidd, D. C., and Castano, E. (2017). Panero et al. (2016): failure to replicate methods caused the failure to replicate results. *J. Pers. Soc. Psychol.* 112, e1–e4. doi: 10.1037/pspa0000072

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *J. Nonverbal Behav.* 20, 165–182. doi: 10.1007/BF02281954

Levin, D. T., Hymel, A. M., and Baker, L. (2013). "Belief, desire, action and other stuff: theory of mind in movies," in *Psychocinematics: Exploring Cognition at the Movies*, ed. A. Shimamura (Oxford: Oxford University Press), 244–266. doi: 10.1093/acprof:oso/9780199862139.003.0013

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol* 58, 259–289. doi: 10.1146/annurev.psych.58.110405.085654

Loschky, L. C., Larson, A. M., Magliano, J. P., and Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PLOS ONE* 10:e0142474. doi: 10.1371/journal.pone.0142474

Mar, R. A., and Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspect. Psychol. Sci.* 3, 173–192. doi: 10.1111/j.1745-6924.2008.00073.x

Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J., and Peterson, J. B. (2006). Bookworms versus nerds: exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *J. Res. Pers.* 40, 694–712. doi: 10.1016/j.jrp.2005.08.002

Meins, E., and Fernyhough, C. (1999). Linguistic acquisitional style and mentalising development: the role of maternal mind-mindedness. *Cogn. Dev.* 14, 363–380. doi: 10.1016/S0885-2014(99)00010-6

Meins, E., Fernyhough, C., and Harris-Waller, J. (2014). Is mind-mindedness trait-like or a quality of close relationships? evidence from descriptions of significant others, famous people, and works of art. *Cognition* 130, 417–427. doi: 10.1016/j.cognition.2013.11.009

Meyer, M. L., and Lieberman, M. D. (2016). Social working memory training improves perspective-taking accuracy. *Soc. Psychol. Pers. Sci.* 7, 1–9. doi: 10.1177/1948550615624143

Mosconi, M. W., Mack, P. B., McCarthy, G., and Pelphrey, K. A. (2005). Taking an "intentional stance" on eye-gaze shifts: a functional neuroimaging study of social perception in children. *Neuroimage* 27, 247–252. doi: 10.1016/j.neuroimage.2005.03.027

Mutz, D. C. (2006). *Effects of "In-Your-Face" Television Discourse on Perceptions of a Legitimate Opposition. Hendricks Symposium–Department of Political Science.* Available at: http://digitalcommons.unl.edu/politicalsciencehendricks/16

Neal, D. T., and Chartrand, T. L. (2011). Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. *Soc. Psychol. Pers. Sci.* 2, 673–678. doi: 10.1177/1948550611406138

Neisser, U. (1988). Five kinds of self-knowledge. *Philos. Psychol.* 1, 35–59. doi: 10.1080/09515088808572924

Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., et al. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *J. Pers. Soc. Psychol.* 111, e46–e54. doi: 10.1037/pspa0000064

Parkinson, B. (1999). Relations and dissociations between appraisal and emotion ratings of reasonable and unreasonable anger and guilt. *Cogn. Emot.* 13, 347–385. doi: 10.1080/026999399379221

Parkinson, B. (2001). "Putting appraisal in context," in *Appraisal Processes in Emotion: Theory, Methods, Research*, eds K. R. Scherer, A. Schorr, and T. Johnstone (New York, NY: Oxford University Press), 173–186.

Peskin, J., and Astington, J. W. (2004) The effects of adding metacognitive language to story texts. *Cogn. Dev.* 19, 253–273. doi: 10.1016/j.cogdev.2004.01.003

Pino, M. C., and Mazza, M. (2016). The use of "Literary Fiction" to promote mentalizing ability. *PLOS ONE* 11:e0160254. doi: 10.1371/journal.pone.0160254

Premack, D. G., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512

Rooney, B., Keyes, H., and Brady, N. (2012). Shared or separate mechanisms for self-face and other-face processing? Evidence from adaptation. *Front. Psychol.* 3:66. doi: 10.3389/fpsyg.2012.00066

Rosenblau, G., Kliemann, D., Heekeren, H. R., and Dziobek, I. (2015). Approximating implicit and explicit mentalizing with two naturalistic video-based tasks in typical development and autism spectrum disorder. *J. Autism Dev. Disord.* 45, 953–965. doi: 10.1007/s10803-014-2249-9

Salt, B. (1992). *Film Style and Technology: History and Analysis*. London: Starword.

Shamay-Tsoory, S. G., and Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia* 45, 3054–3067. doi: 10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132, 617–627. doi: 10.1093/brain/awn279

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., and Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex* 46, 668–677. doi: 10.1016/j.cortex.2009.04.008

Suckfüll, M. (2010). Films that move us: moments of narrative impact in an animated short film. *Projections* 4, 41–63. doi: 10.3167/proj.2010.040204

Tan, E. S. H. (2013). "The empathic animal meets the inquisitive animal in the cinema: notes on a psychocinematics of mind reading," in *Psychocinematics: Exploring Cognition at the Movies*, ed. A. P. Shimamura (New York, NY: Oxford University Press).

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

van Doorn, E. A., van Kleef, G. A., and van der Pligt, J. (2015). Deriving meaning from others' emotions: attribution, appraisal, and the use of emotions as social information. *Front. Psychol.* 6:1077. doi: 10.3389/fpsyg.2015.01077

van Kleef, G. A. (2009). How emotions regulate social life: the emotions as social information (EASI) model. *Curr. Dir. Psychol. Sci.* 18, 184–188. doi: 10.1111/j.1467-8721.2009.01633.x

van Kleef, G. A., De Dreu, C. K. W., and Manstead, A. S. R. (2004). The interpersonal effects of anger and happiness in negotiations. *J. Pers. Soc. Psychol.* 86, 57–76. doi: 10.1037/0022-3514.86.1.57

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraircing function of wrong bekfs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Zaki, J., and Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls, and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085

Corrigendum: Watching More Closely: Shot Scale Affects Film Viewers' Theory of Mind Tendency But Not Ability

Brendan Rooney[1]* and Katalin E. Bálint[2,3]

[1] School of Psychology, University College Dublin, Dublin, Ireland, [2] Tilburg Center for Cognition and Communication, Tilburg University, Tilburg, Netherlands, [3] Institute for Media, Knowledge and Communication, University of Augsburg, Augsburg, Germany

A corrigendum on

Watching More Closely: Shot Scale Affects Film Viewers' Theory of Mind Tendency But Not Ability

by Rooney, B., and Bálint, K. E. (2018). Front. Psychol. 8:2349. doi: 10.3389/fpsyg.2017.02349

In the original article, we referred to Canini et al., 2013. This was an error. It should be Canini et al., 2011. In the reference section the reference was incorrectly written as:

Canini, L., Benini, S., and Leonardi, R. (2013). Classifying cinematographic shot types. *Multimed. Tools Appl.* 62, 51–73. doi: 10.1007/s11042-011-0916-9

The correct reference should be:

Canini, L., Benini, S., and Leonardi, R. (2011). "Affective analysis on patterns of shot types in movies," in *Proceedings of the 7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011).*

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way.

The original article has been updated.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

It Is Not Just in Faces! Processing of Emotion and Intention from Biological Motion in Psychiatric Disorders

Łukasz Okruszek *

Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

Social neuroscience offers a wide range of techniques that may be applied to study the social cognitive deficits that may underlie reduced social functioning—a common feature across many psychiatric disorders. At the same time, a significant proportion of research in this area has been conducted using paradigms that utilize static displays of faces or eyes. The use of point-light displays (PLDs) offers a viable alternative for studying recognition of emotion or intention inference while minimizing the amount of information presented to participants. This mini-review aims to summarize studies that have used PLD to study emotion and intention processing in schizophrenia (SCZ), affective disorders, anxiety and personality disorders, eating disorders and neurodegenerative disorders. Two main conclusions can be drawn from the reviewed studies: first, the social cognitive problems found in most of the psychiatric samples using PLD were of smaller magnitude than those found in studies presenting social information using faces or voices. Second, even though the information presented in PLDs is extremely limited, presentation of these types of stimuli is sufficient to elicit the disorder-specific, social cognitive biases (e.g., mood-congruent bias in depression, increased threat perception in anxious individuals, aberrant body size perception in eating disorders) documented using other methodologies. Taken together, these findings suggest that point-light stimuli may be a useful method of studying social information processing in psychiatry. At the same time, some limitations of using this methodology are also outlined.

Keywords: biological motion, schizophrenia, affective disorders, eating disorders, anxiety disorders, neurodegenerative diseases, social neuroscience, emotion recognition

INTRODUCTION

It has recently been highlighted that the field of social neuroscience offers a number of techniques that can be effectively used for studying the processes that may underlie reduced functioning of psychiatric patients (Cacioppo et al., 2014; Fett et al., 2015; Ibáñez et al., 2016). Social cognitive deficits are found in various psychiatric populations (Samamé et al., 2012; Savla et al., 2013; Plana et al., 2014; Weightman et al., 2014) and may be of great importance for patients' functional capacity (Fett et al., 2011). Although a wide range of techniques can be used to examine emotion recognition and theory of mind in patients, a substantial proportion of studies have examined the processing of social information conveyed by static displays of human faces or eyes (Savla et al., 2013). While the use of these types of stimuli is well-established in social cognitive studies, the static nature of the

stimuli limits the ecological validity of this measurement method, given the dynamic nature of social cognitive processes. To overcome this problem, one may utilize videoed vignettes of actions and/or interactions of real-life social agents for social cognitive examination (McDonald et al., 2003; Dziobek et al., 2006). However, for such complex stimuli to be correctly processed a wide range of both verbal and non-verbal signals (facial and bodily movements, gaze direction, prosody, proximity between the agents) must be taken into consideration. Thus, patients' inability to process these types of stimuli correctly reflects a wide variety of underlying social cognitive problems. Furthermore, perception of either static or dynamic full displays of real-life actors may be affected by numerous confounding factors, e.g., likeability of the agent presented or cultural factors (Mehta et al., 2011).

Minimalistic, dynamic, point-light displays (PLDs) may be a viable alternative for presenting social information while avoiding the problems that can afflict studies that use static or dynamic full displays of agents. Since the introduction of point-light motion methodology to the field of experimental psychology, by Swedish psychologist Gunnar Johansson (Johansson, 1973), numerous researchers have used it to show that the human visual system is finely tuned to decipher information on the gender, physical characteristics, affective state, or intention of the person presented (see Troje, 2013) for a review of studies on biological motion perception in healthy individuals). Furthermore, the presentation of whole-body motion that is visually downgraded to several point-lights attached to the main joints and limbs of the body, may be a culturally unbiased way to study social information processing (Pica et al., 2011).

In addition, the pattern of neural activity and connectivity during the processing of PLDs may be, to some extent, similar to that observed when processing other forms of social agent presentation (faces, animated shapes; Dasgupta et al., 2017). Processing of the whole-body motion from PLDs is strongly linked to the posterior superior temporal sulcus (pSTS) activation, which is mostly lateralized to the right hemisphere (Van Overwalle and Baetens, 2009). At the same time, face-processing network includes occipital and fusiform face areas, posterior and anterior STS, as well as amygdala and insula (Duchaine and Yovel, 2015). Furthermore, while processing of both types of the stimuli strongly engage pSTS, (Deen et al., 2015) observed that, despite significant overlap, pSTS responses to faces and PLDs may differ reliably, with face-sensitive pSTS region being placed slightly anterior to region responding to biological motion.

While a large body of research was devoted to the study of various aspects of face perception across psychiatric disorders, knowledge of emotion or intention processing on the basis of biological motion processing in patients is relatively scarce. This may be a little surprising, especially given the amount of attention that biological motion processing received in the field of neurodevelopmental disorders (Pavlova, 2012, 2017). Thus, this article aims to provide a review of findings on recognition of emotion or intention from biological motion across psychiatric disorders.

METHODS

A PubMed search using terms "("biological motion" or "point-light motion") and ("emotion" or "intention")" was performed to identify studies for the current mini-review. Additionally the search was supplemented by relevant articles found by reviewing the references provided in the identified articles. Relevant studies have been grouped accordingly to major categories from ICD-10 "Mental Behavioral and Neurodevelopmental disorders" section. Given that the studies on biological motion processing in autism spectrum disorders and other developmental disorders were reviewed in Pavlova (2012), and more recently in Pavlova (2017), findings from this areas are not discussed in the current review. Additionally, description of the commonly used PLD tasks has been provided in the "*Tasks*" section.

TASKS

Most of the studies which examined processing of emotion from PLDs in psychiatric populations (Schizophrenia: Bigelow et al., 2006; Couture et al., 2010; Henry et al., 2010; Brittain et al., 2012; Kern et al., 2013; Vaskinn et al., 2016; Bipolar Disorder: Vaskinn et al., 2017; MDD: Loi et al., 2013; Eating Disorders: Zucker et al., 2013; Lang et al., 2015; Dapelo et al., 2017; Alzheimer's Dementia: Henry et al., 2012) utilized stimuli developed by Heberlein et al. (2004). During the Emotion from the Biological Motion task (EBM) participant observes a single point-light agent walking across the screen and his/her task is to select the alternative which best describes agent's affective state (happiness, sadness, fear, anger, neutral). Another set of stimuli for investigating emotion recognition in dyadic and monadic PLDs was developed by Lorey et al. (2012) and effectively applied to investigate social cognitive processes in psychiatric populations (Kaletsch et al., 2014a,b).

Two tasks were used to investigate intentions inference from PLDs across psychiatric populations. During the Gesture Perception Task (GPT; Jaywant et al., 2016b) participant is presented with single PLD (Zaini et al., 2013) and has to: (i) classify gesture performed by PLD as either communicative or non-communicative; and (ii) verbally describe each action. Alternatively, Communicative Interaction Database Five Alternative Forced Choice task (CID-5; Manera et al., 2015) present 21 dyadic PLDs and requires participant to: (i) decide if agents communicated or acted independently; and (ii) to identify the correct action description among the five alternatives.

NEURODEGENERATIVE DISORDERS

Two studies examined the ability to recognize emotion from PLDs in patients with Alzheimer's dementia (Henry et al., 2012; Insch et al., 2015). The first (Henry et al., 2012) observed that while deficits in facial emotion recognition can be found both in patients with AD and in individuals with mild cognitive impairment (MCI), deficient EBM performance

was observed only in patients with AD. This observation was further corroborated by Insch et al. (2015), who found decreased performance in emotion recognition from PLDs in older adults, which was further reduced in patients with AD.

Another line of studies (Jaywant et al., 2016a,b) examined biological motion processing in individuals with Parkinson's disease (PD). Interestingly, patients with PD demonstrated reduced sensitivity to biological motion (Jaywant et al., 2016a) and recognition of non-communicative, object-oriented gestures (Jaywant et al., 2016b), but did not differ from healthy controls when describing communicative gestures in GPT (Jaywant et al., 2016b).

SCHIZOPHRENIA

Patients with schizophrenia (SCZ) present deficits across multiple domains of the biological motion processing, including biological vs. scrambled motion discrimination (Kim et al., 2005, 2011, 2013; Kern et al., 2013; Jahshan et al., 2015) and detection of masked biological motion (Hastings et al., 2013; Spencer et al., 2013; Matsumoto et al., 2015, 2017). For a detailed discussion of the behavioral and neural correlates of biological motion processing in SCZ, please refer to our recent systematic review and meta-analysis of studies in this area (Okruszek and Pilecka, 2017). A sub-meta-analysis of six studies that assessed EBM performance (Bigelow et al., 2006; Couture et al., 2010; Henry et al., 2010; Brittain et al., 2012; Kern et al., 2013; Vaskinn et al., 2016) revealed moderate to large ($d = 0.61$) deficits in SCZ. Thus, while still impaired, this domain of social cognition differentiates SCZ from healthy controls to a lesser extent than does facial emotion identification ($d = 0.89$; Kohler et al., 2010) or emotional prosody processing ($d = 1.24$; Hoekert et al., 2007). Furthermore, links have been found between recognition of emotion from biological motion and higher-order social perception (Brittain et al., 2012), facial emotion identification and empathic accuracy (Olbert et al., 2013), and neurocognition and functional capacity (Engelstad et al., 2017).

Furthermore, we have shown that SCZ display reduced ability to explicitly categorize actions of dyadic PLDs as either communicative or individual in CID-5 (Okruszek et al., 2015). However, we have recently observed that despite biological motion processing deficits, SCZ are still able to use information carried by a communicative action of one agent to predict the action of the other agent ("interpersonal predictive coding"; Okruszek et al., 2018). Furthermore, similar perceptual biases were elicited in SCZ and in healthy controls by observing communicative gestures of one agent during PLD-based simultaneous masking detection task (Okruszek et al., 2017a). These findings, suggesting intact interpersonal predictive coding in SCZ were congruent with our recent functional neuroimaging results (Okruszek et al., 2017b): reduced activity and functional connectivity of the right pSTS, but similar action observation network activity were observed in SCZ compared with healthy controls during processing of communicative interactions vs. individual actions of dyadic PLDs (Okruszek et al., 2017b).

AFFECTIVE DISORDERS

While recognition of biological motion appears intact in patients with major depressive disorder (Kaletsch et al., 2014b), studies of emotion recognition from PLDs have revealed the same mood-congruent biases in patients when processing biological motion using other types of social stimuli, i.e., faces (Bourke et al., 2010) or verbal prosody (Péron et al., 2011; Loi et al., 2013). Using EBM, Loi et al. (2013) found that patients with depression exhibit a deficit in the recognition of happiness, but not of anger, sadness, fear, or neutral states, compared with both patients with depression in remission and healthy controls with no history of depression. On the other hand, Kaletsch et al. (2014b) observed that patients with MDD rate negative (but not positive) dyadic interactions presented in PLDs as more negative and more intense than do healthy controls.

Recently, a small but significant ($d = 0.40$) impairment in EBM was documented in patients with bipolar disorder (Vaskinn et al., 2017). No differences were observed between patients with type I and type II BD, or between patients with and without a history of psychosis. Furthermore, unlike the MDD group, patients with BD showed a similar extent of impairment for all emotions and no mood-congruent biases, and no association between impairments and either depressive or manic symptomatology.

ANXIETY DISORDERS

It has been documented that depth-ambiguous displays of biological motion are more often interpreted as being oriented toward rather than away from the viewer, even when both interpretations are equally plausible (Vanrie et al., 2004). This effect was termed "facing-the-viewer bias" and is usually explained by the preposterous consequences associated with mistaking an approaching agent for a retreating one, and thus may be interpreted as the impact of top-down factors (e.g., attribution of hostile intentions) on perception. One of the factors that has been shown to affect susceptibility to facing-the-viewer bias during the perception of a bistable point-light walker is the level of anxiety in an individual (Van de Cruys et al., 2013; Heenan and Troje, 2014; Heenan et al., 2014). Furthermore, facing-the-viewer bias has been found to be reduced by physical exercise and an anxiety-reducing task (progressive muscle relaxation; Heenan and Troje, 2014). Interestingly, the opposite bias (interpreting the walker as facing away from the observer) was observed in individuals with high levels of social anxiety, which can be interpreted in terms of "wishful seeing" and protecting oneself (Van de Cruys et al., 2013). Facing-the-viewer bias has also been found to be mediated by inhibitory abilities in individuals with high social anxiety (Heenan and Troje, 2015).

EATING DISORDERS

The main focus of studies using biological motion stimuli to study social perception in eating disorders has been abilities

associated with processing the weight or BMI of the agent. Individuals with either anorexia nervosa (AN; Phillipou et al., 2016) or bulimia nervosa (BN; Vocks et al., 2007) were shown to display abnormal processing of the body size of PLDs. When it comes to emotion processing, two studies examined EBM performance in individuals with AN (Zucker et al., 2013; Lang et al., 2015). Zucker et al. (2013) found overall worse recognition of emotion from biological motion by patients with AN compared with both healthy controls and weight-restored (≥12 months) individuals with AN. Deficient EBM performance was associated with symptom severity as measured by self-reported dietary restraint in patients. Moreover, analyses of the recognition of specific emotions revealed that individuals with AN attributed more anger and less sadness to the PLDs than controls and weight-restored individuals with AN. No differences were found, however, for the remaining categories (fear, happiness, neutral). These results were partially replicated by Lang et al. (2015), who found decreased recognition of sadness from PLDs in a well-powered (*n* = 97) sample of females with AN compared with healthy controls. Furthermore, overall worse recognition of emotion conveyed by biological motion was observed in adolescent individuals with AN compared with demographically matched controls (Lang et al., 2015). Finally, emotion recognition from faces and point-light motion was recently compared in individuals with AN and BN by Dapelo et al. (2017), who found specific impairment in processing emotion from faces in both groups of individuals with eating disorders, but no differences in EBM performance between patients and healthy controls.

PERSONALITY DISORDERS

Reduced recognition of emotion from whole-body motion was recently documented in healthy participants with elevated levels of traits associated with positive schizotypy syndrome (Blain et al., 2017). At the same time, no differences were found between patients with borderline personality disorder (BPD) and healthy controls in recognition of affective states from PLDs (Kaletsch et al., 2014a).

CONCLUSION

This review focused on the application of biological motion methodology to the study of emotion or intention inference in patients with psychiatric disorders. Two main conclusions may be drawn from the current review. First, the social cognitive problems found in most of the psychiatric samples using PLDs were of smaller magnitude than that found for other methods of social stimuli presentation (e.g., face, voice; SCZ: Okruszek and Pilecka, 2017; BD: Vaskinn et al., 2017; AN/BN: Dapelo et al., 2017; BPD: Kaletsch et al., 2014a). It has been suggested that the contribution of body motion to processing information about a person may be particularly important when viewing conditions are suboptimal or the person is viewed at a

distance (Yovel and O'Toole, 2016). Correct identification of a person's affective state or intention prior to a close proximity encounter may be crucial for one's survival, thus the extraction of such information from biological motion may be one of our most basic and evolutionarily oldest social cognitive abilities. Furthermore, given the extensive neural networks that mediate processing of the human face (Haxby et al., 2000), recognition of emotion or intention from biological motion may be less affected by abnormal brain functioning in patients, compared with the processing of social information coming from other modalities. Direct support for this suggestion comes from the neuropsychological observations of body-face dissociation in emotion recognition in patients with limbic lesions, who were shown to be able to correctly recognize whole-body expressions, even despite alterations in facial affect processing (Sprengelmeyer et al., 2010; Atkinson et al., 2012). Additionally, while decreased facial emotion recognition was observed in both MCI and AD, decreased emotion processing from PLDs is observed only in patients with fully developed AD (Henry et al., 2012). Additionally, even though numerous studies documented decreased intention attribution in psychiatric patients (Fett et al., 2015), intact recognition of communicative interactions from both single (Jaywant et al., 2016b) and dyadic (Okruszek et al., 2015) PLDs was found in patients. Furthermore, intact interpersonal predictive coding was observed in SCZ with paradigms presenting dyadic PLDs (Okruszek et al., 2017a, 2018). Thus, studies that aim to examine the mechanisms associated with the processing of social information in psychiatric disorders may benefit from combining standard methodologies (e.g., recognition of emotion or intention from static displays of faces) and dynamic PLD-based tasks.

The second main conclusion of the current review is the fact that specific social cognitive biases that have previously been observed using other methods (e.g., mood-congruent bias in MDD, Loi et al., 2013; Kaletsch et al., 2014b), increased threat perception in individuals with elevated anxiety (Heenan and Troje, 2015), aberrant body size perception in eating disorders (Vocks et al., 2007; Phillipou et al., 2016) can also be found in studies using PLDs. Thus, even though the information presented in PLDs is extremely limited, the stimuli are sufficient to elicit disorder-specific, social cognitive biases. Recognition of basic emotions conveyed by biological motion has been found to be relatively unaffected by cultural factors (Parkinson et al., 2017), thus PLDs may be effectively employed to study cross-cultural factors affecting social functioning in psychiatric populations (Mohan et al., 2016).

Taken together, these observations suggest that PLDs may be used as an additional source of information on social cognitive processes, especially when combined with other forms of social information presentation. One way to accomplish this may be by using multimodal stimuli that combine PLDs with auditory stimuli (Piwek et al., 2015). Furthermore, a wide variety of PLD tasks is readily available, some of which have already been shown to have satisfactory psychometric values (Kern et al., 2013). Finally, Shi et al. (2017) recently presented a Kinect-based method that allows one to produce PLDs without having access to a full motion-capture laboratory. In this way, point-light

stimuli can be tailored to the specific needs of a study using low-cost and user-friendly methods.

While the benefits of using PLDs have been listed above, some drawbacks of this approach should also be mentioned. First, PLD-based tasks may have limited test-retest reliability, thus may not be suitable for longitudinal assessments (Kern et al., 2013). Second, none of the abovementioned tasks has undergone a standardization procedure, which limits their usefulness for clinical practice. Finally, knowledge of the neural markers of biological motion processing abnormalities in psychiatric populations is severely limited, especially when compared with the extensive literature on facial affect processing in psychiatric disorders.

AUTHOR CONTRIBUTIONS

ŁO reviewed the literature and wrote the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Atkinson, A. P., Vuong, Q. C., and Smithson, H. E. (2012). Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. *Neuroimage* 59, 1700–1712. doi: 10.1016/j.neuroimage.2011.08.073

Bigelow, N. O., Paradiso, S., Adolphs, R., Moser, D. J., Arndt, S., Heberlein, A., et al. (2006). Perception of socially relevant stimuli in schizophrenia. *Schizophr. Res.* 83, 257–267. doi: 10.1016/j.schres.2005.12.856

Blain, S. D., Peterman, J. S., and Park, S. (2017). Subtle cues missed: impaired perception of emotion from gait in relation to schizotypy and autism spectrum traits. *Schizophr. Res.* 183, 157–160. doi: 10.1016/j.schres.2016.11.003

Bourke, C., Douglas, K., and Porter, R. (2010). Processing of facial emotion expression in major depression: a review. *Aust. N Z J. Psychiatry* 44, 681–696. doi: 10.3109/00048674.2010.496359

Brittain, P. J., Ffytche, D. H., and Surguladze, S. A. (2012). Emotion perception and functional outcome in schizophrenia: the importance of negative valence and fear. *Psychiatry Res.* 200, 208–213. doi: 10.1016/j.psychres.2012.07.027

Cacioppo, J. T., Cacioppo, S., Dulawa, S., and Palmer, A. A. (2014). Social neuroscience and its potential contribution to psychiatry. *World Psychiatry* 13, 131–139. doi: 10.1002/wps.20118

Couture, S. M., Penn, D. L., Losh, M., Adolphs, R., Hurley, R., and Piven, J. (2010). Comparison of social cognitive functioning in schizophrenia and high functioning autism: more convergence than divergence. *Psychol. Med.* 40, 569–579. doi: 10.1017/s003329170999078x

Dapelo, M. M., Surguladze, S., Morris, R., and Tchanturia, K. (2017). Emotion recognition in face and body motion in bulimia nervosa. *Eur. Eat. Disord. Rev.* 25, 595–600. doi: 10.1002/erv.2554

Dasgupta, S., Tyler, S. C., Wicks, J., Srinivasan, R., and Grossman, E. D. (2017). Network connectivity of the right STS in three social perception localizers. *J. Cogn. Neurosci.* 29, 221–234. doi: 10.1162/jocn_a_01054

Deen, B., Koldewyn, K., Kanwisher, N., and Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb. Cortex* 25, 4596–4609. doi: 10.1093/cercor/bhv111

Duchaine, B., and Yovel, G. (2015). A revised neural framework for face processing. *Annu. Rev. Vis. Sci.* 1, 393–416. doi: 10.1146/annurev-vision-082114-035518

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., et al. (2006). Introducing MASC: a movie for the assessment of social cognition. *J. Autism Dev. Disord.* 36, 623–636. doi: 10.1007/s10803-006-0107-0

Engelstad, K. N., Sundet, K. S., Andreassen, O. A., and Vaskinn, A. (2017). Body language reading of emotion in schizophrenia: associations with symptoms and functional outcome. *Scand. J. Psychol.* 58, 359–366. doi: 10.1111/sjop.12381

Fett, A. K. J., Shergill, S. S., and Krabbendam, L. (2015). Social neuroscience in psychiatry: unravelling the neural mechanisms of social dysfunction. *Psychol. Med.* 45, 1145–1165. doi: 10.1017/s0033291714002487

Fett, A.-K. J., Viechtbauer, W., Dominguez, M.-G., Penn, D. L., van Os, J., and Krabbendam, L. (2011). The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. *Neurosci. Biobehav. Rev.* 35, 573–588. doi: 10.1016/j.neubiorev.2010.07.001

Hastings, C. N. M., Brittain, P. J., and Ffytche, D. H. (2013). An asymmetry of translational biological motion perception in schizophrenia. *Front. Psychol.* 4:436. doi: 10.3389/fpsyg.2013.00436

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0

Heberlein, A. S., Adolphs, R., Tranel, D., and Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *J. Cogn. Neurosci.* 16, 1143–1158. doi: 10.1162/0898929041920423

Heenan, A., Best, M. W., Ouellette, S. J., Meiklejohn, E., Troje, N. F., and Bowie, C. R. (2014). Assessing threat responses towards the symptoms and diagnosis of schizophrenia using visual perceptual biases. *Schizophr. Res.* 159, 238–242. doi: 10.1016/j.schres.2014.07.024

Heenan, A., and Troje, N. F. (2014). Both physical exercise and progressive muscle relaxation reduce the facing-the-viewer bias in biological motion perception. *PLoS One* 9:e99902. doi: 10.1371/journal.pone.0099902

Heenan, A., and Troje, N. F. (2015). The relationship between social anxiety and the perception of depth-ambiguous biological motion stimuli is mediated by inhibitory ability. *Acta Psychol.* 157, 93–100. doi: 10.1016/j.actpsy.2015.02.012

Henry, J. D., Thompson, C., Rendell, P. G., Phillips, L. H., Carbert, J., Sachdev, P., et al. (2012). Perception of biological motion and emotion in mild cognitive impairment and dementia. *J. Int. Neuropsychol. Soc.* 18, 866–873. doi: 10.1017/S1355617712000665

Henry, J. D., Von Hippel, C., Ruffman, T., Perry, Y., and Rendell, P. G. (2010). Threat perception in schizophrenia-spectrum disorders. *J. Int. Neuropsychol. Soc.* 16, 805–812. doi: 10.1017/S1355617710000640

Hoekert, M., Kahn, R. S., Pijnenborg, M., and Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: review and meta-analysis. *Schizophr. Res.* 96, 135–145. doi: 10.1016/j.schres.2007.07.023

Ibáñez, A., García, A. M., Esteves, S., Yoris, A., Muñoz, E., Reynaldo, L., et al. (2016). Social neuroscience: undoing the schism between neurology and psychiatry. *Soc. Neurosci.* 13, 1–39. doi: 10.1080/17470919.2016.1245214

Insch, P. M., Slessor, G., Phillips, L. H., Atkinson, A., and Warrington, J. (2015). The impact of aging and Alzheimers disease on decoding emotion cues from bodily motion. *AIMS Neurosci.* 2, 139–152. doi: 10.3934/neuroscience.2015.3.139

Jahshan, C., Wynn, J. K., Mathis, K. I., and Green, M. F. (2015). The neurophysiology of biological motion perception in schizophrenia. *Brain Behav.* 5, 75–84. doi: 10.1002/brb3.303

Jaywant, A., Shiffrar, M., Roy, S., and Cronin-Golomb, A. (2016a). Impaired perception of biological motion in Parkinson's disease. *Neuropsychology* 30, 720–730. doi: 10.1037/neu0000276

Jaywant, A., Wasserman, V., Kemppainen, M., Neargarder, S., and Cronin-Golomb, A. (2016b). Perception of communicative and non-communicative motion-defined gestures in Parkinson's disease. *J. Int. Neuropsychol. Soc.* 22, 540–550. doi: 10.1017/s1355617716000114

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi: 10.3758/bf03212378

Kaletsch, M., Krüger, B., Pilgramm, S., Stark, R., Lis, S., Gallhofer, B., et al. (2014a). Borderline personality disorder is associated with lower

confidence in perception of emotional body movements. *Front. Psychol.* 5:1262. doi: 10.3389/fpsyg.2014.01262

Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., Stark, R., et al. (2014b). Major depressive disorder alters perception of emotional body movements. *Front. Psychiatry* 5:4. doi: 10.3389/fpsyt.2014.00004

Kern, R. S., Penn, D. L., Lee, J., Horan, W. P., Reise, S. P., Ochsner, K. N., et al. (2013). Adapting social neuroscience measures for schizophrenia clinical trials, Part 2: trolling the depths of psychometric properties. *Schizophr. Bull.* 39, 1201–1210. doi: 10.1093/schbul/sbt127

Kim, J., Doop, M. L., Blake, R., and Park, S. (2005). Impaired visual recognition of biological motion in schizophrenia. *Schizophr. Res.* 77, 299–307. doi: 10.1016/j.schres.2005.04.006

Kim, J., Norton, D., McBain, R., Ongur, D., and Chen, Y. (2013). Deficient biological motion perception in schizophrenia: results from a motion noise paradigm. *Front. Psychol.* 4:391. doi: 10.3389/fpsyg.2013.00391

Kim, J., Park, S., and Blake, R. (2011). Perception of biological motion in schizophrenia and healthy individuals: a behavioral and FMRI study. *PLoS One* 6:e19971. doi: 10.1371/journal.pone.0019971

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., and Moberg, P. J. (2010). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophr. Bull.* 36, 1009–1019. doi: 10.1093/schbul/sbn192

Lang, K., Dapelo, M. M., Khondoker, M., Morris, R., Surguladze, S., Treasure, J., et al. (2015). Exploring emotion recognition in adults and adolescents with anorexia nervosa using a body motion paradigm. *Eur. Eat. Disord. Rev.* 23, 262–268. doi: 10.1002/erv.2358

Loi, F., Vaidya, J. G., and Paradiso, S. (2013). Recognition of emotion from body language among patients with unipolar depression. *Psychiatry Res.* 209, 40–49. doi: 10.1016/j.psychres.2013.03.001

Lorey, B., Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., et al. (2012). Confidence in emotion perception in point-light displays varies with the ability to perceive own emotions. *PLoS One* 7:e42169. doi: 10.1371/journal.pone.0042169

Manera, V., Ianì, F., Bourgeois, J., Haman, M., Okruszek, Ł. P., Rivera, S. M., et al. (2015). The multilingual CID-5: a new tool to study the perception of communicative interactions in different languages. *Front. Psychol.* 6:1724. doi: 10.3389/fpsyg.2015.01724

Matsumoto, Y., Takahashi, H., Miyata, J., Sugihara, G., Murai, T., and Takahashi, H. (2017). Neural basis of altered earlier attention and higher order biological motion processing in schizophrenia. *Soc. Neurosci.* doi: 10.1080/17470919.2017.1366363 [Epub ahead of print].

Matsumoto, Y., Takahashi, H., Murai, T., and Takahashi, H. (2015). Visual processing and social cognition in schizophrenia: relationships among eye movements, biological motion perception and empathy. *Neurosci. Res.* 90, 95–100. doi: 10.1016/j.neures.2014.10.011

McDonald, S., Flanagan, S., Rollins, J., and Kinch, J. (2003). TASIT: a new clinical tool for assessing social perception after traumatic brain injury. *J. Head Trauma Rehabil.* 18, 219–238. doi: 10.1097/00001199-200305000-00001

Mehta, U. M., Thirthalli, J., Gangadhar, B. N., and Keshavan, M. S. (2011). Need for culture specific tools to assess social cognition in schizophrenia. *Schizophr. Res.* 133, 255–256. doi: 10.1016/j.schres.2011.09.007

Mohan, S. N., Mukhtar, F., and Jobson, L. (2016). Protocol for a between-group experimental study examining cultural differences in emotion processing between Malay and Caucasian adults with and without major depressive disorder. *BMJ Open* 6:e012774. doi: 10.1136/bmjopen-2016-012774

Okruszek, Ł., and Pilecka, I. (2017). Biological motion processing in schizophrenia—Systematic review and meta-analysis. *Schizophr. Res.* 190, 3–10. doi: 10.1016/j.schres.2017.03.013

Okruszek, Ł., Haman, M., Kalinowski, K., Talarowska, M., Becchio, C., and Manera, V. (2015). Impaired recognition of communicative interactions from biological motion in schizophrenia. *PLoS One* 10:e0116793. doi: 10.1371/journal.pone.0116793

Okruszek, Ł., Piejka, A., Wysokiński, A., Szczepocka, E., and Manera, V. (2017a). The second agent effect: interpersonal predictive coding in people with schizophrenia. *Soc. Neurosci.* doi: 10.1080/17470919.2017.1415969 [Epub ahead of print].

Okruszek, Ł., Wordecha, M., Jarkiewicz, M., Kossowski, B., Lee, J., and Marchewka, A. (2017b). Brain correlates of recognition of communicative

interactions from biological motion in schizophrenia. *Psychol. Med.* doi: 10.1017/s0033291717003385 [Epub ahead of print].

Okruszek, Ł., Piejka, A., Wysokiński, A., Szczepocka, E., and Manera, V. (2018). Biological motion sensitivity, but not interpersonal predictive coding is impaired in schizophrenia. *J. Abn. Psychol.* doi: 10.1037/abn0000335 [Epub ahead of print].

Olbert, C. M., Penn, D. L., Kern, R. S., Lee, J., Horan, W. P., Reise, S. P., et al. (2013). Adapting social neuroscience measures for schizophrenia clinical trials, Part 3: fathoming external validity. *Schizophr. Bull.* 39, 1211–1218. doi: 10.1093/schbul/sbt130

Parkinson, C., Walker, T. T., Memmi, S., and Wheatley, T. (2017). Emotions are understood from biological motion across remote cultures. *Emotion* 17, 459–477. doi: 10.1037/emo0000194

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cereb. Cortex* 22, 981–995. doi: 10.1093/cercor/bhr156

Pavlova, M. A. (2017). Sex and gender affect the social brain: beyond simplicity. *J. Neurosci. Res.* 95, 235–250. doi: 10.1002/jnr.23871

Péron, J., El Tamer, S., Grandjean, D., Leray, E., Travers, D., Drapier, D., et al. (2011). Major depressive disorder skews the recognition of emotional prosody. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 987–996. doi: 10.1016/j.pnpbp.2011.01.019

Phillipou, A., Rossell, S. L., Gurvich, C., Castle, D. J., Troje, N. F., and Abel, L. A. (2016). Body image in anorexia nervosa: body size estimation utilising a biological motion task and eyetracking. *Eur. Eat. Disord. Rev.* 24, 131–138. doi: 10.1002/erv.2423

Pica, P., Jackson, S., Blake, R., and Troje, N. F. (2011). Comparing biological motion perception in two distinct human societies. *PLoS One* 6:e28391. doi: 10.1371/journal.pone.0028391

Piwek, L., Pollick, F., and Petrini, K. (2015). Audiovisual integration of emotional signals from others' social interactions. *Front. Psychol.* 9:116. doi: 10.3389/fpsyg.2015.00611

Plana, I., Lavoie, M.-A., Battaglia, M., and Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders. *J. Anxiety Disord.* 28, 169–177. doi: 10.1016/j.janxdis.2013.09.005

Samamé, C., Martino, D. J., and Strejilevich, S. A. (2012). Social cognition in euthymic bipolar disorder: systematic review and meta-analytic approach. *Acta Psychiatr. Scand.* 125, 266–280. doi: 10.1111/j.1600-0447.2011.01808.x

Savla, G. N., Vella, L., Armstrong, C. C., Penn, D. L., and Twamley, E. W. (2013). Deficits in domains of social cognition in schizophrenia: a meta-analysis of the empirical evidence. *Schizophr. Bull.* 39, 979–992. doi: 10.1093/schbul/sbs080

Shi, Y., Ma, X., Ma, Z., Wang, J., Yao, N., Gu, Q., et al. (2017). Using a Kinect sensor to acquire biological motion: toolbox and evaluation. *Behav. Res. Methods* doi: 10.3758/s13428-017-0883-9 [Epub ahead of print].

Spencer, J. M. Y., Sekuler, A. B., Bennett, P. J., and Christensen, B. K. (2013). Contribution of coherent motion to the perception of biological motion among persons with Schizophrenia. *Front. Psychol.* 4:507. doi: 10.3389/fpsyg.2013.00507

Sprengelmeyer, R., Atkinson, A. P., Sprengelmeyer, A., Mair-Walther, J., Jacobi, C., Wildemann, B., et al. (2010). Disgust and fear recognition in paraneoplastic limbic encephalitis. *Cortex* 46, 650–657. doi: 10.1016/j.cortex.2009.04.007

Troje, N. F. (2013). "What is biological motion? definition, stimuli and paradigms," in *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*, eds M. D. Rutherford and V. A. Kuhlmeier, (Cambridge, MA: MIT Press), 13–36.

Van de Cruys, S., Schouten, B., and Wagemans, J. (2013). An anxiety-induced bias in the perception of a bistable point-light walker. *Acta Psychol.* 144, 548–553. doi: 10.1016/j.actpsy.2013.09.010

Van Overwalle, F., and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage* 48, 564–584. doi: 10.1016/j.neuroimage.2009.06.009

Vanrie, J., Dekeyser, M., and Verfaillie, K. (2004). Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception* 33, 547–560. doi: 10.1068/p5004

Vaskinn, A., Lagerberg, T. V., Bjella, T. D., Simonsen, C., Andreassen, O. A., Ueland, T., et al. (2017). Impairment in emotion perception from body

movements in individuals with bipolar I and bipolar II disorder is associated with functional capacity. *Int. J. Bipolar Disord.* 5:13. doi: 10.1186/s40345-017-0083-7

Vaskinn, A., Sundet, K., Østefjells, T., Nymo, K., Melle, I., and Ueland, T. (2016). Reading emotions from body movement: a generalized impairment in schizophrenia. *Front. Psychol.* 6:2058. doi: 10.3389/fpsyg.2015.02058

Vocks, S., Legenbauer, T., Rüddel, H., and Troje, N. F. (2007). Static and dynamic body image in bulimia nervosa: mental representation of body dimensions and biological motion patterns. *Int. J. Eat. Disord.* 40, 59–66. doi: 10.1002/eat.20336

Weightman, M. J., Air, T. M., and Baune, B. T. (2014). A review of the role of social cognition in major depressive disorder. *Front. Psychiatry* 5:179. doi: 10.3389/fpsyt.2014.00179

Yovel, G., and O'Toole, A. J. (2016). Recognizing people in motion. *Trends Cogn. Sci.* 20, 383–395. doi: 10.1016/j.tics.2016.02.005

Zaini, H., Fawcett, J. M., White, N. C., and Newman, A. J. (2013). Communicative and noncommunicative point-light actions featuring high-resolution representation of the hands and fingers. *Behav. Res. Methods* 45, 319–328. doi: 10.3758/s13428-012-0273-2

Zucker, N., Moskovich, A., Bulik, C. M., Merwin, R., Gaddis, K., Losh, M., et al. (2013). Perception of affect in biological motion cues in anorexia nervosa. *Int. J. Eat. Disord.* 46, 12–22. doi: 10.1002/eat.22062

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Check for
updates

Children Facial Expression Production: Influence of Age, Gender, Emotion Subtype, Elicitation Condition and Culture

Charline Grossard[1,2]*, Laurence Chaby[2,3], Stéphanie Hun[4], Hugues Pellerin[1],
Jérémy Bourgeois[4], Arnaud Dapogny[2], Huaxiong Ding[5], Sylvie Serret[4], Pierre Foulon[6],
Mohamed Chetouani[2], Liming Chen[5], Kevin Bailly[2], Ouriel Grynszpan[2] and
David Cohen[1,2]*

[1] Service de Psychiatrie de l'Enfant et de l'Adolescent, GHU Pitie-Salpetriere Charles Foix, Assistance Publique – Hôpitaux de Paris, Paris, France, [2] Institut des Systèmes Intelligents et de Robotique (ISIR), CNRS UMR 7222, Sorbonne Université, Paris, France, [3] Institut de Psychologie, Université Paris Descartes, Sorbonne Paris Cité University, Paris, France, [4] Cognition Behaviour Technology (CoBTeK), EA 7276, University of Nice Sophia Antipolis, Nice, France, [5] Laboratoire d'Informatique en Image et Systèmes d'Information (LIRIS), Ecole Centrale de Lyon, CNRS, UMR 5205, 69134, Villeurbanne, France, [6] Groupe Genious Healthcare, Montpellier, France

The production of facial expressions (FEs) is an important skill that allows children to share and adapt emotions with their relatives and peers during social interactions. These skills are impaired in children with Autism Spectrum Disorder. However, the way in which typical children develop and master their production of FEs has still not been clearly assessed. This study aimed to explore factors that could influence the production of FEs in childhood such as age, gender, emotion subtype (sadness, anger, joy, and neutral), elicitation task (on request, imitation), area of recruitment (French Riviera and Parisian) and emotion multimodality. A total of one hundred fifty-seven children aged 6–11 years were enrolled in Nice and Paris, France. We asked them to produce FEs in two different tasks: imitation with an avatar model and production on request without a model. Results from a multivariate analysis revealed that: (1) children performed better with age. (2) Positive emotions were easier to produce than negative emotions. (3) Children produced better FE on request (as opposed to imitation); and (4) Riviera children performed better than Parisian children suggesting regional influences on emotion production. We conclude that facial emotion production is a complex developmental process influenced by several factors that needs to be acknowledged in future research.

Keywords: emotion, production, facial expression, development, children

INTRODUCTION

From an early age and throughout one's lifespan, emotional skills are essential to communicate our emotions to others and to modulate and adapt our behavior according to both our internal feelings and the reaction of others (Saarni, 1999; Halberstadt et al., 2001). The ability to understand what we feel, to deal with our own emotion and that of others, and to show emotional empathy are factors of

integration in the society at all ages of life. Although our experience of the world is multimodal (we see objects, hear sounds, feel texture, smell odors, and taste flavors), visual signals and languages are key social signals in humans (Adolphs, 2003). Among visual signals, facial expressions (FE) are crucial components of emotional signals. They allow people to understand and express not only emotions (Izard, 1971; Izard, 2001) but also social motivation (Fridlund, 1997).

Facial expressions recognition has been investigated in numerous studies, showing that many variables can influence the interpretation of FEs: (i) FE recognition increases during childhood with the age of the perceiver (Herba et al., 2006; Lawrence et al., 2015) and declines for older adults compared to young adults (see Ruffman et al., 2008). (ii) Modality influences emotion recognition, and multimodal supports are easier to recognize than unimodal supports (Castellano et al., 2008; Luherne-du Boullay et al., 2014). (iii) The condition of presentation from static or dynamic support is also important (Biele and Grabowska, 2006; Trautmann et al., 2009). (iv) FE are more easily recognized when the producer is younger rather than older (Fölster et al., 2014). (v) Girls are generally more efficient in identifying emotion (Hall et al., 2000; Lawrence et al., 2015) but not all studies support this conclusion (Herba et al., 2006). Some differences in methodology could explain these differences, as the choice of the intensity of the expressions (Hoffmann et al., 2010). (vi) Emotion recognition is higher when emotions were both recognized and expressed by members of the same regional group (Elfenbein and Ambady, 2002). Moreover, majority group members are poorer at judging minority members than the reverse. (vii) The context in which FE is produced can also contribute to emotion recognition (Wallbott, 1988; Mobbs et al., 2006). (viii) The different emotional FEs themselves are not equally identified: joy appears to be one of the easiest FE to be recognized (Lawrence et al., 2015).

Facial expressions production has received less attention than FE recognition in the literature. There are mainly three methods to evaluate FE production. The first is the measure approach which describes and measures objectively observable and measurable changes of facial components. The most widely used method is the Facial Action Coding System (FACS, Ekman et al., 2002) which requires a trained expert to rate. The second and the most commonly used in the establishment of a dataset is the judgment approach introduced by Darwin (1872) which is based on the fact that everyone can relate a FE to an emotion. This method consists of presenting FE to a sample of judges, and the accuracy of the FE is inferred thanks to their rating. In most previous studies (Egger et al., 2011; Dalrymple et al., 2013), researchers recorded individuals when they produced a FE. Then, blind annotators had to rate the video in two steps: first, they had to first identify which emotion was produced and then had to rate its intensity. Few studies try to rate the quality of the emotion, and the way to do it is not consensual. In studies of children, Egger et al. (2011) asked the judges how well the emotion was portrayed. Mazurski and Bond (1993) looked at the certainty of the judge that the emotion he recognized was the good one. In studies of adults, such as the GEMEP (Bänziger et al., 2012), the judges had to rate the authenticity and the plausibility of the FE. The third method to assess FE is based on algorithmic automatic assessments trained on large datasets that provide a normed FE material (Zeng et al., 2009). However, this method requires the algorithm to be previously trained on a dataset already rated by human judges.

To date, most of the datasets describing a large dataset of FE concern adult FE. In the most recent studies, the datasets propose both static and dynamic sequences with different face orientations (Pantic et al., 2005), multimodal production (Bänziger et al., 2012) as well as played (e.g., professional actors) or natural facial productions (Zhang et al., 2014). But very few datasets concern FE of children (see **Table 1**). Moreover, most of them include only static 2D supports (mainly photographs). The Facewarehouse dataset is the only one made of 3D video recordings of FE, but it does not include just children nor does it indicate how many children are involved (Cao et al., 2014).

Most studies regarding FE production were conducted in adulthood. Ekman et al. (1987) defined six emotions as universal

TABLE 1 | Databases that include children facial expressions.

Databases	Population	Emotions	Support
NIMH-ChEFS (Egger et al., 2011)	39 girls and 20 boys from 10 to 17 years old	Fear, anger, happiness, sadness, and neutral	482 photographs
Dartmouth database (Dalrymple et al., 2013)	40 caucasian girls and 40 caucasian boys from 6 to 16 years old	Neutral, satisfaction, happiness, sadness, anger, fear, surprise, and disgust	Photographs
Facewarehouse (Cao et al., 2014)	150 people from 7 to 80 years old (proportion of children unknown)	Mouth stretch, smile, brow lower, brow raiser, anger, jaw left, jaw right, jaw forward, mouth left, mouth right, dimpler, chin raiser, lip puckerer, lip funneler, sadness, lip roll, grin, cheek blowing, and eyes closed	3D Vidéos
Japanese database (Komatsu and Hakoda, 2012)	53 boys et 54 girls from 11 to 13 years old	Neutral, happiness, surprise, anger, and sadness	535 photographs
Slides depicting facial expression of affect (Mazurski and Bond, 1993)	3 boys (9 to 11 years old) et 3 girls (8 to 12 years old)	Anger, disgust, fear, happiness, surprise, and neutral	Photographs
CAFE (LoBue and Thrasher, 2014)	90 girls and 64 boys from racially and ethnically diverse group between 2 and 8 years old	Anger, fear, sadness, happiness, neutral, surprise, and disgust	1192 photographs

(sadness, happiness, anger, surprise, fear, disgust, also combined with contempt), common among all humans, independently of culture or origin. Nowadays, this theory is questioned. If it is generally accepted that these six emotions are innate for a part, new studies show that culture can modulate FE production (Elfenbein et al., 2007). Moreover, other factors influence FE production. Women are described as more expressive than men (Brody and Hall, 2000). They tend to produce more positive emotions while males express more anger. FE production is also influenced by the context around the producer. FE of a participant is better recognized if he produces it in presence of a friend than in presence of a stranger (Wagner and Smith, 1991). People produce more easily FE of happiness in pleasant situations with people but tend to hide negative FE in unpleasant situations with people around them (Lee and Wagner, 2002).

In terms of development, it appears that most of the facial components of human expression can be observed shortly after birth like expression of enjoyment and interest that are present from the opening days of life (Sullivan and Lewis, 2003). Researcher first thought that infant FEs corresponded to adults FEs (see Differential emotion theory in Izard and Malatesta, 1987), but it's now known that FEs in infancy are not present like their adult-counterparts (Oster, 2005). The first reason is that emotion in infancy cannot be compared to emotion in adulthood. Sroufe (1996) described precursor emotions in infancy which do not involve some degree of cognitive evaluation like for emotions in adults. He described wariness and frustration that are similarly manifested in crying and distress. This observation concurs with the study of Camras et al. (2007) that do not find different FEs for fear and anger at 11 months. Another reason of differences between adult and infant FEs could be linked to the motor structure of infant face. Camras et al. (1996) noted that infants may produce FEs in a non-related situation because of an enlarged recruitment among facial muscles during movement. For example, infants of 5 and 7 months raise their brows as they open their mouth, producing an expression of surprise.

Holodynski and Friedlmeier (2006) proposed that infants learned adult-like expressions thanks to a sociocultural based internalization model; caregivers reproduced infant expressions in a selective and exaggerated form, allowing children to learn the concordance between their emotion and a given FE.

However, the apparition of adult-like expressions is not well known (Oster, 2005). Bennett et al. (2005) showed that the organization of facial expressivity increases during infancy. 12-month infant showed more specific expression to a situation than 4-month infants. In response to tickle, the number of infants exhibiting joy expression increased and the number exhibiting other expressions (like surprise or interest) decreased. It seems that children continue to learn how to produce FE even in late childhood. Ekman et al. (1980) showed that the ability to produce FE improves between 5 and 13 years. However, they do not perfectly produce all FE. In the same way, Gosselin et al. (2011) showed that children between 5 and 9 years old activated unexpected action components when they were asking to produce sadness and joy.

The subtype of emotion can also influence productions of children. Brun (2001) studied the FE in children between 3 and 6 years old. The children had to evoke the FE from a sound link to an emotion. The production of FE depends on age and the targeted emotion: joy is already well produced at 3 years old while anger, sadness and surprise are still not mastered at 6 years old. Field and Walden (1982) also found that positive emotions are easier to produce than negative emotions. However, LoBue and Thrasher (2014) asked children to imitate FE of an adult and found no effects of age or emotion subtype on the production of FE for children between 2 and 8 years old.

Most studies assessed the effect of gender on emotion production with girls that produce more positive FE and boys more negative FE. During adolescence, gender differences have been reported with (i) judges rating girls' positive expressions stronger than boys' productions, and boys' expressions of anger, sadness, and surprise stronger than girls' expressions (Komatsu and Hakoda, 2012); and (ii) with girls smiling more often than boys (LaFrance et al., 2003). However, LoBue and Thrasher (2014) found no effect of gender on FR production for children between 2 and 8 years old. Effectively, the effect of gender seems to be modulating by other factors. Chaplin and Aldao's (2013) meta-analytic review confirmed the interaction between gender, age and type of emotion during FE. They found no gender difference in infancy and preschoolers. However, they found that children and adolescent girls express more positive emotion than boys. Conversely, a small effect of gender appears in infancy, preschoolers and childhood but disappears in adolescence for the production of internalizing emotions (such as sadness or sympathy) with more accuracy for girls. For externalizing emotions (like anger), they found no difference in infancy. But boys were better than girls in production during childhood. Unexpectedly, the differences reverse in adolescence with better productions of externalizing emotions for girls than for boys.

As in adults, ethnicity and culture seems to influence FE production. Comparing four groups of 3 year old girls (European–American, Chinese girls adopted in a European–American family, non-adopted Chinese–American girls and Chinese girls living in mainland China), Camras et al. (2006) found that European–American girls were more expressive than Chinese–American girls and mainland Chinese girls. Adopted Chinese girls generally fell between the European–American group and the 2 other Chinese groups. They differed significantly from the 2 other Chinese groups for disgust. The influence of ethnicity is also shown by Louie et al. (2013). They found that preschooler of Asian American parents and from Korean parents tend to be less expressive than preschoolers from European American family for sadness and exuberance. These findings showed that ethnicity can influence the production of emotion but also that culturally based family environment modulates the effect of ethnicity. Moreover, this effect seems to appear in the 1st year of life (Camras et al., 2007; Muzard et al., 2017).

So far, very few studies have proposed to study spontaneous production of FE (e.g., Sato and Yoshikawa, 2007). Most of the time, the targeted population produces FE on request (e.g., Egger et al., 2011; Dalrymple et al., 2013). However, FE can be produced while imitating a model (e.g., a picture, a drawing, a video of a virtual agent or another human like in LoBue and Thrasher, 2014). In the current paper, we will call this type of tasks

"imitation" as opposed to FE production "on request" (e.g., an oral or writing order, or pictures or oral contexts without model).

Also, few research targeted FE in children. They supposed that many variables could influence children's productions as gender, culture, emotion subtype, but data are missing to understand the effects of these variables through age. Open questions remain regarding typical child performances in producing FE between 6 and 11 years old. Moreover, the influence of the type of tasks and the modality in which they are presented are not well documented. The first aim of our work is to explore the quality of the FEs of children between 6 and 11 years old. We tested the capacities of typical children to produce FE on demand and the several moderating variables such as age, gender, type of emotion, condition of production (visual vs. bimodal), context of elicitation (imitation vs. acting on request) and region (Parisian vs. French Riviera) that could influence their productions. We hypothesized performance to increase with age, girls to perform better than boys, positive emotions to be easier to produce than negative emotions, bimodal presentation to make FE easier to produce than visual unimodal presentation, imitation to make FE easier to produce than acting on request, and Mediterranean children to perform better than Parisian children.

The current work enters into the larger project, JEMImE, intended to improve FE of children with ASD. Children with ASD have difficulties to identify and produce adapted FE (Uljarevic and Hamilton, 2013; Gordon et al., 2014). The JEMImE project aims to create a serious game to stimulate children with ASD to produce adapted FE in context. To reach this goal the game inspired by JeStimule, that aims to train emotion recognition in children with ASD (Serret et al., 2014), will automatically score online children's FE production to help the child (or the therapist) to monitor his production. In order to provide this feedback an algorithm that is able to recognize in real time the production of the player will be integrated into the game. To deal with the lack of extended datasets with children producing FE, we had to record a large dataset. The second aim of our work is so to capture and rate a large dataset of children's FEs in order to train the algorithm (Grossard et al., 2017).

MATERIALS AND METHODS

Participants

Children were recruited in two French public schools, one in Paris, one in Nice, from January 2015 to January 2016. The two schools were not located in areas known to be recruiting a high rate of children with socio-economic or developmental risk[1]. We only recruited native French children. In total, 157 children aged between 6 and 11 years old (boys, $N = 52\%$; girls $N = 48\%$) were enrolled in the study. Origins were varied but we included more Caucasian children (77.1%), and fewer African children (8.3%), Asian children (7%) and Maghreb children (7%). The percentage of Caucasian children was higher in Nice (89.7%) than in Paris (58.7%). Before inclusion, written consents were obtained after proper information from school directors, parents

---

[1]http://www.education.gouv.fr/cid187/l-education-prioritaire.html

and children. Each child was met alone during approximately 40 min to complete the protocol. The study was approved by the ethical committee of Nice University (*Comité de Protection des Personnes Sud Méditerranée*) under the number 15-HPNCL-02.

Tasks

Two tasks (demands of FE production on request and by imitation) were proposed. The two tasks were chosen in order to collect productions with and without a model (here an avatar) and thus to compare facial production in the two different tasks. Children had to produce four FEs: joy, anger, sadness, and neutral.

In the imitation task, the child must imitate the facial productions (visual modality) and the facial and vocal productions (audiovisual modality) of an avatar presented on his screen in short videos of 3–4 s. Two avatars (1 boy/1 girl) were created for this tool in order to counteract a possible gender effect of the model on FE recognition. These avatars were first tested with 20 adults who had to recognize the emotion produced and reach a recognition rate above 80%. Each of the avatars produced the four emotions. The avatars and the FEs were presented in a random order. The audiovisual condition combines FEs with emotional noises (such as crying for sadness, rage for anger or pleasure for joy, a/a/ held for neutral emotion). These sounds were extracted from an audio dataset validated in adults (Belin et al., 2008).

In the production on request, the child had to produce a FE (visual modality) or a facial and vocal expression (audiovisual modality) on request. The name of the emotion was displayed on the computer screen and read by the clinician. The order of presentation of emotions within this task was also random.

Design and Recording

Each child produced each emotion twice on request and four times in imitation (**Figure 1**). We doubled the imitation condition in order to have enough trials with avatars of both genders. The two tasks were first proposed in visual condition alone, then in audiovisual condition (facial and vocal). For each modality, they were proposed in a random order to avoid a learning effect (**Figures 1A,B**) and the modality presentation (visual modality vs. audiovisual modality) was counterbalanced. Each of this order was balanced according to gender and age (**Table 2**).

Each child was video recorded for 2–3 s using a 2D/3D video camera. Each video contained one FE. During the recording children had their own screen and the examiner had another. The examiner was seated in front of them in order to avoid that children turn their head out of the screen (**Figure 1**).

Imitation Task Instruction

The following instructions were given:

- [visual modality]: "You will see an animated face on the screen. It will produce an emotion with his face, like joy for example. You'll have to do the same thing with your face."
- [audiovisual modality]: "You will see an animated face on the screen. It will produce an emotion with his face and his

FIGURE 1 | Design and recording of the FE tasks. **(A)** Installation during the recording; **(B)** children screen showing two avatars showing two different FE; **(C)** Examiner control screen. Written informed consent was obtained from the participant for the publication of this image.
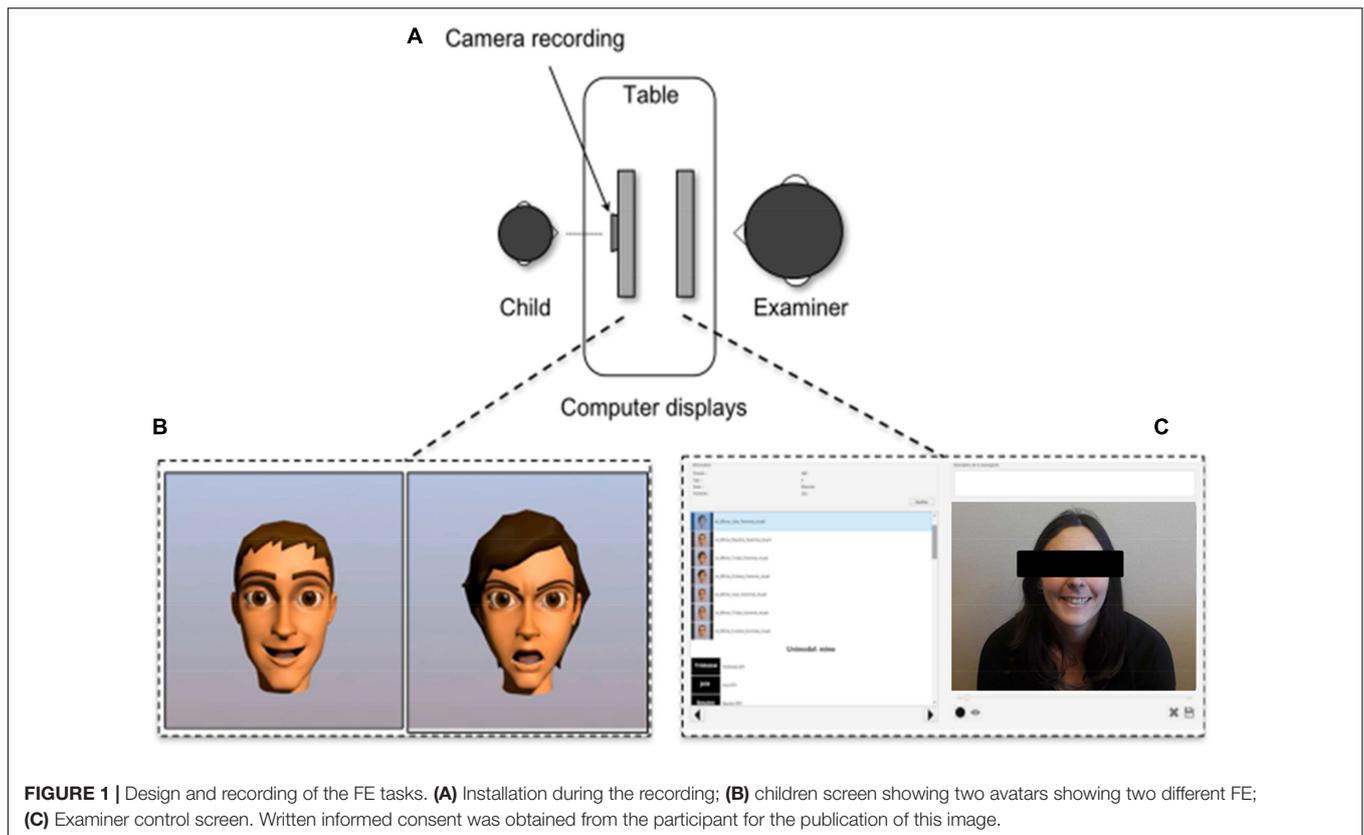
TABLE 2 | Repartition of children according to age, gender, site and order of presentation.

Age	Sex	6–7 years			7–8 years			8–9 years			9–10 years			10–11 years			Total		
Site		Nice	Paris	All	Nice	Paris	All	Nice	Paris	All	Nice	Paris	All	Nice	Paris	All	Nice	Paris	All
	Girls	5	1	6	7	1	8	3	1	4	2	1	3	2	1	3	19	5	24
Order 1	Boys	6	1	7	7	3	10	3	2	5	2	2	4	2	2	4	20	10	30
	Girls	3	1	4	1	2	3	2	1	3	3	2	5	2	1	3	11	7	18
Order 2	Boys	1	2	3	1	3	4	2	1	3	2	3	5	0	1	1	6	10	16
	Girls	3	0	3	0	1	1	2	2	4	3	1	4	3	1	4	11	5	16
Order 3	Boys	1	1	2	1	4	5	2	1	3	2	2	4	2	2	4	8	10	18
	Girls	3	1	4	1	1	2	2	2	4	2	2	4	3	1	4	11	7	18
Order 4	Boys	1	2	3	1	3	4	2	2	4	2	1	3	2	1	3	8	9	17
	Girls	14	3	17	9	5	14	9	6	15	10	6	16	10	4	14	52	24	76
	Boys	9	6	15	10	13	23	9	6	15	8	8	16	6	6	12	42	39	81
Total	Children	23	9	32	19	18	37	18	12	30	18	14	32	16	10	26	94	63	157

voice, like joy for example. You'll have to do the same thing with your face and your voice." We collected 16 videos per child.

On Request Task Instruction

The following instructions were given: "I will tell you a word which expresses an emotion when we feel something:

– [visual modality]: Could you show with your face what you do when you feel sadness/joy/anger/nothing?"

– "[audiovisual modality]: Could you show with your face and your voice what you do when you feel sadness/joy/anger/nothing?" We collected eight videos per child.

Coding

To analyze the productions of the children, all the videos recorded needed to be annotated. For our purpose we chose to keep a more naturalistic way of rating emotion. Indeed, the serious game JEMImE is aimed at teaching children with ASD how to produce adapted FE in the most natural way. We had

TABLE 3 | Emotion production as a function of age, gender, order, modality, elicitation task, emotion and sites: results from the GLMM model.

Variable	β estimate	Standard error	p
Age	0.131	0.04	0.001
Gender (boys vs. girls)	0.066	0.120	0.584
Order	−0.005	0.053	0.918
Modality (visual vs. audiovisual)	0.098	0.076	0.198
Elicitation task (on request vs. imitation)	0.536	0.083	<0.001
Emotion (happiness vs. sadness)	1.434	0.107	<0.001
Emotion (neutral vs. sadness)	1.684	0.111	<0.001
Emotion (anger vs. sadness)	0.909	0.100	<0.001
Site (Nice vs. Paris)	0.283	0.124	0.022

to look for how to judge the quality of an FE, which is not consensual in the literature. To construct our coding tools, we decided to consider the quality of an FE like a combination of recognizing and credibility. By postulating that if the emotion cannot be recognized it cannot be credible, it is possible to create a continuum between recognition and credibility. Indeed, we decided to create a scale from 0 to 10 where 0 corresponds to the absence of the expression, 5 to the recognition of the emotion but it does not seem credible and 10 to an emotion that is recognized and credible. Like the other tools, this scale allows to judge the presence of the emotion (0 = no recognition vs. 5 = recognition) and its quality (5 = recognition without credibility vs. 10 = recognizing and credible emotion). For each video, the judges had to complete four scales (one for each emotion: happiness, sadness, anger, and neutral). This method allows the judge to annotate one to four emotions for an expression. Indeed, a perfect production of happiness would be rated 10 in the scale for happiness and 0 on the three other scales. But for a less-specific expression (such as when children laugh while trying to produce anger), the judges would annotate multiple emotions for a unique expression (like anger 5 and joy 5). In terms of algorithmic purposes this may be of interest.

We asked three judges to annotate all the videos. The judges were French Caucasian adults (2 women and 1 man) aged 25, 34, and 40 years. They were all cognitive or developmental psychologists. The videos were blindly rated thanks to a special tool created for that purpose. In order to assess the reliability of the tool and the rating method, we asked two judges to independently annotate 10 children (240 videos in total). Children were chosen according to age, gender and presentation order of the tasks. Inter-agreement was assessed using intraclass correlation coefficients. We found excellent rates between the two judges for Happiness (ICC = 0.93), Anger (ICC = 0.92), Sadness (ICC = 0.93), and Neutral (ICC = 0.93).

Statistical Analysis

The data of the present study were analyzed using the statistical program R, version 3.3.1 (R Foundation for Statistical Computing), with two-tailed tests (see Supplementary Data Sheet S1). The variable to be explained was the FE rating score of the expected emotion. The distribution was not normal and

followed mainly a bimodal distribution with two peaks: the first peak was close to zero and the second close to 10 and only 23% of all coding scores were between 3 and 7. All attempts to transform FE rating score into a variable reaching normal distribution failed. Therefore, we transformed the FE rating score into a binary variable: failure for all scores < 5 and success for all score ≥ 5. We first explored whether each variable [gender, age, and emotion (joy, neutral, anger, or sadness), presentation order, sex of the avatar, presentation modality (visual vs. bimodal), elicitation task (imitation vs. on request), and sites (Paris vs. Nice)] was associated or not with FE rating score with bivariate analysis. Then we used a Generalized Linear Mixed Model (GLMM; lme4 and lmerTest packages) to explore the data. Given the number of observations, all variables were included in the multivariate model with the exception of the support, which was strongly dependant on the elicitation task. A binomial family was specified in the GLMM model to estimate the log-odds ratio for the corresponding factors in the model. Factors included could be gender (boy vs. girl), age, emotion (joy, neutral, anger, or sadness), presentation order, sex of the avatar, presentation modality (visual vs. bimodal), elicitation task (imitation vs. on request), and sites (Paris vs. Nice).

Finally, we also tested interactions between age, gender, and emotion as exploratory analysis given the previous results in the literature (see section "Introduction").

RESULTS

Emotion Production According to Age, Gender, and Tasks

Figures 2, **3** show mean rating scores of children emotion production according to age and gender for imitation (**Figure 2**) and on request tasks (**Figure 3**). Bivariate analyses showed that there was a significant effect for age with higher scores for older children (β = 0.131, standard error = 0.04, $p < 0.001$) but no effect of gender (β = 0.066, standard error = 0.120, $p = 0.584$). There was no significant effect for the order of presentation (β = −0.005, standard error = 0.053, $p = 0.918$), for the visual modality vs. the audiovisual modality (β = 0.098, standard error = 0.076, $p = 0.198$). However, we found several effects for elicitation task, with the on request elicitation showing higher rating scores than imitation (β = 0.53, standard error = 0.083, $p < 0.001$), for emotion with the best scores obtained with neutral, then happiness, then anger and finally sadness (neutral vs. sadness: β = 1.68, standard error = 0.111, $p < 0.001$; happiness vs. sadness: β = 1.43, standard error = 0.107, $p < 0.001$; anger vs. sadness: β = −0.909, standard error = 0.1, $p < 0.001$), and for sites with children from Nice showing higher scores than Parisian children (β = 0.28, standard error = 0.12, $p = 0.022$).

Multivariate Analysis

We kept in the GLMM the following explanatory variables: age, gender (boys vs. girls), order, modality (visual vs. audiovisual), emotion (joy, neutral, anger, or sadness), elicitation task (imitation vs. on request), and sites (Paris vs. Nice) (**Table 3**). The model formulation became: number of successes for the expected

FIGURE 2 | Mean emotion production scoring during the imitation task according to age and gender. Error bars are 95% bootstrapped confidence intervals.
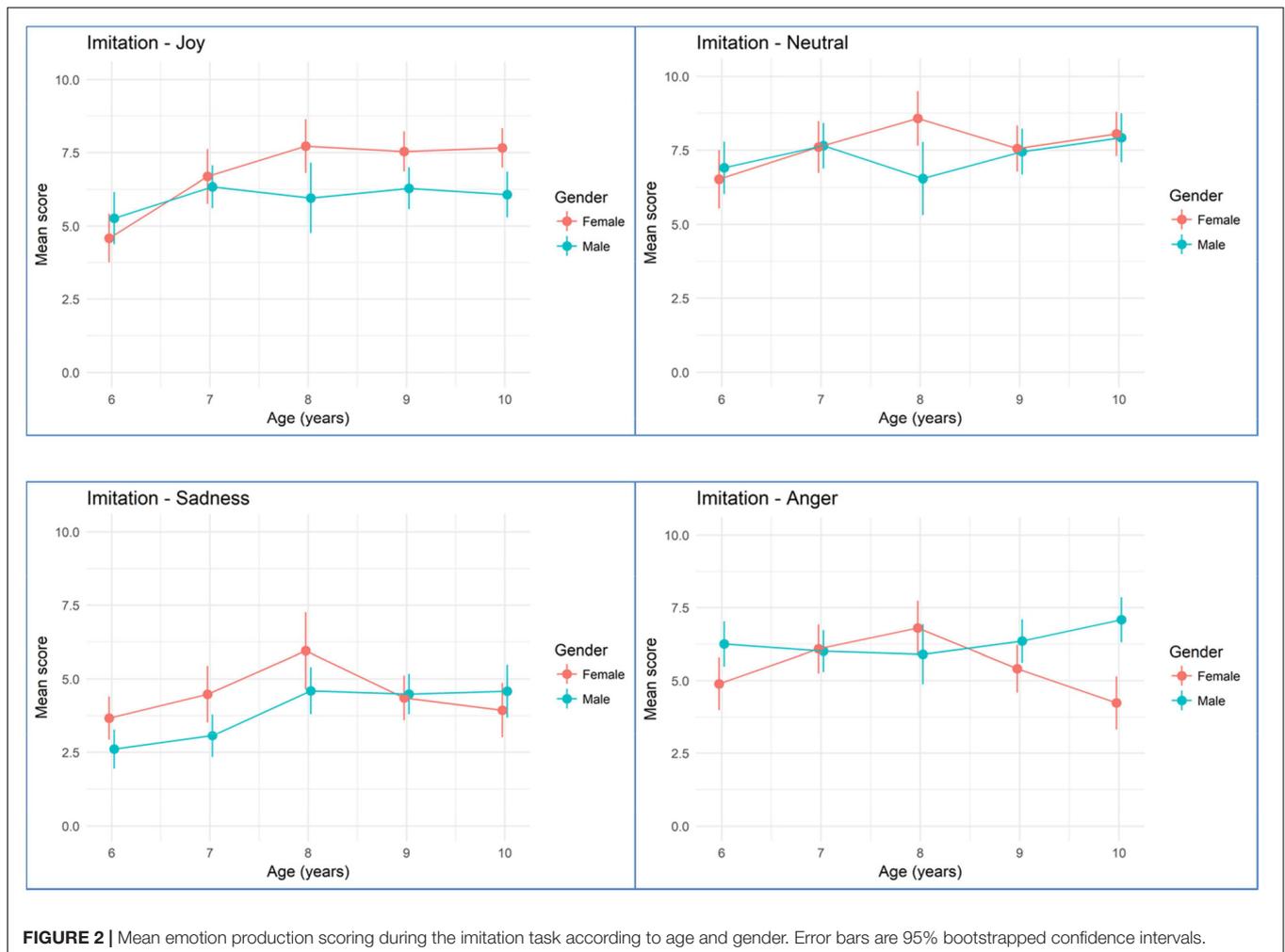
emotion ~ Age + Gender + Order + Modality + Elicitation task + Emotion + Sites + (1/child name). Emotion production significantly increased with age, was easier during the on request elicitation task (as opposed to the imitation elicitation task), was easier for positive emotion than negative emotions and within negative emotion easier for anger than sadness, and finally was easier for children from Nice than from Paris. Since the most difficult emotion to produce appeared to be sadness, we calculated the model adjusted odd ratios with sadness as the referential emotion. Emotion rating score significantly increased with a factor 1.14 when the child's age increases by 1 year. During on request elicitation task, emotion rating score significantly increased by a factor 1.71 compared to the imitation task. Emotion rating score significantly increased by a factor 5.39 for neutral, by a factor 4.20 for happiness, and by a factor 2.48 for anger compared to sadness. Finally, emotion rating score significantly increased by a factor 1.33 for Mediterranean participants compared to Parisian ones.

Finally, we tested interaction between age, gender, and emotion. Two way interactions were estimated from two models run separately. The model formulations became: number of

successes for the expected emotion ~ Elicitation task + Order + Modality + Age + Emotion*Gender + Sites + (1/child name); and number of successes for the expected emotion ~ Elicitation task + Order + Modality + Age*Gender + Emotion + Sites + (1/child name). Three way interactions were estimated from another model run separately. The model formulation became: number of successes for the expected emotion ~ Elicitation task + Order + Modality + Age*Emotion*Gender + Sites + (1/child name). Two and three way interactions are summarized in **Table 4** with sadness as the referential emotion. We did not find a significant interaction between age and gender. FE expression did not increase faster with age in boys or girls (adjusted odd ratio = 1.03). We found a significant interaction between anger (as opposed to sadness) and gender. Compared to the productions of anger for girls, emotion rating increased by a factor 1.68 for boys (adjusted odd ratio). Finally, we found two significant interactions between age and gender and emotion subtypes. For the production of joy (as opposed to sadness), we found a negative interaction with age and gender. The production decreased by a factor 0.56 for boys and age (adjusted odd ratio) meaning that age increases girls ability to produce joy compared to boys by a
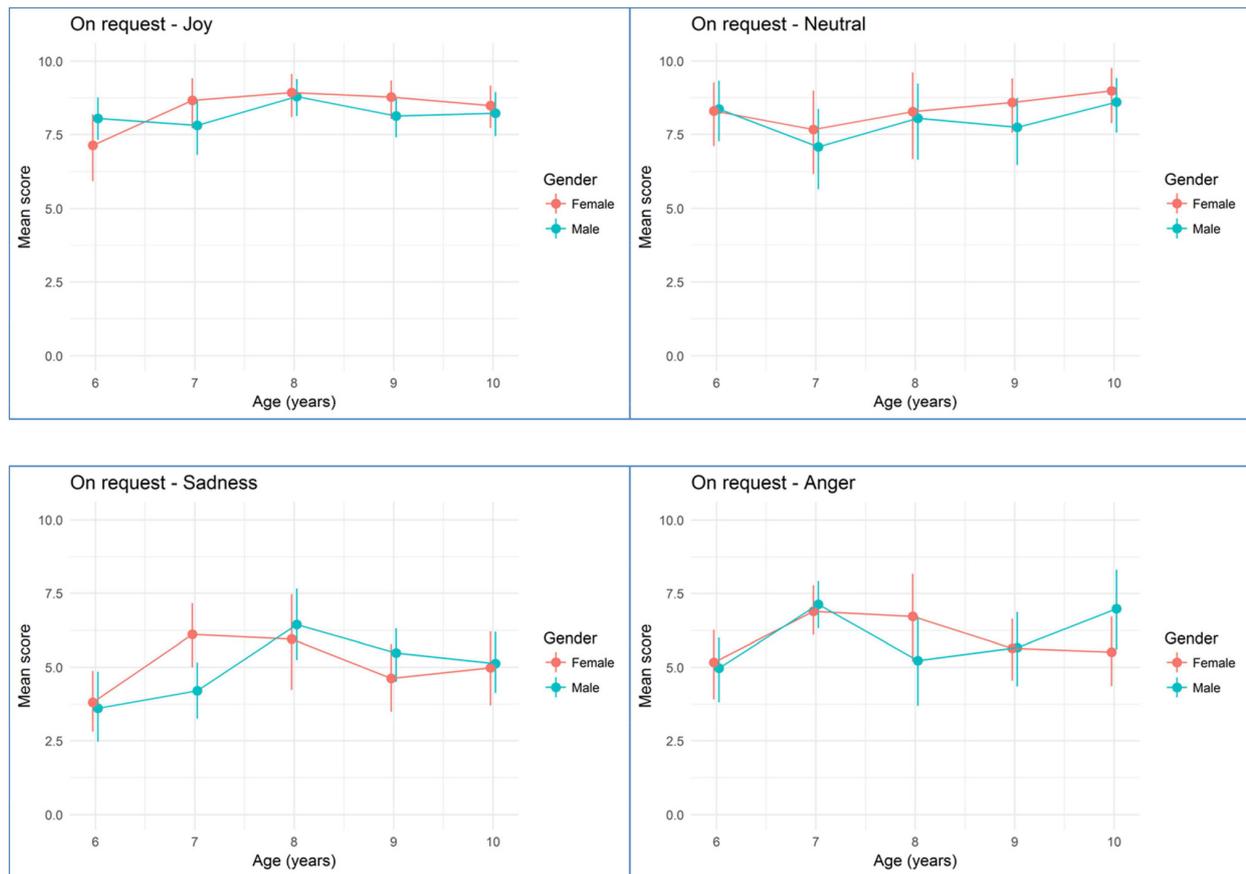
FIGURE 3 | Mean emotion production scoring during the on request task according to age and gender. Error bars are 95% bootstrapped confidence intervals.

factor 1.79 (1/0.56). Note that it doesn't mean that girls produce joy better than boys. A similar interaction was found between the production of neutral FE (as opposed to sadness) and age and gender. The production decreased by a factor 0.72 for boys and age.

DISCUSSION

The aim of this study was to evaluate the quality of the production of FE by children on demand, the development of this ability and some factors that could influence it. Recognition of FE is well documented and the six emotions described by Ekman et al. (2013) are well recognized between 6 and 11 years. However, few studies have analyzed the production of FE in childhood. This lack of data can be explained by the difficulty to implement a protocol adapted to children, to recruit a large population, to collect the data (especially video recordings which need specific material and installation) and to rate them appropriately. Thanks to our protocol, we recorded 3875 short videos of 157 children between 6 and 11 years of age producing FEs of joy, anger, sadness and neutral expressions and rated them in terms of recognition quality and credibility. This dataset will be used to train an algorithm to recognize in real time the FE of children when

TABLE 4 | Interaction model between age, gender and emotion with sadness as the referential emotion modality.

Variable	β estimate	Standard error	p
Model with 2-way interaction (age*gender)			
Gender (boys vs. girls) * Age	0.028	0.08	0.728
Model with 2-way interaction (emotion*gender)			
Emotion (joy) * Gender (boys vs. girls)	−0.141	0.212	0.505
Emotion (neutral) * Gender (boys vs. girls)	−0.013	0.221	0.954
Emotion (anger) * Gender (boys vs. girls)	0.516	0.199	0.010
Model with 3-way interaction (age*emotion*gender)			
Emotion (joy) * Gender (boys vs. girls) * Age	−0.584	0.149	<0.001
Emotion (neutral) * Gender (boys vs. girls)* Age	−0.325	0.151	0.031
Emotion (anger) * Gender (boys vs. girls) * Age	−0.158	0.137	0.247

playing with the serious game JEMImE computed to train FE and recognition in social contexts (Grossard et al., 2017). It will allow them to adjust their productions thanks to real time feedbacks.

As expected, the accuracy of FE emotional production increased with age. Whatever the other moderators, the FEs are best produced in older children. But it is important to note that children did not produce FE perfectly well, even for the oldest children (e.g., mean score at 10 years old is 6.5/10).

Other significant moderators of the quality of FE include the targeted emotion. For example, the score for the production of anger oscillate between 5 and 7.5 (for a maximum of 10), whatever the task. We expected that positive emotions would be easier to produce than negative emotions. Effectively, joy is produced with more accuracy than anger or sadness. Neutral emotion remains the state the most easily produced. However, in the on request task, joy is produced as well as neutral, even by young children (**Figure 3**). These findings concur with the observation of Brun (2001) demonstrating that joy is the emotion the most quickly mastered by children. Sadness is the emotion produced with less accuracy. These differences between positive and negative emotions may also come from the context of the signing. In adulthood, Lee and Wagner (2002) found that participants tend to hide their negative emotion when there are people around. In our protocol, some children tend to laugh when they had to produce negative emotion, because they appear embarrassed. Thereby, the important differences between positive and negative emotion in our study could be related to social rules already integrated in young children.

Based on previous studies, we expected that girls would produce positive FE with better quality than boys, and that boys would produce negative FE with better quality than girls (LaFrance et al., 2003; Komatsu and Hakoda, 2012; Chaplin and Aldao, 2013). We did find a significant interaction between gender and anger FE. Boys are better for producing anger than girls. Girls did not significantly produced joy with more quality than boys. However, we also found a significant interaction between age, gender and emotion subtype for joy, sadness, and neutral meaning that the differences between boys and girls may change according to age. Our results join the results of Chaplin and Aldao (2013) who also found a significant interaction between age, gender and emotion. We also looked at the effect of avatars gender on the productions of FE but found no significant effect. Boys and girls produced FE in a similar way, whatever the gender of the avatar. However, the quality of the children's production may depend of the quality of avatars. The fact that these avatars were previously rated by adults rather than children may bias the validity of the stimuli material when used on children.

We also expected that children would be helped by the bimodality. However, we found no effect of the modality on the productions of FE. Specifically, the presence of sound did not support the children's productions. In the bimodality, it appears that sometimes children can produce a correct sound the FE does not concur with the emotion targeted. In these cases, the annotator tends to pay more attention to the FE than the sound for two reasons: (i) FE are social signals that convey more strongly the information of the emotion felt than sound, (ii) the dataset was created to design an algorithm for automated facial recognition to be integrated in a serious game for ASD (Grossard et al., 2017). As a consequence, it is possible that raters considered that the most important information to rate was the facial signal. This tendency to pay more attention to FE than sound could modulate the effect of the modality.

We also expected an effect of the task on the children's productions. We proposed two different tasks, (i) one task of production with a model, the imitation task, (ii) one task of production without model, the on request task. We expected that children would perform better in imitation task because the model could help children in their productions. However, children significantly produced FE of better quality in the on request task than in imitation task. In fact, during the imitation task, children tried to stick as well as possible to the model. They did not need to understand the played emotion and tended to just analyze the placement of the elements on the avatar's face. Indeed, the productions were not always credible but also sometimes not well recognizable. In contrast, in the on request task, children had to themselves represent what the emotion triggers in order to produce the correct FE. This conscious control due to representation of the emotion requested to the child may be reparable because for somehow, they have a more important latency before starting their productions (subjective impression of raters but not objectively measured). Thereby, their productions tended to be closer to a real spontaneous expression, and also more credible.

The worse results in the imitation task could also come from our choice to use avatars instead of real persons to support the productions of the children. We choose avatars because of the interest of people with ASD for virtual environment (Boucenna et al., 2014). In a future work, we will propose our protocol to children with ASD and will compare their results to the results of typical developing children.

We also studied the effect of the site on the productions of the children's FEs. We found a significant effect between the two locations, in favor of children from Nice. This effect is subtle, as the size effect is not large. There are two ways to interpret this result. (i) The site effect is likely due to cultural factors as people in the south of France and the Mediterranean coast in general tend to be known as more expressive than those from Parisian. These findings concur with the literature that reports an effect of social environment on the production of FE (Camras et al., 2006). (ii) As the annotators were Caucasian and there were more Caucasian children recruited in Nice (89.7%) than in Paris (58.7%), judges might have been more accurate in recognizing FE on Caucasian children. These observations concur with the in-group advantage in emotion recognition (Elfenbein and Ambady, 2002).

Finally, the way to rate the productions of typical children was adapted to the requirements of the game as well as the design of the algorithm that will be implemented in the serious game. The choice of rating the credibility and the use of four scales at a time may have influenced the ratings. However, we obtained an excellent agreement between judges who rated the videos and our results are in accordance with the literature. Moreover, our coding procedure mixed recognition and credibility. Thinking of neutral emotion, what a credible neutral expression is may be odd to understand (e.g., no movement, only opening mouth). Since we are working on an algorithm that should recognize

emotional and neutral FE we had to keep the same scoring for all FE. However, this limitation is more theoretical than empirical, since we had very few ambiguous neutral FE (10% scores between 3 and 7) in the dataset.

CONCLUSION

In this study, we evaluated the effect of different moderators on the productions of FEs in children between 6 and 11 years old. We found that age, emotion, task and cultural environment modulate their productions. Also, production on request was easier than production imitating an avatar model. Taking into account these variables is necessary for the evaluation of competences of typical children but also comparison with a pathological population. In a future research, we plan to propose this protocol to children with ASD in order to characterize and compare their productions to those of typical children. We will also use the dataset to train classification algorithms for FE recognition in order to integrate it into the serious game JEMImE.

REFERENCES

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4, 165–178. doi: 10.1038/nrn1056

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827

Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* 40, 31–539. doi: 10.3758/BRM.40.2.531

Bennett, D. S., Bendersky, M., and Lewis, M. (2005). Does the organization of emotional expression change over time? Facial expressivity from 4 to 12 months. *Infancy* 8, 167–187. doi: 10.1207/s15327078in0802_4

Biele, C., and Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Exp. Brain Res.* 171, 1–6. doi: 10.1007/s00221-005-0254-0

Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., et al. (2014). Information communication technology (ICT) and autism: overview and focus on early developmental issues and social robotics. *Cogn. Comput.* 6, 722–740. doi: 10.1007/s12559-014-9276-x

Brody, L. R., and Hall, J. A. (2000). "Gender, emotion, and expression," in *Handbook of Emotions*, 2nd Edn, eds M. Lewis and J. M. Haviland-Jones (New York: Guilford Press), 338–349.

Brun, P. (2001). Psychopathologie de l'émotion chez l'enfant: l'importance des données développementales typiques. *Enfance* 53, 281–291. doi: 10.3917/enf.533.0281

Camras, L. A., Chen, Y., Bakeman, R., Norris, K., and Cain, T. R. (2006). Culture, ethnicity, and children's facial expressions: a study of European American, mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion* 6, 103–114. doi: 10.1037/1528-3542.6.1.103

Camras, L. A., Lambrecht, L., and Michel, G. F. (1996). Infant "surprise" expressions as coordinative motor structures. *J. Nonverbal Behav.* 20, 183–195. doi: 10.1007/BF02281955

Camras, L. A., Oster, H., Ujiie, T., Campos, J. J., Bakeman, R., and Meng, Z. (2007). Do infants show distinct negative facial expressions for fear and anger? Emotional expression in 11-month-old European American, Chinese, and Japanese infants. *Infancy* 11, 131–155. doi: 10.1111/j.1532-7078.2007.tb00219.x

Cao, C., Weng, Y., Zhou, S., Tong, Y., and Zhou, K. (2014). Facewarehouse: a 3d facial expression database for visual computing. *IEEE Trans. Vis. Comput. Graph.* 20, 413–425. doi: 10.1109/TVCG.2013.249

Castellano, G., Kessous, L., and Caridakis, G. (2008). "Emotion recognition through multiple modalities: face, body gesture, speech," in *Affect and Emotion in Human-Computer Interaction. Lecture Notes in Computer Science*, Vol. 4868, eds C. Peter and R. Beale (Berlin: Springer), 92–103.

Chaplin, T. M., and Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychol. Bull.* 139, 735–765. doi: 10.1037/a0030737

Dalrymple, K. A., Gomez, J., and Duchaine, B. (2013). The Dartmouth Database of Children's Faces: acquisition and validation of a new face stimulus set. *PLoS One* 8:e79131. doi: 10.1371/journal.pone.0079131

Darwin, C. R. (1872). *The Expression of the Emotions in Man and Animals*, 1st Edn. London: John Murray. doi: 10.1037/10001-000

Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., et al. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *Int. J. Methods Psychiatr. Res.* 20, 145–156. doi: 10.1002/mpr.343

Ekman, P., Friesen, W. V., and Ellsworth, P. (2013). *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Oxford: Pergamon Press.

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City: A Human Face.

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *J. Pers. Soc. Psychol.* 53, 712–717. doi: 10.1037/0022-3514.53.4.712

Ekman, P., Roper, G., and Hager, J. C. (1980). Deliberate facial movement. *Child Dev.* 51, 886–891. doi: 10.2307/1129478

Elfenbein, H. A., and Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol. Bull.* 128, 203–235. doi: 10.1037/0033-2909.128.2.203

Elfenbein, H. A., Beaupré, M., Lévesque, M., and Hess, U. (2007). Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion* 7, 131–146. doi: 10.1037/1528-3542.7.1.131

Field, T. M., and Walden, T. A. (1982). Production and discrimination of facial expressions by preschool children. *Child Dev.* 53, 1299–1311. doi: 10.2307/1129020

Fölster, M., Hess, U., and Werheid, K. (2014). Facial age affects emotional expression decoding. *Front. Psychol.* 5:30. doi: 10.3389/fpsyg.2014.00030

Fridlund, A. J. (1997). "The new ethology of human facial expressions," in *The Psychology of Facial Expression*, eds J. A. Russell and J. Fernandez-Dols (Cambridge: Cambridge University Press), 103–129. doi: 10.1017/CBO9780511659911.007

AUTHOR CONTRIBUTIONS

CG, SH, JB, AD, and HD: conception, acquisition, and interpretation of data, drafting the work. LaC, SS, PF, MC, LiC, KB, OG, and DC: conception, interpretation of data and revising the work. HP: analysis and interpretation of data, drafting the work.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00446/full#supplementary-material

Gordon, I., Pierce, M. D., Bartlett, M. S., and Tanaka, J. W. (2014). Training facial expression production in children on the autism spectrum. *J. Autism Dev. Disord.* 44, 2486–2498. doi: 10.1007/s10803-014-2118-6

Gosselin, P., Maassarani, R., Younger, A., and Perron, M. (2011). Children's deliberate control of facial action units involved in sad and happy expressions. *J. Nonverbal Behav.* 35, 225–242. doi: 10.1007/s10919-011-0110-9

Grossard, C., Hun, S., Serret, S., Grynszpan, O., Foulon, P., Dapogny, A., et al. (2017). Rééducation de l'expression émotionnelle chez l'enfant avec trouble du spectre autistique grâce aux supports numériques: le projet JEMImE. *Neuropsychiatr. Enfance Adolesc.* 65, 21–32. doi: 10.1016/j.neurenf.2016.12.002

Halberstadt, A. G., Denham, S. A., and Dunsmore, J. C. (2001). Affective social competence. *Soc. Dev.* 10, 79–119. doi: 10.1111/1467-9507.00150

Hall, J. A., Carter, J. D., and Horgan, T. G. (2000). "Gender differences in nonverbal communication of emotion," in *Studies in Emotion and Social Interaction. Second Series. Gender and Emotion: Social Psychological Perspectives*, ed. A. H. Fischer (New York, NY: Cambridge University Press), 97–117. doi: 10.1017/CBO9780511628191.006

Herba, C. M., Landau, S., Russell, T., Ecker, C., and Phillips, M. L. (2006). The development of emotion-processing in children: effects of age, emotion, and intensity. *J. Child Psychol. Psychiatry* 47, 1098–1106. doi: 10.1111/j.1469-7610.2006.01652.x

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., and Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: women recognize only subtle facial emotions better than men. *Acta Psychol.* 135, 278–283. doi: 10.1016/j.actpsy.2010.07.012

Holodynski, M., and Friedlmeier, W. (2006). *Development of Emotions and Their Regulation: A Socioculturally Based Internalization Model.* Boston, MA: Kluwer Academic.

Izard, C. (2001). Emotional intelligence or adaptive emotions? *Emotion* 1, 249–257. doi: 10.1037/1528-3542.1.3.249

Izard, C. E. (1971). *The Face of Emotion.* New York: Appleton-CenturyCrofts.

Izard, C. E., and Malatesta, C. Z. (1987). "Perspectives on emotional development I: differential emotions theory of early emotional development," in *The First Draft of This Paper was Based on an Invited Address to the Eastern Psychological Association.* New York, NY: Wiley-Interscience.

Komatsu, S., and Hakoda, Y. (2012). Construction and evaluation of a facial expression database of children. *Shinrigaku kenkyu* 83, 217–224. doi: 10.4992/jjpsy.83.217

LaFrance, M., Hecht, M. A., and Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. *Psychol. Bull.* 129, 305–334. doi: 10.1037/0033-2909.129.2.305

Lawrence, K., Campbell, R., and Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Front. Psychol.* 6:761. doi: 10.3389/fpsyg.2015.00761

Lee, V., and Wagner, H. (2002). The effect of social presence on the facial and verbal expression of emotion and the interrelationships among emotion components. *J. Nonverbal Behav.* 26, 3–25. doi: 10.1023/A:1014479919684

LoBue, V., and Thrasher, C. (2014). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Front. Psychol.* 5:1532. doi: 10.3389/fpsyg.2014.01532

Louie, J. Y., Oh, B. J., and Lau, A. S. (2013). Cultural differences in the links between parental control and children's emotional expressivity. *Cultur. Divers. Ethnic Minor. Psychol.* 19, 424–434. doi: 10.1037/a0032820

Luherne-du Boullay, V., Plaza, M., Perrault, A., Capelle, L., and Chaby, L. (2014). Atypical crossmodal emotional integration in patients with gliomas. *Brain Cogn.* 92, 92–100. doi: 10.1016/j.bandc.2014.10.003

Mazurski, E. J., and Bond, N. W. (1993). A new series of slides depicting facial expressions of affect: a comparison with the pictures of facial affect series. *Aust. J. Psychol.* 45, 41–47. doi: 10.1080/00049539308259117

Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., and Frith, C. D. (2006). The Kuleshov Effect: the influence of contextual framing on emotional attributions. *Soc. Cogn. Affect. Neurosci.* 1, 95–106. doi: 10.1093/scan/nsl014

Muzard, A., Kwon, A. Y., Espinosa, N., Vallotton, C. D., and Farkas, C. (2017). Infants' emotional expression: differences in the expression of pleasure and discomfort between infants from Chile and the United States. *Infant Child Dev.* 26:e2033. doi: 10.1002/icd.2033

Oster, H. (2005). The repertoire of infant facial expressions: an ontogenetic perspective. *Emot. Dev.* 261–292.

Pantic, M., Valstar, M., Rademaker, R., and Maat, L. (2005). "Web-based database for facial expression analysis," in *Proceedings of the IEEE International Conference on Multimedia and Expo, 2005* (Amsterdam: IEEE). doi: 10.1109/ICME.2005.1521424

Ruffman, T., Henry, J. D., Livingstone, V., and Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging. *Neurosci. Biobehav. Rev.* 32, 863–881. doi: 10.1016/j.neubiorev.2008.01.001

Saarni, C. (1999). *The Development of Emotional Competence.* New York, NY: Guilford Press.

Sato, W., and Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1–18. doi: 10.1016/j.cognition.2006.05.001

Serret, S., Hun, S., Iakimova, G., Lozada, J., Anastassova, M., Santos, A., et al. (2014). Facing the challenge of teaching emotions to individuals with low-and high-functioning autism using a new Serious game: a pilot study. *Mol. Autism* 5, 37. doi: 10.1186/2040-2392-5-37

Sroufe, L. A. (1996). *Emotional Development.* New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511527661

Sullivan, M. W., and Lewis, M. (2003). Emotional expressions of young infants and children: a practitioner's primer. *Infants Young Child.* 16, 120–142. doi: 10.1097/00001163-200304000-00005

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Uljarevic, M., and Hamilton, A. (2013). Recognition of emotions in autism: a formal meta-analysis. *J. Autism Dev. Disord.* 43, 1517–1526. doi: 10.1007/s10803-012-1695-5

Wagner, H. L., and Smith, J. (1991). Facial expression in the presence of friends and strangers. *J. Nonverbal Behav.* 15, 201–214. doi: 10.1007/BF00986922

Wallbott, H. G. (1988). In and out of context: influences of facial expression and context information on emotion attributions. *Br. J. Soc. Psychol.* 27, 357–369. doi: 10.1111/j.2044-8309.1988.tb00837.x

Zeng, Z., Pantic, M., Roisman, G. I., and Huang, T. S. (2009). A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 39–58. doi: 10.1109/TPAMI.2008.52

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., et al. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image Vis. Comput.* 32, 692–706. doi: 10.1016/j.imavis.2014.06.002

Jumping for Joy: The Importance of the Body and of Dynamics in the Expression and Recognition of Positive Emotions

Marcello Mortillaro[1]* and Daniel Dukes[1,2]

[1] *Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland,* [2] *Psychology Research Institute, University of Amsterdam, Amsterdam, Netherlands*

The majority of research on emotion expression has focused on static facial prototypes of a few selected, mostly negative emotions. Implicitly, most researchers seem to have considered all positive emotions as sharing one common signal (namely, the smile), and consequently as being largely indistinguishable from each other in terms of expression. Recently, a new wave of studies has started to challenge the traditional assumption by considering the role of multiple modalities and the dynamics in the expression and recognition of positive emotions. Based on these recent studies, we suggest that positive emotions are better expressed and correctly perceived when (a) they are communicated simultaneously through the face and body and (b) perceivers have access to dynamic stimuli. Notably, we argue that this improvement is comparatively more important for positive emotions than for negative emotions. Our view is that the misperception of positive emotions has fewer immediate and potentially life-threatening consequences than the misperception of negative emotions; therefore, from an evolutionary perspective, there was only limited benefit in the development of clear, quick signals that allow observers to draw fine distinctions between them. Consequently, we suggest that the successful communication of positive emotions requires a stronger signal than that of negative emotions, and that this signal is provided by the use of the body and the way those movements unfold. We hope our contribution to this growing field provides a new direction and a theoretical grounding for the many lines of empirical research on the expression and recognition of positive emotions.

Keywords: emotion, positive emotions, dynamics, facial expression, bodily expression, emotion expression, emotion recognition

INTRODUCTION

The last 15 years have seen unprecedented interest in positive emotions, sustained, presumably, by the development of fields like positive psychology (Fredrickson and Joiner, 2002) and emotional intelligence (Quoidbach et al., 2010; Nelis et al., 2011). Before then, emotion research had largely focused on a set of almost entirely negative emotions that had been identified by Ekman (1992, 1993). In fact, Ekman's original set of basic emotions featured only one positive emotion – joy or happiness – and, consequently, several authors considered joy-happiness as the *only* positive emotion in their early studies (e.g., Oatley and Johnson-Laird, 1987). Conceiving of positive

emotion in this way led to them being treated as one, single, undifferentiated class of events, and this naturally became an obstacle toward progress in positive emotion research. In perhaps the clearest sign that the field has since matured, a recent and comprehensive review by Shiota et al. (2017) argues that positive emotions may be differentiated based on distinct autonomic nervous system signatures, different effects on cognition and judgment, and specific non-verbal behaviors.

In this article, we focus on the non-verbal behaviors associated with positive emotions. We offer a new perspective as to why the quest for the identification of specific signals of positive emotions needs to be redirected beyond static prototypical faces. We are aware that the positive vs. negative distinction could be debated and that emotional communication is a more complex process than the simple perception of emotion categories – as we have discussed elsewhere (Mortillaro et al., 2013; Scherer et al., 2013, 2018; Reschke et al., 2017). However, this paper is about the signals that can be used for the accurate communication of pleasant emotional states (e.g., a smile that signals embarrassment is not one of them) and does not assume that these signals are exclusive to genuine emotion signaling (a polite smile is a pure social signal). The reader should be aware that this is a brief perspective paper and not an attempt at an exhaustive review. We therefore focus on the most relevant literature for our argument and highlight what is novel and worthwhile about our perspective. Furthermore, we decided to focus on why this quest should include the dynamics of facial movements and the body, although a similar case could be made to include the voice (Sauter and Scott, 2007; Sauter, 2017), the context (Hassin et al., 2013; Aviezer et al., 2017), and even autonomic signals like pupil dilation (Kret, 2015).

We begin with an overview of the standard accounts of facial emotion expression and recognition, before providing a justification for why we feel a change in direction for empirical studies of positive emotion is necessary.

ENJOYMENT SMILE: THE ONLY SIGN FOR ALL POSITIVE EMOTIONS?

Research in non-verbal behavior in emotion has traditionally concentrated on the face and, following the approach used by Ekman to identify basic emotions, has aimed at identifying prototypical configurations of facial expression. However, this approach has not proved very successful for positive emotions.

Progress was initially hampered by an implicit consensus that all positive emotions were essentially expressed in the same way. Notably, the enjoyment smile [the result of the action of the *zygomaticus major* muscle and the contraction of the *orbicularis oculi pars lateralis* muscle (Ekman and Friesen, 1978)] was originally held to be the only (and ubiquitous) sign of positive emotions. In a quote from 1992 that not only outlines the problem but also offers a possible solution, Ekman wrote, "One of the questions remaining about smiles is whether the different positive emotions (e.g., amusement, contentment, relief,

etc.) have distinctive forms of smiling, or if the variety of positive emotions share one signal and can be inferred only from other behavioral or contextual cues. I presume that all of these forms of enjoyment share the musculature described by Duchenne, and are distinguished by their dynamics, not their morphology" (Ekman, 1992, p. 67).

Several studies have since then shown that there are various types of smiles, with different interpersonal functions (for example, Rychlowska et al., 2017), and that most smiles are social signals and not simple reflections of inner feelings (Fridlund, 1997). However, even when signs other than the smile are included, the pool of positive emotions linked to particular static expressions remains very limited, and there are only a few studies that have explicitly compared multiple positive emotion expressions (e.g., Hofmann et al., 2017). In one notable exception, Campos et al. (2013) confirmed the critical role of the Duchenne smile across several positive emotions. The authors identified associations between each positive emotion and some facial action units, but the resulting configurations were not entirely different while it was the inclusion of head and upper body movements that made the emotions more distinguishable. For example, facial expressions of pride and contentment can be differentiated only by their associated head position.

In a recent review, Sauter (2017) suggests a more complex version of Ekman's view of positive emotion as a family of 'forms of enjoyment.' In fact, Sauter suggests four families of positive emotions – 'epistemological,' 'prosocial,' 'savoring,' and 'agency-approach.' Based on her review, only epistemological emotions (amusement, awe, interest, and relief) and pride appear to have distinct recognizable facial and/or vocal displays. It is worth noting, however, that the prototypical expression of pride also includes bodily movements aimed at postural expansion, which involves, for example, pulling the shoulders back and raising the head.

All in all, there is only weak evidence for the differentiation between positive emotions based on static facial features. We hypothesize that the expressive elements that differentiate positive emotions most clearly reside in the dynamics of facial expression and in the body.

HYPOTHESIS: FACIAL DYNAMICS AND BODY REPRESENTATIONS ARE CRITICAL FOR DISTINGUISHING NON-VERBAL DISPLAYS OF POSITIVE EMOTION

From a functional perspective, there is an enduring debate about whether emotion expressions are direct reflections of inner-states (I smile because I am happy), or whether emotions are expressed as social signals (I smile at you to show you I am happy; see Parkinson, 2005). From an evolutionary perspective, this debate is often drawn along the lines of whether the emotional expression is made for the benefits of the expresser (such as when someone widens his/her eyes in states of fear

to increase the perceptual uptake in order to prepare his/her escape from danger) which may serve as an emotional cue to observers, or, alternatively, whether the expression may be used intentionally to communicate something to observers (for a discussion, see Schmidt and Cohn, 2001; Kret and Straffon, 2018).

In order to demonstrate our argument, we will focus on what the observer picks up from the expression rather than the processes that produce the expression (Frijda and Tcherkassof, 1997). In evolutionary terms, negative emotions (e.g., fear and anger) are more critical for survival than positive emotions (e.g., pride and interest) because they are more likely to be understood as signs of potentially life-threatening situations that require an immediate response. There is an element of urgency that is not present in the case of positive emotions and that requires the signal to be understood quickly, clearly, and very specifically. These are the benefits of prototypical facial expressions; they have a "snapshot" quality that makes them rapidly recognizable and the emotions effectively identifiable (Ekman, 1993). Consequently, it makes sense that signals have evolved to rapidly and effectively communicate the potential dangers in the environment to conspecifics and that skills have evolved to recognize that threat. In a recent study, Gold et al. (2013) found that participants could recognize the traditional six basic emotions (including joy as the only positive emotion) with comparable accuracy regardless of whether they viewed the expressions as naturally evolving, temporally reversed, temporally randomized expressions, or as a single snapshot. This result supports the hypothesis that dynamic information is not necessary for the correct recognition of basic negative emotions.

The fact the positive emotions are less critical for survival is not to deny the importance of their social functions. Positive emotions are involved in affiliation and cooperation and therefore important for adaptation (Campos et al., 2015). Different positive emotions have specific functions – respond to material opportunities or social stimuli, facilitate playing new skills, encode novel information – that require distinct expressive signals to be effectively communicated (Shiota et al., 2014). However, as mentioned previously, it appears that static faces do not provide a clear enough signal. While static facial expressions are sufficient for distinguishing negative emotions in most circumstances, we argue that the distinction between positive emotions critically requires additional information that is provided by the dynamics and body representations.

Dynamic representations of emotion expressions evidently contain more information than static ones, but they do not always increase the rate at which emotions are recognized (Scherer et al., 2011). In fact, it is not the sum of static cues that explains why dynamic stimuli are better recognized in some conditions, but rather the specific information that is conveyed by the movement (Ambadar et al., 2005). Interestingly, Jack et al. (2014) suggest that the perception process is temporally driven and that dynamic facial expressions transmit an evolving hierarchy of signals over time, from biologically basic (approach/avoidance) to social information, such as emotion categories. Similarly, the increase in information

provided by adding bodily information to facial expressions does not automatically increase the rate at which emotions are correctly recognized. Studies show that the interaction between bodies and faces is more complex than simply aggregating the information from each modality (Aviezer et al., 2008, 2012).

App et al. (2011) suggest that the body promotes social-status emotions, that the face promotes survival emotions, and that touch promotes intimate emotions. Elsewhere, Martinez et al. (2016) found that for the standard set of six basic emotions, five of which are negative, the face was significantly better than the body in conveying emotional information. Again, these two studies provide indirect support for our hypothesis that the face is critical and sufficient for the communication of basic, survival-related emotions, but not for other types of emotions.

It seems then that good evolutionary, social and functional justifications can be found for arguing that positive emotions need to be signaled more "loudly" in order to be correctly identified and recognized than negative emotions. We turn now to recent empirical studies that seem to support our argument.

Evidence About Dynamic Facial Expressions of Positive Emotions

Researchers mostly used – and still use – static prototypical facial expressions in their studies (Scherer et al., 2011). Recently, however, there is a growing trend toward the use of dynamic expressions that do not fully correspond to the traditional prototypes (Bänziger et al., 2012; O'Reilly et al., 2016; Krumhuber et al., 2017). This methodological choice allows emotions to be studied that are not found in the standard basic set (as there is no fixed, pre-defined prototype to be portrayed) and to compare subtly different emotions.

In a recent review concerning the role of dynamics in emotion recognition, Krumhuber et al. (2013, p. 42) wrote that motion "…confers particular benefits when static information is inefficient or unavailable." Given the absence of prototypical facial configurations, it is therefore not surprising that the study of positive emotions has benefited from the inclusion of dynamic stimuli. Indeed, movement dynamics are an integral part of the emotion perception process, and it is used by perceivers to differentiate deliberate and genuine smiles (that is when the smiles are spontaneous and reflect a felt positive emotional state) or to judge the naturalness of the emotion expression tout court (Sato and Yoshikawa, 2004; Krumhuber and Kappas, 2005; Schmidt et al., 2006). In one pioneering study using synthetic facial expressions, Wehrle et al. (2000) and Kaiser and Wehrle (2001) found that positive emotional states such as pleasure, happiness, and elation, could be distinguished by their facial expressions when dynamic stimuli were presented. In a more recent study, Mortillaro et al. (2011) showed that joy, interest, pride, and sensory pleasure could only be distinguished when the dynamic properties of the expressions were taken into account. It was not the presence or the absence of certain facial movements that could be used to reliably differentiate these emotions, but rather the duration of the movements and their frequency within

one emotion expression. Similarly, Fujimura and Suzuki (2010) found that two out of the three positive emotions that they included in their study were significantly better recognized in the dynamic than in the static presentation mode, while only one out of the five non-positive emotions (fearful) showed the same significant advantage when presented dynamically.

Other studies have demonstrated the special role of dynamic movements for specific positive emotions. For example, while the search for a prototypical static facial expression of interest has proven inconclusive, emotional expressions of interest can be well recognized when it presented in a dynamic fashion (Dukes et al., 2017). Furthermore, Nelson and Russell (2014) have shown that different types of pride can only be differentiated when dynamically presented. Similarly, Namba et al. (2017) found a different dynamic pattern of movements in posed and spontaneous expressions of amusement – a difference that did not appear in static expressions.

Overall, it appears that the dynamic representation of positive emotions may be critical for them to be readily identified and differentiated (for a similar position, see Fujimura et al., 2012).

Evidence About the Bodily Expression of Positive Emotions

The expression of emotions through body movements and gestures has been understudied in comparison to facial and vocal expressions [for a general discussion of the neurological basis of the perception of emotions from the body and for the reasons to consider bodily expressions in affective science, please see the works of de Gelder (2006, 2009)]. Nevertheless, results of a number of studies showed that emotions can be recognized from bodies (e.g., de Gelder and Van den Stock, 2011) and even from very limited information like point-light body displays (Atkinson et al., 2004). A full review of this literature is beyond the scope of a perspective article and therefore, we will only discuss studies that investigated the bodily expression of several positive emotions.

In one of the largest studies available on the bodily expression of emotions, Dael et al. (2012) identified patterns of body movements that were specific to positive emotions. Even more importantly, they showed that positive emotions could be correctly discriminated from their bodily movements alone, even more so than the negative emotions. On average the positive emotions were correctly classified 63.3% of the time on the basis of bodily movements (when chance level was 8.33%), while the negative emotions were only correctly classified 46.7% of the time.

Similarly, App et al. (2011) found that pride and love were better recognized in the body than in the face, while happiness and sympathy were recognized at the same level in the two modalities. Dael et al. (2013) studied the dynamic properties of arm movements. Even though they did not explicitly compare the six positive emotions, substantial differences among them are clear in most, if not all, the parameters they reported. This corroborates our hypothesis that bodily movements are critical for distinguishing between positive emotions.

The effects of bodily representations on expressing specific positive emotions also tend to support our argument. The clearest case comes from research on pride for which there is general consensus about a prototypical expression involving a particular posture and specific gestures (Tracy and Robins, 2004). Another positive emotion for which the body seems to carry important information is interest. Dukes et al. (2017) found that facial expressions alone were not able to reliably communicate interest; however, when the face was paired with the body, the recognition accuracy for interest more than doubled, and interest became as easily recognized as Ekman's six basic emotions.

There is sufficient empirical evidence to suggest that the identification and recognition of positive emotions is made comparatively easier by the inclusion of bodily representations whereas, similarly to the inclusion of dynamic information, this seems less important for negative emotions.

CONCLUSION

In this paper, we briefly reviewed some of the most recent and relevant literature on the expression of positive emotions. The results consistently indicate that the research of purely facial static prototypes is likely inconclusive. If specific (or typical) expressions for positive emotions exist, they are more likely to be found in expressions that include dynamic and bodily elements, like body posture and gesture. It is more than 10 years since the prototypical expressions of pride were established and, so far, only a few scholars have pointed out that it is the body and posture or the dynamic representation of these expressions that sets them apart from those of joy. It is now time to accept that static facial expressions are useful, but that they do not capture the whole richness of real-life emotion communication. Future studies, especially when positive emotions are considered, should only use multimodal, dynamic expressions.

AUTHOR CONTRIBUTIONS

MM conceived and wrote the first draft of the manuscript. DD contributed to the conception of the article and helped revise the first and subsequent drafts.

FUNDING

REFERENCES

Ambadar, Z., Schooler, J. W., and Conn, J. F. (2005). Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

App, B., McIntosh, D. N., Reed, C. L., and Hertenstein, M. J. (2011). Nonverbal channel use in communication of emotion: how may depend on why. *Emotion* 11, 603–617. doi: 10.1037/a0023164

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33, 717–746. doi: 10.1068/p5096

Aviezer, H., Ensenberg, N., and Hassin, R. R. (2017). The inherently contextualized nature of facial emotion perception. *Curr. Opin. Psychol.* 17, 47–54. doi: 10.1016/j.copsyc.2017.06.006

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., et al. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception: research article. *Psychol. Sci.* 19, 724–732. doi: 10.1111/j.1467-9280.2008.02148.x

Aviezer, H., Trope, Y., and Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338, 1225–1229. doi: 10.1126/science.1224313

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827

Campos, B., Schoebi, D., Gonzaga, G. C., Gable, S. L., and Keltner, D. (2015). Attuned to the positive? Awareness and responsiveness to others' positive emotion experience and display. *Motiv. Emot.* 39, 780–794. doi: 10.1007/s11031-015-9494-x

Campos, B., Shiota, M. N., Keltner, D., Gonzaga, G. C., and Goetz, J. L. (2013). What is shared, what is different? Core relational themes and expressive displays of eight positive emotions. *Cogn. Emot.* 27, 37–52. doi: 10.1080/02699931.2012.683852

Dael, N., Goudbeek, M., and Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception* 42, 642–657. doi: 10.1068/p7364

Dael, N., Mortillaro, M., and Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion* 12, 1085–1101. doi: 10.1037/a0025737

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nat. Rev. Neurosci.* 7, 242–249. doi: 10.1038/nrn1872

de Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3475–3484. doi: 10.1098/rstb.2009.0190

de Gelder, B., and Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Front. Psychol.* 2:181. doi: 10.3389/fpsyg.2011.00181

Dukes, D., Clément, F., Audrin, C., and Mortillaro, M. (2017). Looking beyond the static face in emotion recognition: the informative case of interest. *Vis. Cogn.* 25, 575–588. doi: 10.1080/13506285.2017.1341441

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1016/j.concog.2017.10.008

Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392. doi: 10.1037/0003-066X.48.4.384

Ekman, P., and Friesen, W. V. (1978). *The Facial Action Coding System*. Palo Alto, CA: Consulting Psychologist Press.

Fredrickson, B. L., and Joiner, T. (2002). Positive emotions trigger upward spirals toward emotional well-being. *Psychol. Sci.* 13, 172–175. doi: 10.1111/1467-9280.00431

Fridlund, A. J. (1997). "The new ethology of human facial expressions," in *New Directions in the Study of Facial Expressions*, eds J. A. Russell and J. M. F. Dols (Cambridge: Cambridge University Press), 103–129.

Frijda, N. H., and Tcherkassof, A. (1997). "Facial expressions as modes of action readiness," in *The Psychology of Facial Expression*, eds J. A. Russell and J. M. Fernandez-Dols (Cambridge: Cambridge University Press), 78–102. doi: 10.1017/CBO9780511659911.006

Fujimura, T., Sato, W., and Okanoya, K. (2012). Subcategories of positive emotion. *Psychologia* 55, 1–8. doi: 10.2117/psysoc.2012.1

Fujimura, T., and Suzuki, N. (2010). Effects of dynamic information in recognising facial expressions on dimensional and categorical judgments. *Perception* 39, 543–552. doi: 10.1068/p6257

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., et al. (2013). The efficiency of dynamic and static facial expression recognition. *J. Vis.* 13:23. doi: 10.1167/13.5.23.

Hassin, R. R., Aviezer, H., and Bentin, S. (2013). Inherently ambiguous: facial expressions of emotions, in context. *Emot. Rev.* 5, 60–65. doi: 10.1177/1754073912451331

Hofmann, J., Platt, T., and Ruch, W. (2017). Laughter and smiling in 16 positive emotions. *IEEE Trans. Affect. Comput.* 8, 495–507. doi: 10.1109/TAFFC.2017.2737000

Jack, R. E., Garrod, O. G., and Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* 24, 187–192. doi: 10.1016/j.cub.2013.11.064

Kaiser, S., and Wehrle, T. (2001). "Facial expressions as indicators of appraisal processes," in *Appraisal Processes in Emotions: Theory, Methods, Research*, eds K. R. Scherer, A. Schorr, and T. Johnstone (New York, NY: Oxford University Press), 285–300.

Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Front. Psychol.* 6:711. doi: 10.3389/fpsyg.2015.00711

Kret, M. E., and Straffon, L. M. (2018). Reply to Crivelli et al.: the different faces of fear and threat. Evolutionary and cultural insights. *J. Hum. Evol.* doi: 10.1016/j.jhevol.2017.11.006 [Epub ahead of print].

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonverbal Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Krumhuber, E., Skora, P., Küster, D., and Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Martinez, L., Falvello, V. B., Aviezer, H., and Todorov, A. (2016). Contributions of facial expressions and body language to the rapid perception of dynamic emotions. *Cogn. Emot.* 30, 939–952. doi: 10.1080/02699931.2015.1035229

Mortillaro, M., Mehu, M., and Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. *Soc. Psychol. Pers. Sci.* 2, 262–271. doi: 10.1177/1948550610389080

Mortillaro, M., Mehu, M., and Scherer, K. R. (2013). "The evolutionary origin of multimodal synchronization and emotional expression," in *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (New York, NY: Oxford University Press), 3–25.

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., and Nakao, T. (2017). Spontaneous facial expressions are different from posed facial expressions: morphological properties and dynamic sequences. *Curr. Psychol.* 36, 593–605. doi: 10.1007/s12144-016-9448-9

Nelis, D., Kotsou, I., Quoidbach, J., Hansenne, M., Weytens, F., Dupuis, P., et al. (2011). Increasing emotional competence improves psychological and physical well-being, social relationships, and employability. *Emotion* 11, 354–366. doi: 10.1037/a0021554

Nelson, N. L., and Russell, J. A. (2014). Dynamic facial expressions allow differentiation of displays intended to convey positive and hubristic pride. *Emotion* 14, 857–864. doi: 10.1037/a0036789

Oatley, K., and Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cogn. Emot.* 1, 29–50. doi: 10.1080/02699938708408362

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., et al. (2016). The EU-emotion stimulus set: a validation study. *Behav. Res. Methods* 48, 567–576. doi: 10.3758/s13428-015-0601-4

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Pers. Soc. Psychol. Rev.* 9, 278–311. doi: 10.1207/s15327957pspr0904_1

Quoidbach, J., Berry, E. V., Hansenne, M., and Mikolajczak, M. (2010). Positive emotion regulation and well-being: comparing the impact of eight savoring and dampening strategies. *Pers. Individ. Dif.* 49, 368–373. doi: 10.1016/j.paid.2010.03.048

Reschke, P. R., Walle, E. A., and Dukes, D. (2017). Interpersonal development in infancy: the interconnectedness of emotion understanding and social cognition. *Child Dev. Perspect.* 11, 173–183. doi: 10.1111/cdep.12230

Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., and Niedenthal, P. M. (2017). Functional smiles: tools for love, sympathy, and war. *Psychol. Sci.* 28, 1259–1270. doi: 10.1177/0956797617706082

Sato, W., and Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Sauter, D. A. (2017). The nonverbal communication of positive emotions: an emotion family approach. *Emot. Rev.* 9, 222–234. doi: 10.1177/1754073916667236

Sauter, D. A., and Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motiv. Emot.* 31, 192–199. doi: 10.1007/s11031-007-9065-x

Scherer, K. R., Clark-Polner, E., and Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *Int. J. Psychol.* 46, 401–435. doi: 10.1080/00207594.2011.626049

Scherer, K. R., Mortillaro, M., and Mehu, M. (2013). Understanding the mechanisms underlying the production of facial expression of emotion: a componential perspective. *Emot. Rev.* 5, 47–53. doi: 10.1177/1754073912451504

Scherer, K. R., Mortillaro, M., Rotondi, I., Sergi, I., and Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion inference. *J. Pers. Soc. Psychol.* 114, 358–379. doi: 10.1037/pspa0000107

Schmidt, K. L., Ambadar, Z., Cohn, J. F., and Reed, L. I. (2006). Movement differences between deliberate and spontaneous facial expressions: zygomaticus major action in smiling. *J. Nonverbal Behav.* 30, 37–52. doi: 10.1007/s10919-005-0003-x

Schmidt, K. L., and Cohn, J. F. (2001). Human facial expressions as adaptations: evolutionary questions in facial expression research. *Am. J. Phys. Anthropol.* 116, 3–24. doi: 10.1002/ajpa.20001

Shiota, M. N., Campos, B., Oveis, C., Hertenstein, M. J., Simon-Thomas, E., and Keltner, D. (2017). Beyond happiness: building a science of discrete positive emotions. *Am. Psychol.* 72, 617–643. doi: 10.1037/a0040456

Shiota, M. N., Neufeld, S. L., Danvers, A. F., Osborne, E. A., Sng, O., and Yee, C. I. (2014). Positive emotion differentiation: a functional approach. *Soc. Pers. Psychol. Compass* 8, 104–117. doi: 10.1111/spc3.12092

Tracy, J. L., and Robins, R. W. (2004). Show your pride: evidence for a discrete emotion expression. *Psychol. Sci.* 15, 194–197. doi: 10.1111/j.0956-7976.2004.01503008.x

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Pers. Soc. Psychol.* 78, 105–119. doi: 10.1037//0022-3514.78.1.105

Dynamic Displays Enhance the Ability to Discriminate Genuine and Posed Facial Expressions of Emotion

Shushi Namba[1]*, Russell S. Kabir[1], Makoto Miyatani[2] and Takashi Nakao[2]

[1] Graduate School of Education, Hiroshima University, Hiroshima, Japan, [2] Department of Psychology, Hiroshima University, Hiroshima, Japan

Accurately gauging the emotional experience of another person is important for navigating interpersonal interactions. This study investigated whether perceivers are capable of distinguishing between unintentionally expressed (genuine) and intentionally manipulated (posed) facial expressions attributed to four major emotions: amusement, disgust, sadness, and surprise. Sensitivity to this discrimination was explored by comparing unstaged dynamic and static facial stimuli and analyzing the results with signal detection theory. Participants indicated whether facial stimuli presented on a screen depicted a person *showing* a given emotion and whether that person was *feeling* a given emotion. The results showed that genuine displays were evaluated more as felt expressions than posed displays for all target emotions presented. In addition, sensitivity to the perception of emotional experience, or discriminability, was enhanced in dynamic facial displays, but was less pronounced in the case of static displays. This finding indicates that dynamic information in facial displays contributes to the ability to accurately infer the emotional experiences of another person.

Keywords: spontaneous facial expressions, posed facial expressions, dynamics, facial expressions, emotion

INTRODUCTION

Facial expressions provide a signature of the emotional state of an interlocutor to indicate behaviors that are appropriate in an interpersonal situation (Keltner and Haidt, 2001; Ekman, 2003). However, not all facial displays reflect emotional experiences that are actually being felt by the expresser, and can even be co-opted. Humans have been shown to be able to feign facial expressions of felt emotions as a form of intentional deception to gain social advantages (Krumhuber and Manstead, 2009) and to stage displays that are meant to solicit the help of others (Ekman, 2001). Staged or posed facial expressions display an emotion that an expresser ostensibly intends to convey, whereas unstaged or genuine expressions are thought to portend the sense of authenticity that accompanies the spontaneity of felt emotional expressions. The endogenous nature of emotional experiences is posited to increase the trustworthiness of the expresser by emboldening the need to embark upon and ensure a successful social interaction. For example, Johnston et al. (2010) showed that genuine smiles could make perceivers opt for cooperative behavior more than posed smiles. On the other side of the spectrum, pretending to be sad is an expressive strategy that leads to loss consequences for the perceiver when an expresser

feigns sadness to take advantage of a perceiver's reciprocal kindness or compensatory behavior in response (Reed and DeScioli, 2017). Thus, the ability to differentiate genuine displays of emotional experiences from posed ones can be important for dealing with day-to-day social interactions.

Recent work has been conducted on whether people can distinguish between genuine and posed displays of emotion (e.g., McLellan et al., 2010; Douglas et al., 2012; Dawel et al., 2015). McLellan et al. (2010) showed that adults are capable of differentiating posed and genuine facial displays of happiness, sadness, and fear. Dawel et al. (2015) also replicated the finding that adults could discriminate the authenticity of happy and sad displays. Moreover, a neuroimaging study showed that the perception of genuine and posed non-verbal behaviors occurs through different neural activation processes (McLellan et al., 2012; McGettigan et al., 2013). Although there have been few studies that investigate this ability, most prior research suggests that people can make a distinction when judging genuine and posed facial displays.

Nevertheless, previous research has suffered from two major shortcomings: (1) the presence of "staged" contamination in genuine displays due to a lack of accounting for the possible effects of intentional manipulation, and (2) a failure to include dynamic aspects when preparing facial stimuli for experimental investigations. First, research methodologies have mainly relied upon the proprietary facial stimuli created by McLellan et al. (2010), which recruited participants who were expressly informed of the purpose of the study as one to investigate the feasibility of creating stimulus material. The experimenters then proceeded recording the facial expressions of participants as they were evoked by emotion elicitation pictures, sounds, and imagined scenarios. While the experimenters selected genuine displays based on databases of affective picture stimuli and other established experimental techniques from empirical studies, the fact that participants were made aware of the purpose of the facial stimuli ahead of the experiment might have allowed for the confounding effects of intentional manipulation to occur in genuine facial displays as they unfolded. This raises an issue as it is thought that such intentional influences might inhibit spontaneous facial reactions (Smoski and Bachorowski, 2003; Kunzmann et al., 2005). Furthermore, selection of genuine stimuli in the study relied heavily on criteria undertaken for intended facial expressions made by actors (Gosselin et al., 1995; Suzuki and Naitoh, 2003), as several findings have shown actors' expressions to be relatively similar to spontaneous expressions (e.g., Carroll and Russell, 1997; Scherer and Ellgring, 2007; Gosselin et al., 2010). While it is indeed the case that expressions made by professional actors might encompass some experiences of felt emotion in the process, they are ultimately designed to emphasize a message through intentional or strategic manipulation (Buck and VanLear, 2002). This suggests that facial stimuli used in previous studies could have been biased from being subject to intentional manipulation by participants themselves or through selection criteria that was based on the staged facial expressions of actors. Indeed, McLellan et al. (2010) tagged the cheek raising found in the expression of happiness as a property that distinguishes genuine and posed smiles, but

other studies have shown that the presence of cheek raising more likely reflects expressive intensity rather than pleasant experience (Krumhuber and Manstead, 2009; Guo et al., 2018). In other words, previous studies might actually be tapping differences in expressive intensity rather than an underlying ability to tell the difference between posed versus genuine expressions. Recent work by Dawel et al. (2017) also showed that observers did not regard the McLellan et al. (2010) genuine faces as actual genuine displays. Thus, it is clear that to better understand the ability for individuals to differentiate genuine displays containing emotional experiences from posed ones, unintentionally manipulated displays that are most frequently expressed in strong evocations of genuine emotional situations should be implemented.

Second, previous experiments have employed static facial stimuli and largely ignored the dynamic aspects of facial expressions. Dynamic information in facial expressions for various emotions has been increasingly recognized as an important aspect in the phenomenon of emotion perception (Krumhuber et al., 2013) and the recognition of crowd valence (Ceccarini and Caudek, 2013). Ceccarini and Caudek (2013) found that dynamic over static facial information captures the attention of perceivers attending to threatening stimuli. Furthermore, Krumhuber and Manstead (2009) showed that observers can differentiate spontaneous and posed smiles when rating the genuineness and amusement of dynamic displays, but not static ones. Although the importance of dynamic information in differentiating facial expressions has been put forth, not all emotions have been accounted for. Given the evidence from previous studies that have underscored the dynamic aspects of facial expression for emotion perception (e.g., Wehrle et al., 2000; Sato and Yoshikawa, 2007), operationalizing dynamic displays as stimuli for other emotions like surprise, disgust, and sadness, in addition to amusement, would allow for sensitivity in the perception of emotional experience to be evaluated. Taken together, it remains unclear whether people can differentiate genuine from posed facial displays because there is a possibility that the genuine displays used in previous studies are different from spontaneous facial reactions to emotional experiences. Moreover, it is necessary to consider dynamic information that might affect this discriminability beyond the emotion of amusement through investigations of other emotions like surprise, disgust, and sadness.

Thus, the current study re-investigated hypotheses related to the ability for perceivers to distinguish genuine from posed facial expressions by critically implementing facial display stimuli generated in the absence of intentional manipulation. This effort aimed to eliminate the influence of intentional effects in genuine facial stimuli as much as possible to test the assumption in the literature that people can differentiate between genuine and posed facial expressions (McLellan et al., 2010; Douglas et al., 2012; Dawel et al., 2015). Furthermore, this study explored whether the presence of dynamic information in facial stimuli strengthens this genuine-posed discriminability or not in the case of negative emotions in addition to amusement. Considering the findings of Krumhuber and Manstead (2009), it was assumed that sensitivity to this discrimination would be increased for dynamic

displays as compared to static ones, and that the evidence base for the phenomenon would be extended beyond amusement to surprise, disgust, and sadness. To further control for the effects of expressive intent as much as possible, the current study utilized the spontaneous facial data obtained in a previous study (Namba et al., 2017a). Spontaneous and posed facial expressions for the emotions of amusement, disgust, surprise, and sadness were recorded to compare morphological aspects in that study, where video clips of secretly recorded facial behaviors as expressers experienced a strong emotion in a room by themselves were used as genuine displays. Posed facial stimuli were derived from the same data of expressers intentionally generating facial expressions according to explicit instructions (for further detail, see Namba et al., 2017c).

MATERIALS AND METHODS

Participants

Fifty-eight participants (35 female, 23 male; *M* age = 23.98, *SD* = 1.67) were recruited from Hiroshima University and the local community via e-mail and advertisements, and were compensated with 500 yen after the experiment.

Participants were randomly assigned to one of two groups: (a) *dynamic presentation* (12 female, 18 male; *M* age = 24.00, *SD* = 1.49), and (b) *static presentation* (11 female, 17 male; *M* age = 23.96, *SD* = 1.86). This assignment resulted in 30 individuals designated to the *dynamic presentation* group, and 28 individuals designated to the *static presentation* group. All participants were native Japanese speakers with normal or corrected-to-normal vision. There was no evidence of the presence of a neurological or psychiatric disorder. Written informed consent was obtained from each participant before the investigation, in line with a protocol approved by the Ethical Committee of the Graduate School of Education, Hiroshima University.

Stimuli

Clips of spontaneous and posed facial actions induced without expressive intent recorded in Namba et al. (2017c) were used as genuine and posed facial displays. Genuine facial displays were elicited in an individual environment with emotion elicitation films (Gross and Levenson, 1995), while posed facial displays were expressed in accordance with the explicit instruction "to express the target emotion." Namba et al. (2017c) picked only the four emotion types of amusement, surprise, disgust and sadness

FIGURE 1 | A depiction of the experimental flow for each trial.
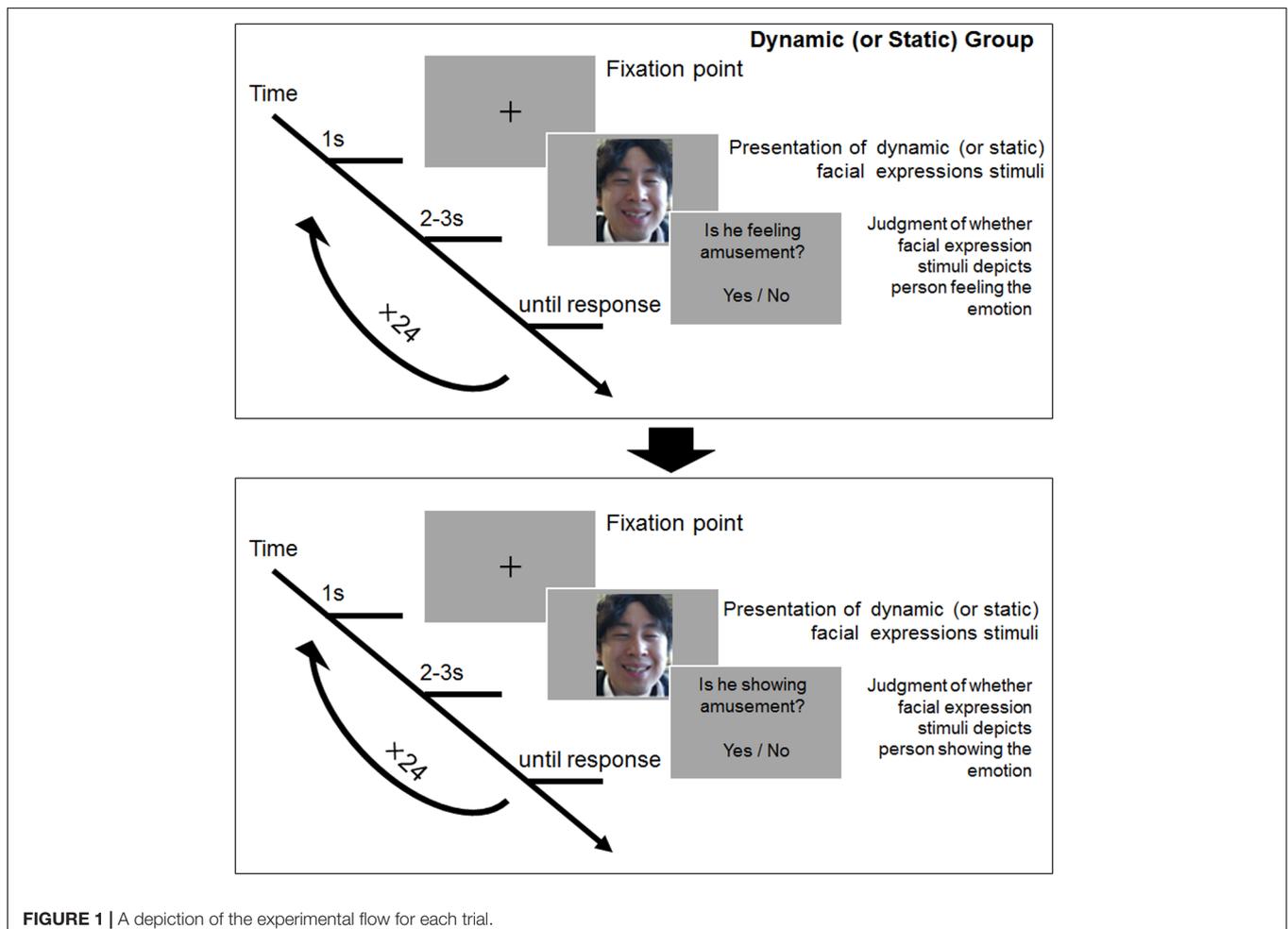
that were confirmed by a previous study to elicit target emotions in Japanese adults viewing emotion elicitation films (Sato et al., 2007). After recording their genuine expressions, participants were debriefed about their candid recordings in line with protocols set by the Ethical Committee of the Graduate School of Education, Hiroshima University, to which data collection was affirmed or denied if the participant consented to the use of their recordings for analysis, and in the event that consent was not given, the recorded data was deleted in front of the participant (Namba et al., 2017c). Among these facial displays, the parts of the clips to be used as stimuli were selected based on the following criteria: (1) the spontaneous and posed facial expressions contained the most frequently expressed and representative properties among expressers (Namba et al., 2017c), (2) the spontaneous facial expression contained facial components related to target emotional experiences in other empirical studies (Namba et al., 2017a,b), and (3) the same expresser was present in both the spontaneous and posed facial expressions in order to avoid inter-target differences. Additionally, dynamic and static presentations were created using these clips. In dynamic presentations, facial displays were played continuously from onset to peak display of facial expression. In static presentations, facial displays were edited such that only one peak frame of a facial expression was presented. Two expressers were assigned to each emotion including a neutral state representing no emotion. Consequently, 2 (expresser) × 4 (emotion: amusement, disgust, surprise and sadness) × 3 (display: genuine, posed and neutral) × 2 (presentation style: dynamic and static) clips were used, resulting in 48 total clips and 24 clips per presentation style. For dynamic presentation, the mean duration of unfolding genuine facial displays was 2.88 s (SD = 2.03), whereas those of posed and neutral ones were 2.50 and 2.38 s (SDs = 1.07 and 1.30). Welch's two sample t-test revealed that the durations among all displays were not different (uncorrected ps > 0.57). The overall durations were 2.58 s (SD = 1.47), and for static presentation all durations were set to 2.5 s. Furthermore, we checked the perceived intensity of expressions as a preliminary analysis. Seven individuals (3 female, 4 male) evaluated the intensity of facial clips on an 8-point scale ranging from 0 (not at all) to 7 (the strongest). One-way analysis of variance revealed that the perceived intensity was different among three displays [$F(2,110) = 128.69$, $p < 0.001$]. Multiple comparisons also showed differences between neutral (M = 0.41, SD = 0.61) and genuine (M = 3.52, SD = 2.13) or posed (M = 3.88, SD = 1.81; ps < 0.001), but no significant difference was found between genuine and posed displays (p = 0.08).

Procedure

The procedure of experimental tasks was conducted in line with the design implemented by McLellan et al. (2010). The task program was created using Visual C#. Each facial clip was presented on the screen of a laptop computer. Two groups of participants were assigned a facial stimuli presentation style: dynamic or static. The task program presented each trial into a block by culling the stimulus to be presented from a pool of 24 dynamic facial stimuli and 24 posed facial stimuli. We asked participants to perform two types of judgment tasks for

the perception of emotional states via facial displays. The first was a *show condition* to judge whether the specific emotion was being depicted (e.g., "Is he showing sadness?"), and the second was a *feel condition* to judge whether the specific emotion was being experienced by the target (e.g., "Is she feeling happiness?"). Participants gave a yes-or-no answer to sort the show and feel conditions. The order of facial stimuli was randomized, and the blocks for the show and feel conditions were counterbalanced using a Latin Square design. **Figure 1** depicts the experimental flow.

Upon arrival at the laboratory and before doing the experimental tasks, participants were given careful instructions about the concept of genuine and posed facial displays and their requirements as participants. The instructions were as follows: "People sometimes express genuine facial displays caused by actual emotional experiences, while some people can express posed facial displays of emotion by intentional manipulation. In this study, we aim to understand whether people have the ability to detect these two types expressions accurately or not. There are two tasks we would like you to do. The first is to decide whether or not the expressions presented to you are showing each emotion, and the second is to decide whether or not the person depicted is feeling each emotion."

After completing the instructions, all participants did a practice trial with facial stimuli not used in the main trial (semi-spontaneous anger, fear and posed anger, fear and a neutral stimulus). The facial stimuli for this rehearsal were made by a research assistant who was unaffiliated with the study. When participants completed the practice trial, the research assistant confirmed that participants understood the task. If there were no problems, the main trial was initiated. However, if there were issues understanding the task, participants were reminded of the instructions and asked to practice the trial again.

Statistical Analysis

Although McLellan et al. (2010) conducted two analyses for the sensitivity between genuine, posed, and neutral facial displays utilizing only stimuli of posed displays, our study focused only on the comparison between genuine and posed displays as the target phenomenon for experiment, as well as for the sake of clarity. Yes-or-no answers to the facial displays were analyzed using a signal detection method that allows for separate modeling of the sensitivity and response criterion. Additionally, population-level sensitivity and the response criterion were estimated using a Bayesian hierarchical model (Rouder et al., 2007; Vuorre, 2017). In the vein of a generalized linear mixed model (Wright and London, 2009), our model (including a predictor) can be described as follows:

$$y_{ij} \sim \text{Bernoulli}(p_{ij})$$

$$\Phi(p_{ij}) = B_{0j} + B_{1j} * \text{Display}_{ij} + B_2 * \text{Presentation}_{ij}$$
$$+ B_3 * \text{Display}_{ij} * \text{Presentation}_{ij}$$

The outcomes y_{ij} were 1 if participant j responded "Yes" on trial i, and 0 if they responded "No". Also, the outcomes for participant j and trial i were Bernoulli distributed with

probability p_{ij}. The probability was transformed into z-scores with Φ which represented the cumulative normal density function. B_0 described the response criterion that corresponded to the tendency to answer "Yes" or "No", and B_1 described the sensitivity to facial displays. B_2 described the difference in response criterion between dynamic and static presentations, and B_3 described sensitivity. The sensitivity of the feel condition could be interpreted as the discriminability of emotional experiences in facial displays. Also, due to the assumed shortage of signal to be detected, B_1 in the show condition could be interpreted

as the frequency of emotional concept recognition for genuine versus posed facial displays. To estimate the population-levels parameters for B_0 and B_1, multivariate normal distribution with means and a covariance matrix for the parameters are described in the following expression:

$$\begin{bmatrix} B_{0j} \\ B_{1j} \end{bmatrix} \sim N(\begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}, \ \Sigma)$$

The means μ_1 and μ_2 can be interpreted as the population levels response criterion and sensitivity, respectively. In the following results, analysis was performed in R (3.3.3, R Core Team, 2016) using the *brms* packages (Bürkner, 2017). All iterations were set to 2,000 and burn in samples were set to 1,000, with the number of chains set to four. The value of Rhat for all parameters equalled 1.0, indicating convergence across the four chains.

RESULTS

Table 1 shows the percentage of Yes responses by judgment condition, presentation style, and facial displays for all emotions

TABLE 1 | List of the percentage of Yes responses that emerged in judgment conditions and facial displays.

Display type	Show condition (% yes)		Feel condition (% yes)	
	Dynamic	Static	Dynamic	Static
All emotions				
Neutral	2	12	10	17
Posed	88	78	31	59
Genuine	78	70	75	66
Amusement				
Neutral	0	2	0	0
Posed	98	91	38	80
Genuine	95	95	85	86
Surprise				
Neutral	3	2	7	2
Posed	95	70	23	47
Genuine	73	46	75	52
Disgust				
Neutral	5	32	13	46
Posed	90	86	32	75
Genuine	97	96	75	66
Sadness				
Neutral	0	13	18	18
Posed	67	66	32	34
Genuine	48	43	63	59

TABLE 2 | Estimated parameters on each condition for all emotions using a signal detection model.

Parameter	MAP	95%CI[]
Show condition		
Response criteria (Beta1)	1.06	[0.86, 1.25]
Sensitivity to display (Beta2)	−0.38	[−0.57, −0.16]
Response criteria between presentations (Beta3)	0.46	[0.07, 0.82]
Sensitivity to display between presentations (Beta4)	−0.13	[−0.56, 0.25]
Feel condition		
Response criteria (Beta1)	−0.12	[−0.25, −0.01]
Sensitivity to display (Beta2)	0.66	[0.50, 0.85]
Response Criteria between presentations (Beta3)	−0.72	[−0.99, −0.49]
Sensitivity to display between presentations (Beta4)	1.03	[0.64, 1.35]

MAP stands for Maximum a Posteriori estimate. 95% CI represents 95% credible intervals.
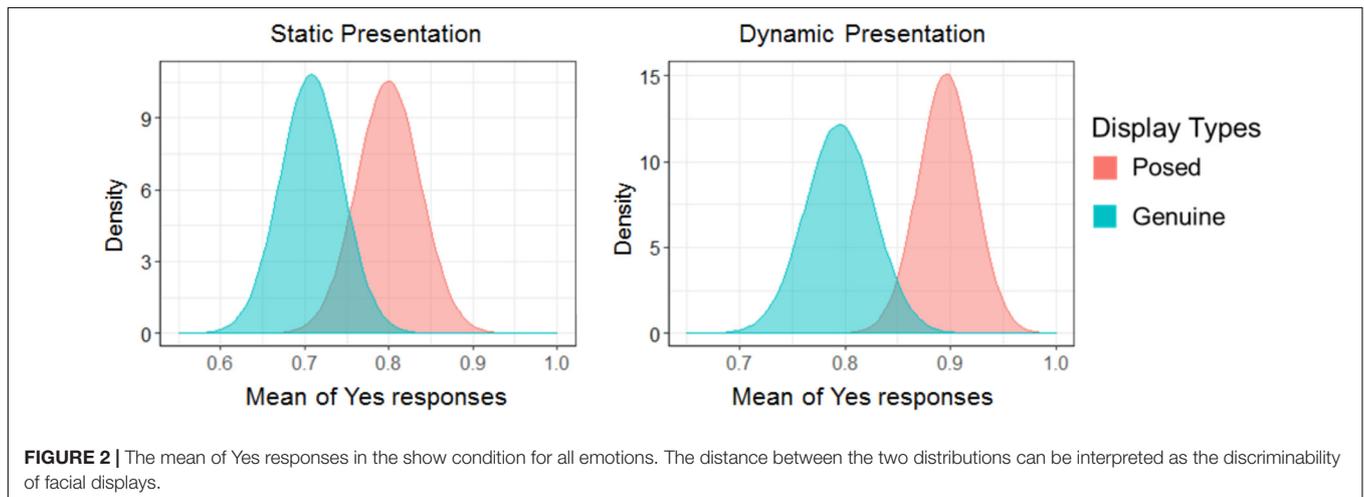
FIGURE 2 | The mean of Yes responses in the show condition for all emotions. The distance between the two distributions can be interpreted as the discriminability of facial displays.

FIGURE 3 | The mean of Yes responses in the feel condition for all emotions. The distance between the two distributions can be interpreted as the discriminability of facial displays.
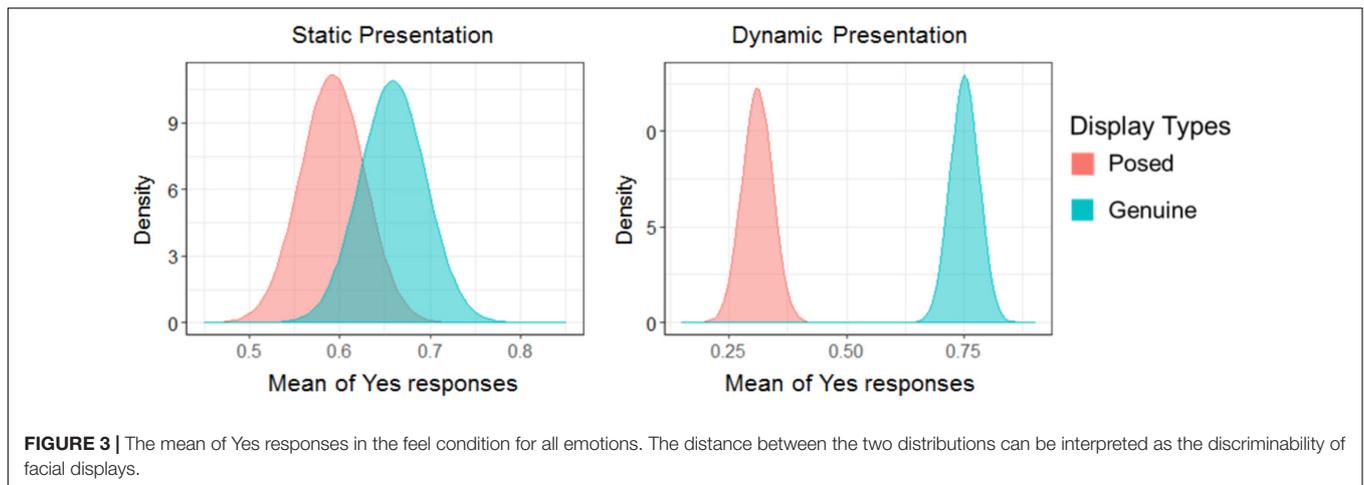
in total, as well as separated by each emotion. The following results were expected to be found according to our hypotheses: (1) genuine displays would be aligned with an answer of "Yes" in both the show and feel conditions, (2) posed displays would be answered with "Yes" for the show condition, but not the feel condition, and (3) neutral displays would be responded with "No" in both conditions. Comparisons using **Table 1** indicated several observations. For example, static presentations decreased the percentage of Yes responses in the show condition for all emotions. In the case of the feel condition, dynamic presentation promoted discriminability for all emotions. Hierarchical signal detection theory was applied in order to confirm these observations. Although results for the response criteria were also estimated, only the results for the sensitivity to displays are reported below to avoid redundancy.

The Show Condition Path to All Emotions

Figure 2 describes the percentage of Yes responses in the show condition by the type of facial displays for all emotions and presentation styles. Furthermore, results of a hierarchical signal detection method to estimate parameters for the show condition can be seen in **Table 2**. If the 95% credible interval of the parameters does not include zero, it can be inferred that there is a significant effect as in classical statistical hypothesis testing. **Table 2** shows that a negative value for the sensitivity to displays emerged, which indicates that participants responded "Yes" more frequently to posed displays than genuine displays ($\beta_1 = -0.37$ [$-0.59, -0.16$]). In other words, participants were able to differentiate genuine facial displays from posed ones. Specifically, participants judged posed displays as the facial display showing a specific target emotion more frequently than the genuine displays.

The Feel Condition Path to All Emotions

The percentage of Yes responses for the feel condition to all emotions is presented in **Figure 3**. Also, **Table 2** provides estimated parameters for the feel condition. The results for

the sensitivity to displays indicated that genuine displays cause Yes responses on the feel condition to occur more frequently than posed ones ($B_1 = 0.68$ [$0.49, 0.85$]). Moreover, the results for the sensitivity to displays between presentation styles indicated that when the presentations style was dynamic, the sensitivity to differentiate between genuine and posed ones was higher than when it was static ($B_3 = 0.98$ [$0.63, 1.34$]). Taken together, perceivers could distinguish genuine from posed facial expressions and their sensitivity was higher under the conditions that facial displays were presented dynamically, rather than statically.

TABLE 3 | Estimated parameters on show condition across each emotion using a Bayesian signal detection model.

Parameters	MAP	95%CI[]
Amusement		
Response criteria (Beta1)	1.73	[1.38, 2.36]
Sensitivity to display (Beta2)	−0.10	[−0.80, 0.48]
Response criteria between presentations (Beta3)	0.81	[0.01, 1.94]
Sensitivity to display between presentations (Beta4)	−0.80	[−2.15, 0.29]
Surprise		
Response criteria (Beta1)	1.10	[0.79, 1.44]
Sensitivity to display (Beta2)	−0.78	[−1.25, −0.44]
Response criteria between presentations (Beta3)	1.15	[0.55, 1.84]
Sensitivity to display between presentations (Beta4)	−0.34	[−1.25, 0.30]
Disgust		
Response criteria (Beta1)	1.18	[0.90, 1.49]
Sensitivity to display (Beta2)	0.69	[0.19, 1.24]
Response criteria between presentations (Beta3)	0.13	[−0.39, 0.82]
Sensitivity to display between presentations (Beta4)	−0.13	[−1.34, 0.86]
Sadness		
Response criteria (Beta1)	0.43	[0.19, 0.65]
Sensitivity to display (Beta2)	−0.53	[−0.86, −0.21]
Response criteria between presentations (Beta3)	0.06	[−0.46, 0.50]
Sensitivity to display between presentations (Beta4)	0.13	[−0.55, 0.78]

MAP stands for Maximum a Posteriori estimate. 95% CI represents 95% credible intervals.

The Show Condition Across Emotions

Next, to consider the specific characteristics across different types of emotions, we investigated data from the show condition for each emotion. **Figure 4** shows the percentage of Yes responses in the show condition across emotions. In this case, we conducted a simple signal detection model that did not include a hierarchical structure to avoid model complexity and to stabilize the convergence. The estimated parameters are described in **Table 3**. For amusement, a result for the sensitivity was not found. For surprise, the value of the sensitivity to displays was negative ($\beta_1 = -0.78$ [-1.25, -0.44]). The results of sadness indicated that the value of the sensitivity to displays was negative ($B_1 = -0.53$ [-0.86, -0.21]). For disgust, the results indicated that the value of the sensitivity to displays was positive ($B_1 = 0.69$ [0.19, 1.24]). In sum, posed displays of surprise and sadness were consistent with the results for all emotions, but disgust was found to be in the opposite direction for the showing condition.

The Feel Condition Across Emotions

Finally, we provided estimated parameters using data on the feel condition across emotions. **Figure 5** shows the marginal effects on the feel condition across emotions, and **Table 4** lists the estimated parameters. For amusement, the result for the sensitivity indicated the same directions as the parameters for the feel condition and all emotions ($B_1 = 0.80$ [0.42, 1.15]; $B_3 = 1.13$ [0.39, 1.87]). For surprise, the results were consistent with the parameters in the path to all emotions ($B_1 = 0.80$ [0.41, 1.09]; $B_3 = 1.30$ [0.61, 1.96]). For disgust, the results indicated that the values of the two types of sensitivity to displays were positive ($B_1 = 0.45$ [0.12, 0.78]; $B_3 = 1.40$ [0.71, 2.08]). The results for sadness indicated that the sensitivity to displays was positive ($B_1 = 0.72$ [0.41, 1.06]). Subsequently, all results across emotions found that participants judged the genuine displays as the facial display where the person on-screen was experiencing the specific target emotion, rather than posed displays. Furthermore, when participants differentiated the genuine and posed facial

FIGURE 4 | The mean of Yes responses on the show condition across each emotion. The distance between the two distributions can be interpreted as the discriminability of facial displays. Comparisons are shown for each specific emotion: **(A)** amusement, **(B)** surprise, **(C)** disgust, and **(D)** sadness.
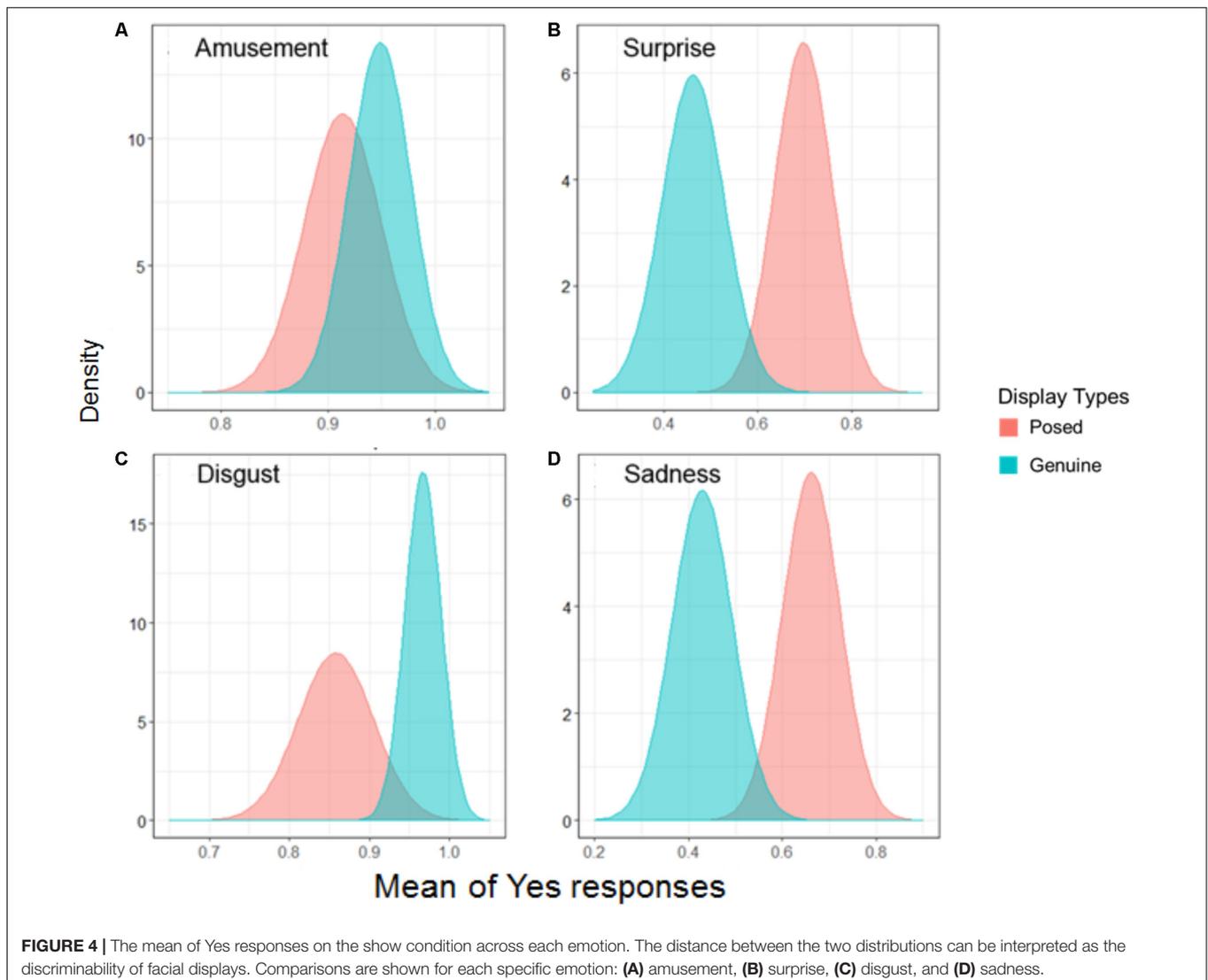
TABLE 4 | Estimated parameters on feel condition across each emotion using a Bayesian signal detection model.

Parameters	MAP	95%CI[]
Amusement		
Response criteria (Beta1)	0.26	[0.04, 0.52]
Sensitivity to display (Beta2)	0.80	[0.42, 1.15]
Response criteria between presentations (Beta3)	−1.16	[−1.66, −0.66]
Sensitivity to display between presentations (Beta4)	1.13	[0.39, 1.87]
Surprise		
Response criteria (Beta1)	−0.40	[−0.64, −0.15]
Sensitivity to display (Beta2)	0.80	[0.41, 1.09]
Response criteria between presentations (Beta3)	−0.68	[−1.14, −0.19]
Sensitivity to display between presentations (Beta4)	1.30	[0.61, 1.96]
Disgust		
Response criteria (Beta1)	0.10	[−0.14, 0.34]
Sensitivity to display (Beta2)	0.42	[0.12, 0.78]
Response criteria between presentations (Beta3)	−1.16	[−1.64, −0.65]
Sensitivity to display between presentations (Beta4)	1.42	[0.71, 2.08]
Sadness		
Response criteria (Beta1)	−0.46	[−0.69, −0.20]
Sensitivity to display (Beta2)	0.72	[0.41, 1.06]
Response criteria between presentations (Beta3)	−0.08	[−0.54, 0.42]
Sensitivity to display between presentations (Beta4)	0.16	[−0.48, 0.84]

MAP stands for Maximum a Posteriori estimate. 95% CI represents 95% credible intervals.

displays in terms of the existence of emotional experiences for amusement, surprise, and disgust, dynamic presentations notably increased the sensitivity to displays compared to static ones.

DISCUSSION

The present study investigated whether or not people can distinguish between genuine and posed facial displays of emotion by focusing on dynamic or static presentation styles. The results indicated three key findings. First, people judged posed displays as showing surprise and sadness more than the genuine displays. Second, the results of the feel condition disambiguated that people distinguish between genuine and posed facial displays of emotion in terms of their estimation that the experiences were authentically felt. Finally, the study found that perceivers are more capable of differentiating whether expressers are having a felt emotional experience when dynamic facial display processes are present over static ones.

Judging Whether the Specific Emotion Was Being Shown

This study clarified the characteristics of genuine and posed displays, with the latter being recognized as the facial display showing a specific target emotion (described in **Figure 2**). This result is consistent with several previous studies in which the percentages of observers matching the predicted emotion to posed facial displays were considerably higher than spontaneous ones (e.g., Motley and Camden, 1988;

Naab and Russell, 2007; Calvo and Nummenmaa, 2016). This result suggests that posed facial expressions are vital to the process of conveying an emotion, but that their utility does not manifest itself evenly for all emotions. For amusement, there were no differences between spontaneous and posed displays when it came to whether the target emotion was being shown. Motley and Camden (1988) suggested that only spontaneous facial expressions of positive emotions and not negative ones were recognizable above chance level, as is similar to the recognition of posed faces. In this case, it could be suggested that the perceptual information used to show amusement is not different between spontaneous and posed displays. For disgust, the results of the present study indicated that when judging the show condition for a target emotion genuine displays did so more frequently than posed displays, as described in **Figure 4**. Facial expressions of disgust function to convey potential threats like biological factors directly linked to death to an interlocutor (Tybur et al., 2013), and it is therefore possible that spontaneous expressions might contain the perceptual information to convey disgust more clearly than posed expressions.

Judging Whether the Specific Emotion Was Being Felt

The current study revealed that perceivers possess a sensitivity to facial displays that is related to the accurate inference of the emotional experiences from genuine, but not posed, facial expressions. As shown in section "The Feel Condition Across Emotions," this study observed no difference in this discriminability across emotions. Considering that there was a difference among emotions in show condition, this result is impressive. The ability to detect emotional experiences in facial expressions might be more important or more general for successful social interactions than the ability to detect the mere showing of an emotion. Both genuine and posed facial expressions can be regarded as means to express the internal state of the person signaling, that in turn directs the behavior of the observer, establishes a representation of the world for the expresser to draw from, and allows them to commit to future courses of action (Scarantino, 2017; Van Kleef, 2017). The difference between the two expressions is the endogenous nature of emotional experiences, which can be connected to the trustworthiness of the message in facial displays. From the perspective of the biological and evolutionary function of social emotions, people respond sensitively to signals with high credibility and emotional salience (Niedenthal and Brauer, 2012). Therefore, the results of this study extend the literature from previous studies consistent with the hypothesis that people can discern genuine and posed facial displays (McLellan et al., 2010; Dawel et al., 2015). However, there are small differences between previous findings and our results. Previous studies suggested that the sensitivity for emotional experiences to facial displays was specific across each emotion rather than a generalized skill, but we found that specificity for the types of emotion disappeared when non-social spontaneous facial expressions were used as genuine facial stimuli. Therefore, our results offer evidence that people might have a general

FIGURE 5 | The mean of the Yes response on the feel condition across each emotion. The distance between the two distributions can be interpreted as the discriminability of facial displays. Comparisons are shown for each specific emotion: **(A)** amusement, **(B)** surprise, **(C)** disgust, and **(D)** sadness.
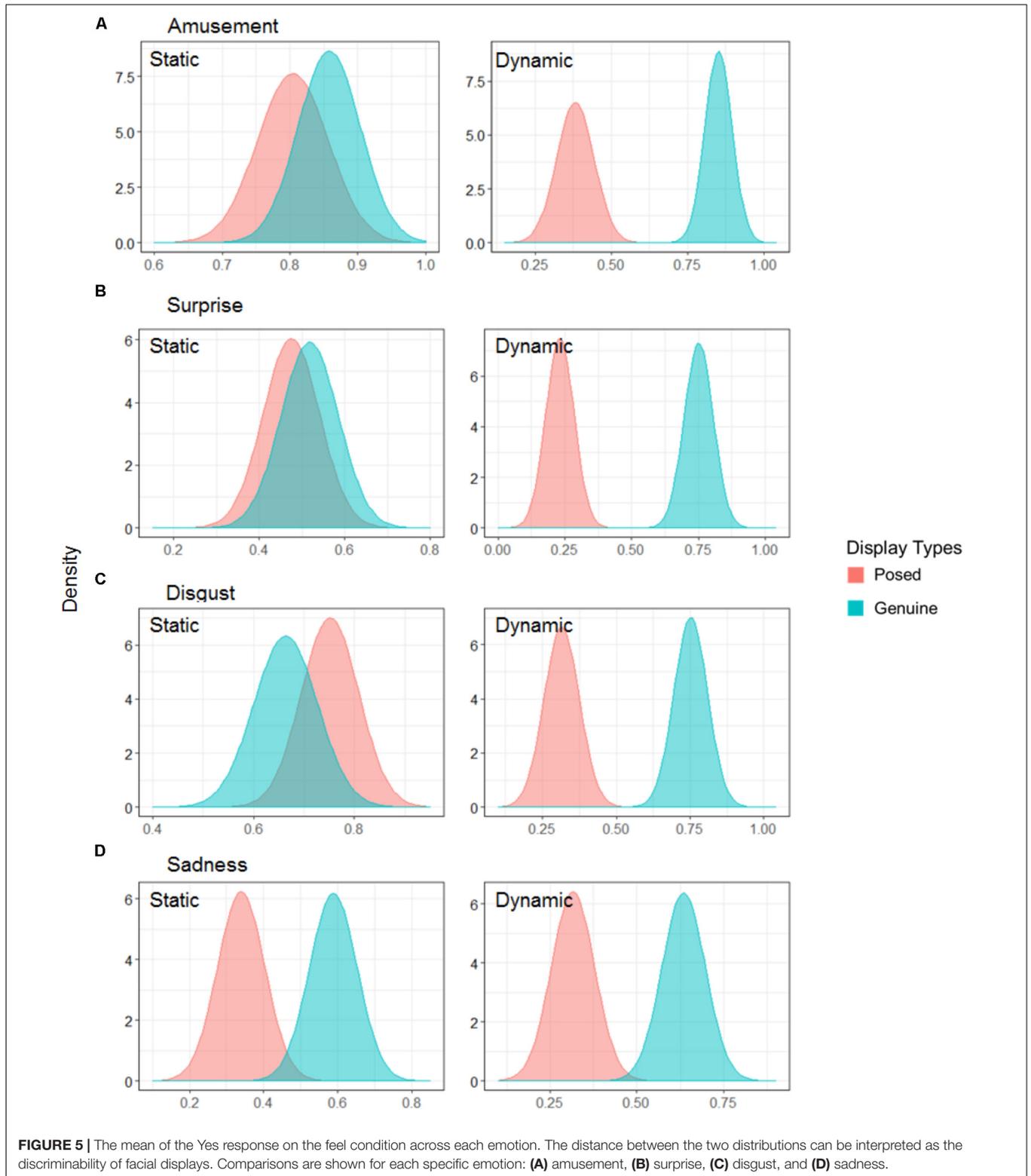
discriminability that allows them to differentiate between genuine and posed displays when it comes to perceiving felt emotional content in an expresser. Moreover, the facial stimuli presented in this study were morphologically distinct between genuine and posed facial displays, as suggested by Namba et al. (2017c). The accurate inference of emotional experience may be due to differences in morphological features, but not intensity.

Dynamic Information Related to the Sensitivity to Facial Displays

Interestingly, the signal detection model in the present study provides empirical support for the idea that sensitivity for the perception of emotional experiences to displays depends upon whether the presentation style is dynamic or static. As suggested in previous studies, this finding indicates that dynamic facial displays simply offer more information for a perceiver to parse the emotional experience of the expresser (Krumhuber and Manstead, 2009; Krumhuber et al., 2013), due to a tradeoff in the amount of information available in dynamic interactions as compared to static interactions. Ambadar et al. (2005) also showed the advantage of dynamic presentation in an emotion recognition task as one that captures the intrinsic temporal quality of an unfolding expression rather than mere increases in static facial frames. Our study did not compare dynamic and multi-static stimuli, but did show that non-linear motion of spontaneous expressions might raise ecological validity, and suggested that such situations could increase the discriminability of the expresser's experiences of emotions like surprise and disgust. Our findings could also imply that further research related to the perception of emotional experiences in facial expressions, such as those in the realm of emotional contagion (Hatfield et al., 2014), might benefit from using dynamic genuine facial expressions as stimuli because the standardized practice of presenting static stimuli may play a role in the lack of detection of emotional experiences from facial displays.

Limitations and Future Studies

While this study showed that people can distinguish between genuine and posed facial displays of emotion and that this sensitivity depends on whether the facial displays unfolded dynamically or not, several limitations should be noted. First, a signal detection model using binary reactions allowed for the provision of response criteria in addition to sensitivity. However, Dawel et al. (2017) indicated that the yes-or-no response provides far less information than a rating scale about the relative perceived genuineness of different stimuli. Therefore, additional studies should be conducted using rating scales, such as a neutral midpoint scale (e.g., perceived genuineness: $-7 = completely fake$; $0 = don't know$; $+7 = completely genuine$, Dawel et al., 2017).

Next, the results of this study should be interpreted for only the four emotions investigated: amusement, surprise, disgust, and sadness. Although fearful displays have typically been used in previous studies (McLellan et al., 2010; Douglas et al., 2012; Dawel et al., 2015), the current study did not examine fear due to the lack of evidence in the domain of spontaneous facial expressions of fear. Also, other emotions such as happiness (McLellan et al., 2010), anger and contempt (Fischer and Roseman, 2007) should be considered to extend the evidence base of these findings to future studies.

In addition, while a Bayesian probit regression procedure-based signal detection model was able to produce these results, a larger sample size of study participants and facial stimuli could provide for a more robust understanding of the effects and allow for separate analyses of each emotion of interest through a generalized linear model. The data from the present study will be appended as Supplementary Material so that researchers can access it as open data and further examine or build upon the evidence base in future collaborative research projects or novel statistical approaches.

Finally, we used spontaneous facial expressions that were secretly recorded to avoid the effects of intentional manipulation. Although these facial stimuli can allow for fine-grained understanding of the sensitivity to facial displays to be explored, such stimuli cannot control other subtle non-verbal cues like head or eye movements. We avoided stimuli that included these features as much as possible during the facial stimuli selection stage, but it is difficult to control for these subtle actions in a non-social experimental environment designed to capture genuine facial stimuli. To overcome these barriers, further studies might consider the use of computerized facial expressions (Jack et al., 2014; Krumhuber and Scherer, 2016), as it may be possible to conduct research controlling small movements in facial stimuli by letting an avatar load the genuine displays.

CONCLUSION

The current study revealed that people are capable of distinguishing genuine from posed facial expressions by judging whether the target emotion was being shown and felt by the expresser. Specifically, posed displays were more frequently judged as the facial expressions showing specific emotions of surprise and sadness than genuine displays, whereas genuine displays were evaluated as the felt expressions of a target emotion in the case of amusement, surprise, disgust and sadness. Additionally, variability in the discriminability of authentic experiences was examined and found to depend on whether the facial display was dynamically or statically presented. The sensitivity to detect emotional experiences of amusement, surprise, and disgust was lower in the statically presented facial expressions, whereas dynamic information enhanced the discriminability for observers to detect the emotional experiences of others depicted in facial displays. Still, as the perception of facial expressions depends heavily on the surrounding context, it will be necessary to corroborate these findings with data from many other investigations. We hope that these distinctions on the type of stimuli presented and their characteristics can be taken into consideration by future researchers interested in the domain of emotional facial expressions and their properties.

ETHICS STATEMENT

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee.

AUTHOR CONTRIBUTIONS

SN conducted the research, statistical analysis, and served as the primary author of the manuscript. RK offered revisions,

summary, and literature review. MM and TN contributed to confirmation of the research protocol, further review of methods and analysis, and feedback on the manuscript.

REFERENCES

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Buck, R., and VanLear, C. A. (2002). Verbal and nonverbal communication: distinguishing symbolic, spontaneous, and pseudo-spontaneous non-verbal behavior. *J. Commun.* 52, 522–541. doi: 10.1111/j.1460-2466.2002.tb02560.x

Bürkner, P. C. (2017). brms: an R package for bayesian multilevel models using Stan. *J. Stat. Softw.* 80, 1–28. doi: 10.18637/jss.v080.i01

Calvo, M. G., and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Carroll, J. M., and Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *J. Pers. Soc. Psychol.* 72, 164–176. doi: 10.1037/0022-3514.72.1.164

Ceccarini, F., and Caudek, C. (2013). Anger superiority effect: the importance of dynamic emotional facial expressions. *Vis. Cogn.* 21, 498–540. doi: 10.1080/13506285.2013.807901

Dawel, A., Palermo, R., O'Kearney, R., and McKone, E. (2015). Children can discriminate the authenticity of happy but not sad or fearful facial expressions, and use an immature intensity-only strategy. *Front. Psychol.* 6:462. doi: 10.3389/fpsyg.2015.00462

Dawel, A., Wright, L., Irons, J., Dumbelton, R., Palermo, R., O'Kearney, R., et al. (2017). Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behav. Res.* 49, 1539–1562. doi: 10.3758/s13428-016-0813-2

Douglas, K. M., Porter, R. J., and Johnston, L. (2012). Sensitivity to posed and genuine facial expressions of emotion in severe depression. *Psychiatry Res.* 196, 72–78. doi: 10.1016/j.psychres.2011.10.019

Ekman, P. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage.* New York, NY: W. W. Norton & Co.

Ekman, P. (2003). *Emotions Revealed.* New York, NY: Times Books.

Fischer, A. H., and Roseman, I. J. (2007). Beat them or ban them: the characteristics and social functions of anger and contempt. *J. Pers. Soc. Psychol.* 93, 103–115. doi: 10.1037/0022-3514.93.1.103

Gosselin, P., Kirouac, G., and Doré, F. Y. (1995). Components and recognition of facial expression in the communication of emotion by actors. *J. Pers. Soc. Psychol.* 68, 83–96. doi: 10.1037/0022-3514.68.1.83

Gosselin, P., Perron, M., and Beaupré, M. (2010). The voluntary control of facial action units in adults. *Emotion* 10, 266–271. doi: 10.1037/a0017748

Gross, J. J., and Levenson, R. W. (1995). Emotion elicitation using films. *Cogn. Emot.* 9, 87–108. doi: 10.1080/02699939508408966

Guo, H., Zhang, X. H., Liang, J., and Yan, W. J. (2018). The dynamic features of lip corners in genuine and posed smiles. *Front. Psychol.* 9:202. doi: 10.3389/fpsyg.2018.00202

Hatfield, E., Bensman, L., Thornton, P. D., and Rapson, R. L. (2014). New perspectives on emotional contagion: a review of classic and recent research on facial mimicry and contagion. *Interpers. Int. J. Pers. Relat.* 8, 159–179. doi: 10.5964/ijpr.v8i2.162

Jack, R. E., Garrod, O. G., and Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* 24, 187–192. doi: 10.1016/j.cub.2013.11.064

Johnston, L., Miles, L., and Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *Br. J. Soc. Psychol.* 49, 107–127. doi: 10.1348/014466609X412476

Keltner, D., and Haidt, J. (2001). "Social functions of emotions," in *Emotions and Social Behavior. Emotions: Current Issues and Future Directions*, eds T. J. Mayne and G. A. Bonanno (New York, NY: Guilford Press), 192–213.

Krumhuber, E. G., Kappas, A., and Manstead, S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., and Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 807–820. doi: 10.1037/a0017844

Krumhuber, E. G., and Scherer, K. R. (2016). The look of fear from the eyes varies with the dynamic sequence of facial actions. *Swiss J. Psychol.* 75, 5–14. doi: 10.1024/1421-0185/a000166

Kunzmann, U., Kupperbusch, C. S., and Levenson, R. W. (2005). Behavioral inhibition and amplification during emotional arousal: a comparison of two age groups. *Psychol. Aging* 20, 144–158. doi: 10.1037/0882-7974.20.1.144

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., et al. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cereb. Cortex* 25, 246–257. doi: 10.1093/cercor/bht227

McLellan, T. L., Johnston, L., Dalrymple-Alford, J., and Porter, R. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cogn. Emot.* 24, 1277–1292. doi: 10.1016/j.neulet.2012.10.039

McLellan, T. L., Wilcke, J. C., Johnston, L., Watts, R., and Miles, L. K. (2012). Sensitivity to posed and genuine displays of happiness and sadness: a fMRI study. *Neurosci. Lett.* 531, 149–154. doi: 10.1016/j.neulet.2012.10.039

Motley, M. T., and Camden, C. T. (1988). Facial expression of emotion: a comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *West. J. Speech Commun.* 52, 1–22. doi: 10.1080/10570318809389622

Naab, P. J., and Russell, J. A. (2007). Judgments of emotion from spontaneous facial expressions of New Guineans. *Emotion* 7, 736–744. doi: 10.1037/1528-3542.7.4.736

Namba, S., Kabir, R. S., Miyatani, M., and Nakao, T. (2017a). Spontaneous facial actions map onto emotional experiences in a non-social context: toward a component-based approach. *Front. Psychol.* 8:633. doi: 10.3389/fpsyg.2017.00633

Namba, S., Kagamihara, T., Miyatani, M., and Nakao, T. (2017b). Spontaneous facial expressions reveal new action units for the sad experiences. *J. Nonverbal Behav.* 41, 203–220. doi: 10.1007/s10919-017-0251-6

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., and Nakao, T. (2017c). Spontaneous facial expressions are different from posed facial expressions: morphological properties and dynamic sequences. *Curr. Psychol.* 36, 593–605. doi: 10.1007/s12144-016-9448-9

Niedenthal, P. M., and Brauer, M. (2012). Social functionality of human emotion. *Annu. Rev. Psychol.* 63, 259–285. doi: 10.1146/annurev.psych.121208.131605

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Reed, L. I., and DeScioli, P. (2017). The communicative function of sad facial expressions. *Evol. Psychol.* 15, 1–9. doi: 10.1177/1474704917700418

Rouder, J. N., Lu, J., Sun, D., Speckman, P., Morey, R. D., and Naveh-Benjamin, M. (2007). Signal detection models with random participant and item effects. *Psychometrika* 72, 621–642. doi: 10.1007/s11336-005-1350-6

Sato, W., Noguchi, M., and Yoshikawa, S. (2007). Emotion elicitation effect of films in a Japanese sample. *Soc. Behav. Pers. Int. J.* 35, 863–874. doi: 10.2224/sbp.2007. 35.7.863

Sato, W., and Yoshikawa, S. (2007). Enhanced experience of emotional arousal in response to dynamic facial expressions. *J. Nonverbal Behav.* 31, 119–135. doi: 10.1007/s10919-007-0025-7

Scarantino, A. (2017). Do emotions cause actions, and if so how? *Emot. Rev.* 9, 326–334. doi: 10.1177/1754073916679005

Scherer, K. R., and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* 7, 113–130. doi: 10.1037/1528-3542.7.1.11

Smoski, M., and Bachorowski, J.-A. (2003). Antiphonal laughter between friends and strangers. *Cogn. Emot.* 17, 327–340. doi: 10.1080/02699930302296

Suzuki, K., and Naitoh, K. (2003). Useful information for face perception is described with FACS. *J. Nonverbal Behav.* 27, 43–55. doi: 10.1023/A: 1023666107152

Tybur, J. M., Lieberman, D., Kurzban, R., and DeScioli, P. (2013). Disgust: evolved function and structure. *Psychol. Rev.* 120, 65–84. doi: 10.1037/a0030778

Van Kleef, G. A. (2017). The social effects of emotions are functionally equivalent across expressive modalities. *Psychol. Inq.* 28, 211–216. doi: 10.1080/1047840X. 2017.1338102

Vuorre, M. (2017). *Bayesian Estimation of Signal Detection Models, Part 2.* Available at: https://vuorre.netlify.com/post/2017/bayesian-estimation-of-signal-detection-theory-models-part-2/

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Pers. Soc. Psychol.* 78, 105–119. doi: 10.1037/0022-3514.78.1.105

Wright, D. B., and London, K. (2009). Multilevel modelling: beyond the basic applications. *Br. J. Math. Stat. Psychol.* 62, 439–456. doi: 10.1348/ 000711008X327632

Brain Responses to Dynamic Facial Expressions: A Normative Meta-Analysis

Oksana Zinchenko [1], Zachary A. Yaple [1,2] and Marie Arsalidou [3,4]*

[1] Centre for Cognition and Decision Making, Institute for Cognitive Neuroscience, National Research University Higher School of Economics, Moscow, Russia, [2] Department of Psychology, National University of Singapore, Singapore, Singapore, [3] Department of Psychology, National Research University Higher School of Economics, Moscow, Russia, [4] Department of Psychology, York University, Toronto, ON, Canada

Identifying facial expressions is crucial for social interactions. Functional neuroimaging studies show that a set of brain areas, such as the fusiform gyrus and amygdala, become active when viewing emotional facial expressions. The majority of functional magnetic resonance imaging (fMRI) studies investigating face perception typically employ static images of faces. However, studies that use dynamic facial expressions (e.g., videos) are accumulating and suggest that a dynamic presentation may be more sensitive and ecologically valid for investigating faces. By using quantitative fMRI meta-analysis the present study examined concordance of brain regions associated with viewing dynamic facial expressions. We analyzed data from 216 participants that participated in 14 studies, which reported coordinates for 28 experiments. Our analysis revealed bilateral fusiform and middle temporal gyri, left amygdala, left declive of the cerebellum and the right inferior frontal gyrus. These regions are discussed in terms of their relation to models of face processing.

Keywords: dynamic faces, fMRI meta-analysis, activation likelihood estimate, social cognition, facial expressions

INTRODUCTION

Effective face processing is essential for perceiving and recognizing intentions, emotion and mental states in others. Facial expressions have traditionally been investigated by utilizing static pictures of faces as opposed to dynamic moving faces (i.e., short video clips). Faces elicit activity in an established set of brain areas that includes the fusiform gyri associated with face perception, amygdala associated with processing affect and fronto-temporal regions associated with knowledge of a person (Fusar-Poli et al., 2009 for meta-analyses). Some suggest that dynamic faces compared to static faces are more ecologically valid (Bernstein and Yovel, 2015), and facilitate recognition of facial expressions (Ceccarini and Caudek, 2013). O'Toole et al. (2002) explain that when both static and dynamic identity information are available, people tend to rely primarily on static information for face recognition (i.e., supplemental information hypothesis), whereas dynamic information such as motion contributes to the quality of the structural information accessible from a human face (representation enhancement hypothesis). This dynamic information plays a key role in social interactions when evaluating the mood or intentions of others (Langton et al., 2000; O'Toole et al., 2002). The brain areas that respond to dynamic faces are not fully characterized with up-to-date meta-analysis methods and findings in the field. The purpose of this study is to examine concordance in brain regions associated with dynamic facial expressions using quantitative meta-analysis.

Functional magnetic resonance imaging (fMRI) studies investigating face perception typically reveal activation within the fusiform gyrus and occipital gyrus, areas part of the core regions of face processing, which mediate visual analysis of faces (O'Toole et al., 2002; Gobbini and Haxby, 2007). The extended system associated with extracting meaning from faces includes the inferior frontal cortex and amygdalae (Haxby et al., 2000). Notably, compared to static faces, much fewer fMRI studies use dynamic face stimuli, likely due to methodological and practical challenges in using dynamic faces. Specifically, short videos of faces need to be standardized in terms of presentation speed (i.e., how fast a neutral face transforms to an emotional expression), as this requires consistency across emotions. Similarly, morphed faces are modified to transform a static photo from a neutral to an emotional expression in a series of frames. Thus, adopting a protocol for using dynamic facial expressions (e.g., videos and morphs) requires more computational processing and in turn more time to prepare.

These additional efforts, however, have been found to be beneficial in populations that have an altered sensitivity to faces. For example, research shows that regions related to visual properties (i.e., the core system) and emotional/cognitive processing of faces (i.e., the extended system) are hypoactive in patients with autism spectrum disorders (Hadjikhani et al., 2007; Bookheimer et al., 2008; Nomi and Uddin, 2015 for review). Dynamic changes in facial expressions were used to show that individuals with and without autism spectrum disorders elicit equivalent activity in occipital regions, and differential activity in the fusiform gyrus, amygdala and superior temporal sulcus, suggesting a dysfunction in the relational and affective processing of faces (Pelphrey et al., 2007). Thus, in practice, usage of dynamic stimuli would be advantageous when studying populations with difficulties in processing faces and emotions.

A recent review of the face perception literature adopted the model of core and extended systems to explain processing of dynamic faces in typical adults (Bernstein and Yovel, 2015). This review provides support for a dorsal stream that encompasses the superior temporal sulcus, and encodes low-frequency information such as face motion, head rotation and processing of moving facial parts (O'Toole et al., 2002; Peyrin et al., 2004, 2005, 2010; Saxe, 2006), and a ventral stream that comprises bilateral inferior occipital cortex and fusiform gyrus, and processes high-frequency information such as facial expressions and face parts (e.g., Eger et al., 2004; Iidaka et al., 2004; Corradi-Dell'Acqua et al., 2014). Since the dorsal stream processes more information about movement of faces, dynamic facial expressions should involve more activation of the superior temporal lobe.

An early meta-analysis analyzed coordinates from 11 experiments on dynamic facial expressions and identified concordance in temporal, parietal, and frontal cortices (Arsalidou et al., 2011). Since then, there has been an increase in the number of fMRI studies that examine brain responses to dynamic faces. Critically, there have been methodological advances to the activation likelihood estimation (ALE) method (Turkeltaub et al., 2012) and documented implementation errors in the old ALE software that have since been corrected (Eickhoff et al., 2017); ALE software developers recommend

re-analyses and evaluation of current and past meta-analyses. Thus, the purpose of the current paper was to examine brain areas associated with processing of dynamic facial expressions in healthy adults and establish their implication above and beyond to brain areas responding to static faces and other control tasks.

METHODS

Literature Search and Article Selection
A literature search was performed using Web of Science (http://apps.webofknowledge.com/) on October, 6th, 2017, keywords ("dynamic faces" OR "facial motion" AND "fMRI"), years 1995–2017, yielding a total of 114 articles. **Figure 1** shows the steps taken to identify eligible articles. Specifically, we excluded articles that: (1) reported no fMRI data; (2) studies that did not report whole brain analysis; (3) reported no data on healthy adults; (4) did not report fMRI coordinates and, (5) articles with irrelevant tasks. Articles surviving these criteria underwent a full text review by two researchers independently (O.Z. and Z.Y.). The remaining articles included healthy adults; reported stereotaxic coordinates in Talairach or Montreal Neurological Institute (MNI) space from random effects whole-brain analysis, which reported a contrast (i.e., experiment) comparing dynamic with static faces. Articles from a previous meta-analysis and an eligible study within it (Arsalidou et al., 2011) resulted in 7 additional articles. All relevant experiments from each article were included in the analysis because the most recent algorithm uses a correction to avoid summation of within-group effects and provides increased power (Turkeltaub et al., 2012). **Table 1** shows participant demographics and details from a total 28 experiments from 14 articles, sorted by 15 separate subject groups, which were included in the meta-analysis. The number of experiments we included in the analysis adheres to current recommendations ($n = 17$–20) for achieving sufficient statistical power (Eickhoff et al., 2017).

Meta-Analysis
The meta-analysis was performed using GingerALE software (2.3.6), which relies on ALE, a coordinate-based meta-analytic method (Eickhoff et al., 2009, 2017) available at http://www.brainmap.org/ale/. Foci from different articles were used to create a probabilistic map that compares the likelihood of activation compared to random spatial distribution. MNI coordinates were converted to Talairach space using the Lancaster et al. (2007) transformation. Significance was assessed using a cluster-level threshold for multiple comparisons at $p = 0.05$ with a cluster-forming threshold set to $p = 0.001$ (Eickhoff et al., 2012, 2017). GingerALE software does not provide an option for estimating replicability of the data, however, based on simulations of ALE analyses that have been performed to test sensitivity, number of incidental clusters and statistical power (Eickhoff et al., 2016), a recommended minimum number of experiments ($N = 17$–20) has been proposed (Eickhoff et al., 2017). Moreover, a cluster-level threshold sets the cluster minimum volume such that only, for example, 5% of the simulated data clusters exceed this size, minimizing the possibility that an ALE peak could be driven by only one study.

FIGURE 1 | PRISMA flowchart for eligibility of articles (Template by Moher et al., 2009).
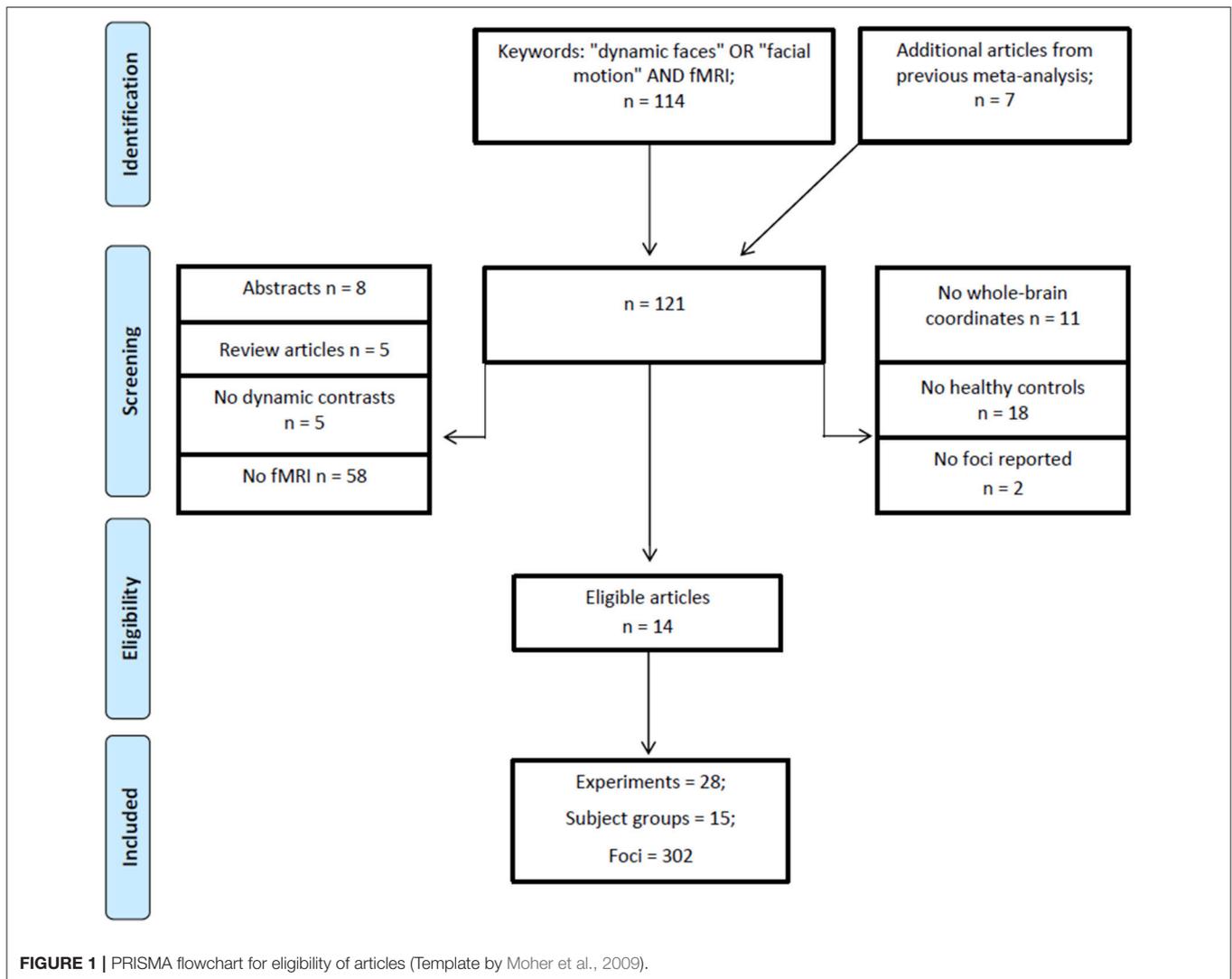
The majority of studies used tasks where participants were instructed to passively observe facial stimuli (Sato et al., 2004; Trautmann et al., 2009; Pentón et al., 2010; Arsalidou et al., 2011) or to perform a simple target detection task (Pelphrey et al., 2007; Robins et al., 2009; Lee et al., 2010; Sato et al., 2015). Two studies asked to rank the presented emotional expressions (Grosbras and Paus, 2006; Sarkheil et al., 2013); three studies instructed the participants to make a decision about the gender of face stimuli (Hurlemann et al., 2008; Pentón et al., 2010; Ceccarini and Caudek, 2013); one study asked to rank the meaningfulness of moving faces and judge the fluidity of facial motions (Schultz et al., 2013); in another study participants were told to identify the category of face stimuli (LaBar et al., 2003); and in another study participants performed a one-back matching task (Schultz and Pilz, 2009). Five articles reported experiments related to dynamic > static in various emotions: anger (LaBar et al., 2003; Grosbras and Paus, 2006), fear (Sato et al., 2004), and happiness (Sato et al., 2004; Trautmann et al., 2009; Arsalidou et al., 2011). Six articles presented participants with dynamic > static faces after

subtracting neutral from emotional faces in one (Hurlemann et al., 2008), several (Pelphrey et al., 2007; Robins et al., 2009; Schultz and Pilz, 2009), or no emotional component (Lee et al., 2010; Pentón et al., 2010). One article reported experiments regarding the morph intensity effect in dynamic faces (Sarkheil et al., 2013), and two articles contrasted dynamic faces to mosaic stimuli (Sato et al., 2015; we note that this study reported fMRI coordinates using magnetic encephalography-fMRI data reconstruction) or scrambled faces (Schultz et al., 2013).

RESULTS

Analyses included data from 216 right-handed participants (27.24 ± 9.02 years; 39.81% men, **Table 1** for details).

ALE Map

The largest cluster with the highest ALE value was found in the right hemisphere and extended from the inferior temporal and occipital, to fusiform and superior temporal gyri (**Figure 2,**

TABLE 1 | Descriptive information of studies and contrasts used in the meta-analyses.

Article	n	Male	Handedness	Age range	Task	Contrast	Foci	p-value[c]
Arsalidou et al., 2011	15	2	N/A	26.3 ± 4.5	Dynamic and static happy and neutral[a]	Happy: dynamic > static	2	P < 0.01 using cluster level threshold p = 0.05
						Dynamic > static	5	P < 0.01, using cluster level threshold p = 0.05
Grosbras and Paus, 2006	20	10	R	19–46	Angry and neutral movements of faces[b]	Neutral: dynamic > control	28	P < 0.05 using Gaussian random-field theory to correct for multiple comparisons
						Angry: dynamic > control	27	P < 0.05 using Gaussian random-field theory to correct for multiple comparisons
Hurlemann et al., 2008	14	7	R	25.04 ± 2.4	Dynamic happy and angry facial animations[b]	Dynamic:emotional > neutral	10	P < 0.001, uncorrected
						Dynamic: angry > neutral	3	
						Dynamic: happy > neutral	17	
LaBar et al., 2003	10	5	R	21–30	Dynamic and static; angry and fearful[b]	Anger morph > static	6	P < 0.001, uncorrected
						Fear morph > static	16	
						Identity morph > static neutral	16	
						Emotion morph > static emotion	17	
Lee et al., 2010	17	7	R	24.94 ± 4.16	Dynamic and static turning head[a]	Turning heads > static heads	9	P < 0.05, cluster corrected,
Pelphrey et al., 2007	8	6	All R	24.1 ± 5.6	Dynamic and static; angry and fearful[a]	Dynamic emotions > static emotions (normal group)	6	P < 0.05, uncorrected
Pentón et al., 2010	13	8	N/A	19–55	Static and dynamic, neutral and fearful faces[b]	Dynamic > static	21	P< 0.05, FDR corrected
Robins et al., 2009	10	3	N/A	22.3 ± 4.6	Dynamic angry, happy, fearful, and neutral[b]	Dynamic emotion > neutral	5	P < 0.001
Sarkheil et al., 2013	20	9	R	20–42	Angry and happy morph face stimuli[b]	Intensity effect (more > less)	8	P < 0.05, cluster-size thresholding
Sato et al., 2015	15	9	R	26.9 ± 3.9	Fearful, happy, and neutral dynamic and static faces and mosaics[a]	Dynamic facial > dynamic mosaics; time 150–200	13	P < 0.05 corrected for multiple comparisons with a height threshold of P < 0.01 (uncorrected)
						Dynamic facial > dynamic mosaics; time 200–250	3	
						Dynamic facial > dynamic mosaics; time 250–300	4	
						Dynamic facial > dynamic mosaics; time 300–350	6	
						Dynamic facial > dynamic mosaics; time 350–400	5	
Sato et al., 2004	11	*	R	26.5	Dynamic fearful and neutral faces[a]	Fear: dynamic > static	18	P < 0.05
	11	*				Happy: dynamic > static	12	
Schultz and Pilz, 2009	10	6	N/A	N/A	Dynamic and static; angry and surprised[b]	Dynamic faces > static faces	6	P < 0.05, FDR-corrected and cluster-wise corrected
Schultz et al., 2013	26	14	R	22–39	Video recordings of moving faces, static faces and scrambled order of dynamic faces[b]	Movies with ordered frames > movies with scrambled frames	3	P < 0.001, uncorrected
						Original 25 Hz movies > static faces	4	

(Continued)

TABLE 1 | Continued

Article	n	Male	Handedness	Age range	Task	Contrast	Foci	p-value[c]
Trautmann et al., 2009	16	0	R	21.6 ± 2.3	Dynamic and static; happy and disgust[a]	Dynamic faces (happy > neutral) > static faces (happy > neutral)	14	P < 0.001, uncorrected
						Dynamic faces (disgust > neutral) > static faces (disgust > neutral)	18	

n = sample size; * = 22 participants (10 males) participated in two studies, gender assignment was not specified; N/A, not available; R, all right handed; [a] studies that instruct participants to passively view faces; [b] studies that instruct participants to make judgments about faces, [c] thresholding settings reported in paper.
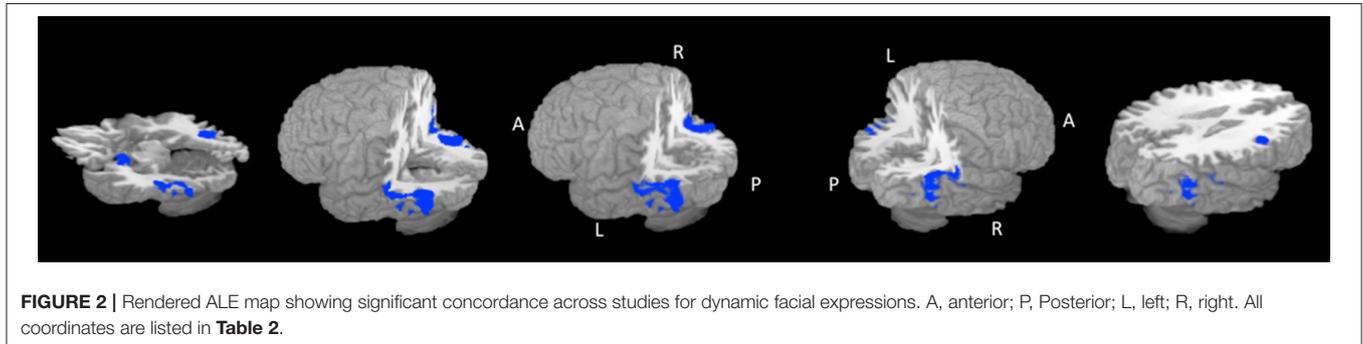
FIGURE 2 | Rendered ALE map showing significant concordance across studies for dynamic facial expressions. A, anterior; P, Posterior; L, left; R, right. All coordinates are listed in **Table 2**.

Table 2). The second cluster was found in left hemisphere and extended from the middle occipital and temporal gyri to the fusiform gyrus and cerebellum. Other areas included the left amygdala, and right inferior frontal gyrus.

DISCUSSION

We examined concordance across studies in brain areas responding more to dynamic facial expressions. We report concordance in: (a) areas associated with the core visual system of processing faces such as fusiform gyrus and posterior parts of the superior temporal gyrus, (b) areas associated with the extended system for processing faces such as the left amygdala, inferior frontal gyrus, and anterior parts of the superior temporal gyrus and (c) a cluster within the cerebellar declive, a region previously not highlighted in models of facial cognition. We build on previous models of face processing and discuss possible roles of these areas during the processing of dynamic faces.

In comparison with the previous meta-analysis on dynamic faces (Arsalidou et al., 2011); the current analysis yields similar brain regions, however the output resulted in less clusters that were larger in size and carried higher ALE values. When comparing the top clusters, the amygdala and cerebellar declive are found in the left hemisphere for both the current and previous analyses. Clusters in right precuneus (BA 7) and cuneus, and left hypothalamus, previously found to be concordant (Arsalidou et al., 2011), were not observed in the current meta-analysis; these areas had both lower ALE scores and smaller cluster volumes. We note three methodological choices that may account for differences in the current and previous meta-analyses; (a) the number of experiments included in the current meta-analyses

is larger, which provide increased power, (b) the GingerALE algorithm, which allows for controlling for within-group effects and provides increased power (Turkeltaub et al., 2012) and (c) the thresholding approach follows cluster-level threshold for controlling for multiple comparisons, which is more suitable for ALE meta-analyses (Eickhoff et al., 2016, 2017). Critically, the current meta-analysis shows that the overall size of clusters in occipito-temporal regions is similar in the right and left hemisphere, suggesting bilateral engagement.

Specifically, bilateral occipito-temporal gyri comprise of the fusiform and superior temporal gyri, areas are most associated with face processing; the fusiform gyri are implicated in configuring relations among visual features and relying on high-spatial-frequency to form face percepts as a whole (e.g., Vuilleumier et al., 2003; Iidaka et al., 2004; Sabatinelli et al., 2011), or in part (e.g., Rossion et al., 2003; Nichols et al., 2010; Yaple et al., 2016). This is consistent with models that classify the fusiform gyrus as part of the core visual processing system for faces (Gobbini and Haxby, 2007), and as part of the ventral stream of face processing (e.g., Bernstein and Yovel, 2015).

Moreover, we observe concordance in posterior and more dorsal parts of the superior temporal gyri. The superior temporal gyri are known for their involvement in the analysis of low-spatial frequency information (i.e., global facial information) such as gaze direction and motion associated with interpreting social signals (Allison et al., 2000; Taylor et al., 2009; Wegrzyn et al., 2015). According to the face perception model by Haxby and colleagues posterior parts of the superior temporal sulcus are part of the core visual face processing system responsible for basic visual analyses of faces, whereas adjacent more anterior parts of the superior temporal gyri are part of the extended system

TABLE 2 | Concordant brain regions associated with dynamic facial expressions.

Volume mm³	ALE Value	Talairach			Brain Area
		x	y	z	
9,256	0.0395	42	−64	2	Right Inferior Temporal Gyrus
	0.0366	44	−60	−6	Right Inferior Temporal Gyrus BA 19
	0.0282	40	−56	−16	Right Fusiform Gyrus BA 37
	0.0256	54	−44	4	Right Middle Temporal Gyrus BA 22
	0.0138	56	−42	18	Right Superior Temporal Gyrus BA 13
	0.0130	40	−78	−2	Right Inferior Occipital Gyrus BA 19
8,480	0.0389	−42	−70	2	Left Middle Occipital Gyrus BA 37
	0.0216	−50	−44	4	Left Middle Temporal Gyrus BA 22
	0.0199	−44	−46	−16	Left Fusiform Gyrus BA 37
	0.0197	−40	−64	−8	Left Fusiform Gyrus BA 19
	0.0190	−38	−56	−18	Left Cerebellum, Declive
	0.0185	−50	−58	4	Left Middle Temporal Gyrus BA 37
	0.0135	−46	−72	−14	Left Fusiform Gyrus BA 19
1,160	0.0188	−16	−6	−12	Left Amygdala
840	0.0234	48	4	24	Right Inferior Frontal Gyrus BA 9

that is responsible for further processing of personal information (Haxby et al., 2000; Gobbini and Haxby, 2007). Our data are also consistent with the more recent interpretation of a dorsal face processing pathway proposed by Bernstein and Yovel (2015). Importantly, consistent with the representation enhancement hypothesis (O'Toole et al., 2002) we propose that dynamic faces may show increased implication in superior temporal cortices because they provide richer input for the brain to interpret.

As part of the left occipito-temporal cluster we observed concordance in the cerebellar declive, an area not highlighted as part of face processing models. Traditionally, the cerebellum was known for its involvement in motor functioning. However, its role in cognitive and affective processing has been discussed (e.g., Brooks, 1984; Paulin, 1993; Doya, 2000; Stoodley and Schmahmann, 2010) and a generic role in timing mechanisms has been proposed (e.g., Ivry and Spencer, 2004). Past meta-analyses identify concordance in the cerebellum for static facial expressions (Fusar-Poli et al., 2009), however its role in social cognition remains unclear. In relation to social processes some have shown that the cerebellum is associated with mirroring and mentalizing motor actions (Van Overwalle et al., 2014, 2015). We suggest that the cerebellum may play a role in tracking the sequences for conveying the signal and updating the information about perceptual features in a face to predict possible changes, similar to its involvement in the motor system.

Concordance in the left amygdala and right inferior frontal gyrus is respectively associated with emotional and cognitive processing of faces. The amygdala responds to all sorts of emotional stimuli such as fear processing and fear conditioning (LeDoux, 2003), reward and punishment (Gupta et al., 2011). Growing evidence suggests that amygdala activation is not specific to fearful expressions or any particular emotion (van der Gaag et al., 2007), but rather it processes salient information

of faces (Fitzgerald et al., 2006). It has been suggested that the amygdala contribute to social-emotional recognition (Adolphs et al., 2002; Adolphs and Spezio, 2006) and processing of salient face stimuli during unpredictable situations (Adolphs, 2010). Some have emphasized the evolutionary significance of the amygdalae, suggesting it plays a role in detecting relevant stimuli (Sander et al., 2003) and signaling potentially significant consequential events (Fitzgerald et al., 2006). Thus, based on past findings, perhaps the processing of dynamic faces requires increased amygdala activation due to an increased vigilance in observing the dynamically changing salient features of faces.

The inferior frontal gyrus, a part of the ventrolateral prefrontal cortex, is associated with all sorts of cognitive functions including response inhibition (Aron et al., 2003; Hampshire et al., 2009, 2010), working memory (Yaple and Arsalidou, in press), negative priming (Yaple and Arsalidou, 2017) and mental attention (Arsalidou et al., 2013). A hierarchical model of the prefrontal cortex suggests that the inferior frontal gyri would be responsible for simple, non-abstract judgments (Christoff et al., 2009). The majority of studies asked participants to make simple judgments about gender, emotion, or motion of faces congruent with this hypothesis. Regarding right lateralization, relevant to social interactions, the right inferior frontal gyrus is active when processing social information such as cooperative interaction (Liu et al., 2015) and interpersonal interactions (Liu et al., 2016). It has been shown that bilateral inferior frontal gyrus as a part of the dorsomedial network (Bzdok et al., 2013), which is involved in contemplation of others' mental states (Mar, 2011 for meta-analysis). Alternatively, based on a trade-off between task difficulty and the mental-attentional capacity of the individual, the right hemisphere is hypothesized to be favored in simple, automatized processes (Pascual-Leone, 1989; Arsalidou et al., 2018 for details). Overall, right inferior frontal gyrus's activation during face perception may be associated with cognitive processing of social information processing or maintaining with simple task requirements.

LIMITATIONS

Data presented here represent concordance across fMRI studies that investigated dynamic vs. static facial expressions and across different emotional states. ALE methodological limitations have been discussed elsewhere (Zinchenko and Arsalidou, 2018; Yaple and Arsalidou, in press) and include lack of control of statistical methodologies adopted by original articles and consideration only of peak coordinates. A shortcoming of the current study is data we report here are in majority based on female participants as original articles favored recruiting female participants who may show a greater response to faces.

CONCLUSION

A coordinate-based meta-analysis was performed to assess the concordance of brain activations derived from experiments that identified more activity in dynamic compared to static faces and

other control tasks. We observed concordance across studies in brain areas well established in the face processing literature, as well as the cerebellum, which is not discussed in models associated with face processing. The observed results suggest that dynamic faces require increased resources in the brain to process complex, dynamically changing features of faces. The current data provide a stereotaxic set of brain regions that underlie dynamic facial expression in typical adults. Practically, these normative data can serve as a benchmark for future studies with atypical populations, such as individuals with autism spectrum disorder. Theoretically, these findings provide further support for an extended set of areas that support processing of dynamic facial expression. Overall, our present findings can inform current models and help guide future studies on dynamic facial expressions.

AUTHOR CONTRIBUTIONS

OZ helped collect and analyze data and prepared the first draft of the manuscript. ZY helped collect and analyze data and contributed to manuscript preparation. MA conceptualized research and contributed to manuscript preparation.

ACKNOWLEDGMENTS

REFERENCES

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Ann. N. Y. Acad. Sci.* 1191, 42–61. doi: 10.1111/j.1749-6632.2010.05445.x

Adolphs, R., Baron-Cohen, S., and Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. *J. Cogn. Neurosci.* 14, 1264–1274. doi: 10.1162/089892902760807258

Adolphs, R., and Spezio, M. (2006). Role of the amygdala in processing visual social stimuli. *Prog. Brain Res.* 156, 363–378. doi: 10.1016/S0079-6123(06)56020-0

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/S1364-6613(00)01501-1

Aron, A. R., Bullmore, E. T., Sahakian, B. J., and Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat. Neurosci.* 6, 115–117. doi: 10.1038/nn1003

Arsalidou, M., Morris, D., and Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain Topogr.* 24, 149–163. doi: 10.1007/s10548-011-0171-4

Arsalidou, M., Pascual-Leone, J., Johnson, J., Morris, D., and Taylor, M. J. (2013). A balancing act of the brain: activations and deactivations driven by cognitive load. *Brain Behav.* 3, 273–285. doi: 10.1002/brb3.128

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., and Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: meta-analyses of fMRI studies. *Dev. Cogn. Neurosci.* 30, 239–250. doi: 10.1016/j.dcn.2017.08.002

Bernstein, M., and Yovel, G. (2015). Two neural pathways of face processing: a critical evaluation of current models. *Neurosci Biobehav Rev.* 55, 536–546. doi: 10.1016/j.neubiorev.2015.06.010

Bookheimer, S. Y., Wang, A. T., Scott, A., Sigman, M., and Dapretto, M. (2008). Frontal contributions to face processing differences in autism: evidence from fMRI of inverted face processing. *J. Int. Neuropsychol. Soc.* 14, 922–932. doi: 10.1017/S135561770808140X

Brooks, V. B. (1984). Cerebellar functions in motor control. *Hum. Neurobiol.* 2, 251–260.

Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., et al. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Front. Hum. Neurosci.* 7:232. doi: 10.3389/fnhum.2013.00232

Ceccarini, F., and Caudek, C. (2013). Anger superiority effect: The importance of dynamic emotional facial expressions. *Vis. Cogn.* 21, 498–540. doi: 10.1080/13506285.2013.807901

Christoff, K., Keramatian, K., Gordon, A. M., Smith, R., and Mädler, B. (2009). Prefrontal organization of cognitive control according to levels of abstraction. *Brain Res.* 1286, 94–105. doi: 10.1016/j.brainres.2009.05.096

Corradi-Dell'Acqua, C., Schwartz, S., Meaux, E., Hubert, B., Vuilleumier, P., and Deruelle, C. (2014). Neural responses to emotional expression information in high-and low-spatial frequency in autism: evidence for a cortical dysfunction. *Front. Hum. Neurosci.* 8:189. doi: 10.3389/fnhum.2014.00189

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr. Opin. Neurobiol.* 10, 732–739. doi: 10.1016/S0959-4388(00)00153-7

Eger, E., Schyns, P. G., and Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. *Neuroimage* 22, 232–242. doi: 10.1016/j.neuroimage.2003.12.028

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., and Fox, P. T. (2012). Activation likelihood estimation revisited. *Neuroimage* 59, 2349–2361. doi: 10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Laird, A. R., Fox, P. M., Lancaster, J. L., and Fox, P. T. (2017). Implementation errors in the GingerALE Software: description and recommendations. *Hum. Brain Mapp.* 38, 7–11. doi: 10.1002/hbm.23342

Eickhoff, S. B., Laird, A., Grefkes, C., Wang, L., Zilles, K., and Fox, P. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* 30, 2907–2926. doi: 10.1002/hbm.20718

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., et al. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage* 137, 70–85. doi: 10.1016/j.neuroimage.2016.04.072

Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., and Phan, K. L. (2006). Beyond threat: amygdala reactivity across multiple expressions of facial affect. *Neuroimage* 30, 1441–1448. doi: 10.1016/j.neuroimage.2005.11.003

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J. Psychiatry Neurosci.* 34, 418–432.

Gobbini, M. I., and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32–41. doi: 10.1016/j.neuropsychologia.2006.04.015

Grosbras, M. H., and Paus, T. (2006). Brain networks involved in viewing angry hands or faces. *Cereb. Cortex* 16, 1087–1096. doi: 10.1093/cercor/bhj050

Gupta, R., Koscik, T. R., Bechara, A., and Tranel, D. (2011). The amygdala and decision making. *Neuropsychologia* 49, 760–766. doi: 10.1016/j.neuropsychologia.2010.09.029

Hadjikhani, N., Joseph, R. M., Snyder, J., and Tager-Flusberg, H. (2007). Abnormal activation of the social brain during face perception in autism. *Hum. Brain Mapp.* 28, 441–449. doi: 10.1002/hbm.20283

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., and Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* 50, 1313–1319. doi: 10.1016/j.neuroimage.2009.12.109

Hampshire, A., Thompson, R., Duncan, J., and Owen, A. M. (2009). Selective tuning of the right inferior frontal gyrus during target detection. *Cogn. Affect. Behav. Neurosci.* 9, 103–112. doi: 10.3758/CABN.9.1.103

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0

Hurlemann, R., Rehme, A. K., Diessel, M., Kukolja, J., Maier, W., Walter, H., et al. (2008). Segregating intra-amygdalar responses to dynamic facial emotion with cytoarchitectonic maximum probability maps. *J. Neurosci. Methods* 172, 13–20. doi: 10.1016/j.jneumeth.2008.04.004

Iidaka, T., Yamashita, K., Kashikura, K., and Yonekura, Y. (2004). Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. An investigation with event-related fMRI. *Cogn. Brain Res.* 18, 196–204. doi: 10.1016/j.cogbrainres.2003.10.005

Ivry, R. B., and Spencer, R. M. (2004). The neural representation of time. *Curr. Opin. Neurobiol.* 14, 225–232. doi: 10.1016/j.conb.2004.03.013

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., and McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cereb. Cortex*, 13, 1023–1033. doi: 10.1093/cercor/13.10.1023

Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., et al. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM?152 brain template. *Hum. Brain Mapp.* 28, 1194–1205. doi: 10.1002/hbm.20345

Langton, S. R., Watt, R. J., and Bruce, I. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends Cogn. Sci.* 4, 50–59. doi: 10.1016/S1364-6613(99)01436-9

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cell Mol. Neurobiol.* 23, 727–738. doi: 10.1023/A:1025048802629

Lee, L. C., Andrews, T. J., Johnson, S. J., Woods, W., Gouws, A., Green, G. G., et al. (2010). Neural responses to rigidly moving faces displaying shifts in social attention investigated with fMRI and MEG. *Neuropsychologia* 48, 447–490. doi: 10.1016/j.neuropsychologia.2009.10.005

Liu, T., Saito, H., and Oi, M. (2015). Role of the right inferior frontal gyrus in turn-based cooperation and competition: a near-infrared spectroscopy study. *Brain Cogn.* 99, 17–23. doi: 10.1016/j.bandc.2015.07.001

Liu, T., Saito, H., and Oi, M. (2016). Obstruction increases activation in the right inferior frontal gyrus. *Soc. Neurosci.* 11, 344–352. doi: 10.1080/17470919.2015.1088469

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* 62, 103–134. doi: 10.1146/annurev-psych-120709-145406

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Prisma Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

Nichols, D. F., Betts, L. R., and Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Front. Psychol.* 1:28. doi: 10.3389/fpsyg.2010.00028

Nomi, J. S., and Uddin, L. Q. (2015). Face processing in autism spectrum disorders: from brain regions to brain networks. *Neuropsychologia* 71, 201–216. doi: 10.1016/j.neuropsychologia.2015.03.029

O'Toole, A. J., Roark, D. A., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* 6, 261–266. doi: 10.1016/S1364-6613(02)01908-3

Pascual-Leone, J. (1989). "An organismic process model of Witkin's field-dependence-independence," in *Human Development, Vol. 3, Cognitive Style and Cognitive Development*, eds T. Globerson and T. Zelniker (Westport, CT: Ablex Publishing), 36–70.

Paulin, M. G. (1993). The role of the cerebellum in motor control and perception. *Brain Behav. Evol.* 41, 39–50. doi: 10.1159/000113822

Pelphrey, K. A., Morris, J. P., McCarthy, G., and Labar, K. S. (2007). Perception of dynamic changes in facial affect and identity in autism. *Soc. Cogn. Affect. Neurosci.* 2, 140–149. doi: 10.1093/scan/nsm010

Pentón, L. G., Fernández, A. P., León, M. A., Ymas, Y. A., García, L. G., Iturria-Medina, Y., et al. (2010). Neural activation while perceiving biological motion in dynamic facial expressions and point-light body action animations. *Neural Regen. Res.* 5, 1076–1083. doi: 10.3969/j.issn.1673-5374.2010.14.007

Peyrin, C., Baciu, M., Segebarth, C., and Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage* 23, 698–707. doi: 10.1016/j.neuroimage.2004.06.020

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top–down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *J. Cogn. Neurosci.* 22, 2768–2780. doi: 10.1162/jocn.2010.21424

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., and Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *Neuroimage* 28, 464–473. doi: 10.1016/j.neuroimage.2005.06.006

Robins, D. L., Hunyadi, E., and Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain Cogn.* 69:269–278. doi: 10.1016/j.bandc.2008.08.007

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., and Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain* 126, 2381–2395. doi: 10.1093/brain/awg241

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., et al. (2011). Emotional perception: meta-analyses of face and natural scene processing. *Neuroimage* 54, 2524–2533. doi: 10.1016/j.neuroimage.2010.10.011

Sander, D., Grafman, J., and Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Rev. Neurosci.* 14, 303–316. doi: 10.1515/REVNEURO.2003.14.4.303

Sarkheil, P., Goebel, R., Schneider, F., and Mathiak, K. (2013). Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. *Soc. Cogn. Affect. Neurosci.* 8, 950–957. doi: 10.1093/scan/nss092

Sato, W., Kochiyama, T., and Uono, S. (2015). Spatiotemporal neural network dynamics for the processing of dynamic facial expressions. *Sci. Rep.* 5:12432. doi: 10.1038/srep12432

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., and Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cogn. Brain Res.* 20, 81–91. doi: 10.1016/j.cogbrainres.2004.01.008

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi: 10.1016/j.conb.2006.03.001

Schultz, J., Brockhaus, M., Bülthoff, H. H., and Pilz, K. S. (2013). What the human brain likes about facial motion. *Cereb. Cortex* 23, 1167–1178. doi: 10.1093/cercor/bhs106

Schultz, J., and Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Exp. Brain Res.* 194, 465–475. doi: 10.1007/s00221-009-1721-9

Stoodley, C. J., and Schmahmann, J. D. (2010). Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. *Cortex* 46, 831–844. doi: 10.1016/j.cortex.2009.11.008

Taylor, M. J., Arsalidou, M., Bayless, S. J., Morris, D., Evans, J. W., and Barbeau, E. J. (2009). Neural correlates of personally familiar faces: parents, partner and own faces. *Hum. Brain Mapp.* 30, 2008–2020. doi: 10.1002/hbm.20646

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., and Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp.* 33, 1–13. doi: 10.1002/hbm.21186

van der Gaag, C., Minderaa, R. B., and Keysers, C. (2007). The BOLD signal in the amygdala does not differentiate between dynamic facial expressions. *Soc. Cogn. Affect. Neurosci.* 2, 93–103. doi: 10.1093/scan/nsm002

Van Overwalle, F., Baetens, K., Mariën, P., and Vandekerckhove, M. (2014). Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. *Neuroimage* 86, 554–572. doi: 10.1016/j.neuroimage.2013.09.033

Van Overwalle, F., D'aes, T., and Mariën, P. (2015). Social cognition and the cerebellum: a meta-analytic connectivity analysis. *Hum. Brain Mapp.* 12, 5137–5154. doi: 10.1002/hbm.23002

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631. doi: 10.1038/nn1057

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., et al. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* 69, 131–140. doi: 10.1016/j.cortex.2015.05.003

Yaple, Z. A., and Arsalidou, M. (in press). N-back working memory task. Meta-analysis of normative fMRI studies with children. *Child Dev.* doi: 10.1111/cdev.13080. [Epub ahead of print].

Yaple, Z., and Arsalidou, M. (2017). Negative priming: a meta-analysis of fMRI studies. *Exp. Brain Res.* 235, 3367–3374. doi: 10.1007/s00221-017-5065-6

Yaple, Z. A., Vakhrushev, R., and Jolij, J. (2016). Investigating emotional top down modulation of ambiguous faces by single pulse TMS on early visual cortices. *Front. Neurosci.* 10:305. doi: 10.3389/fnins.2016.00305

Zinchenko, O., and Arsalidou, M. (2018). Brain responses to social norms: meta-analyses of fMRI studies. *Hum. Brain Mapp.* 39, 955–970. doi: 10.1002/hbm.23895

Incongruence Between Observers' and Observed Facial Muscle Activation Reduces Recognition of Emotional Facial Expressions From Video Stimuli

Tanja S. H. Wingenbach[1,2,3]*, Mark Brosnan[1], Monique C. Pfaltz[3], Michael M. Plichta[3,4] and Chris Ashwin[1]

[1] Centre for Applied Autism Research, Department of Psychology, University of Bath, Bath, United Kingdom, [2] Social and Cognitive Neuroscience Laboratory, Centre of Biology and Health Sciences, Mackenzie Presbyterian University, São Paulo, Brazil, [3] Department of Consultation-Liaison Psychiatry and Psychosomatic Medicine, University Hospital Zurich, Zürich, Switzerland, [4] Department of Psychiatry, Psychosomatic Medicine, and Psychotherapy, University Hospital Frankfurt, Frankfurt, Germany

According to embodied cognition accounts, viewing others' facial emotion can elicit the respective emotion representation in observers which entails simulations of sensory, motor, and contextual experiences. In line with that, published research found viewing others' facial emotion to elicit automatic matched facial muscle activation, which was further found to facilitate emotion recognition. Perhaps making congruent facial muscle activity explicit produces an even greater recognition advantage. If there is conflicting sensory information, i.e., incongruent facial muscle activity, this might impede recognition. The effects of actively manipulating facial muscle activity on facial emotion recognition from videos were investigated across three experimental conditions: (a) explicit imitation of viewed facial emotional expressions (stimulus-congruent condition), (b) pen-holding with the lips (stimulus-incongruent condition), and (c) passive viewing (control condition). It was hypothesised that (1) experimental condition (a) and (b) result in greater facial muscle activity than (c), (2) experimental condition (a) increases emotion recognition accuracy from others' faces compared to (c), (3) experimental condition (b) lowers recognition accuracy for expressions with a salient facial feature in the lower, but not the upper face area, compared to (c). Participants (42 males, 42 females) underwent a facial emotion recognition experiment (ADFES-BIV) while electromyography (EMG) was recorded from five facial muscle sites. The experimental conditions' order was counter-balanced. Pen-holding caused stimulus-incongruent facial muscle activity for expressions with facial feature saliency in the lower face region, which reduced recognition of lower face region emotions. Explicit imitation caused stimulus-congruent facial muscle activity without modulating recognition. Methodological implications are discussed.

Keywords: facial emotion recognition, imitation, facial muscle activity, facial EMG, embodiment, videos, dynamic stimuli, facial expressions of emotion

INTRODUCTION

Embodied cognition accounts postulate that there are interrelations between bodily actions (e.g., body posture, gestures) and cognitions. When we acquire memory, we store all information of the specific situation (i.e., context, affect, behaviour, etc.) together in a representation of the situation also containing embodiments (Barsalou, 2008). When we experience an aspect of this initial situation, the remaining memory stored in the representation can get activated (Niedenthal, 2007). For example, observing a smile can activate a representation of a situation that contained smiling (e.g., receiving positive news). This representation can include both the accompanying affect (e.g., feeling happy) and its physical components, including physiological responses and facial muscle activations. In support of this idea, observing facial emotional expressions within a laboratory setting has been found to lead to congruency between observers' and observed facial muscle activation (e.g., Dimberg, 1982; Dimberg and Thunberg, 1998; Dimberg et al., 2000; Hess and Blairy, 2001; Sato and Yoshikawa, 2007; Achaibou et al., 2008; Likowski et al., 2012). This phenomenon of an observer showing implicit facial muscle activation congruent with the muscle activation in the observed emotional face is generally termed 'facial mimicry' (for a literature review, see Hess and Fischer, 2014). Such implicit facial mimicry involves unconscious mechanisms (Dimberg et al., 2000), as muscle activations occur automatically and outside of awareness when healthy people perceive emotional facial expressions (Dimberg, 1982). This automatic muscle activation is different to explicit imitation, which involves the deliberate intention to explicitly imitate the expression of another person and awareness about the activity. Based on embodied cognition accounts, the representation of the emotional expression produced in the observer should facilitate facial emotion recognition of the observed expression due to the stimulus-congruency in facial muscle activations.

Support for the idea that stimulus-congruent facial muscle activation facilitates facial emotion recognition comes from a study investigating the effects of actively manipulating facial movements in observers on facial emotion recognition in others. Oberman et al. (2007) compared recognition rates for happiness, disgust, fear, and sadness using an experimental condition where facial mimicry was 'blocked' by having participants actively bite on a pen without the lips touching it. Even though the word 'blocked' was used, the manipulation actually created constant muscular activity, which served to produce a non-specific steady state of muscle activity interfering with facial mimicry. Oberman et al. (2007) reported reduced recognition of images displaying disgust and happiness from hindering the observer's facial mimicry by pen-holding, compared to a condition where no facial movement manipulation was performed. Since recognition was impaired for two out of four investigated facial emotion expressions, Oberman et al. (2007) concluded that facial emotion recognition can be selectively impaired when facial mimicry is hindered. The published literature generally supports the link between automatic stimulus-congruent facial muscle activation in observers and facilitated facial emotion recognition in others (Wallbott, 1991; Stel and van Knippenberg, 2008;

Neal and Chartrand, 2011; Sato et al., 2013; but see also Blairy et al., 1999; Rives Bogart and Matsumoto, 2010). Many conclude that being able to engage in facial mimicry facilitates recognition based on the congruency between the facial muscle activation in the stimulus and the observer.

Another explanation for diminished recognition accuracy when participants' facial movements are actively manipulated (e.g., biting on pen) is that active manipulations themselves induce muscle feedback. Considering what is known from the literature on embodied cognition, it should be noted that such facial muscle feedback itself can have an effect on social processes such as facial emotion recognition. When mouth movement is actively manipulated, the activation in the observer's face does not align with the activation in the observed expression, instead of being stimulus-congruent as during facial mimicry. This conflicting facial muscle activation could be causing interference during the decoding of the expression leading to decreased recognition accuracy. Ponari et al. (2012) investigated the specific effects of facial muscle manipulation location on facial emotion recognition. These authors manipulated participants' movement of the lower and upper facial muscles and tested the effects on recognition accuracy of individual emotions. In their study, one group of participants bit on a chopstick horizontally without the lips touching it to fix facial movement in the lower face region (and hinder facial mimicry). The other group in the study had two small stickers attached at the inner edge of the eyebrows and were instructed to push the stickers together to fix facial movement in the upper face region. The inducement of steady facial muscle activation in observers (in the lower and upper face region) diminished recognition of the facial emotional expressions with facial feature saliency in the lower and upper face region, respectively. It is thus possible that the effects on facial emotion recognition in the studies by Oberman et al. (2007) and Ponari et al. (2012) were not the result of hindered facial mimicry. Instead, it is possible that the diminished recognition of certain emotional expressions resulted from the stimulus-incongruent muscle feedback induced by the facial muscle manipulations. This effect could result particularly when the facial region of the salient facial feature in the observed emotional expression is being affected by the facial muscle manipulation in the observer and the resulting facial muscle activity in observers is incongruent with the observed facial muscle activation. Further research is needed investigating this stimulus-incongruency interpretation experimentally.

However, if automatic stimulus-congruent facial muscle feedback in observers (i.e., facial mimicry) facilitates facial emotion recognition, it is plausible that more intense and deliberate muscle activation could facilitate decoding of the observed facial expression of emotion even further (e.g., from explicit imitation of observed facial expression). This assumption is supported by the results of a study by Conson et al. (2013). The study showed better facial emotion recognition performance in actors who explicitly imitated the observed facial emotional expressions and used the resulting generated feeling for decoding emotions (in line with embodiment), compared to actors who used contextual information and thus a more knowledge-based approach. Based on this study, it

seems that explicit stimulus-congruent facial muscle activation in observers facilitates facial emotion recognition. However, it is unknown whether the two actor groups differed in their facial muscle activity. The usage of facial EMG allows to investigate differences in facial muscle activity between the various experimental conditions that are assumed to affect facial emotion recognition and is thus indicated. Further, participants were actors with specialised training in nonverbal communication, which includes expressing emotions. Thus, further investigation of explicit imitation and its effect on facial emotion recognition in more general population samples is necessary.

A study considering these factors was conducted by Schneider et al. (2013), who investigated facial emotion recognition in a sample of undergraduate students and applied facial movement manipulations while measuring facial EMG. Results showed that explicit imitation of observed facial expression led to earlier accurate recognition in a morphed sequence of emotional expressions compared to a condition where participants were instructed to suppress their own facial expressions. This suppression condition was intended to hinder participants in producing stimulus-congruent facial muscle activation. In the same study, the condition with free facial movement also led to earlier correct emotion recognition than the expression suppression condition. However, explicit imitation did not lead to a significant advantage over the free facial movement condition. These results suggest that suppression of facial muscle activation in observers diminishes facial emotion recognition rather than that explicit imitation enhances recognition. However, the effectiveness of the instruction to suppress any facial muscle is questionable. Indeed, the EMG results showed no difference in facial muscle activation during the expression suppression condition compared to the free facial movement condition. It is possible that the suppression instruction had a recognition-impairing effect due to other mechanisms like cognitive load. Thus, it might be better to actively manipulate facial muscles to being stimulus-incongruent. With the results on explicit imitation from Schneider et al. (2013) being in contrast to reports by Conson et al. (2013), it still remains to be answered whether explicit stimulus-congruent facial muscle activation in observers facilitates facial emotion recognition or a lack of stimulus-congruency diminishes facial emotion recognition.

Published research has included either an explicit imitation condition or a condition where participants held a pen in their mouth, alongside a condition without any facial movement manipulation. Much of the previous research testing the effects of facial muscle manipulations on emotion recognition ability has used either static images or morphed image sequences, which are limited in ecological validity compared to other types of stimuli. Many previous studies have only used a limited number of basic emotion categories, along only two or three muscle sites in the face to measure muscle activity, which limits the measures about emotion processing and activity in the face. The present study is the first report the authors are aware of to include all three experimental conditions in one experiment to assess how facial emotion recognition is affected by explicit facial muscle activation: (1) an Explicit Imitation condition where participants

were told to exactly imitate the expressions they saw while they viewed video stimuli of others displaying various emotional expressions, (2) a Pen-Holding condition where participants held a pen tightly with the lips of their mouth while they watched the videos, and (3) a Passive Viewing control condition where participants just passively viewed the videos. The present study also increased the number of emotion categories included (i.e., anger, disgust, fear, sadness, surprise, happiness, embarrassment, contempt, pride) and measured EMG from five different muscle sites (corrugator supercilii, zygomaticus major, levator labii, depressor anguli oris, and lateral frontalis).

The aim of the present study was to induce explicit facial muscle activation and to investigate the effects of actively manipulating facial muscle activity to being stimulus-congruent and stimulus-incongruent on subsequent facial emotion recognition accuracy based on more ecologically valid stimuli. There were three hypotheses: (1) Enhanced facial muscle activity throughout the face was expected to result from the Explicit Imitation condition, and in the muscles of the lower face region from the Pen-Holding condition, compared to the Passive Viewing control condition. (2) It was hypothesised that enhanced congruency of facial muscle activity between the stimuli and observers (Explicit Imitation condition) would facilitate recognition of emotion compared to the Passive Viewing control condition. (3) It was further hypothesised that the Pen-Holding condition would induce stimulus-incongruent facial muscle activity in observers' mouth region, resulting in poorer recognition of facial emotional expressions with salient facial features in the lower face region compared to the Passive Viewing control condition.

MATERIALS AND METHODS

Participants

A total of 86 university students (43M/43F; Mean age = 19.6, SD = 3.6) were recruited through Campus advertising at the University of Bath and represented both Humanities and Science Departments (54 from Humanities and 32 from Sciences). Technical equipment failure resulted in the loss of data for two participants, resulting in a final sample of 84 participants (41M/43F; Mean age = 19.6, SD = 3.6). Based on a power analyses using G*Power (Faul et al., 2007) for the planned analyses to test the main hypotheses (i.e., two-tailed paired samples t-tests), a sample size of 84 retrieves 0.78 power with an alpha level of 5% and a small effect size of dz = 0.3. The majority of participants in the final sample were undergraduate students (n = 82), with one participant enrolled in a Master's Programme and another in a Ph.D. Programme. Two participants reported about a diagnosis of Major Depression and one participant reported about a diagnosis of an Anxiety Disorder. These participants reported to be on medication and not to experience any symptoms of their mental disorders at the time of participation. Thus, these participants were included in the analyses[1]. All participants had

---

[1] Analyses on the accuracy data were also conducted excluding these three participants, which had no effect on the outcome of the results.

normal or corrected-to-normal vision. Ethical approval for the current study was granted by the Psychology Ethics Committee at the University of Bath.

Material
Facial Emotion Videos
The facial emotion recognition experiment included videos from the validated Amsterdam Facial Expression Set – Bath Intensity Variations (ADFES-BIV; Wingenbach et al., 2016), which is an adaptation from the ADFES (van der Schalk et al., 2011). The ADFES-BIV set contains 360 videos: 12 different encoders (7 male, 5 female) each displaying 10 expressions (anger, disgust, fear, sadness, surprise, happiness, contempt, embarrassment, pride, and neutral/blank stare) across 3 expression intensities (low, intermediate, high). The ADFES-BIV includes 10 more videos of one additional female encoder displaying each of the 10 expression categories once for practise trials. An example image for each emotion category can be found in van der Schalk et al. (2011). Each video is 1040 ms in length. For more detail on the ADFES-BIV (see Wingenbach et al., 2016).

Electromyography (EMG) Recording
The BIOPAC MP150 System with the Acqknowledge software (Version 4, Biopac Systems, Inc., Goleta, CA, United States) and EMG110C units for each of the five facial muscle sites (corrugator supercilii, zygomaticus major, levator labii, depressor anguli oris, and lateral frontalis) were used for recording of the EMG data. Pairs of shielded surface silver–silver chloride (Ag–AgCl) electrodes (EL254S) filled with conductive gel (saline based Signa Gel) and with a contact area of 4mm diameter were used. The EMG signal was amplified by 2000 and online bandpass filtering of 10 Hz and 500 Hz was applied. Grounding was achieved through the VIN- of the TSD203 (GSR), the data of which is not reported in this paper. The sampling rate was 1000 Hz throughout the experiment.

Procedure
Participants were tested in a quiet testing laboratory at the University of Bath, and written consent was obtained prior to study participation. Participants were seated approximately 60 cm from the PC monitor. Before EMG electrode attachment, participants' faces were cleaned with alcohol swabs. The 10 face EMG electrodes were then placed in pairs over the respective muscle sites on the left side of the face, which was done according to the guidelines by Fridlund and Cacioppo (1986). The electrodes of each pair of electrodes were placed in close proximity to each other using double-stick adhesive rings, with the distance being about 1 cm between the electrode centres. EMG was recorded from five different face muscle sites during the whole duration of the testing session. Participants were kept blind about the true purpose of the study of assessing the effect of facial muscle activity on facial emotion recognition. Thus, participants were told that the electrodes would be measuring pulse and sweat response to facial emotional expressions. After all the electrodes were placed on the face, the participants initially watched a short neutral-content video clip lasting 4 min 18 s, in order to facilitate settling into the research session and to reduce any strong feelings they might have had before the testing session (see Wingenbach et al., 2016). Participants then passively watched 90 videos of the ADFES-BIV to assess facial mimicry without a cognitive load; those results will be presented elsewhere (Wingenbach et al., n.d.). Afterwards, participants underwent the facial emotion recognition task, the data of which is presented in this manuscript. The study included all videos from the ADFES-BIV. However, the facial emotion recognition task of the study presented within this manuscript comprised 280 trials including 10 practise videos. Each of the experimental conditions included equal representations for each of the encoders in the videos, the emotion categories, and the expression intensity levels. There were six different versions of the facial emotion recognition experiment, with each one representing a different order of the three experimental conditions. Participants were pseudo-randomly assigned to one of the six conditions, with the sex ratios being balanced across the versions. Counter-balancing the order of the experimental conditions was important, because performance (e.g., accuracy of response) often increases over the course of the experiment (see section "Results").

There were 90 trials within each of the three different conditions in the facial emotion recognition experiment: (1) Explicit Imitation, (2) Pen-Holding, and (3) Passive Viewing control condition. During the Explicit Imitation condition, participants were instructed to exactly imitate the facial expressions they observed in the videos (including the blank stare in the neutral expression) as soon as they perceived them. For the Pen-Holding condition, participants were told to hold a pen tightly with their lips, with one end of the pen sticking straight out of their mouth, with pressure applied by the lips (but not the teeth). This manipulation aimed to actively induce facial muscle activity, which also would be incongruent with the emotional expressions included in the study with facial feature saliency in the lower part of the face. The experimenter demonstrated to each participant how the pen was to be held in the mouth, and only after the experimenter was satisfied with the pen-holding technique, the experiment was started. The instruction for the Passive Viewing control condition was to simply watch the videos. Each trial started with a blank screen presented for 500 ms, which was followed by a fixation cross for 500 ms appearing in the centre of the screen. Immediately after the disappearance of the cross the stimulus appeared, followed by a blank screen for 500 ms before the answer screen appeared. The methods used for the facial emotion recognition task were the same as those reported in a previous study using the stimulus set and task (Wingenbach et al., 2016). The answer screen contained 10 labels (neutral and the nine emotion categories) included in the experiment distributed evenly across the screen in two columns and alphabetical order. The participant used a mouse-click to choose their answer, and the mouse-click triggered the next trial. The mouse position was variable. Participants were instructed to choose an emotion label promptly. No feedback was provided about the correctness of the answer. (For more detail about the task procedure, see Wingenbach et al., 2016). After completion of the computer-task, participants were debriefed and compensated with either course credit or GBP 7.

EMG Data Preparation

Several participants (who did not undergo the Explicit Imitation condition as last condition) verbally self-reported after the testing session that they were unable to stop themselves from imitating the observed facial expressions in subsequent conditions. Thus, the raw EMG data of all participants was visually inspected at trial level to identify participants whose EMG activity pattern suggested explicit imitation in other experimental conditions. Imitative activity on a trial basis during the Passive Viewing control condition was clearly visible in the raw data. **Figure 1** displays the raw data of two selected participants across the whole experiment. Visually comparing the activity in the Passive Viewing control condition from **Figures 1A,B** clearly shows that the participant from **Figure 1B** explicitly imitated in the Passive Viewing control condition. The EMG activity of this participant was, for many trials, as intense in the Passive Viewing control condition as in the Explicit Imitation condition, whereas it should have been similar to the corrugator and

frontalis channel during the pen-holding. Eighteen participants were subsequently identified to have shown explicit imitation in conditions other than the Explicit Imitation condition. Looking through the raw EMG data, a further two participants were identified who did not show constant elevated EMG activity in the muscles of the lower part of the face in the Pen-Holding condition, consistent with tightly holding a pen in their mouth (see **Figure 1** as example for the distinctive EMG activation in the first three channels: zygomaticus, depressor, levator). Another participant misunderstood the instructions and did the Explicit Imitation condition twice, so no data on the Passive Viewing control condition exists for this participant. Consequently, the EMG data of these participants for the experimental conditions where the instructions were not fully complied with were excluded from EMG data analyses. The same approach was taken for the accuracy of response data. In addition, there were errors for recording EMG from certain muscles for some participants, which meant the EMG data for

FIGURE 1 | Raw electromyography (EMG) signal from two participants as recorded for the five facial muscles investigated across the three experimental conditions of the study. It was zoomed in at trial level considering stimulus on- and offsets for identification of experimental conditions per participant where task instructions were not fully complied with and thus to exclude from analyses. **(A)** A participant's EMG activity in compliance with the three experimental conditions. **(B)** Explicit imitation by a participant in the Passive Viewing control condition. Spikes in the EMG signal in the Passive Viewing control condition of similar height as during the Explicit Imitation condition demonstrate explicit imitation instead of passive viewing in the control condition.

some participants was not complete. Again, these participants were still included, but the EMG data of the muscles where problems occurred were excluded from the EMG data analyses. The resulting sample sizes per muscle in each analysis are reported in the respective results section. Participants were not fully excluded from analyses in order to retain enough power for the analyses.

EMG Data Processing

The Autonomic Nervous System Laboratory 2.6 (ANSLAB; Wilhelm and Peyk, 2005) was used for offline filtering of the EMG data. The EMG signals were 50 Hz notch filtered, 28 Hz high-pass filtered, and the rectified signal was smoothed with a moving average width of 50 ms. A duration of 2.6 s from stimulus onset (excluding the pre-stimulus baseline) was used as the event window, and mean values were calculated and extracted for the event period averaged across all trials with MATLAB (MATLAB 2016b, The MathWorks); this was done for each muscle within each of the three experimental conditions. To assure that the imitation activity was captured within these means, we added 1.5 s to the stimulus offset; a figure demonstrating this necessity based on the activation timings can be found in the Supplementary Figure S1.

DATA ANALYSES AND RESULTS

Accuracy Changes Across the Experiment Cheque

When participants complete a task consisting of many trials or repeatedly do conditions of a new task, this produces learning effects and the participant's performance will improve over time. Foroughi et al. (2017) showed that participants shifted from a more effortful approach during a task (which included 48 trials) to a more automatic approach. The faster participants completed a trial and the more trials participants completed, the smaller their pupil dilation became, indicating automatic processing. In the current study, the order of the experimental conditions was counter-balanced for the participants to counter within-task improvements. The accuracy of response data was investigated for the expected within-task improvements over the course of the experiment. This analysis was necessary despite the counter-balancing of the order of the experimental conditions, because the accuracy of response data from the experimental conditions where the instructions were not fully complied with by individual participants (as identified through the EMG data inspection described in section "EMG Data Preparation") were excluded from further analyses. The elimination of specific conditions for some participants led to unequal numbers of data points per experimental condition. Consequently, the eliminations combined with an increase of accuracy of response over the course of experiments could potentially bias the results. The resulting means for each condition will be inflated for the experimental condition with more data points where this experimental condition was the last condition. Conversely, sample means will be deflated for the experimental condition where more data points factor in

from when the experimental condition was undertaken first. Such biases could affect results for any within-subject analyses. It was not foreseeable before data collection that the instruction to explicitly imitate facial emotional expressions would have long-lasting effects on some participants in that they carried over the explicit imitation to subsequent conditions (as described in section "EMG Data Preparation"). Thus, the current study was planned with a within-subject experimental design and respective analyses.

To test for within-task improvements, the individual consecutive trials of the facial emotion recognition task were split into three equal 'blocks,' and accuracy of response was calculated for each block in order of their presentation for each participant (i.e., first 90 trials, second 90 trials, third 90 trials). Then, difference scores were calculated between the accuracy of response from the first and second block and between the second and third block. The two resulting difference scores were tested for a significant change using one-sample t-tests to test for a significant increase in accuracy of response over the course of the experiment. The alpha-level of 5% was Bonferroni-corrected to account for multiple comparisons; the resulting p-values were compared to a p-value of 0.025 ($p = 0.05/2$) for significance determination. The within-task improvements analysis was conducted on a sample of $N = 83$ (84 minus one female who did the Explicit Imitation condition twice, as there was no data for this person's Passive Viewing control condition). Cohen's d is presented as effect size measure. If the results show significant increase in accuracy of response from the first to the second and to the third Block, then this has important implications for the analyses. That is, between-subject analyses with only the first experimental condition each participant completed will be necessary instead of the planned within-subject analyses.

The one-sample t-tests showed that there was a significant increase in accuracy of response for participants from the first to the second Block [$M = 3.69$, $SD = 6.06$, $t(82) = 5.55$, $p < 0.001$, Cohen's $d = 0.609$], and from the second to the third Block [$M = 1.72$, $SD = 5.15$, $t(82) = 3.05$, $p = 0.003$, Cohen's $d = 0.335$]. Since accuracy scores increased significantly over the course of the experiment, between-subject analyses needed to be conducted to test the hypotheses of the current study. (The results from the within-subject analyses are presented in the Supplementary Figure S2).

Hypotheses Testing: Facial Muscle Activity Manipulation

To test the effectiveness of the experimental manipulations, the EMG data was statistically examined using generalised linear models for each muscle separately with Experimental Condition included as a factor for each analysis with its three levels (Explicit Imitation, Passive Viewing, and Pen-Holding). Due to the right-skewed nature of the EMG data, gamma distribution and log link function were specified in the conducted analyses. Pairwise comparisons were used to follow up significant main effects of Experimental Condition. Due to the necessary data eliminations described in Section "EMG Data Preparation," the sample sizes for the EMG data per experimental condition varied. The resulting n per comparison are presented with the results.
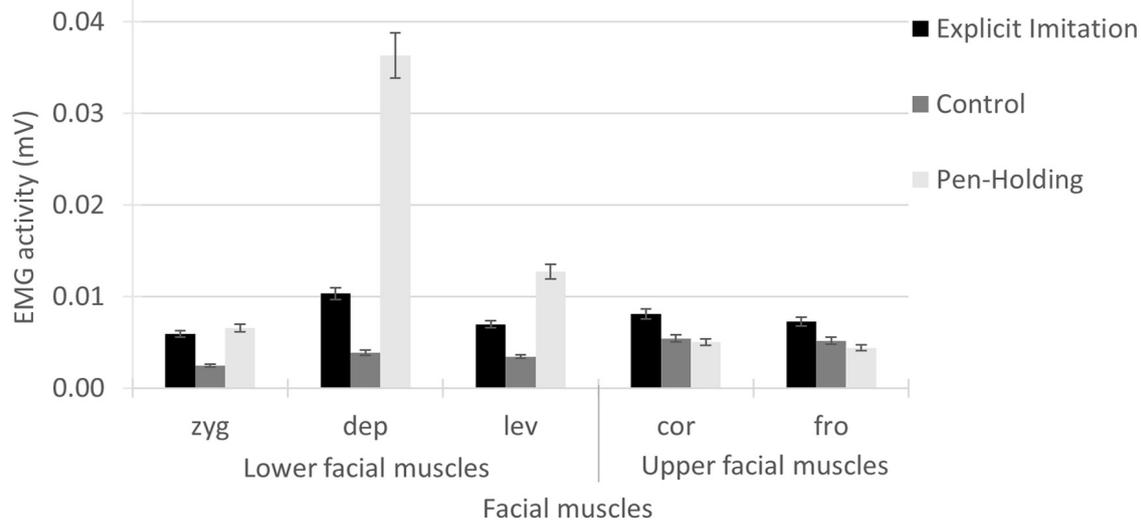
FIGURE 2 | Electromyography activity per facial muscle for each of the experimental conditions. In the Explicit Imitation condition (EI), participants imitated the observed expressions. Participants were holding a pen in the mouth with their lips in the Pen-Holding condition (PH) and passively viewed facial emotional expressions in a Passive Viewing control condition (C). Error bars represent standard errors of the means. zyg, zygomaticus (EI > C < PH); dep, depressor (EI > C < PH); lev, levator (EI > C < PH); cor, corrugator (EI > C = PH); fro, frontalis (EI > C = PH).

Generalised linear model results for the EMG activity in the *zygomaticus* muscle showed a significant main effect of Experimental Condition [Wald $\chi^2(2) = 141.79$, $p < 0.001$]; see **Figure 2**. Pairwise comparisons showed that the EMG activity in the zygomaticus was significantly higher in the Explicit Imitation condition ($N = 79$, $M = 0.0059$, $SD = 0.0031$) than in the Passive Viewing control condition [$N = 68$, $M = 0.0025$, $SD = 0.0023$, $\beta = -0.88$, Wald $\chi^2(1) = 1.43$, $p < 0.001$], but was not significantly different from the Pen-Holding condition [$N = 69$, $M = 0.0066$, $SD = 0.0036$, $\beta = 0.11$, Wald $\chi^2(1) = 98.74$, $p < 0.234$]. The EMG activity in the zygomaticus during the Pen-Holding condition was significantly higher than during the Passive Viewing control condition ($p < 0.001$).

Generalised linear model results for the EMG activity in the *depressor* muscle showed a significant main effect of Experimental Condition [Wald $\chi^2(2) = 538.87$, $p < 0.001$]; see **Figure 2**. Pairwise comparisons showed that the EMG activity in the depressor was significantly lower in the Explicit Imitation condition ($N = 80$, $M = 0.0103$, $SD = 0.0050$) than in the Pen-Holding condition [$N = 70$, $M = 0.0363$, $SD = 0.0243$, $\beta = 1.26$, Wald $\chi^2(1) = 181.33$, $p < 0.001$] and significantly higher in the Explicit Imitation condition than in the Passive Viewing control condition [$N = 69$, $M = 0.0039$, $SD = 0.0028$, $\beta = -0.98$, Wald $\chi^2(1) = 109.86$, $p < 0.001$]. The EMG activity in the depressor during the Pen-Holding condition was significantly higher than during the Passive Viewing control condition ($p < 0.001$).

Generalised linear model results for the EMG activity in the *levator* showed a significant main effect of Experimental Condition [Wald $\chi^2(2) = 222.39$, $p < 0.001$]; see **Figure 2**. Pairwise comparisons showed that the EMG activity in the levator was significantly lower in the Explicit Imitation condition

($N = 76$, $M = 0.0070$, $SD = 0.0033$, $p < 0.001$) than in the Pen-Holding condition [$N = 67$, $M = 0.0127$, $SD = 0.0083$, $\beta = 0.60$, Wald $\chi^2(1) = 50.44$, $p < 0.001$] and significantly higher in the Explicit Imitation condition than in the Passive Viewing control condition [$N = 66$, $M = 0.0034$, $SD = 0.0019$, $\beta = -0.71$, Wald $\chi^2(1) = 68.62$, $p < 0.001$]. The EMG activity in the levator during the Pen-Holding condition was significantly higher than during the Passive Viewing control condition ($p < 0.001$).

Generalised linear model results for the EMG activity in the *corrugator* showed a significant main effect of Experimental Condition [Wald $\chi^2(2) = 27.62$, $p < 0.001$]; see **Figure 2**. Pairwise comparisons showed that the EMG activity in the corrugator was significantly higher in the Explicit Imitation condition ($N = 76$, $M = 0.0081$, $SD = 0.0038$) than in the Passive Viewing control condition [$N = 65$, $M = 0.0054$, $SD = 0.0039$, $\beta = -0.40$, Wald $\chi^2(1) = 16.43$, $p < 0.001$] and the Pen-Holding condition [$N = 67$, $M = 0.0050$, $SD = 0.0036$, $\beta = -0.48$, Wald $\chi^2(1) = 23.48$, $p < 0.001$]. The EMG activity in the corrugator during the Pen-Holding condition was not significantly different than during the Passive Viewing control condition ($p = 0.466$).

Generalised linear model results for the EMG activity in the *frontalis* showed a significant main effect of Experimental Condition [Wald $\chi^2(2) = 26.22$, $p < 0.001$]; see **Figure 2**. Pairwise comparisons showed that the EMG activity in the frontalis was significantly higher in the Explicit Imitation condition ($N = 81$, $M = 0.0072$, $SD = 0.0072$) than in the Pen-Holding condition [$N = 71$, $M = 0.0044$, $SD = 0.0047$, $\beta = -0.50$, Wald $\chi^2(1) = 24.81$, $p < 0.001$] and the Passive Viewing control condition [$N = 70$, $M = 0.0052$, $SD = 0.0044$, $\beta = -0.34$, Wald $\chi^2(1) = 11.35$, $p = 0.001$]. The EMG activity in the frontalis during the Pen-Holding condition was not significantly different from the Passive Viewing control condition ($p = 0.125$).

Hypotheses Testing: Facial Muscle Activity Manipulation and Emotion Recognition Accuracy

Since it was hypothesised that the pen-holding would affect recognition of emotional expressions with facial feature saliency in the lower part of the face but not the upper part of the face, respective variables for the recognition scores were created. The 'lower face saliency' variable included accuracy scores for disgust, happiness, embarrassment, contempt, and pride. The 'upper face saliency' variable included accuracy scores for anger, fear, sadness, and surprise. This categorisation was based on the location of the facial features that are characteristic for each expression (and the number thereof) in the face stimulus set used; a table listing all facial features per emotion category is printed in van der Schalk et al. (2011). A mean accuracy score was calculated across the emotions included in the lower and upper face saliency variables resulting in a maximum accuracy score of nine (i.e., 100%) each, as there were nine trials per emotion category. Since it was hypothesised that explicit imitation of observed emotional expressions would facilitate recognition of all emotions, the two categories (lower and upper face saliency) were combined to retrieve a recognition score across 'all emotion categories.' The maximum possible accuracy score for the latter variable was 18 (i.e., 100%). Whereas analyses were conducted with the accuracy scores, the accuracy scores of the three variables were transformed into percentages in the figures presenting the results to facilitate interpretation.

Only the first experimental condition a participant underwent was included in the between-subject analyses, as the first condition naturally could not have been influenced by former instructions. This between-subject approach decreased the sample size to $n = 28$ for the Passive Viewing control condition (14 male, 14 female) and the Explicit Imitation condition (13 male, 15 female). The sample size was 26 (13 male, 13 female) for the Pen-Holding condition. Three comparisons were conducted using independent samples t-tests to test the hypotheses of the current study. The accuracy scores of the variable 'all emotion categories' from the Explicit Imitation condition were compared to the accuracy scores from the Passive Viewing control condition to test whether enhanced stimulus-congruent facial muscle activation facilitated recognition. To test whether stimulus-incongruent facial muscle activation impeded recognition, the accuracy scores of the variables 'lower face saliency' and 'upper face saliency' from the Pen-Holding were compared to the Passive Viewing control condition. The alpha-level of 5% was Bonferroni-corrected to account for multiple comparisons. The resulting p-values were compared to a p-value of 0.017 ($p = 0.05/3$) for significance determination. Cohen's d is presented as effect size measure.

The independent samples t-test comparing accuracy of response across 'all emotion categories' included in the task from the Explicit Imitation condition ($M = 11.53$, $SD = 1.83$) to the Passive Viewing control condition ($M = 11.57$, $SD = 1.91$) showed no significant difference between the two experimental conditions [$t(54) = -0.79$, $p = 0.938$, Cohen's $d = -0.021$]; see **Figure 3A**.

FIGURE 3 | Mean accuracy of response in percentages for the three variables investigated from the experimental conditions as compared. Each panel visualises the results from one of the three conducted comparisons using independent samples t-tests. **(A)** Accuracy of response from the Explicit Imitation condition and the Passive Viewing control condition *across all emotion categories*. **(B)** Accuracy of response from the Passive Viewing control condition and the Pen-Holding condition for the *emotion categories with saliency in the lower part of the face*. **(C)** Accuracy of response from the Passive Viewing control condition and the Pen-Holding condition for the *emotion categories with saliency in the upper part of the face*. Error bars represent standard errors of the means. *p-value significant

Comparing the accuracy rates of the 'lower face saliency' emotion category using independent samples t-tests showed that the accuracy rates were significantly higher in the Passive Viewing control condition ($M = 5.26$, $SD = 0.93$) than in the Pen-Holding condition ($M = 4.62$, $SD = 0.93$, $t(52) = 2.53$, $p = 0.014$) with a medium to large effect size (Cohen's $d = 0.688$); see **Figure 3B**.

Comparing the accuracy rates of the 'upper face saliency' emotion category using independent samples *t*-tests showed that the accuracy rates from the Passive Viewing control condition ($M = 6.30$, $SD = 1.28$) were not significantly different from the Pen-Holding condition [$M = 6.17$, $SD = 0.84$, $t(52) = 0.44$, $p = 0.663$, Cohen's $d = 0.123$]; see **Figure 3C**.

DISCUSSION

The current study investigated the effects of active facial muscle manipulations in observers on their ability to recognise emotions from others' faces. Results showed that facial muscle manipulations effectively changed observers' facial muscle activity. Holding a pen in the mouth increased the activity of facial muscles in the lower face region compared to a control condition with no facial movement manipulation, while explicit imitation of observed facial emotion produced enhanced facial muscle activity across the face compared to the control condition. In line with the facial muscle manipulation, holding a pen in the mouth was found to produce lower accuracy for recognising facial displays of emotion when the most salient facial feature was in the lower face region compared to passively viewing emotional expressions. In contrast, explicitly imitating the emotional expression seen in others did not result in greater recognition of these emotional expressions compared to passive viewing of the videos. The current findings provide support for embodied cognition accounts, but only when the experimental condition involved stimulus-incongruent facial muscle activity while observing emotional expressions in others, and not when the condition involved stimulus-congruent facial muscle activity. The methodological implications for investigations like the current research with a within-subject study design are discussed.

Based on embodied cognition accounts, it was hypothesised that explicit facial muscle activity that is congruent with the observed facial expression would increase recognition rates compared to passive viewing. While explicitly imitating the perceived facial expressions of emotion by others in videos resulted in higher facial muscle activity compared to when they passively viewed the facial expressions, results showed the explicit imitation of others emotions had no facilitating effect on facial emotion recognition. These results are in line with those by Schneider et al. (2013), who similarly reported EMG results showing differences in facial muscle activation between the Explicit Imitation condition and their other two experimental conditions, but no corresponding increase in emotion recognition compared to passive viewing. It was assumed that if automatic subtle stimulus-congruent facial muscle activation facilitated facial emotion recognition (e.g., Oberman et al., 2007), then increasing muscle intensity (i.e., explicit imitation) should increase recognition even more when comparing to a control condition. Though, a study by Hess and Blairy (2001) investigated the intensity of facial mimicry in relation to facial emotion recognition and did not find evidence for a facilitating effect on decoding accuracy due to increased intensity of stimulus-congruent automatic facial muscle activation in observers. Together, these results imply

that increased intensity of observers' stimulus-congruent facial muscle activation does not facilitate recognition. In this case, it is even possible that congruent facial muscle activation in general does not facilitate facial emotion recognition, as reported by Rives Bogart and Matsumoto (2010) based on absent stimulus-congruent facial muscle activity in individuals with face paralysis (i.e., Moebius syndrome) and no different performance at facial emotion recognition compared to non-paralysed controls.

As expected, holding a pen in the mouth caused increased EMG activity in the muscles of the lower face region, especially the depressor muscle. Further in line with the predictions, accuracy scores were significantly lower in the Pen-Holding condition compared to the Passive Viewing control condition when recognising emotional expressions with feature saliency in the lower face region. Effects for mouth movement manipulations on recognition of emotions with saliency in the lower face region are in line with previous studies. For example, disgust and happiness recognition are impaired when mouth movements are manipulated with a pen compared to passive viewing without facial movement manipulation (Oberman et al., 2007; Ponari et al., 2012). For both emotions, the salient facial feature of the corresponding facial expression is situated in the lower face region (mouth and nose, respectively) (Leppänen and Hietanen, 2007; Calvo and Nummenmaa, 2008; Khan et al., 2012). Oberman et al. (2007) interpreted their finding of disgust and happiness recognition being diminished in the Pen-Holding condition compared to the Passive Viewing condition as facial mimicry being a necessary component of facial emotion recognition based on the hindrance of facial mimicry during the Pen-Holding. This explanation does not align with the finding from the current study that stimulus-congruent facial muscle activation did not facilitate recognition.

An alternative interpretation is that facial muscle activations, as achieved through pen-holding, induce facial muscle feedback that is incongruent with the muscle activation underlying the observed facial expression. Since embodiments also include the typical facial expressions of emotions, it was proposed that facial muscle feedback in an observer that is in conflict with the perceived visual information might hamper recognition (Wood et al., 2016). Stimulus-incongruency in facial muscle activation can be determined anatomically. Whereas smiling (through zygomaticus activation for happiness expression) and nose wrinkling (through levator activation for disgust expression) are upward movements, holding a pen in the mouth is an action in the opposite direction, indicating antagonist muscle activation. Importantly, it should be noted that antagonist muscles initiate movement in opposing directions and can thus not be activated simultaneously; this is anatomically impossible (Stennert, 1994). The EMG data from the current study showed that the pen in the mouth induced the greatest muscle activity in the depressor, which indeed is the antagonist muscle to the levator (which itself is a synergist to the zygomaticus). As antagonist muscle, depressor activation produces muscle feedback that is incompatible with smiling/nose wrinkling. The incongruency in facial muscle activation from the pen-holding could have interfered with the embodied representation of the emotions involving facial feature saliency in the lower face region.

Observing a facial expression with facial feature saliency in the mouth region (e.g., happiness) would elicit the representation of that emotion, but with a pen in the mouth (i.e., depressor activation), there would be a contradiction in the incoming sensory information. This is because concurrent depressor activation would elicit an association with an emotion whose facial expression involves the depressor. The conflicting muscle activations and the resulting muscle feedback could potentially make recognition of emotional expressions with facial feature saliency in the lower face more difficult. This interpretation aligns with an EEG study that demonstrated that the understanding of facial emotion (i.e., semantic retrieval demand) with facial feature saliency in the lower face region is impaired by active manipulation of muscle activity around the mouth (Davis et al., 2017). Together, these results suggest that recognition is diminished when there is interference between visual and motor information, in line with the wider literature on action-perception matching based on representations (Wohlschläger, 2000; Brass et al., 2001; for a review article see Blakemore and Decety, 2001).

Limitations, Methodological Considerations, and Future Research

The Explicit Imitation condition and the Pen-Holding condition required additional action from the participants as opposed to the Passive Viewing condition. It could be argued that the results from the current study are based on the additional cognitive load the experimental conditions imposed rather than specific effects of the manipulations. However, Tracy and Robins (2008) demonstrated across two studies that participants are able to accurately recognise emotions, even more complex emotions like pride and embarrassment, under cognitive load. It seems thus unlikely that the findings from the current study are the result of cognitive load. There was a different number of emotional categories included in this study with saliency in the upper part of the face (4) compared to those in the lower face region (5), and this difference could have affected the results.

The current study manipulated the muscles of the lower face region, but not the muscles of the upper face region. Future research should systematically test the effects of stimulus-incongruent muscle activity across the entire face on facial emotion recognition. Researchers have attempted to fix facial muscles in the upper face region by instructing participants to perform certain facial movements (e.g., Ponari et al., 2012). It is likely that such performed facial action (e.g., drawing eyebrows together) is associated with a specific emotional facial expression even if only partially. To overcome this limitation, it could be instructed that participants activate a specific muscle and the effect on recognition of emotional expressions that involve mainly other muscles could be investigated. For example, participants could be asked to smile, frown, wrinkle their nose, etc. each across a set amount of trials displaying varying emotional expressions. Then it could be investigated if stimulus-incongruent facial movements decrease recognition compared to stimulus-congruent expressions. This approach would allow to identify for which muscle interference has the greatest impact on the recognition of specific emotions. These

results could have implications for individuals receiving Botox treatments.

Results from the within-subject analyses of the current study showed that the accuracy rates from the Pen-Holding condition were comparable to the Passive Viewing control condition, against the expectation for the emotions with saliency in the lower face region. This finding can, however, be explained by a combination of two occurrences. The first occurrence was the necessary data eliminations, which lead to uneven numbers of participants for the six versions of the experiment. More participants underwent the Passive Viewing control condition first in the experiment sequence than last, while the number of participants per order in the Pen-Holding condition was similar. The second occurrence was the increase in recognition accuracy over the course of the experiment producing higher recognition rates in the last experimental condition a participant underwent. Combining these two occurrences resulted in lower mean accuracy scores for the Passive Viewing control condition, making the mean similar to the mean from the Pen-Holding condition instead of higher. The small albeit non-significant increase in facial emotion recognition when explicitly imitating observed facial expressions compared to the Passive Viewing control condition from the current study can also be explained by the combination of necessary data eliminations and an increase in recognition accuracy over the course of the experiment, as most participants included in the analyses underwent the Explicit Imitation condition last in the experiment. Consequently, theoretical interpretation of the findings from the within-subject analyses of the current study is problematic.

The advantage of a within-subject design is usually that the found effects are the result of the experimental manipulations and not due to potential differences between samples as can be the case in between-subject designs, thereby reducing the error variance. However, the instruction to explicitly imitate the observed facial expressions turned out to have a lasting effect on more than a few participants in the current study. Those participants showed a similar pattern of facial muscle activation in the Pen-Holding condition and Passive Viewing control condition as during the Explicit Imitation condition when the Explicit Imitation condition preceded these conditions. This occurrence indicates that explicit imitation was carried out in the other experimental conditions as well (and led to data loss in the current study). This occurrence is very important to consider for researchers who are intending to conduct research similar to the current study. To avoid data eliminations and potential resulting data confounding effects (see next paragraph), it is advisable to apply a between-subject design. Nonetheless, the instruction to explicitly imitate facial emotional expressions having such a long-lasting effect constitutes an interesting finding in itself. The question why some people automatically keep imitating expressions against the task instructions should gain further attention in future research of this type. Example research questions to address could be: Are these individuals more likely to experience emotion contagion? Do those individuals possess higher empathy?

Further noteworthy is the continuous increase in accuracy of response over the course of the experiment in the current study,

independent of the instructions given to participants for the various experimental conditions. The resulting methodological implication is the importance to balance the order of presentation of the experimental conditions when using a within-subject design (as done with the current study) or to apply a between-subject design. The latter option is recommendable when it is likely that unequal amounts of participants will be excluded per order of experimental condition. Nonetheless, that accuracy rates do increase even without the explicit feedback about the correctness of the response is interesting. It indicates some sort of underlying learning processes and it is possible that focussing attention on decoding of facial emotion also outside the laboratory in everyday social interactions might lead to improvements in facial emotion recognition, which could be particularly relevant for clinical populations with impairments in facial emotion recognition.

CONCLUSION

Taken together, the current study showed that explicit stimulus-incongruent facial muscle activations in observers hamper recognition compared to passively viewing expressions. It was further demonstrated that explicit stimulus-congruent facial muscle activation does not lead to a facial emotion recognition advantage compared to passively viewing expressions. This latter finding is peculiar since awareness was added to the stimulus-congruent facial muscle activations and the facial muscle activation was explicit (i.e., explicit imitation). Nonetheless, the results from the current study imply that stimulus-congruent facial muscle activations in observers have no facilitating effect on facial emotion recognition and that only stimulus-incongruent facial muscle activations hamper recognition. Given that observing facial emotion might elicit an emotion representation, incongruency between an observed emotion and the facial activity in the observer's face might disrupt the encoding process due to the embodiment of facial emotional expressions, in line with embodied cognition accounts of emotion.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the University of Bath Psychology Ethics Committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Bath Psychology Ethics Committee.

AUTHOR CONTRIBUTIONS

TW conceptualised, designed, and carried out the study. MB, CA, and MoP supervised TW. MiP and TW prepared the data for analyses. MoP provided the means for data preparation. TW analysed the data and interpreted the results. TW wrote the manuscript. All authors edited the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00864/full#supplementary-material

REFERENCES

Achaibou, A., Pourtois, G., Schwartz, S., and Vuilleumier, P. (2008). Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia* 46, 1104–1113. doi: 10.1016/j.neuropsychologia.2007.10.019

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Blairy, S., Herrera, P., and Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *J. Nonverbal Behav.* 23, 5–41. doi: 10.1023/A:1021370825283

Blakemore, S.-J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567. doi: 10.1038/35086023

Brass, M., Bekkering, H., and Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychol.* 106, 3–22. doi: 10.1016/S0001-6918(00)00024-X

Calvo, M. G., and Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *J. Exp. Psychol. Gen.* 137, 471–494. doi: 10.1037/a0012771

Conson, M., Ponari, M., Monteforte, E., Ricciato, G., Sarà, M., Grossi, D., et al. (2013). Explicit recognition of emotional facial expressions is shaped by expertise: evidence from professional actors. *Front. Psychol.* 4:382. doi: 10.3389/fpsyg.2013.00382

Davis, J. D., Winkielman, P., and Coulson, S. (2017). Sensorimotor simulation and emotion processing: impairing facial action increases semantic retrieval demands. *Cogn. Affect. Behav. Neurosci.* 17, 652–664. doi: 10.3758/s13415-017-0503-2

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology* 19, 643–647. doi: 10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U., and Thunberg, M. (1998). Rapid facial reactions to emotional facial expressions. *Scand. J. Psychol.* 39, 39–45. doi: 10.1111/1467-9450.00054

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Foroughi, C. K., Sibley, C., and Coyne, J. T. (2017). Pupil size as a measure of within-task learning. *Psychophysiology* 54, 1436–1443. doi: 10.1111/psyp.12896

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Hess, U., and Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *Int. J. Psychophysiol.* 40, 129–141. doi: 10.1016/S0167-8760(00)00161-6

Hess, U., and Fischer, A. (2014). Emotional mimicry: why and when we mimic emotions. *Soc. Personal. Psychol. Compass* 8, 45–57. doi: 10.1111/spc3.12083

Khan, R. A., Meyer, A., Konik, H., and Bouakaz, S. (2012). "Human vision inspired framework for facial expressions recognition," in *Proceedings of the 19th Conference on Image Processing 2012* (Orlando, FL: IEEE International), 2593–2596.

Leppänen, J. M., and Hietanen, J. K. (2007). Is there more in a happy face than just a big smile? *Vis. Cogn.* 15, 468–490. doi: 10.1080/13506280600765333

Likowski, K. U., Mühlberger, A., Gerdes, A. B., Wieser, M. J., Pauli, P., and Weyers, P. (2012). Facial mimicry and the mirror neuron system: simultaneous acquisition of facial electromyography and functional magnetic resonance imaging. *Front. Hum. Neurosci.* 6:214. doi: 10.3389/fnhum.2012.00214

Neal, D. T., and Chartrand, T. L. (2011). Embodied emotion perception: amplifying and dampening facial feedback modulates emotion perception accuracy. *Soc. Psychol. Personal. Sci.* 2, 673–678. doi: 10.1177/1948550611406138

Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316, 1002–1005. doi: 10.1126/science.1136930

Oberman, L. M., Winkielman, P., and Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Soc. Neurosci.* 2, 167–178. doi: 10.1080/17470910701391943

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., and Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion* 12, 1398–1403. doi: 10.1037/a0028588

Rives Bogart, K., and Matsumoto, D. (2010). Facial mimicry is not necessary to recognize emotion: facial expression recognition by people with Moebius syndrome. *Soc. Neurosci.* 5, 241–251. doi: 10.1080/17470910903395692

Sato, W., Fujimura, T., Kochiyama, T., and Suzuki, N. (2013). Relationships among facial mimicry, emotional experience, and emotion recognition. *PLoS One* 8:e57889. doi: 10.1371/journal.pone.0057889

Sato, W., and Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, 104, 1–18. doi: 10.1016/j.cognition.2006.05.001

Schneider, K. G., Hempel, R. J., and Lynch, T. R. (2013). That "poker face" just might lose you the game! The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion. *Emotion* 13, 852–866. doi: 10.1037/a0032847

Stel, M., and van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychol. Sci.* 19, 984–985. doi: 10.1111/j.1467-9280.2008.02188.x

Stennert, E. (1994). Why does the frontalis muscle "never come back"? functional organization of the mimic musculature. *Facial Nerve* 91–95. doi: 10.1007/978-3-642-85090-5_25

Tracy, J. L., and Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion* 8, 81–95. doi: 10.1037/1528-3542.8.1.81

van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

Wallbott, H. G. (1991). Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *Br. J. Soc. Psychol.* 30, 207–219. doi: 10.1111/j.2044-8309.1991.tb00939.x

Wilhelm, F. H., and Peyk, P. (2005). *ANSLAB: Autonomic Nervous System Laboratory (Version 2.6)*. Available at: http://www.sprweb.org

Wingenbach, T. S., Ashwin, C., and Brosnan, M. (2016). Validation of the Amsterdam dynamic facial expression set - bath intensity variations (ADFES-BIV): a set of videos expressing low, intermediate, and high intensity emotions. *PLoS One* 11:e0147112. doi: 10.1371/journal.pone.0147112

Wingenbach, T. S. H., Brosnan, M., Pfaltz, M. C., Peyk, P., and Ashwin, C. (n.d.). Distinct facial mimicry patterns for discrete emotions.

Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Res.* 40, 925–930. doi: 10.1016/S0042-6989(99)00239-4

Wood, A., Rychlowska, M., Korb, S., and Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends Cogn. Sci.* 20, 227–240. doi: 10.1016/j.tics.2015.12.010

Dynamics Matter: Recognition of Reward, Affiliative, and Dominance Smiles From Dynamic vs. Static Displays

Anna B. Orlowska[1]*, Eva G. Krumhuber[2], Magdalena Rychlowska[3] and Piotr Szarota[1]

[1] Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland, [2] Department of Experimental Psychology, University College London, London, United Kingdom, [3] School of Psychology, Queen's University Belfast, Belfast, United Kingdom

Smiles are distinct and easily recognizable facial expressions, yet they markedly differ in their meanings. According to a recent theoretical account, smiles can be classified based on three fundamental social functions which they serve: expressing positive affect and rewarding self and others (reward smile), creating and maintaining social bonds (affiliative smile), and negotiating social status (dominance smiles) (Niedenthal et al., 2010; Martin et al., 2017). While there is evidence for distinct morphological features of these smiles, their categorization only starts to be investigated in human faces. Moreover, the factors influencing this process – such as facial mimicry or display mode – remain yet unknown. In the present study, we examine the recognition of reward, affiliative, and dominance smiles in static and dynamic portrayals, and explore how interfering with facial mimicry affects such classification. Participants (N = 190) were presented with either static or dynamic displays of the three smile types, whilst their ability to mimic was free or restricted via a pen-in-mouth procedure. For each stimulus they rated the extent to which the expression represents a reward, an affiliative, or a dominance smile. Higher than chance accuracy rates revealed that participants were generally able to differentiate between the three smile types. In line with our predictions, recognition performance was lower in the static than dynamic condition, but this difference was only significant for affiliative smiles. No significant effects of facial muscle restriction were observed, suggesting that the ability to mimic might not be necessary for the distinction between the three functional smiles. Together, our findings support previous evidence on reward, affiliative, and dominance smiles by documenting their perceptual distinctiveness. They also replicate extant observations on the dynamic advantage in expression perception and suggest that this effect may be especially pronounced in the case of ambiguous facial expressions, such as affiliative smiles.

Keywords: smile, facial expression, emotion, dynamic, mimicry

INTRODUCTION

A smile can be simply described as a contraction of the *zygomaticus major* - a facial muscle which pulls the lip corners up toward the cheekbones (Ekman and Friesen, 1982), named by Duchenne de Boulogne (1862/1990) "a muscle of joy." This unique movement makes it an easily recognizable facial expression. However, smiles can also be confusing in their meanings and functions they

serve. Despite the association between smiles and positive feelings and intentions (Ekman et al., 1990), trust (Krumhuber et al., 2007) and readiness to help (Vrugt and Vet, 2009), smiles can also be displayed during unpleasant experiences, i.e., to hide negative feelings (Ekman and Friesen, 1982), and be perceived as a signal of lower social status (Ruben et al., 2015) or intelligence (Krys et al., 2014). Smiling is therefore used in a wide variety of situations, depending on the context and social norms learned through socialization and experience. Not only can the use of smiles and their social function vary considerably (e.g., Szarota et al., 2010; Rychlowska et al., 2015), but the very expression of a smile comes in many forms. This is because the contraction of the *zygomaticus major* muscle [defined as Action Unit (AU) 12 in the Facial Action Coding System; Ekman et al., 2002] – the core feature of any smile expression – often involves the activation of other facial muscles, creating a range of possible variations. Ekman (2009), for example, identified and described 18 types of smiles, differentiated in terms of their appearance and the situation in which they are likely to occur. Moreover, AU12 can be accompanied by the presence of other AUs and thus convey emotions such as disgust or surprise (Du et al., 2014; Calvo et al., 2018).

Despite its variability, the most commonly used smile typology is the distinction between 'true'/genuine and 'fake'/false smiles, with the former being sincere displays of joy and amusement, and the latter being produced voluntarily, possibly to increase others' trust and cooperation (Frank, 1988). True and false smiles can be distinguished on the basis of their morphology: the presence of supposedly involuntary eye constriction (AU6 – the contraction of the *orbicularis oculi* muscle), a classic criterion based on early studies by Duchenne de Boulogne (1862/1990). Although the true vs. false smile typology is parsimonious and extensively documented in the literature, it is not without shortcomings. Specifically, contemporary empirical evidence reveals that people are able to deliberately show Duchenne smiles (Krumhuber and Manstead, 2009; Gunnery et al., 2013), thereby limiting the usefulness of this criterion. More importantly, however, the binary nature of the typology fails to account for the variability of smiles produced in everyday life. People smile in many situations, involving diverse emotions or very little emotion. Some expressions undeniably convey more positive affect than others. However, the assertion that all smiles which fail to reflect joy and amusement must be false and potentially manipulative, seems oversimplifying. It is at least theoretically possible that an enjoyment smile is just one among many true smiles.

An alternative theoretical account proposes that smiles can be classified in accordance to how they affect people's behavior in the service of fundamental tasks of social living (Niedenthal et al., 2010; Martin et al., 2017). This typology defines three physically distinct smiles of *reward, affiliation,* and *dominance,* which serve the main function of social communication and interaction (Niedenthal et al., 2013). Reward smiles communicate positive emotions and sensory states such as happiness or amusement, thereby potentially rewarding both the sender and the perceiver. Affiliative smiles communicate positive social motives and are used to create and maintain social bonds. A person displaying an affiliative smile intends to be perceived as friendly and polite. Finally, dominance smiles are used to impose and maintain higher social status. The person displaying this type of smile intends to be perceived as superior. Recent research by Rychlowska et al. (2017, Study 1) explored the physical appearance of reward, affiliative, and dominance smiles, including a description of the facial characteristics of each category, suggesting that the three functional smiles are indeed morphologically different. In a subsequent experiment (Rychlowska et al., 2017, Study 2), computer-generated animations of reward, affiliative, and dominance smiles were categorized by human observers and a Bayesian classifier. Despite the generally high categorization accuracy for all three smile types, human and Bayesian performance was lowest for the affiliative smiles, arguably because of their similarity to the reward smiles, as both expressions convey positive social signals and they both involve a symmetrical movement of the *zygomaticus major* muscle.

Given the multiple types of smiles, the diversity of situations in which they appear, and the varying display rules governing their production, the understanding of these facial expressions is a complex process which can rely on multiple mechanisms – such as a perceptual analysis of the expresser's face, conceptual knowledge about the expresser and the situation, and sensorimotor simulation (Niedenthal et al., 2010; Calvo and Nummenmaa, 2016). This last construct involves the recreation of smile-related feelings and neural processes in the perceiver, and is closely related to facial mimicry, which is defined as a spontaneous rapid imitation of other people's expressions (Dimberg et al., 2000). As sensorimotor simulation involves a complex sequence of motor, neural, and affective processes (see Wood et al., 2016, for review), it is more costly than other forms of facial expression processing. Hence, it may be preferentially used for the interpretation of expressions that are important for the observer or non-prototypical, and thus hard to classify (Niedenthal et al., 2010).

Existing literature suggests that facial mimicry, often used to index sensorimotor simulation of emotion expressions, is sensitive to social and contextual factors. Its occurrence may depend on the type of expression observed (Hess and Fischer, 2014), but also on the social motivation (Fischer and Hess, 2017), attitudes toward the expresser (e.g., Likowski et al., 2008), and group status (e.g., Sachisthal et al., 2016). Furthermore, it can be experimentally altered or restricted in laboratory settings using various pen-in-mouth procedures, stickers, chewing-gum, or sports mouthguards. In these cases, preventing mimicry responses has been shown to impair observers' ability to accurately recognize happiness and disgust (Oberman et al., 2007; Ponari et al., 2012) and discriminate between false and genuine smiles (Maringer et al., 2011; Rychlowska et al., 2014).

Parallel to these findings, the results of other studies investigating the role of facial mimicry in emotion recognition were not conclusive (e.g., Blairy et al., 1999; Korb et al., 2014). Several factors could explain such inconsistencies: First, measuring rather than blocking facial mimicry may not necessarily show its involvement in expression recognition. Also, facial mimicry could be more implicated in recognition tasks that are especially difficult, i.e., when classifying low-intensity facial

expressions or judging subtle variations between different types of a given facial expression (Hess and Fischer, 2014). This makes the interpretation of smiles an especially useful paradigm for studying the role of facial mimicry.

Another potential explanation for disparate research findings could be related to the way in which the stimuli are presented. Previous studies using facial electromyography (EMG; e.g., Sato et al., 2008; Rymarczyk et al., 2016) reveal that dynamic video stimuli lead to enhanced mimicry in comparison with static images. In particular, higher intensities of AU12 and AU6 – the core smile movements – have been reported when participants watched dynamic rather than static expressions of happiness. Dynamic materials have higher ecological validity (Krumhuber et al., 2013, 2017), given that in everyday social encounters facial expressions are moving and rapidly changing depending on the situation. As emotion processing is not only based on the perception of static configurations of facial muscles, but also on how the facial expression unfolds (Krumhuber and Scherer, 2016), dynamic displays provide additional information which is not present in static images. Furthermore, past research reveals better recognition and higher arousal ratings of emotions when they are shown in dynamic than static form (e.g., Hyniewska and Sato, 2015; Calvo et al., 2016). Dynamic displays may therefore provide relevant cues which facilitate the decoding of facial expressions.

The present work focuses on the distinction between the three functional smiles of reward, affiliation, and dominance (Niedenthal et al., 2010). Instead of using computer-generated faces as done by Rychlowska et al. (2017), we employed static images and dynamic videos of human actors displaying the three types of smiles. Our experiment extends previous research (Rychlowska et al., 2017; Martin et al., 2018) in three ways by testing (1) how accurately naïve observers can discriminate between the three functional smiles, (2) whether the capacity to classify these smiles is affected by facial muscle restriction that prevents mimicry responses, and (3) whether the type of display (static vs. dynamic) influences smile recognition, thereby moderating the potential effects of muscle restriction. In line with previous findings (Rychlowska et al., 2017), we predict that observers should be able to accurately classify the three functional smiles, with affiliative smiles being more ambiguous than reward and dominance smiles. We also anticipate that, consistent with previous work (Maringer et al., 2011; Rychlowska et al., 2014), facial muscle restriction should disrupt participants' ability to interpret the three smile types. Finally, we hypothesize that impairments in smile classification in the muscle restriction condition should be especially strong in the static, rather than dynamic condition, given the relative smaller amount of information provided by stimuli of static nature.

MATERIALS AND METHODS

Participants and Design

The study had a three-factorial experimental design with the stimulus display (dynamic vs. static) and muscle condition (free vs. restricted) as between-subject variables, and smile type (reward, affiliative, dominance) as within-subject variable. A total of 190 participants, mostly students at University College London, were recruited and voluntarily took part in the study in exchange for a £2 voucher or course credits. One hundred seventy-eight subjects identified themselves as White and 12 as mixed race. Technical failure resulted in the loss of data for two participants, leaving a final sample of 188 participants (137 women), ranging in age between 18 and 45 years ($M = 22.2$ years, $SD = 4.2$). A power analysis using G*Power 3.1 (Faul et al., 2007) for a $3 \times 2 \times 2$ interaction, assuming a medium-sized effect (Cohen's $f = 0.25$) and a 0.5 correlation between measures, indicated that this sample size would be sufficient for 95% power. All participants had normal or corrected-to-normal vision. Ethical approval for the present study was granted by the UCL Department of Psychology Ethics Committee.

Materials

Stimuli were retrieved from a set developed by Martin et al. (2018) and featured eight White actors (four female) in frontal view, expressing the three smile types: reward smile (eight stimuli), affiliative smile (eight stimuli), and dominance (six stimuli) smile. Actors posed each smile type after being coached about its form and accompanying social motivations (see Martin et al., 2017; Rychlowska et al., 2017). In morphological terms (FACS, Ekman et al., 2002), reward smiles consisted of Duchenne smiles that were characterized by symmetrical activation of the Lip Corner Puller (AU12), the Cheek Raiser (AU6), Lips Part (AU25) and/or Jaw Drop (AU26). Affiliative smiles consisted of Non-Duchenne smiles that involved the Lip Corner Puller (AU12), the Chin Raiser (AU17), with or without Brow Raiser (AU1-2). Dominance smiles consisted of asymmetrical Non-Duchenne smiles (AU12L or AU12R), with additional actions, such as Head Up (AU53), Upper Lip Raiser (AU10), and/or and Lips Part (AU25) (see **Figure 1**). We employed both static and dynamic portrayals of each smile expression, netting 22 static and 22 dynamic stimuli. Dynamic stimuli were short videoclips (2.6 s) which showed the face changing from non-expressive to peak emotional display. Static stimuli consisted of a single frame of the peak expression. All stimuli were displayed in color on white backgrounds (size: 960×540 pixels).

Procedure

Participants were tested individually in the laboratory. After providing informed consent, they were randomly assigned to one of the four experimental conditions, resulting in approximately 47 people per cell. Using the Qualtrics software (Provo, UT, United States), participants were instructed that they would view a series of smile expressions. Their task was to classify the smiles into three categories. The following brief definitions of each smile type, informed by previous research Rychlowska et al. (2017), were provided: (a) *reward smile*: "a smile displayed when someone is happy, content or amused by something," (b) *affiliative smile*: "a smile which communicates positive intentions, expresses a positive attitude to another person or is used when someone wants to be polite," and (c) *dominance smile*: "a smile displayed when someone feels superior, better and more

FIGURE 1 | Example of a reward smile **(A)**, affiliative smile **(B)**, and dominance smile **(C)** at the peak intensity of the display.

competent or wants to communicate condescension toward another person."

In addition to these smile descriptions, participants were given examples of situations in which each type of expression was likely to occur: (a) *reward smile*: "being offered a dream job or seeing a best friend, not seen for a long time," (b) *affiliative smile:* "entering a room for a job interview or greeting a teacher," (c) *dominance smile:* "bragging to a rival about a great job offer, meeting an enemy after winning an important prize." Situational descriptions were pre-tested in a pilot study, in which participants (N = 33) were asked to choose amongst the three functional smiles the expression that best matched a particular situation (from a pool of 13 situational descriptions). For the present study, we selected the situation that was judged to be the most appropriate for each type of smile expression (selection frequency: reward: 94%, affiliative, 93%, dominance: 75%).

During the muscle restriction condition, participants were informed that people were more objective in their judgments of emotions when their facial movements were restrained. A similar cover story has been used by Maringer et al. (2011). In order to inhibit the relevant facial muscles, participants were to hold a pencil sideways, using both lips and teeth, without exerting any pressure (for a similar procedure see Niedenthal et al., 2001; Maringer et al., 2011). The experimenter demonstrated the correct way of holding the pen in the mouth, and only after the experimenter was satisfied with the pen holding technique, the experiment was started. There was no additional instruction in the free muscle condition.

After some comprehension checks of the three types of smile expressions, participants were presented with static or dynamic versions of the 22 stimuli, shown in a random sequence at the center of the screen. Dynamic sequences were played in their entire length; static photographs were displayed for the same length as the videos (2.6 s). For each stimulus, participants rated their confidence (from 0 to 100%) about the extent to which the expression was a reward, an affiliative, or a dominance smile. If they felt that more than one category applied, they could respond using multiple sliders to choose the exact confidence levels for each response category. Ratings across the three response categories had to sum up to 100%. We defined classification accuracy as the likelihood of correctly classifying a smile expression in line with the predicted target label (reward, affiliation, dominance). After completion of the experiment, participants were debriefed and thanked.

RESULTS

Smile Classification

To test whether the three functional smiles are correctly classified by naïve observers, we calculated the mean confidence ratings for correct (i.e., function-consistent) answers for each smile type (accuracy rates). A 2 (stimulus display: static, dynamic) $\times$ 2 (muscle condition: free, restricted) $\times$ 3 (smile type: reward, affiliative, dominance) ANOVA, with smile type as within-subjects variable, and classification accuracy as the dependent measure yielded significant main effects of smile type, $F(2,368) = 17.41$, $p < 0.001$, $\eta_p^2 = 0.09$, and stimulus display, $F(1,184) = 13.51$, $p < 0.001$, $\eta_p^2 = 0.07$. The two main effects were qualified by a significant interaction between smile type and display, $F(2,368) = 3.99$, $p = 0.021$, $\eta_p^2 = 0.02$. The main effect of muscle condition, $F(1,184) = 0.89$, $p = 0.348$, $\eta_p^2 = 0.01$, the smile type by muscle condition interaction $F(2,368) = 2.71$, $p = 0.070$, $\eta_p^2 = 0.01$, the display by muscle condition interaction $F(1,184) = 1.90$, $p = 0.170$, $\eta_p^2 = 0.01$, and the smile type, display and muscle condition interaction $F(2,368) = 0.16$, $p = 0.845$, $\eta_p^2 = 0.001$, were not significant.

The main effect of smile type revealed that reward smiles ($M = 66.25$, $SD = 16.37$) and dominance smiles ($M = 64.47$, $SD = 17.98$) were recognized more accurately than affiliative smiles ($M = 57.70$, $SD = 15.75$, $ps < 0.001$, Bonferroni-corrected). The difference in recognition rates between reward and dominance smiles was not significant ($p = 0.29$, Bonferroni-corrected). The main effect of stimulus display revealed that recognition rates of the three smile types were higher in the dynamic ($M = 65.80$, $SD = 9.92$) than static condition ($M = 59.98$, $SD = 11.71$). However, decomposing the significant interaction between smile type and display with simple effects analyses revealed that affiliative smiles were recognized more accurately in the dynamic ($M = 63.06$, $SD = 13.04$) than static condition ($M = 52.30$, $SD = 16.47$), $F(1,184) = 24.32$, $p < 0.001$, $\eta_p^2 = 0.12$. No significant differences between the dynamic and static condition emerged for the recognition of reward smiles, $F(1,184) = 0.94$, $p = 0.335$, $\eta_p^2 = 0.01$, and dominance smiles, $F(1,184) = 2.87$, $p = 0.092$, $\eta_p^2 = 0.02$ (see **Figure 2**).

Smile Confusions

The confusion matrix in **Table 1** provides a detailed overview of true (false) positives and true (false) negatives in smile
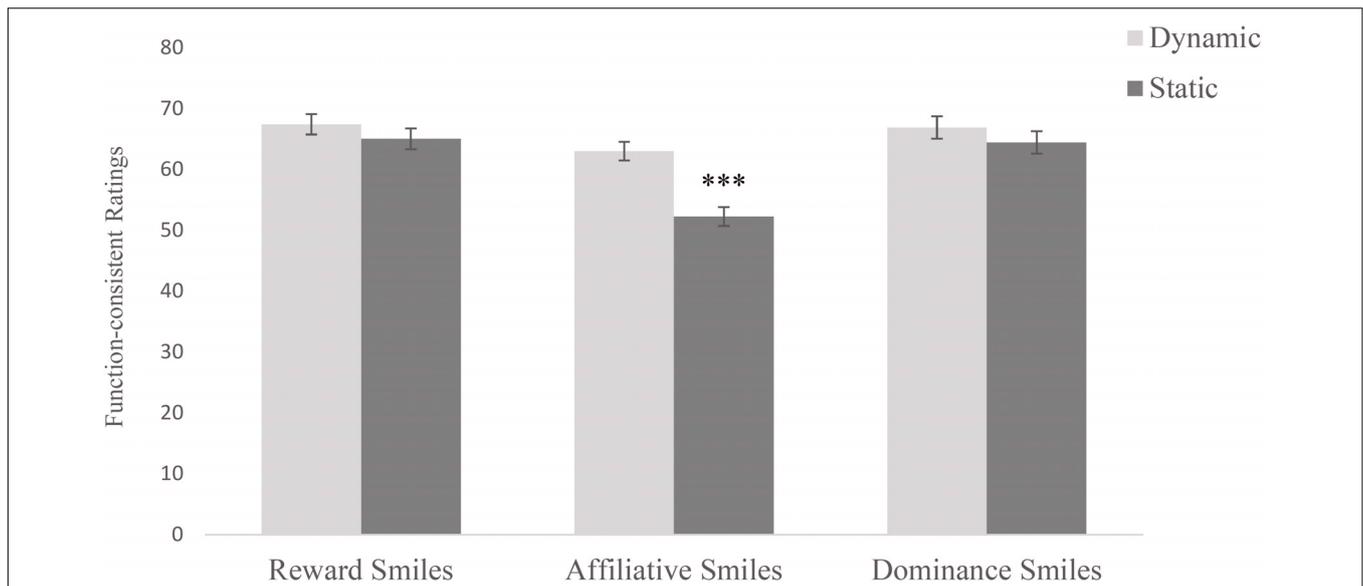
FIGURE 2 | Function-consistent mean ratings (accuracy rates) of the three smile types in the dynamic and static condition. Error bars represent standard errors of the mean. The asterisks indicate a significant difference in the mean ratings between static and dynamic condition ($p < 0.001$).

classification. In order to analyze the type of confusions within a smile type, we followed established procedures (see Calvo and Lundqvist, 2008) and submitted function-consistent and function-inconsistent ratings of the smile expressions to a 2 (stimulus display: static, dynamic) $\times$ 2 (muscle condition: free, restricted) $\times$ 3 (smile type: reward, affiliative, dominance) $\times$ 3 (response: reward, affiliative, dominance) ANOVA, with smile type and response as within-subjects factors. The results revealed a significant main effect of smile type, $F(2,368) = 867.54$, $p < 0.001$, $\eta_p^2 = 0.83$, and response, $F(2,368) = 36.11$, $p < 0.001$, $\eta_p^2 = 0.16$, as well as a significant interaction between the two factors, $F(4,736) = 726.55$, $p < 0.001$, $\eta_p^2 = 0.80$. The interaction between smile type, response, and stimulus display was also significant $F(4,736) = 9.20$, $p < 0.001$, $\eta_p^2 = 0.05$. The response by stimulus display interaction, $F(2,368) = 1.76$, $p = 0.177$, $\eta_p^2 = 0.01$, the smile type, response, and muscle condition interaction $F(4,736) = 2.28$, $p = 0.080$, $\eta_p^2 = 0.01$, as well as the interaction between smile type, response, stimulus display and

muscle condition $F(4,736) = 1.37$, $p = 0.252$, $\eta_p^2 = 0.01$, were not significant.

To decompose the three-way interaction, we examined the interactive effect of response and display separately for each smile type. The interaction of response (reward, affiliative, dominance) and display (static, dynamic) was not significant for the confusions of reward smiles, $F(2,372) = 0.78$, $p = 0.459$, $\eta_p^2 = 0.004$, and dominance smiles, $F(2,372) = 2.8$ $p = 0.062$, $\eta_p^2 = 0.02$, suggesting that the classification of these smiles was similar in both display conditions.

However, the interaction of response and display was significant for the confusion of affiliative smiles, $F(2,372) = 18.57$, $p < 0.001$, $\eta_p^2 = 0.09$. Overall, these smiles were rated higher on affiliation ($M = 57.70$, $SD = 15.75$) than dominance ($M = 32.47$, $SD = 15.39$) and reward ($M = 9.83$, $SD = 9.58$, $ps < 0.001$), but they were also more likely to be confused with dominance than reward smiles, $F(2,372) = 408.03$, $p < 0.001$, $\eta_p^2 = 0.69$. Simple effects analyses revealed that affiliative smiles were equally likely to be classified as reward smiles in both display conditions (static: $M = 10.44$, $SD = 10.88$, dynamic: $M = 9.23$, $SD = 8.13$, $p = 0.386$). However, affiliative smiles were also less likely to be accurately classified as affiliative in the static ($M = 52.27$, $SD = 16.50$) than in the dynamic condition ($M = 63.03$, $SD = 52.27$, $p < 0.001$). This difference results from participants rating affiliative smiles as more dominant in the static condition ($M = 37.29$, $SD = 15.70$) than in the dynamic condition ($M = 27.75$, $SD = 13.58$, $p < 0.001$) (see **Table 1**).

DISCUSSION

The purpose of the present work was to test the extent to which the functional smiles of reward, affiliation, and dominance are

TABLE 1 | Smile type confusions in the static and dynamic condition.

	Display	Reward ratings	Affiliative ratings	Dominance ratings
Reward smiles	Dynamic	67.44	26.37	6.18
	Static	65.04	27.98	6.97
Affiliative Smiles	Dynamic	9.23	63.03***	27.75***
	Static	10.44	52.27***	37.29***
Dominance Smiles	Dynamic	9.16	23.93	66.91
	Static	10.58	27.44	61.98

***$p < 0.001$, significant difference in the mean ratings between static and dynamic display.

distinct and recognizable facial expressions. We also aimed to explore the role of facial muscle restriction and presentation mode in moderating smile classification rates. The results reveal that participants were able to accurately categorize reward, affiliative and dominance smiles. This supports the assumption that diverse morphological characteristics of smiles are identified in terms of their social communicative functions (Niedenthal et al., 2010; Martin et al., 2017). The use of naturalistic human face stimuli, rather than computer-generated faces, extends existing work (Rychlowska et al., 2017), thereby achieving greater ecological validity.

Our results reveal that classification accuracy was significantly lower for affiliative smiles than reward and dominance smiles. This is in line with previous findings by Rychlowska et al. (2017) who showed that human observers and a Bayesian classifier were less accurate in categorizing affiliative smiles compared to reward and dominance smiles (using a binary yes/no classification approach to indicate whether a given expression was – or was not – an instance of a given smile type). The present research used continuous confidence ratings that were not mutually exclusive, thus replicating their findings with human-realistic stimuli and a different response format. Moreover, a closer inspection of participants' ratings reveals that, whereas affiliative smiles were relatively unlikely to be classified as reward, reward smiles were often judged as affiliative, consistently with the results of Rychlowska et al. (2017) and Martin et al. (2018). While this finding suggests that reward smiles – similarly to the Duchenne smiles previously described in the literature – may constitute a more homogeneous, less variable category than other smiles (e.g., Frank et al., 1993), it also highlights similarities between reward and affiliative smiles which both convey positive social motivations. It is worth noting that participants in the present study saw smile expressions of White/Caucasian targets without any background information. The only context given in the study was the definition of the three smile types including examples of situations in which they might potentially occur. Recent work by Martin et al. (2018) suggests that the three types of smiles elicit distinct physiological responses when presented in a social-evaluative context. Adding social context to these displays therefore provides a promising avenue for future research, as the salience of specific interpersonal tasks could facilitate the distinction between affiliative smiles and the other two categories.

As predicted, the current study revealed higher recognition rates of the expressions presented in dynamic compared to static mode, and this applied in particular to affiliative smiles. This finding corroborates existing research on the dynamic advantage in emotion recognition (Hyniewska and Sato, 2015; Calvo et al., 2016). The fact that presentation mode is particularly important in the recognition of affiliative smiles confirms the assumption that dynamic features might be especially helpful in the identification of more subtle and ambiguous facial expressions, i.e., non-enjoyment smiles (Krumhuber and Manstead, 2009). As such, fundamental differences in the timing of smiles such as amplitude, total duration, and speed of onset, apex, and offset (Cohn and Schmidt, 2004)

might inform expression classification (Krumhuber and Kappas, 2005).

Contrary to our predictions and to previous findings (Niedenthal et al., 2010; Maringer et al., 2011; Rychlowska et al., 2014), our results did not support the moderating role of people's ability to mimic in smile classification. According to Calvo and Nummenmaa (2016), facial expressions consist of morphological changes in the face and their underlying affective content. Given that participants were instructed to rate each smile on three pre-designed scales (reward, affiliative, dominance smile), it is possible that this procedure induced cognitive, label-driven, rather than affective processing based on embodied simulation. Alternatively, the provision of a clear definition of the three functional smiles might have failed to encourage the social motivation necessary for facial mimicry to occur (Hofman et al., 2012; Hess and Fischer, 2014). It is also possible that other factors, i.e., trait empathy (Kosonogov et al., 2015) or endocrine levels (Kraaijenvanger et al., 2017) impact smile recognition rates as well as modulate the occurrence of mimicry. We think that it is unlikely that the present results are caused by an improper technique for blocking mimicry given that the experimenter closely monitored whether participants held the pencils correctly. In addition, we used a reliable facial muscle restriction technique employed in previous studies which revealed the moderating role of mimicry in emotion perception (Niedenthal et al., 2001; Maringer et al., 2011).

One potential limitation of our study was that we did not measure mimicry during the smile classification task. It is thus impossible to conclude whether participants in the free mimicry condition were really mimicking the smiles or whether mimicry occurred but did not enhance recognition performance in comparison to the restricted condition. We therefore suggest for future research on mimicry blocking to use EMG measurements in order to assess the presence of facial mimicry in the free muscle condition as well as the effectiveness of mimicry blocking in the restricted muscle condition. Finally, the lack of significant effects of the muscle restriction procedure may also reflect the complexity of sensorimotor simulation; a process which does not always involve measurable facial mimicry. Given that generating a motor output is a critical component for sensorimotor simulation more than facial activity *per se* (e.g., Korb et al., 2015; Wood et al., 2016), future studies could investigate the extent to which judgments of functional smiles are impaired by experimental manipulations that involve the production of conflicting facial movements.

In sum, the present research investigated observers' judgments of reward, affiliative, and dominance smiles. While participants were able to accurately categorize each smile type, recognition accuracy was lower for affiliative than for reward and dominance smiles. Although preventing mimicry responses did not appear to influence participants' classification, the use of dynamic versus static stimuli increased recognition accuracy of affiliative smiles. To our knowledge, this is the first study to test the role of muscle restriction and presentation mode in the recognition of reward, affiliative, and dominance smiles. The results highlight the importance of dynamic information, being particularly salient in

the recognition of affiliative smiles which are the most ambiguous among the three smile types. The lack of a significant effect of facial muscle condition on smile classification suggests that the functional smiles can be recognized based on their physical appearance. Our findings contribute to the understanding of the importance of temporal dynamics in the perception of emotional expressions.

AUTHOR CONTRIBUTIONS

AO, EK, PS, and MR conceived and designed the experiments. AO performed the experiments. AO and EK performed the statistical analysis. AO wrote the first draft of the manuscript. EK, MR, and PS wrote sections of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Blairy, S., Herrera, P., and Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *J. Nonverb. Behav.* 23, 5–41. doi: 10.1023/A:1021370825283

Calvo, M. G., Avero, P., Fernández-Martín, A., and Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion* 16, 1186–1200. doi: 10.1037/emo0000192

Calvo, M. G., Gutiérrez-García, A., and Del Líbano, M. (2018). What makes a smiling face look happy? Visual saliency, distinctiveness, and affect. *Psychol. Res.* 82, 296–309. doi: 10.1007/s00426-016-0829-3

Calvo, M. G., and Lundqvist, D. (2008). Facial expressions of emotion (KDEF): identification under different display-duration conditions. *Behav. Res. Methods* 40, 109–115. doi: 10.3758/BRM.40.1.109

Calvo, M. G., and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Cohn, J. F., and Schmidt, K. (2004). The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multiresolut. Inf. Process.* 2, 1–12. doi: 10.1142/S021969130400041X

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221

Du, S., Tao, Y., and Martinez, A. M. (2014). Compound facial expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1454–E1462. doi: 10.1073/pnas.1322355111

Duchenne de Boulogne, G.-B. (1862/1990). *The Mechanism of Human Facial Expression,* ed. and trans. R. A. Cuthbertson. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511752841

Ekman, P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (revised edition)*. New York, NY: W. W. Norton & Company.

Ekman, P., Davidson, R. J., and Friesen, W. V. (1990). The Duchenne smile: emotional expression and brain physiology: II. *J. Pers. Soc. Psychol.* 58, 342–353. doi: 10.1037/0022-3514.58.2.342

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverb. Behav.* 6, 238–252. doi: 10.1007/BF00987191

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *Facial Action Coding System: The manual on CD ROM*. Salt Lake City, UT: Research Nexus.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fischer, A., and Hess, U. (2017). Mimicking emotions. *Curr. Opin. Psychol.* 17, 151–155. doi: 10.1016/j.copsyc.2017.07.008

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of The Emotions.* New York, NY: W. W. Norton & Company.

Frank, M. G., Ekman, P., and Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *J. Pers. Soc. Psychol.* 64, 83–93. doi: 10.1037/0022-3514.64.1.83

Gunnery, S. D., Hall, J. A., and Ruben, M. A. (2013). The deliberate Duchenne smile: individual differences in expressive control. *J. Nonverb. Behav.* 37, 29–41. doi: 10.1007/s10919-012-0139-4

Hess, U., and Fischer, A. (2014). Emotional mimicry: why and when we mimic emotions. *Soc. Pers. Psychol. Compass* 8, 45–57. doi: 10.1016/j.tics.2015.12.010

Hofman, D., Bos, P. A., Schutter, D. J., and van Honk, J. (2012). Fairness modulates non-conscious facial mimicry in women. *Proc. R. Soc. Lond. B* 279, 3535–3539. doi: 10.1098/rspb.2012.0694

Hyniewska, S., and Sato, W. (2015). Facial feedback affects valence judgments of dynamic and static emotional expressions. *Front. Psychol.* 6:291. doi: 10.3389/fpsyg.2015.00291

Korb, S., Malsert, J., Rochas, V., Rihs, T., Rieger, S., Schwab, S., et al. (2015). Gender differences in the neural network of facial mimicry of smiles – An rTMS study. *Cortex* 70, 101–114. doi: 10.1016/j.cortex.2015.06.025

Korb, S., Niedenthal, P., Kaiser, S., and Grandjean, D. (2014). The perception and mimicry of facial movements predict judgments of smile authenticity. *PLoS One* 9:e99194. doi: 10.1371/journal.pone.0099194

Kosonogov, V., Titova, A., and Vorobyeva, E. (2015). Empathy, but not mimicry restriction, influences the recognition of change in emotional facial expressions. *Q. J. Exp. Psychol.* 68, 2106–2115. doi: 10.1080/17470218.2015.1009476

Kraaijenvanger, E. J., Hofman, D., and Bos, P. A. (2017). A neuroendocrine account of facial mimicry and its dynamic modulation. *Neurosci. Biobehav. Rev.* 77, 98–106. doi: 10.1016/j.neubiorev.2017.03.006

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonverb. Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., and Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion* 7, 730–735. doi: 10.1037/1528-3542.7.4.730

Krumhuber, E. G., Kappas, A., and Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., and Manstead, A. S. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 807–820. doi: 10.1037/a0017844

Krumhuber, E. G., and Scherer, K. R. (2016). The look of fear from the eyes varies with the dynamic sequence of facial actions. *Swiss J. Psychol.* 75, 5–14. doi: 10.1024/1421-0185/a000166

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Krys, K., Hansen, K., Xing, C., Szarota, P., and Yang, M. (2014). Do only fools smile at strangers? Cultural differences in social perception of intelligence of smiling individuals. *J. Cross Cult. Psychol.* 45, 314–321. doi: 10.1177/0022022113513922

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., and Weyers, P. (2008). Modulation of facial mimicry by attitudes. *J. Exp. Soc. Psychol.* 44, 1065–1072. doi: 10.1016/j.jesp.2007.10.007

Maringer, M., Krumhuber, E. G., Fischer, A. H., and Niedenthal, P. M. (2011). Beyond smile dynamics: mimicry and beliefs in judgments of smiles. *Emotion* 11, 181–187. doi: 10.1037/a0022596

Martin, J., Rychlowska, M., Wood, A., and Niedenthal, P. (2017). Smiles as multipurpose social signals. *Trends Cogn. Sci.* 21, 864–877. doi: 10.1016/j.tics.2017.08.007

Martin, J. D., Abercrombie, H. C., Gilboa-Schechtman, E., and Niedenthal, P. M. (2018). Functionally distinct smiles elicit different physiological responses in an evaluative context. *Sci. Rep.* 8:3558. doi: 10.1038/s41598-018-21536-1

Niedenthal, P., Rychlowska, M., and Szarota, P. (2013). "Embodied simulation and the human smile: processing similarities to cultural differences," in *Warsaw Lectures in Personality and Social Psychology (Meaning Construction, the Social World, and the Embodied Mind,* Vol. 3, eds D. Cervone, M. Fajkowska, M. W. Eysenck, and T. Maruszewsk (Clinton Corners, NY: Eliot Werner Publications), 143–160.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., and Innes-Ker, A. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cogn. Emot.* 15, 853–864. doi: 10.1080/02699930143000194

Niedenthal, P. M., Mermillod, M., Maringer, M., and Hess, U. (2010). The Simulation of Smiles (SIMS) model: embodied simulation and the meaning of facial expression. *Behav. Brain Sci.* 33, 417–433. doi: 10.1017/S0140525X10000865

Oberman, L. M., Winkielman, P., and Ramachandran, V. S. (2007). Face to face: blocking facial mimicry can selectively impair recognition of emotional expressions. *Soc. Neurosci.* 2, 167–178. doi: 10.1080/17470910701391943

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., and Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion* 12, 1398–1403. doi: 10.1037/a0028588

Ruben, M. A., Hall, J. A., and Schmid Mast, M. (2015). Smiling in a job interview: when less is more. *J. Soc. Psychol.* 155, 107–126. doi: 10.1080/00224545.2014.972312

Rychlowska, M., Cañadas, E., Wood, A., Krumhuber, E. G., Fischer, A., and Niedenthal, P. M. (2014). Blocking mimicry makes true and false smiles look the same. *PLoS One* 9:e90876. doi: 10.1371/journal.pone.0090876

Rychlowska, M., Jack, R. E., Garrod, O. G., Schyns, P. G., Martin, J. D., and Niedenthal, P. M. (2017). Functional smiles: tools for love, sympathy, and war. *Psychol. Sci.* 28, 1259–1270. doi: 10.1177/0956797617706082

Rychlowska, M., Miyamoto, Y., Matsumoto, D., Hess, U., Gilboa-Schechtman, E., Kamble, S., et al. (2015). Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proc. Natl. Acad. Sci. U.S.A.* 112, E2429–E2436. doi: 10.1073/pnas.1413661112

Rymarczyk, K., Zurawski, L., Jankowiak-Siuda, K., and Szatkowska, I. (2016). Do Dynamic compared to static facial expressions of happiness and anger reveal enhanced facial mimicry? *PLoS One* 11:e0158534. doi: 10.1371/journal.pone.0158534

Sachisthal, M. S., Sauter, D. A., and Fischer, A. H. (2016). Mimicry of ingroup and outgroup emotional expressions. *Compr. Results Soc. Psychol.* 1, 86–105. doi: 10.1080/23743603.2017.1298355

Sato, W., Fujimura, T., and Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *Int. J. Psychophysiol.* 70, 70–74. doi: 10.1016/j.ijpsycho.2008.06.001

Szarota, P., Bedynska, S., Matsumoto, D., Yoo, S., Friedlmeier, W., Sterkowicz, S., et al. (2010). Smiling as a masking display strategy: a cross-cultural comparison. I: A. Blachnio, & A. Przepiorka (red: er). *Closer Emot.* 3, 227–238.

Vrugt, A., and Vet, C. (2009). Effects of a smile on mood and helping behavior. *Soc. Behav. Pers. Int. J.* 37, 1251–1258. doi: 10.2224/sbp.2009.37.9.1251

Wood, A., Rychlowska, M., Korb, S., and Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends Cogn. Sci.* 20, 227–240. doi: 10.1016/j.tics.2015.12.010

Dynamic Facial Expressions Prime the Processing of Emotional Prosody

Patricia Garrido-Vásquez [1,2], Marc D. Pell [3], Silke Paulmann [4] and Sonja A. Kotz [2,5]*

[1] Department of Experimental Psychology and Cognitive Science, Justus Liebig University Giessen, Giessen, Germany,
[2] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [3] School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada, [4] Department of Psychology, University of Essex, Colchester, United Kingdom, [5] Department of Neuropsychology and Psychopharmacology, University of Maastricht, Maastricht, Netherlands

Evidence suggests that emotion is represented supramodally in the human brain. Emotional facial expressions, which often precede vocally expressed emotion in real life, can modulate event-related potentials (N100 and P200) during emotional prosody processing. To investigate these cross-modal emotional interactions, two lines of research have been put forward: cross-modal integration and cross-modal priming. In cross-modal integration studies, visual and auditory channels are temporally aligned, while in priming studies they are presented consecutively. Here we used cross-modal emotional priming to study the interaction of dynamic visual and auditory emotional information. Specifically, we presented dynamic facial expressions (angry, happy, neutral) as primes and emotionally-intoned pseudo-speech sentences (angry, happy) as targets. We were interested in how prime-target congruency would affect early auditory event-related potentials, i.e., N100 and P200, in order to shed more light on how dynamic facial information is used in cross-modal emotional prediction. Results showed enhanced N100 amplitudes for incongruently primed compared to congruently and neutrally primed emotional prosody, while the latter two conditions did not significantly differ. However, N100 peak latency was significantly delayed in the neutral condition compared to the other two conditions. Source reconstruction revealed that the right parahippocampal gyrus was activated in incongruent compared to congruent trials in the N100 time window. No significant ERP effects were observed in the P200 range. Our results indicate that dynamic facial expressions influence vocal emotion processing at an early point in time, and that an emotional mismatch between a facial expression and its ensuing vocal emotional signal induces additional processing costs in the brain, potentially because the cross-modal emotional prediction mechanism is violated in case of emotional prime-target incongruency.

Keywords: emotion, priming, event-related potentials, cross-modal prediction, dynamic faces, prosody, audiovisual, parahippocampal gyrus

1. INTRODUCTION

Emotion is conveyed through different communication channels: facial expressions, tone of voice (emotional prosody), gestures, and others. Moreover, emotional communication in everyday life is dynamic, and we need to constantly monitor the emotional expressions of the people we interact with. However, the majority of past research on emotion perception has focused on

single communication channels (e.g., emotional face processing) and on static stimuli rather than dynamic ones—possibly because these stimuli allow for controlled laboratory testing. Recent research has started to tackle the challenges related to multisensory, dynamic emotion processing. The present study follows this important movement and aims to contribute to the field by investigating cross-modal emotional priming with dynamic stimuli.

1.1. Cross-Modal Modulation of Emotion Processing

In cross-modal emotion perception, at least two processes are involved: cross-modal prediction and audiovisual integration (Jessen and Kotz, 2015). Cross-modal prediction is a mechanism by which information from one modality (e.g., a facial expression) helps predict certain characteristics of the signal in another modality that comes into play later (e.g., a vocal expression). Audiovisual integration refers to the process by which modalities are integrated into a coherent percept. Thus, cross-modal emotional priming, in which visual and auditory information is presented consecutively rather than simultaneously, is a tool to investigate cross-modal prediction independent of audiovisual integration.

Cross-modal emotional priming studies suggest that humans use emotional stimuli from one modality to generate predictions about the other. For example, people are faster and more accurate at deciding whether a facial expression truly reflects an emotion or not when the faces are preceded by emotionally congruent rather than incongruent prosody (Pell, 2005; Pell et al., 2011). These congruency effects also show in event-related potentials (ERPs), in which prime-target congruency modulates an N400-like negativity (Paulmann and Pell, 2010).

In real life audiovisual speech processing, we use a speaker's mouth and face movements to generate predictions about ensuing acoustic stimulation (van Wassenhove et al., 2005; Chandrasekaran et al., 2009). In a similar vein, emotional facial expressions commonly precede vocal emotional input, and may therefore drive a cross-modal emotional prediction mechanism (Jessen and Kotz, 2013; Ho et al., 2015; Kokinous et al., 2015).

Priming research has shown that the presentation of a facial expression affects how subsequent vocal emotional targets are processed: Pourtois and colleagues (Pourtois et al., 2000) used angry or sad facial expressions that were followed by a vocal stimulus with angry intonation. Emotional congruency between facial and vocal expressions affected the amplitude of the auditory N100 component. In a mismatch negativity (MMN) study with the same stimuli, incongruent deviants among congruent standards (or vice versa) triggered an enhanced auditory MMN at 178 ms after sound onset, even though the auditory input was held constant, and participants were instructed to ignore it (de Gelder et al., 1999). Static fearful or happy faces followed by fearful or happy prosody elicited a posterior P2b component in the ERPs, which occurred earlier when face and voice were emotionally congruent rather than incongruent (Pourtois et al., 2002). Thus, vocal emotion processing is influenced by preceding facial information at an early point in time, namely within the

first 250 ms of auditory processing. Due to the different nature of these congruency effects and the lack of ERP studies assessing the contextual influence of dynamic face primes on vocal emotion processing, more priming studies are needed to shed light on these processes.

Apart from cross-modal emotional priming studies, researchers have tested audiovisual integration of emotional information, with temporally aligned visual (facial and/or body expressions) and auditory information. This means that the visual information naturally precedes the auditory information, since mouth or body movements are visible first while the auditory information unfolds over time. Comparisons of audiovisual conditions to a purely auditory condition show that emotional prosody is integrated with its preceding visual input within the first 100 ms of vocal emotion processing (Jessen and Kotz, 2011; Kokinous et al., 2015), reflected in an N100 amplitude suppression for audiovisual compared to auditory-only stimuli. This could be due to the visual input leading to predictions about the to-be-expected auditory input, which facilitates auditory processing, for example in terms of temporal predictability of the auditory signal (Vroomen and Stekelenburg, 2010; Jessen and Kotz, 2013; Schröger et al., 2015).

However, temporal predictability of auditory input, based on preceding visual information, is only one type of predictability which modulates the auditory N100 in cross-modal emotion processing. Several ERP studies have manipulated audiovisual emotional congruency within audiovisual integration paradigms while maintaining temporal predictability. In these studies, emotional congruency between face and voice also differentially affected the amplitude of the N100 response (Ho et al., 2015; Kokinous et al., 2015, 2017; Zinchenko et al., 2015). Thus, visual signals are not only used to predict *when* to expect auditory stimulation, but also *what* to expect.

Since these were integration studies in which the visual and auditory tracks were temporally aligned (albeit with visual information available ahead of auditory input), the emotionally incongruent condition implied incongruency of mouth movements and vocal stimulation, in addition to the emotional mismatch between face and voice: Even though the same sounds were used in emotionally congruent and incongruent conditions (e.g., mouth movement for "ah" paired with "ah" sound), mouth movements differ depending on emotion (e.g., the mouth movement while uttering a neutral "ah" is very different from when uttering an angry "ah"). Thus, apart from an emotional mismatch between the visual and auditory channels, there was also a mismatch in mouth movements, which may at least partly account for the reported congruency effects in the N100.

Two studies (Zinchenko et al., 2015, 2017) show that this could in fact be the case: these authors included a comparison between congruent and incongruent audiovisual sounds (e.g., mouth movement for "ah" combined with "ah" sound vs. "oh" sound), while maintaining emotional face-voice congruency. They observed significant congruency effects already in the N100, which shows that a conflict between mouth movement and sound is sufficient to modulate N100 amplitude. It is therefore necessary to complement previous research with priming studies, in which

visual and auditory tracks follow each other rather than being temporally aligned. This will help us isolate the processes due to emotional conflict from those related to other types of conflict between visual and auditory information. Moreover, as outlined above, the priming paradigm allows studying cross-modal emotional prediction independent of multisensory integration.

1.2. Cross-Modal Modulation of the P200

In several studies that reported N100 modulations by emotional face-voice congruency, the effects extended into the P200 ERP component (Pourtois et al., 2000; Ho et al., 2015; Kokinous et al., 2015; Zinchenko et al., 2015). In the study by Ho et al. (2015), emotional congruency effects in the N100 were affected by attentional manipulations, while this was not the case for the P200. Two other integration studies reported that audiovisual congruency affected the N100 as a function of visual context (Kokinous et al., 2015) or target category (Zinchenko et al., 2015), while the P200 was globally modulated by face-voice congruency (Kokinous et al., 2015; Zinchenko et al., 2015). In another audiovisual integration study, emotional congruency effects were observed in the P200 component only, but not in the N100 (Zinchenko et al., 2017). These results show that N100 and P200 may reflect different processes in emotional face-voice interactions. Therefore, the present study will also investigate congruency effects in the P200, in order to test whether these two components can be functionally dissociated in dynamic cross-modal emotional priming.

1.3. Cross-Modal Modulation of Brain Regions

Previous fMRI studies investigating audiovisual emotion processing have reported several audiovisual convergence areas, most notably the posterior superior temporal sulcus and gyrus (STS/STG) (Kreifelts et al., 2007; Robins et al., 2009; Park et al., 2010; Klasen et al., 2011; Watson et al., 2014; Li et al., 2015) and the thalamus (Kreifelts et al., 2007; Klasen et al., 2011).

Some imaging studies have compared emotionally congruent and incongruent audiovisual stimuli. This allows identifying brain regions important for the integration of emotionally congruent signals (Klasen et al., 2012) and also regions associated with higher processing costs due to audiovisual stimulus incongruency. Incongruent emotional stimuli trigger more widespread activations in the brain than congruent ones, which may reflect more effortful processing in the case of emotional incongruency (Klasen et al., 2011; Müller et al., 2011). For example, the cingulate cortex, an area associated with conflict processing, is more activated by incongruent than congruent stimuli (Klasen et al., 2011; Müller et al., 2011).

Because categorizing incongruent emotional stimuli is much harder than categorizing congruent ones (Collignon et al., 2008; Föcker et al., 2011), these activation differences may reflect task difficulty. Therefore, Watson et al. (2013) morphed visual and auditory emotional stimuli independently on an angry-happy continuum in order to manipulate emotion categorization difficulty. They found that after regressing out the variance due to task difficulty, the incongruency effect remained significant in the

right STS/STG (Watson et al., 2013). Thus, enhanced activations for incongruent as compared to congruent stimuli reflect more than just task difficulty—they may point to enhanced processing effort while the brain tries to make sense of two stimuli that do not belong together.

While these imaging studies are particularly informative about which brain regions are implicated in cross-modal emotion processing, they fail to clearly link brain structures to the time course underlying cross-modal emotional interactions. The present study thus utilized ERPs to explore when incongruency effects for dynamic emotional stimuli are first observed and adds the ERP source localization technique to link high temporal resolution with potential brain sources.

1.4. The Present Study

We applied a cross-modal emotional priming paradigm with short video clips of facial expressions as primes and emotional pseudo-speech stimuli as targets. We aimed at testing whether facial expressions elicit early congruency effects in the auditory ERPs. Furthermore, since emotional priming studies using static face primes are inconclusive regarding the time point at which audiovisual congruency effects first emerge in ERPs (de Gelder et al., 1999; Pourtois et al., 2000, 2002), we wanted to shed more light on this issue using dynamic face primes, which are more ecologically valid than static facial expressions. We predicted congruency effects at an early time point in auditory processing, namely in the N100 and P200.

Since fMRI exhibits a very good spatial, but low temporal resolution, we localized the neural sources of ERP differences in the present study to explore underlying neural activity at precise points in time. In line with previous neuroimaging research, we expected that incongruent compared to congruent targets would trigger enhanced activations in right STS/STG region, which is modulated by emotional congruency in audiovisual emotion processing irrespective of task difficulty (Watson et al., 2013).

In audiovisual integration studies, emotional faces and voices of one category were paired with neutral faces and voices to construct the emotionally incongruent experimental condition (Ho et al., 2015; Kokinous et al., 2015, 2017; Zinchenko et al., 2015, 2017). Additionally, two priming studies have used angry or sad facial expressions paired with angry voice targets (de Gelder et al., 1999; Pourtois et al., 2000), and one priming study has paired happy and fearful faces and voices (Pourtois et al., 2002). Thus, there are two types of incongruency: pairings of different emotion categories or pairings of emotional with neutral material. The present study used both in order to compare whether the type of incongruency makes a difference in cross-modal emotion processing. We will refer to the combination of neutral primes with emotional targets as "neutral" condition throughout, while the pairings of emotional primes and targets of opposing valence will be referred to as "incongruent" condition in the present paper.

To sum up, the aims of the present study were as follows: (1) to describe the time course of dynamic cross-modal emotional priming with ERPs, (2) to identify underlying neural sources of significant ERP congruency effects, and (3) to test whether the type of cross-modal emotional incongruency (pairings of

emotional with neutral stimuli or pairings of stimuli of opposing emotional valence) makes a difference for (1) and (2).

Since previous findings show that part of the processing differences between congruent and incongruent audiovisual emotional stimuli may be due to task difficulty (Watson et al., 2013), we used a gender decision task, which also ensured that participants did not need to consciously focus on the emotional content of our stimuli (Pourtois et al., 2005; Paulmann et al., 2009).

2. METHODS

2.1. Participants

Thirty-six individuals took part in the present experiment. Sample size was based on previous ERP studies (Paulmann and Kotz, 2008; Ho et al., 2015). Two participants had to be excluded from the final sample, one due to technical problems during the EEG measurement and one because of strong noise on almost all scalp electrodes. The remaining 34 participants (17 female) had a mean age of 24.97 years ($SD = 2.35$). All participants reported normal hearing, normal or corrected-to-normal vision and were right-handed (Oldfield, 1971). They received financial compensation for taking part in the experiment. All participants provided written informed consent prior to participating in the experiment. The study was approved by the local ethics committee at the University of Leipzig, and the procedures followed the Declaration of Helsinki.

2.2. Stimulus Material

The stimuli consisted of video files without sound, which were used as primes, and audio files, which were used as targets. The videos were black-and-white recordings of four semi-professional actors (two female) showing the face and some surrounding information (hair, neck, etc.; see **Figure 1** for examples). Actors were videotaped while uttering happy, angry, and neutral sentences with emotionally matching semantic content and showing the corresponding expressions, which means that mouth movements were visible. To create the stimuli, we removed the audio track from the recordings. Faces were cropped and/or centered when necessary in order to be at the center of the display and to have approximately the same size on screen. Gaze was always directed toward the observer. We cut fragments of 520 ms duration from the middle of the original videos, such that the full facial expression was visible from the first video frame on. The 520 ms prime duration was based on considerations that prime durations or prime-target SOAs below 300 ms may lead to reversed priming effects (Bermeitinger et al., 2008; Paulmann and Pell, 2010), which we wanted to prevent. Moreover, dynamic facial expressions elicit the strongest ERP responses within the first 500 ms of processing (Recio et al., 2014), and we aimed to avoid an overlap of these with early vocal emotion processing. This prime duration is also roughly comparable with the temporal precedence of visual information in cross-modal emotional integration studies (Ho et al., 2015; Kokinous et al., 2015). Due to the 40 ms frame length, video duration can only be a multiple of 40, which is why we chose the seemingly arbitrary video duration of 520 ms.
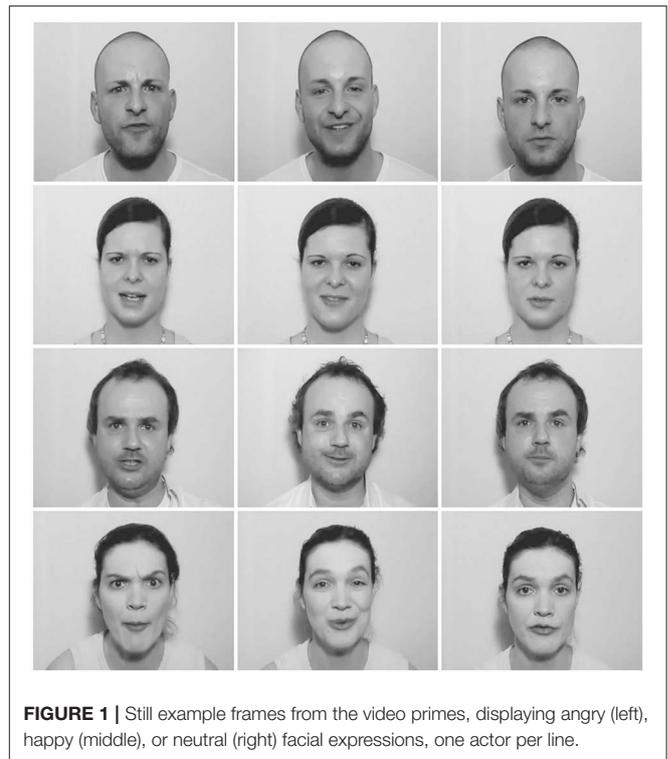
FIGURE 1 | Still example frames from the video primes, displaying angry (left), happy (middle), or neutral (right) facial expressions, one actor per line.

Video selection for the present experiment was based on results from a validation study with 28 participants who were not recruited for the present study (see Garrido-Vásquez et al., 2016 for details). Based on these data, 240 video stimuli (3 emotional categories × 4 actors × 20 videos) were selected, which in the validation study were recognized on average 2.7 times better than chance (chance level: 33%). We used a rather large number of different stimuli to reflect the natural variability inherent in emotion expressions.

The audio files were happy and angry sentences uttered in pseudo-speech by the same four actors who appeared in the videos. Thus, semantic content could not be derived from the sentences, but they nevertheless matched German phonotactic rules and all had the same syntactic structure (e.g., "Hung set das Raap geleift ind nagebrucht."). Duration of these stimuli was approximately 3s. The sentences were digitized at a 16-bit/44.1 kHz sampling rate. They were normalized to peak amplitude to ensure an equal maximum volume for all stimuli. Recognition of these materials was also pre-tested on a different sample of 24 participants, and for each actor and category we selected the 30 highest-ranking stimuli, resulting in 240 happy and angry sentence stimuli to be paired with the videos. Average recognition rates were more than five times higher than chance (chance level: 14%). These stimuli have been used in prior research (e.g., Paulmann et al., 2010; Garrido-Vásquez et al., 2013).

2.3. Procedure

We conducted the EEG experiment in an electrically-shielded and sound-attenuated room. Participants were sitting at a

distance of approximately 100 cm from the computer screen. Videos were presented centrally at an image resolution of 720 × 576 pixels, and the faces subtended a visual angle of approximately three degrees to each side. We used the MPEG-4 codec to optimize timing, and frame rate was 25 frames per second. Auditory stimuli were presented at a constant and comfortable listening level. The experiment was programmed in Presentation (Neurobehavioral Systems, San Francisco, USA).

Each trial started with a black fixation cross on a gray background (1,000 ms), followed by the video prime (520 ms). Immediately after video offset, the fixation cross became visible again and the auditory target played via loudspeakers located left and right to the screen. Identity of the actor in the video and in the audio file always matched within a trial. The fixation cross stayed on screen until the end of the auditory stimulus and was then replaced by a black question mark, prompting participants to indicate whether a female or male speaker had been presented. Answers were provided by means of a button box, and half of the participants pressed the left button for "female" and the right button for "male," while the other half proceeded vice versa. Participants were instructed to answer as fast and accurately as possible. After the button press, a gray blank screen showed up (2,000 ms), and then the next trial began.

The 240 trials were divided into four blocks of 60 trials each and presented in a pseudo-randomized order that differed for each participant. In one third of the trials, prime and target were emotionally congruent (happy-happy or angry-angry), in another third they were incongruent (happy-angry or angry-happy), and yet another third were neutral trials (neutral-happy or neutral-angry). Our randomization allowed for a maximum of three consecutive trials with the same actor, same prime category (happy, angry, or neutral), or the same prime-target relationship (congruent, incongruent, neutral). Unrelated to the current investigation, within the same experimental session we also tested the reverse prime-target order, i.e., with pseudo-sentences as primes and videos as targets (results not reported here). Half of the participants started with the video-as-prime-condition, and the other half with the audio-as-prime condition. Total run-time of the experiment was approximately 60 min including breaks.

2.4. Data Acquisition and Analysis

We recorded the EEG data from 61 scalp electrodes mounted in an elastic cap according to the extended international 10-10 system. Data were referenced to the average of all electrodes online and re-referenced to the mean activity at left and right mastoids offline. Recording was accomplished with a bandpass between DC and 140 Hz, and the data were digitized at 500 Hz. Four electrodes (two horizontal, two vertical) were applied to register eye movements during the measurement, and the ground electrode was placed on the sternum. Electrode resistance was below 5kΩ.

We used FieldTrip (Oostenveld et al., 2011) running on Matlab (The Mathworks, Natick, USA) to further process the EEG data offline. Continuous data were filtered with a highpass filter at a cutoff frequency of 1 Hz (1,762 points, Blackman window, finite impulse response filter). This filter did not only

remove slow drifts, but it also served to replace the baseline, because we were interested in measuring ERPs elicited by the prosodic targets, but obtaining a clean pre-stimulus baseline was not possible due to the prime, which directly preceded the target (see, e.g., Jessen and Kotz, 2011). After cutting the data into epochs of 1,000 ms duration and time-locked to target onset, we first manually inspected all trials for atypical artifacts, which were rejected. Then, the data were subjected to an independent component analysis (ICA) to identify components associated with eye movements or other artifacts (electrocardiographic artifacts or noisy electrodes). These components were removed from the data, and then the ICA-corrected data were inspected manually again in order to reject any trials that still contained artifacts. Furthermore, all trials with incorrect or missing responses were excluded from the data. In total, 21% of all trials were excluded based on these criteria. We applied a 40 Hz lowpass filter on the EEG data for the visual ERP displays.

The clean 1,000 ms epochs were averaged according to target emotion (happy, angry) and congruency with the prime (congruent, incongruent, neutral). Time windows and electrode sites for the N100 and P200 analysis were defined based on the "Collapsed localizers" procedure (Luck and Gaspelin, 2017), which consists of averaging all experimental conditions together and then identifying electrodes and time windows at which the component of interest is maximal. The selected electrodes, at which both N100 and P200 were maximal were: FC3, FCz, FC4, C3, C1, Cz, C2, C4, CP3, CPz, and CP4. Time windows selected according to this procedure were: 80-130 ms post-target onset for the N100 and 180–250 ms post-target onset for the P200. For the ERP amplitude analyses, we averaged the data across the respective time windows and all included electrodes. Furthermore, the selected electrodes and time windows were also used to extract N100 and P200 peak latency for the ERP latency analysis. For both amplitude and latency data, values were submitted to a 3 (congruency) × 2 (emotion) repeated-measures ANOVA. Mauchly's test for sphericity was insignificant for all effects; therefore we used the original degrees of freedom in the ANOVA.

2.5. Source Reconstruction

In case of significant ERP results, we conducted a source reconstruction on the respective ERP time window to uncover neural generators of the effects. These analyses were realized in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Individual electrode locations obtained via digitization were co-registered with SPM's standard template head model in MNI space with a cortical mesh of 8,196 vertices. We constructed the forward model using the Boundary Elements Method implemented in SPM, which is based on realistic head geometry and takes into consideration the different conductor properties of brain tissues. We inverted the data for all conditions and participants together, using the minimum norm estimation algorithm (IID). The results were smoothed with a Gaussian kernel full-width half-maximum (FWHM) of 12 mm.

The six average images (one per condition) for each participant were taken to second-level analyses. We conducted two-sided t-tests for paired samples in order to compute

contrasts between the congruent, incongruent, and neutral conditions collapsed across angry and happy targets. We also computed these contrasts separately according to target emotion (e.g., angry congruent vs. angry incongruent). All contrasts were calculated in both directions (e.g., congruent > incongruent and incongruent > congruent). Results that survived family-wise error correction at an alpha level of $p < 0.05$ were deemed significant.

3. RESULTS

Behavioral data were not further analyzed, since gender decision performance was at ceiling.

3.1. N100 and P200 Amplitude

Figure 2 shows N100 and P200 time-locked to target onset in the three congruency conditions. An overview of means and standard deviations both for ERP amplitude and latency values in all six conditions is provided in **Table 1**.

The ANOVA on the N100 time window yielded a significant main effect of congruency, $F_{(2, 66)} = 4.899$, $p = 0.01$, $\eta^2 p = 0.129$. In the incongruent condition ($M = -3.68$, SD = 1.80) N100 amplitudes were larger than in the congruent condition ($M = -3.40$, SD = 1.73), $t_{(34)} = 2.491$, $p = 0.018$. The same held true when comparing the incongruent to the neutral condition ($M = -3.28$, SD = 1.85), $t_{(34)} = 3.046$, $p = 0.005$. Amplitudes in the congruent and neutral condition did not significantly differ (p

$= 0.407$). Furthermore, we observed a significant main effect of emotion, $F_{(1, 33)} = 12.108$, $p = 0.001$, $\eta^2 p = 0.268$. Angry prosody ($M = -3.66$, SD = 1.66) elicited higher N100 amplitudes than happy prosody ($M = -3.25$, SD = 1.88). The interaction between both factors was insignificant ($p \geq 0.571$).

P200 amplitude was not significantly modulated by congruency or emotion ($ps > 0.150$).

3.2. N100 and P200 Peak Latency

N100 latency was significantly modulated by target congruency, $F_{(2, 66)} = 8.976$, $p < 0.001$, $\eta^2 p = 0.214$. This component peaked later in the neutral condition ($M = 108$ ms, SD = 9.50) than in the congruent ($M = 105$ ms, SD = 9.58), $t_{(34)} = 2.610$, $p = 0.014$, and incongruent ($M = 103$ ms, SD = 10.06), $t_{(34)} = 4.246$, $p < 0.001$ conditions. The latter two did not significantly differ ($p = 0.098$). The main effect of emotion and the congruency x emotion interaction were not significant ($ps \geq 0.299$).

We failed to find any significant main effects or interactions for P200 peak latency ($ps \geq 0.163$).

3.3. Source Reconstruction

Since ERP analyses revealed significant effects only for the N100, we restricted source reconstruction to this component and to the alpha frequency range (Herrmann et al., 2014). Incongruent targets triggered significantly stronger activations in the right parahippocampal gyrus (PHG) than congruent targets. The right PHG was also more active in angry incongruent compared to angry congruent trials. None of the other contrasts survived the threshold of $p < 0.05$ family-wise error corrected. Congruently primed targets did not elicit additional activations when compared to incongruently primed ones, even at a very lenient threshold of $p < 0.01$ (uncorrected). See **Table 2** and **Figure 3** for results of the source reconstruction analysis.
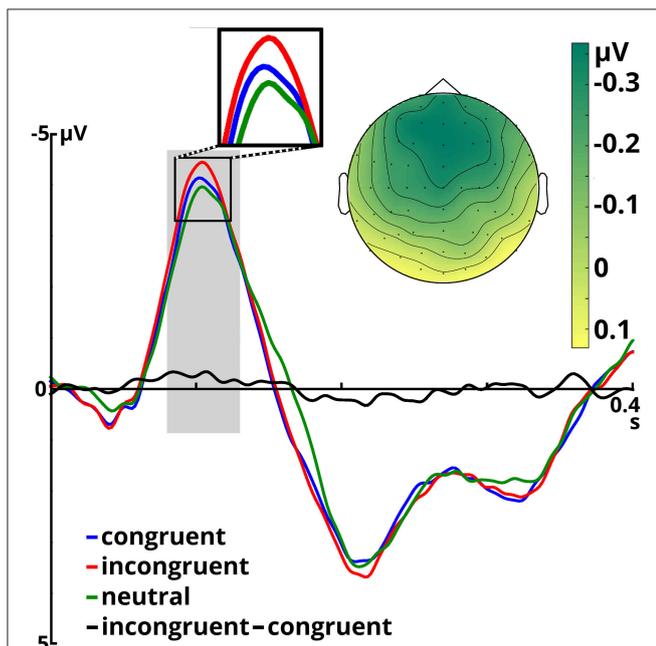
FIGURE 2 | Event-related potentials averaged over all included electrodes for the three congruency conditions and the incongruent - congruent difference, time-locked to target onset. The time window for N100 analysis is shaded in gray. The scalp potential map shows the incongruent - congruent difference in the N100 time window.

TABLE 1 | ERP amplitude and latency results.

N100 MEAN AMPLITUDE

Intonation	Prime	Mean voltage (μV)	SD
Angry	Congruent	-3.51	1.67
Angry	Incongruent	-3.94	1.85
Angry	Neutral	-3.53	1.79
Happy	Congruent	-3.29	1.97
Happy	Incongruent	-3.42	1.98
Happy	Neutral	-3.03	2.11

N100 PEAK LATENCY

Intonation	Prime	Peak latency (ms)	SD
Angry	Congruent	105	11.38
Angry	Incongruent	102	9.74
Angry	Neutral	107	10.10
Happy	Congruent	104	9.42
Happy	Incongruent	104	11.34
Happy	Neutral	108	11.58

TABLE 2 | N100 source reconstruction results.

Contrast	MNI peak voxel	Z	Cluster size	p^a	Region
ic > c	16 -36 -12	4.04	145	0.025	Right PHG
ang ic > ang c	18 -34 -14	3.90	28	0.041	Right PHG

c, congruent; ic, incongruent; ang, anger; PHG, parahippocampal gyrus.
[a]Family-wise error corrected (p-value and cluster size).

FIGURE 3 | Results from the source reconstruction analysis showing significant clusters in the right parahippocampal gyrus. **(A)** Incongruent > congruent, z = -12. **(B)** Anger incongruent > anger congruent, z = -14. Images are thresholded at $p < 0.05$ family-wise error corrected.
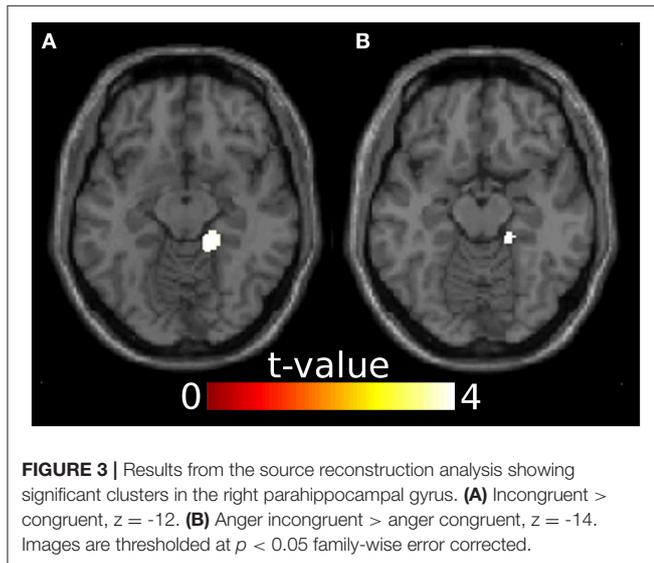
4. DISCUSSION

In the present study we investigated cross-modal emotional priming with videos showing happy, angry, or neutral facial expressions followed by happy or angry emotional prosody. We successfully replicated early audiovisual congruency effects in the N100 ERP component. Building on unimodal and multimodal priming studies with static facial expressions (Pourtois et al., 2000; Werheid et al., 2005), we showed that dynamic emotional face primes successfully establish an emotional context under which subsequent emotional targets are evaluated. By including a neutral prime category in addition to the incongruent one, we were able to show that these two types of prime-target incongruency elicit different processes, which we will discuss in more detail below. Moreover, the right PHG was more activated during the processing of incongruently, rather than congruently primed auditory targets within the N100 time window.

4.1. N100 Enhancement in the Incongruent Condition

Emotional priming affected auditory processing at an early time point, namely in the N100. This is in line with several previous studies, which have shown such early emotional congruency effects (Pourtois et al., 2000; Werheid et al., 2005; Ho et al., 2015; Kokinous et al., 2015, 2017; Zinchenko et al., 2015), with findings from audiovisually presented congruent and incongruent human speech sounds (Zinchenko et al., 2015, 2017), and with studies

comparing unimodal to audiovisual emotion (Jessen and Kotz, 2011; Kokinous et al., 2015) and speech processing (van Wassenhove et al., 2005).

This evidence suggests that information from auditory and visual domains can be combined within the first 100 ms of auditory processing, possibly facilitated through a cross-modal prediction mechanism (van Wassenhove et al., 2005; Jessen and Kotz, 2013; Ho et al., 2015). Commonly, emotional facial expressions temporally precede vocal expressions of emotion in human interactions and thus allow us to predict some characteristics of the ensuing auditory signal, such as its temporal onset and some acoustic properties. However, if for example an angry face precedes a vocal expression of happiness, the prediction is violated, leading to enhanced processing costs. In our study, these were reflected in an enhanced N100 amplitude and right PHG activation, which we will discuss in more detail below.

N100 enhancement in the incongruent condition indicates that emotional significance in the voice could at least partly be extracted already during the first 100 ms of auditory processing. When vocal emotion is presented in isolation (i.e., unimodally), emotional significance is thought to be extracted after approximately 200 ms, in the P200 component (Schirmer and Kotz, 2006; Paulmann and Kotz, 2008; Pell et al., 2015), while earlier steps are associated with sensory processing (Schirmer and Kotz, 2006). However, some studies using unisensory vocal emotional stimuli have also reported emotion effects in the N100 (Liu et al., 2012; Kokinous et al., 2015; Pinheiro et al., 2015), although these may be triggered by low-level features of the stimuli (Schirmer and Kotz, 2006). In the current study low-level features are an insufficient explanation for the N100 modulations, because congruency effects did not differ as a function of target emotion and were modulated only by the prime-target relation *per se*. We could therefore hypothesize that emotional information in the face (e.g., a smile) leads to the prediction that the ensuing vocal stimulus will be of a certain quality (e.g., rather high-pitched) and thereby facilitates auditory processing if this prediction is fulfilled.

4.2. Absence of Significant Congruency Effects in the P200

In contrast to audiovisual emotion studies that reported emotional congruency effects also in the P200 (Pourtois et al., 2000; Ho et al., 2015; Kokinous et al., 2015; Zinchenko et al., 2015) or exclusively in the P200 (Balconi and Carrera, 2011; Yeh et al., 2016; Zinchenko et al., 2017), we failed to observe any significant ERP differences for this component. We argue that in studies showing congruency effects only in the P200, different mechanisms may have shifted emotional congruency effects toward the P200: Balconi and Carrera (2011) used static facial displays whose onset was temporally aligned to their prosodic stimuli; therefore participants may have needed longer than in other studies for combining auditory and visual cues. This is supported by a study by Paulmann et al. (2009), who used static facial expressions whose onset was aligned to (congruent) emotional prosody

onset. They found a P200 amplitude reduction for audiovisual compared to unimodal stimulation, but no N100 effects. Yeh et al. (2016) used bodily expressions, which may be a less reliable predictor of vocal emotional expressions than a face (although these authors did show N100 suppression during audiovisual compared to auditory processing, but irrespective of congruency). Furthermore, identity mismatches between the visual and auditory tracks could have played a role in their study, because the materials came from different stimulus databases. Zinchenko et al. (2017) employed happy and neutral stimuli, and probably the conflict between happy and neutral cues is not big enough to trigger any congruency effects in the N100, but shifts them to the P200. This partly aligns with our study, in which we failed to show N100 differences between congruently and neutrally primed prosodic stimuli. Thus, methodological differences between studies may lead to a temporal shift of cross-modal interactions because participants take longer to process cross-modal emotional congruency.

In the present study, congruency effects started to emerge early, but were rather short-lived. We propose that the lack of P200 effects may follow from the gender decision task we used: As the face was always presented first and identity of the actor in the video and in the audio always matched within a trial, it was sufficient to make the gender decision based on the face only. Even though we did not instruct participants to do this, they may have realized the identity match after a few trials. Thus, it is likely that they rather attended to the face than to the voice in the present experiment, which is supported by the fact that people often prefer emotional information from faces over information from voices or that facial expressions are more difficult to ignore than vocal expressions (Collignon et al., 2008; Klasen et al., 2011; Ho et al., 2015). Moreover, the task we used did not draw attention to the emotional quality of the stimuli. Studies that show both N100 and P200 modulations by cross-modal emotional congruency (Ho et al., 2015; Kokinous et al., 2015; Zinchenko et al., 2015) have at least in part used tasks that draw attention to the emotionality of the voice. It is, however, unclear why congruency effects extended into the P200 in the study by Pourtois et al. (2000), who instructed participants to attend to the faces and ignore the voices, or why cross-modal emotional congruency affected the MMN in the study by de Gelder et al. (1999)—one explanation could be that they used static facial expressions, while the dynamic primes in our study were processed more quickly and efficiently (Mayes et al., 2009), leading only to short-lived congruency effects in the auditory ERPs. Future research manipulating cross-modal emotional congruency should experiment with different task instructions and dynamic vs. static stimuli to shed more light on this issue. In any case, our results are in line with other studies that suggest that N100 and P200 can be functionally dissociated during cross-modal emotion processing (e.g., Ho et al., 2015; Kokinous et al., 2015).

4.3. Role of the Right Parahippocampal Gyrus in Cross-Modal Emotional Priming

Source localization revealed that in incongruent compared to congruent trials, the right PHG was engaged in the N100 time window. This difference was apparently driven by angry target stimuli, because the angry incongruent > angry congruent contrast was significant in the right PHG while the happy incongruent > happy congruent contrast was not.

Two studies comparing bimodal emotional face-voice combinations to unimodal conditions have reported enhanced right PHG activation (Park et al., 2010; Li et al., 2015). PHG is also more active when affective pictures are combined with emotional music as compared to when the pictures are presented in isolation (Baumgartner et al., 2006). These three studies (Baumgartner et al., 2006; Park et al., 2010; Li et al., 2015) used only congruent audiovisual inputs. Thus, the right PHG may be involved in binding emotional information from different modalities, and its enhanced activation for incongruent targets in the N100 window may reflect its stronger recruitment when facial and vocal information mismatch.

There is not much evidence on how the PHG relates to early auditory processing, but it has been associated with auditory deviance detection in oddball tasks during the N100 (Mucci et al., 2007; Karakaş et al., 2009). This evidence is in line with our findings: In the oddball task, in which a sequence of frequent standard stimuli is sometimes interrupted by deviant stimuli, participants will generally expect the standard tone because it occurs with greater likelihood than the deviant. Thus, in case of a deviant the prediction is violated similarly to the emotional prediction in incongruent trials in our experiment. This leads to enhanced processing effort, which engages the PHG. According to a relatively recent account (Aminoff et al., 2013), the PHG codes for contextual associations, and in the context of emotion it facilitates emotion understanding and expectations, which perfectly fits with the current results—if face and voice are emotionally incongruent, then the face is a non-reliable contextual cue, leading to more effortful processing of the voice target in the right PHG.

Interestingly, fMRI studies comparing incongruent to congruent audiovisual emotion processing have not reported right PHG activation (Klasen et al., 2011; Müller et al., 2011). We propose that this could be due to the early nature of these activations, which are potentially hard to capture with fMRI. On the other hand, our source localization results converge with those from fMRI studies in that there are no regions found to be more active in the congruent compared to the incongruent condition (Klasen et al., 2011; Müller et al., 2011).

The question arises as to why source localization failed to reveal any activation foci in the STS/STG, while this region has been reported in the neuroimaging literature comparing incongruent to congruent audiovisual emotion stimuli (Kreifelts et al., 2007; Robins et al., 2009; Park et al., 2010; Klasen et al., 2011; Watson et al., 2013, 2014). One potential reason may be time course. Due to low temporal resolution in fMRI, we do not know when this region comes into play. According to a network analysis of audiovisual emotion processing (Jansma et al., 2014), PHG may modulate STS/STG activity unidirectionally; therefore the distinctive activation patterns based on prime-target congruency in STS/STG may come into play later in time.

4.4. The Neutral Prime Condition: Just Another Incongruent Condition?

In the present study we used both an incongruent condition, with happy primes and angry targets or vice versa, and a neutral condition, in which a neutral facial expression was followed by happy or angry prosody. This allowed testing whether emotional information preceded by neutral information leads to the perception of audiovisual incongruency comparable to the incongruent condition.

N100 amplitude was significantly enhanced in the incongruent compared to the neutral and congruent conditions, which did not significantly differ from each other. Thus, prime-target incongruency may have triggered additional processing effort in the brain, while this was not the case in the neutral prime condition. Moreover, the incongruent condition triggered stronger right PHG activation than the congruent one, which supports the higher processing effort interpretation. Even though this did not apply for the incongruent > neutral contrast, a more liberal threshold of $p < 0.001$ (uncorrected) would yield right PHG activation in this comparison. Thus, we can cautiously state that processing effort in the incongruent condition was also higher than in the neutral condition.

In contrast to our ERP results for the neutral priming condition, other studies reported congruency effects in the auditory N100 when a neutral face preceded angry prosody (Ho et al., 2015; Kokinous et al., 2015, 2017). One potential explanation for this effect could be the task. Kokinous et al. (2015) used an emotion-related task; their participants were asked to indicate whether the prosodic stimulus expressed anger or not. Ho et al. (2015) employed four different tasks: participants judged (1) emotionality in the voice, (2) emotionality in the face, (3) emotional face-voice congruence, or (4) temporal synchrony between face and voice channels. All but the last task were thus emotion-related, and in all but the last task did the authors report that the N100 in response to angry voices was modulated by the fact whether the face was angry or neutral. Thus, the results from the synchrony judgment task Ho et al. (2015) converge with our findings, which we gathered using a gender decision task, a task unrelated to emotion. Attention to the emotional quality of a stimulus may therefore be necessary in order for neutral face primes to trigger congruency effects in the N100. However, we found longer N100 peak latencies for the neutral prime condition compared to the congruent and incongruent conditions. This could mean that an emotional prime speeds up target processing, regardless of congruency, which is in line with previous findings (e.g., Burton et al., 2005).

Generally speaking, neutral stimuli may be less informative for cross-modal prediction because they are not as clear as emotional expressions (Jessen and Kotz, 2015). This is in accord with the results from our pre-test, in which a different set of participants watched and categorized the videos used here. While we obtained very high hit rates for the angry and happy videos (98 and 94%, respectively), neutral videos were recognized with 78% accuracy only. These data support the notion that neutral stimuli are more ambiguous than emotional ones.

The rather small differences between the neutral and congruent conditions in the present study may also be due to our design: If we consider the neutral priming condition an incongruent condition, as has been the case in previous research (e.g., Ho et al., 2015; Kokinous et al., 2015, 2017; Zinchenko et al., 2015, 2017), then two thirds of all trials were incongruent. Due to this imbalance, the neutral trials potentially triggered less conflict than when prime and target were of opposing valence, and the rather small differences between the neutral and congruent conditions in the present study could be attributed to this fact. Moreover, the neutral prime was never paired with a neutral target in the current study and was therefore not suitable to predict ensuing acoustic stimulation. However, it is currently unclear whether prime-target assignments within an experiment can induce transient changes in cross-modal prediction during emotion processing and override long-term associations (Jessen and Kotz, 2013). If these transient changes exist, then the proportion of incongruent among congruent trials in an experiment should influence congruency effects, an issue that still needs to be investigated.

4.5. Limitations

As outlined in the previous paragraph, it is not clear whether the presence of the neutral condition in addition to the congruent and incongruent ones may have affected the current results. Moreover, we tested only two emotional categories (happy and angry), which are furthermore of opposing valence (positive and negative). Thus, we cannot say whether our findings are attributable to the emotions *per se*, or to valence effects. It is also not clear why significant effects in the P200 were absent, and whether quicker and more efficient processing of the dynamic prime stimuli or task effects are a suitable explanation for this observation. These limitations to our study will have to be addressed by future research.

CONCLUSION

The present study employed a cross-modal emotional priming paradigm with dynamic facial expressions. We showed that priming with a dynamic emotional facial expression affects vocal emotion processing already in the N100 ERP component. An enhanced N100 component as well increased right PHG activation to incongruent targets indicate that processing incongruently primed vocal emotional targets was more effortful than when they had been congruently primed, which may be due to the violation of cross-modal predictions. Our data are in line with many ERP studies showing that audiovisual emotional information is already combined within the N100 time window.

AUTHOR CONTRIBUTIONS

PG-V, MP, SP, and SK designed research. PG-V conducted the experiment and analyzed the data. PG-V, MP, SP, and SK wrote the paper.

REFERENCES

Aminoff, E. M., Kveraga, K., and Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends Cogn. Sci.* 17, 379–390. doi: 10.1016/j.tics.2013.06.009

Balconi, M., and Carrera, A. (2011). Cross-modal integration of emotional face and voice in congruous and incongruous pairs: the P2 ERP effect. *Eur. J. Cogn. Psychol.* 23, 132–139. doi: 10.1080/20445911.2011.473560

Baumgartner, T., Lutz, K., Schmidt, C. F., and Jäncke, L. (2006). The emotional power of music: how music enhances the feeling of affective pictures. *Brain Res.* 1075, 151–164. doi: 10.1016/j.brainres.2005.12.065

Bermeitinger, C., Frings, C., and Wentura, D. (2008). Reversing the N400: event-related potentials of a negative semantic priming effect. *Neuroreport* 19, 1479–1482. doi: 10.1097/WNR.0b013e32830f4b0b

Burton, L. A., Rabin, L., Wyatt, G., Frohlich, J., Vardy, S. B., and Dimitri, D. (2005). Priming effects for affective vs. neutral faces. *Brain Cogn.* 59, 322–636 doi: 10.1016/j.bandc.2005.05.006

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi: 10.1371/journal.pcbi.1000436

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Res.* 1242, 126–135. doi: 10.1016/j.brainres.2008.04.023

de Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M., and Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neurosci. Lett.* 260, 133–136. doi: 10.1016/S0304-3940(98)00963-X

Föcker, J., Gondan, M., and Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychol.* 137, 36–47. doi: 10.1016/j.actpsy.2011.02.004

Garrido-Vásquez, P., Pell, M. D., Paulmann, S., Sehm, B., and Kotz, S. A. (2016). Impaired neural processing of dynamic faces in left-onset Parkinson's disease. *Neuropsychologia* 82, 123–133. doi: 10.1016/j.neuropsychologia.2016.01.017

Garrido-Vásquez, P., Pell, M. D., Paulmann, S., Strecker, K., Schwarz, J., and Kotz, S. A. (2013). An ERP study of vocal emotion processing in asymmetric Parkinson's disease. *Soc. Cogn. Affect. Neurosci.* 8, 918–927. doi: 10.1093/scan/nss094

Herrmann, C. S., Rach, S., Vosskuhl, J., and Strüber, D. (2014). Time–frequency analysis of event-related potentials: a brief tutorial. *Brain Topogr.* 27, 438–450. doi: 10.1007/s10548-013-0327-5

Ho, H. T., Schröger, E., and Kotz, S. A. (2015). Selective attention modulates early human evoked potentials during emotional face–voice processing. *J. Cogn. Neurosci.* 27, 798–818. doi: 10.1162/jocn_a_00734

Jansma, H., Roebroeck, A., and Münte, T., F. (2014). A network analysis of audiovisual affective speech perception. *Neuroscience* 256, 230–241. doi: 10.1016/j.neuroscience.2013.10.047

Jessen, S., and Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage.* 58, 665–674. doi: 10.1016/j.neuroimage.2011.06.035

Jessen, S. and Kotz, S. A. (2013). On the role of crossmodal prediction in audiovisual emotion perception. *Front. Hum. Neurosci.* 7:369. doi: 10.3389/fnhum.2013.00369

Jessen, S., and Kotz, S. A. (2015). Affect differentially modulates brain activation in uni-and multisensory body-voice perception. *Neuropsychologia* 66, 134–143. doi: 10.1016/j.neuropsychologia.2014.10.038

Karakaş, H., Karakaş, S., Ceylan, A. Ö., and Tali, E. (2009). Recording event-related activity under hostile magnetic resonance environment: is multimodal EEG/ERP-MRI recording possible? *Int. J. Psychophysiol.* 73, 123–132. doi: 10.1016/j.ijpsycho.2009.03.006

Klasen, M., Chen, Y. H., and Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Rev. Neurosci.* 23, 381–392. doi: 10.1515/revneuro-2012-0040

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., and Mathiak, K. (2011). Supramodal representation of emotions. *J. Neurosci.* 31, 13635–13643. doi: 10.1523/JNEUROSCI.2833-11.2011

Kokinous, J., Kotz, S. A., Tavano, A., and Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Soc. Cogn. Affect. Neurosci.* 10, 713–720. doi: 10.1093/scan/nsu105

Kokinous, J., Tavano, A., Kotz, S. A., and Schröger, E. (2017). Perceptual integration of faces and voices depends on the interaction of emotional content and spatial frequency. *Biol. Psychol.* 123, 155–165. doi: 10.1016/j.biopsycho.2016.12.007

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., and Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37, 1445–1456. doi: 10.1016/j.neuroimage.2007.06.020

Li, Y., Long, J., Huang, B., Yu, T., Wu, W., Liu, Y., et al. (2015). Crossmodal integration enhances neural representation of task-relevant features in audiovisual face perception. *Cereb. Cortex* 25, 384–395. doi: 10.1093/cercor/bht228

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., and Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *Neuroreport* 23, 108–112. doi: 10.1097/WNR.0b013e32834ea757

Luck, S. J., and Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology* 54, 146–157. doi: 10.1111/psyp.12639

Mayes, A. K., Pipingas, A., Silberstein, R. B., and Johnston, P. (2009). Steady state visually evoked potential correlates of static and dynamic emotional face processing. *Brain Topogr.* 22, 145–157. doi: 10.1007/s10548-009-0106-5

Mucci, A., Galderisi, S., Kirkpatrick, B., Bucci, P., Volpe, U., Merlotti, E., et al. (2007). Double dissociation of N1 and P3 abnormalities in deficit and nondeficit schizophrenia. *Schizophr. Res.* 92, 252–261. doi: 10.1016/j.schres.2007.01.026

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., et al. (2011). Incongruence effects in crossmodal emotional integration. *Neuroimage* 54, 2257–2266. doi: 10.1016/j.neuroimage.2010.10.047

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Park, J. Y., Gu, B. M., Kang, D. H., Shin, Y. W., Choi, C. H., Lee, J. M., et al. (2010). Integration of cross-modal emotional information in the human brain: an fMRI study. *Cortex* 46, 161–169. doi: 10.1016/j.cortex.2008.06.008

Paulmann, S., Jessen, S., and Kotz, S. A. (2009). Investigating the multimodal nature of human communication: insights from ERPs. *J. Psychophysiol.* 23, 63–76. doi: 10.1027/0269-8803.23.2.63

Paulmann, S., and Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport* 19, 209–213. doi: 10.1097/WNR.0b013e3282f454db

Paulmann, S., and Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: how much is enough? *Cogn. Affect. Behav. Neurosci.* 10, 230–242. doi: 10.3758/CABN.10.2.230

Paulmann, S., Seifert, S., and Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Soc. Neurosci.* 5, 59–75. doi: 10.1080/17470910903135668

Pell, M. D. (2005). Nonverbal emotion priming: evidence from the facial affect decision task. *J. Nonverb. Behav.* 29, 45–73. doi: 10.1007/s10919-004-0889-8

Pell, M. D., Jaywant, A., Monetta, L., and Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cogn. Emot.* 25, 834–853. doi: 10.1080/02699931.2010.516915

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., and Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biol. Psychol.* 111, 14–25. doi: 10.1016/j.biopsycho.2015.08.008

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., and Gonçalves, Ó. F. (2015). The music of language: an ERP investigation of the effects of musical training on emotional prosody processing. *Brain Lang.* 140, 24–34. doi: 10.1016/j.bandl.2014.10.009

Pourtois, G., de Gelder, B., Bol, A., and Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex* 41, 49–59. doi: 10.1016/S0010-9452(08)70177-1

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., and Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport* 11, 1329–1333. doi: 10.1097/00001756-200004270-00036

Pourtois, G., Debatisse, D., Despland, P. A., and de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Cogn. Brain Res.* 14, 99–105. doi: 10.1016/S0926-6410(02)00064-2

Recio, G., Schacht, A., and Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: specificity and intensity effects in event-related brain potentials. *Biol. Psychol.* 96, 111–125. doi: 10.1016/j.biopsycho.2013.12.003

Robins, D. L., Hunyadi, E., and Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain Cogn.* 69, 269–278. doi: 10.1016/j.bandc.2008.08.007

Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009

Schröger, E., Marzecová, A., and SanMiguel, I. (2015). Attention and prediction in human audition: a lesson from cognitive psychophysiology. *Eur. J. Neurosci.* 41, 641–664. doi: 10.1111/ejn.12816

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102

Vroomen, J., and Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *J. Cogn. Neurosci.* 22, 1583–1596. doi: 10.1162/jocn.2009.21308

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., and Belin, P. (2013). Dissociating task difficulty from incongruence in face-voice emotion integration. *Front. Hum. Neurosci.* 7:744. doi: 10.3389/fnhum.2013.00744

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., and Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face–voice emotional integration. *J. Neurosci.* 34, 6813–6821. doi: 10.1523/JNEUROSCI.4478-13.2014

Werheid, K., Alpay, G., Jentzsch, I., and Sommer, W. (2005). Priming emotional facial expressions as evidenced by event-related brain potentials. *Int. J. Psychophysiol.* 55, 209–219. doi: 10.1016/j.ijpsycho.2004.07.006

Yeh, P. w., Geangu, E., and Reid, V. (2016). Coherent emotional perception from body expressions and the voice. *Neuropsychologia* 91, 99–108. doi: 10.1016/j.neuropsychologia.2016.07.038

Zinchenko, A., Kanske, P., Obermeier, C., Schröger, E., and Kotz, S. A. (2015). Emotion and goal-directed behavior: ERP evidence on cognitive and emotional conflict. *Social Cogn. Affect. Neurosci.* 10, 1577–1587. doi: 10.1093/scan/nsv050

Zinchenko, A., Obermeier, C., Kanske, P., Schröger, E., and Kotz, S. A. (2017). Positive emotion impedes emotional but not cognitive conflict processing. *Cogn. Affect. Behav. Neurosci.* 17, 665–677. doi: 10.3758/s13415-017-0504-1

Detecting Genuine and Deliberate Displays of Surprise in Static and Dynamic Faces

Mircea Zloteanu[1,2]*, Eva G. Krumhuber[2] and Daniel C. Richardson[2]

[1] Department of Computer Science, University College London, London, United Kingdom, [2] Department of Experimental Psychology, University College London, London, United Kingdom

People are good at recognizing emotions from facial expressions, but less accurate at determining the authenticity of such expressions. We investigated whether this depends upon the technique that senders use to produce deliberate expressions, and on decoders seeing these in a dynamic or static format. Senders were filmed as they experienced genuine surprise in response to a jack-in-the-box (Genuine). Other senders faked surprise with no preparation (Improvised) or after having first experienced genuine surprise themselves (Rehearsed). Decoders rated the genuineness and intensity of these expressions, and the confidence of their judgment. It was found that both expression type and presentation format impacted decoder perception and accurate discrimination. Genuine surprise achieved the highest ratings of genuineness, intensity, and judgmental confidence (dynamic only), and was fairly accurately discriminated from deliberate surprise expressions. In line with our predictions, Rehearsed expressions were perceived as more genuine (in dynamic presentation), whereas Improvised were seen as more intense (in static presentation). However, both were poorly discriminated as not being genuine. In general, dynamic stimuli improved authenticity discrimination accuracy and perceptual differences between expressions. While decoders could perceive subtle differences between different expressions (especially from dynamic displays), they were not adept at detecting if these were genuine or deliberate. We argue that senders are capable of producing genuine-looking expressions of surprise, enough to fool others as to their veracity.

Keywords: facial expressions, posed, emotions, genuineness, accuracy, intensity

INTRODUCTION

Facial expressions are an important source of emotional and social information in interpersonal communication. Knowing what another person feels is relevant in predicting someone's psychological state, likely future behavior, and the outcome of social interactions (Johnston et al., 2010). However, not all expressions are truthful reflections of a person's underlying emotions. While genuine emotional expressions may inform about the affective state of a person, deliberate or voluntary expressions reflect the strategic intent of the sender in the absence of felt emotions (Ekman and Rosenberg, 2005). For example, deliberate displays can be used to prevent conflict or escalation, spare feelings, reassure, and gain someone's trust (Ekman and Friesen, 1982). Alternatively, they may be employed to manipulate, deceive, and mask underlying affect or intentions (Ekman and Friesen, 1982). Thus, the ability to discern if someone's emotional display

is genuine or deliberate is of high value in social interaction. The present research explores how different strategies for producing deliberate expressions impact decoders' perception and ability to detect their authenticity.

Research on emotion recognition has consistently found that decoders are adept at recognizing what emotions are indicated by particular facial expressions (Ekman, 2003; Calvo and Nummenmaa, 2015). But, when it comes to judging the authenticity of such facial displays, accuracy rates are markedly lower (Frank and Ekman, 1997; McLellan et al., 2010). When judging deception, for example, they are often at chance levels (Bond and DePaulo, 2006; Porter et al., 2012). This raises questions regarding the role emotions play in communication and social interactions. People regularly produce expressions when they wish to communicate to another person how they feel (Zuckerman et al., 1986). However, the advantage of decoding such expressions hinges on the displays matching the senders' true underlying affect. For instance, liars in real-world high-stakes scenarios have been shown to produce deliberate expressions to aid their deception, which decoders are unable to differentiate from genuine expressions (Porter et al., 2012). This is compounded by the fundamental assumption decoders make that the behavior of others is honest, unless prompted to consider otherwise (DePaulo and DePaulo, 1989). If decoders cannot distinguish deliberate displays from genuine affect these may be used to the advantage of the sender (i.e., lying about one's feelings), leading to misleading or even detrimental inferences. The case may be that senders are capable of producing deliberate expressions that resemble genuine affect sufficiently to fool decoders (Krumhuber and Manstead, 2009). Thus, it is important to understand if human decoders can discriminate genuine and deliberate expressions of emotions.

In the past, much of the emotion perception work attempting to answer this question has focused on a binary distinction between spontaneous (genuine) and posed (deliberate) expressions. To this end, a variety of acted expressions have been considered under the umbrella term of 'posed' displays, thereby glossing over different production methods that may lead to differences in expression and perception. Such voluntary behavior has typically been thought to differ from spontaneous expressions in the neural pathways of cortical and subcortical activation (Rinn, 1984; Morecraft et al., 2001), resulting in marked differences in visual appearance and timing (Cohn and Schmidt, 2004; Namba et al., 2016).

Whilst existing research suggests deliberate displays offer an advantage in emotion recognition tasks (Dawel et al., 2016), their use has been criticized in recent years due to their intentional nature to communicate the desired emotion (see Sauter and Fischer, 2017). Given the prevalence of existing stimulus sets to feature voluntary facial expressions (for a review see Krumhuber et al., 2017), we think it is important to draw a difference between various types of deliberate behavior. For example, the classical 'posed expressions' are voluntarily-produced emotional displays resulting from specific instructions such as those employed in directed facial action tasks (Russell, 1994). 'Portrayed expressions' are spontaneous deliberate expressions that occur in the absence of explicit instructions, but are congruent with the context in

which they occur, such as smiling for a photograph (Vazire et al., 2009). 'Enacted expressions' are expressions voluntarily produced after reliving a congruent past experience of the emotion, often done using method acting techniques (Scherer and Bänziger, 2010). Furthermore, the way in which researchers produce emotional displays for their stimuli vary widely, from using photographic stimuli that senders must imitate (e.g., Field and Walden, 1982), to the direct manipulation of facial muscle activation (e.g., Ekman et al., 1983), or simply using verbal prompts (e.g., Lewis et al., 1987). Thus, a further goal of our research is to shed light on the effect that these different practices may have on how human emotion perception is studied.

Accounting for this large variability in production methods, it seems reasonable to explore the impact of these different types of deliberate displays on expression perception. For this, we focused on the perception of a single emotion: surprise. Surprise is considered a basic emotion, having a distinctive facial configuration that is well recognized cross-culturally (Nelson and Russell, 2013; Namba et al., 2016). It is consistently found to have high recognition rates, second only to happiness (Ekman, 2003). Also, surprise is argued to be a neutral-valence emotion, and one determined by context (Ekman, 2004). In order to elicit surprise spontaneously, we considered the surprise expression to be more closely related to the startle response, i.e., a sudden defensive response to an external aversive stimulus. We therefore used a jack-in-the-box, an approach that in the past has been successful in eliciting a startle response, primarily in infants (e.g., Reissland et al., 2002), due in part to the unpredictable timing and the abrupt appearance of the jack. In addition to genuine expressions of surprise, two types of deliberate expressions were produced either on the basis of a recent emotional experience, or via improvisation based on no/minimal information.

Besides considering expression type, we investigated whether the modality of presentation (static vs. dynamic) can significantly impact authenticity discrimination. While static facial expressions of adequate intensity are sufficient to allow accurate emotion classification, dynamic aspects have been shown to enhance ratings of naturalness (Sato and Yoshikawa, 2004) and intensity (Biele and Grabowska, 2006), leading to stronger facial mimicry (Sato et al., 2008) and brain activation patterns in decoders (Trautmann et al., 2009). Dynamic information also enables better discrimination between genuine and deliberate displays (Krumhuber and Manstead, 2009; Maringer et al., 2011). This may be due to the fact that these are more complex and richer in expressive signal, thereby helping with the processing of emotional information (see Krumhuber et al., 2013). The use of dynamic stimuli may consequently better reflect the true authenticity of an expression.

In the present research, we contrasted genuine expressions of surprise with deliberate expressions produced after seeing an affect-evoking stimulus, i.e., the jack-in-the-box (Rehearsed) or without seeing it (Improvised). Re-enacting a genuine emotional experience is thought to facilitate the production of an authentic-looking deceptive display, as the sender is using the recent affective information of how an emotion feels and makes them behave (Bänziger and Scherer, 2007). This in turn may produce an expression that closely mirrors spontaneous surprise.

Alternatively, improvising an expression by using one's lay beliefs may produce a successful deceptive display (cf. Reisenzein et al., 2006), as the aim is to convey a specific message, which in turn may match the expectations of the decoder (i.e., exaggerated expressions are better recognized; Hess et al., 1997).

We hypothesized differences between the three types of surprise expressions in terms of their perceived genuineness, intensity, and judgmental confidence. Specifically, decoders should be able to accurately and confidently detect genuine surprise (Genuine), but should show poorer performance and less confidence when judging deliberate expressions (Rehearsed and Improvised). Whilst rehearsed surprise might lead to higher ratings of genuineness in comparison to improvised surprise, it is the improvised expressions that are predicted to be perceived as higher in intensity.

These differences in expression perception should be further moderated by the presentation format (static vs. dynamic). Using dynamic stimuli compared to static images stimuli increases ecological validity, allows for subtler elements of an emotion (e.g., onset, timing, duration, and fluidity) to be incorporated into the decoding process, and can improve authenticity discrimination (e.g., Hess and Kleck, 1994; Ambadar et al., 2005; Krumhuber and Kappas, 2005). We therefore predicted that dynamic information enables a better discrimination between genuine and deliberate expressions than what could be achieved with static displays.

MATERIALS AND METHODS

Participants

A total of 120 participants were recruited online through Amazon's Mechanical Turk (MTurk[1]) in exchange for $0.75; MTurk was used due to the benefits offered by online recruitment, and the comparable responses to laboratory samples (see, Casler et al., 2013). After deleting incomplete cases ($N = 31$) the final data encompassed 89 participants (51 men, 38 women), with an age range of 20–54 years ($M = 29.9$, $SD = 8.9$). Informed consent was obtained online prior to their participation. The two-factor experimental design included the presentation format (static vs. dynamic) as between-subjects variable, and expression type (genuine, rehearsed, and improvised) as within-subjects variable. Participants were randomly assigned to one of the two conditions, resulting in 46 people in the static group and 43 people in the dynamic group. A power analysis using G*Power 3.1 (Faul et al., 2007) for an interaction between presentation format (2) and expression type (3), assuming a medium-sized effect (Cohen's $f = 0.18$), determined that this sample size would be sufficient for 95% power. All participants had normal or corrected-to-normal vision. Ethical approval for the present study was granted by the UCL Department of Psychology Ethics Committee.

Stimulus Material

For the production of the stimulus expressions of surprise, 39 university students (12 males, 27 females; $M_{age} = 24.54$, $SD = 5.31$;

[1] www.mturk.com

age range = 19–45 years) were video-recorded under one of the three elicitation conditions:

In the *Genuine* condition, encoders were seated in front of the jack-in-the-box and turned the wheel until the toy "popped out"; a melody played as the wheel was turned prompting the action from the toy. The exact function of the toy was not described prior to the start of the experiment nor was the emotion under investigation explicitly mentioned. A camera was placed at eye-level, and recoded their reaction from the start of the winding action until the end of their behavioral response; the jack was not visible in the videos.

In the *Improvised* condition, encoders turned the wheel, carrying out the same hand action as those in the genuine videos. However, the electronic mechanism that releases the toy was made non-operational. Instead participants watched a video on a tablet positioned in front of the box. The video showed a countdown and played the same melody as the jack-in-the-box. When the word "NOW" appeared on the screen, participants had to act in a surprised manner. The countdown was matched for time and volume with the jack-in-the-box.

In the *Rehearsed* condition, encoders first had the experience of seeing the real jack-in-the-box as those in the genuine condition. The jack's wheel was then disconnected from the releasing mechanism, and the tablet with the countdown video was placed in front of it, as done in the Improvised condition. This time, encoders were asked to reproduce their previous emotional reaction when the word "NOW" appeared on the tablet's screen.

A Panasonic SDR-T50 camcorder was used to record the facial reactions at 25 frames per second. For each condition, there were thirteen exemplars: Genuine (4 men, 9 women), Rehearsed (5 men, 8 women), and Improvised (3 men, 10 women). These produced both static and dynamic portrayals of each expression, netting 39 static and 39 dynamic stimuli. Dynamic stimuli were silent video clips and lasted approximately 10 s. Static stimuli consisted of a single frame of the peak expression taken from each video; defined as the frame before the expression began to relax (see **Figure 1**). All stimuli were displayed in color (size: 1920 pixels × 1080 pixels).

Procedure

The study was conducted using the Qualtrics software (Provo, UT). As mood can affect classification accuracy (Forgas and East, 2008), it was necessary to control for this factor, by asking participants the following question: "How do you feel at this moment?" using a 5-point Likert-type scale
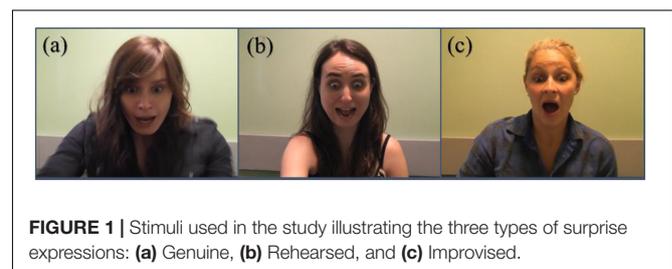
FIGURE 1 | Stimuli used in the study illustrating the three types of surprise expressions: **(a)** Genuine, **(b)** Rehearsed, and **(c)** Improvised.

(1 – *extremely sad*, 5 – *extremely happy*). After obtaining age and gender information, they were instructed to watch each stimulus carefully and rate the facial expression of the sender. It was made clear that some senders were genuinely reacting to a jack-in-the-box, while others never saw the toy puppet popping out and were merely attempting to appear surprised. Participants saw either static or dynamic displays of all 39 stimuli (presentation duration was 10 s in both conditions), in randomized order, and rated the expressions on several dimensions.

The extent to which they perceived the expression as a *genuine* response to seeing the jack-in-the-box was measured using a single item, 5-point Likert-type scale ranging from −2 ('*certain no Jack-in-the-box*'), −1 ('*no Jack-in-the-box*'), midpoint of 0 ('*not sure*'), to +1 ('*with Jack-in-the-box*') and +2 ('*certain with Jack-in-the-box*'), with higher scores indicating greater perceived genuineness. The responses were aggregated across the 13 exemplars of an expression type, yielding a total score ranging from −26 to +26 on perceived genuineness (see Dawel et al., 2016).

Overall *accuracy* of participants' ratings of the expressions were also calculated. A judgment was accurate if participants responded that they thought there was a jack-in-the-box present (with any level of certainty) and indeed the sender was reacting to a jack-in-the-box, or if they responded that there was no jack-in-the-box and, in fact, the sender was only pretending to be surprised. To formulate the measure of accuracy in authenticity discrimination, these responses were compared to the actual conditions of the stimulus, ignoring trials in which the participant responded 'not sure' (see Levine et al., 1999). If there was a match (e.g., rehearsed and improvised expressions were seen as having no jack-in-the-box, and genuine expressions were judged to have a jack-in-the-box), they were coded as accurate (score = 1). If there was a mismatch, it was coded as inaccurate (score = 0), yielding a final total score ranging from 0 to 13 for each expression type. For ease of comprehension, we re-labeled the totals using a percentage scale from 0% (lowest accuracy) to 100% (highest accuracy).

This was followed by participants' *confidence* ratings of their decision (1 – *not at all*, 5 – *very much*) to assess potential discrepancies between accuracy and perceived ability (Vrij and Mann, 2001). Finally, participants were asked to judge the *intensity* of the sender's expression using a 5-point Likert-type scale (1 – *not at all*, 5 – *very much*).

RESULTS

Preliminary analyses indicated no significant differences between male and female participants in their judgment ratings, Fs < 1.95, ps > 0.15. Thus, we collapsed across gender for all subsequent analyses. Adding mood as a covariate did not affect any of the results reported below, ps > 0.30. In both conditions, judgment ratings were averaged across the 13 exemplars within each expression type. A 2 (Format: dynamic vs. static) × 3 (Expression: genuine, improvised, rehearsed) mixed-factorial ANOVA was conducted on each of the four dependent measures. The Greenhouse–Geisser adjustment to the degrees of freedom

was applied when Mauchly's test indicated that the assumption of sphericity had been violated.

Genuineness

There was a significant main effect of Expression, $F(1.81, 157.72) = 39.78$, $p < 0.001$, $\eta_p^2 = 0.314$, but not Format, $F(1,87) = 1.20$, $p = 0.277$, on perceived genuineness. In addition, the interaction between the two factors was significant, $F(1.81, 157.72) = 30.40$, $p < 0.001$, $\eta_p^2 = 0.259$ (see **Figure 2**). To decompose the interaction, the simple main effect of expression was analyzed on each format condition.

The results revealed a significant simple main effect of Expression in the dynamic condition, $F(2,86) = 49.45$, $p < 0.001$, $\eta_p^2 = 0.535$. Pairwise comparisons with Bonferroni correction showed that genuine expressions ($M = 6.86$; $SD = 5.49$) were rated as significantly more genuine than improvised expressions ($M = −3.40$; $SD = 7.60$), $t(42) = 9.68$, $p < 0.001$, 95% CI [8.12, 12.39], $d = 1.48$, and rehearsed expressions ($M = 0.09$; $SD = 7.40$), $t(42) = 6.69$, $p < 0.001$, 95% CI [4.73, 8.81], $d = 1.02$. Improvised expressions were judged to be the least genuine and significantly differed from rehearsed expressions $t(42) = −5.23$, $p < 0.001$, 95% CI [2.14, 4.84], $d = 0.80$.

The analysis also revealed a significant main effect of Expression in the static condition, $F(2,86) = 7.76$, $p = 0.001$, $\eta_p^2 = 0.153$. Pairwise comparisons revealed that genuine expressions ($M = 3.83$; $SD = 8.78$) were rated significantly more genuine than rehearsed expressions ($M = 1.02$; $SD = 8.61$), $t(45) = 3.02$, $p = 0.004$, 95% CI [0.93, 4.67], $d = 0.45$, but no different from improvised expressions ($M = 3.63$; $SD = 8.80$), $t < 1$, $p = 0.839$. Improvised expressions were also judged as significantly more genuine-looking that rehearsed expressions, $t(45) = 3.14$, $p = 0.003$, 95% CI [4.28, 0.94], $d = 0.47$.

When considering differences in genuineness ratings between formats, simple effects analyses showed that improvised
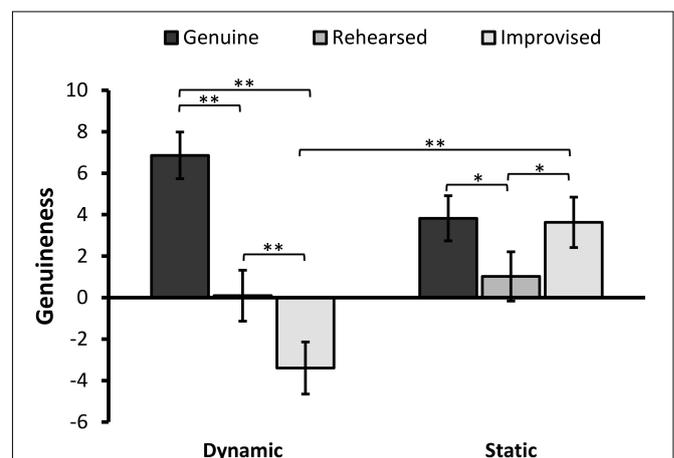
FIGURE 2 | Mean ratings for perceived genuineness of facial expressions (error bars ±1 SE). Positive values indicate that expressions were perceived as more genuine, while negative values indicate that they were perceived as more fake. The asterisks represent a significant difference at *$p < 0.005$ and **$p < 0.001$.

expressions were judged as significantly less genuine-looking when they were presented in dynamic than static format, $F(1,87) = 16.14$, $p < 0.001$, $\eta_p^2 = 0.156$. This difference did not occur in the context of genuine, $F(1,87) = 3.76$, $p = 0.056$, $\eta_p^2 = 0.041$ or rehearsed expressions, $F < 1$, $p > 0.59$.

Accuracy

The ANOVA showed a significant main effect of Expression, $F(1.23,106.94) = 22.08$, $p < 0.001$, $\eta_p^2 = 0.202$, and Format, $F(1,87) = 10.70$, $p = 0.002$, $\eta_p^2 = 0.109$. Overall, accuracies in authenticity discrimination were higher in the dynamic than static condition ($M_{diff} = 8.34$, $SD_{diff} = 2.55$). Also, genuine expressions ($M = 57.92$, $SD = 20.85$) were rated more accurately than both rehearsed ($M = 37.92$, $SD = 21.69$), $t(88) = 5.23$, $p < 0.001$, 95% CI [1.16, 3.11], $d = 0.55$, and improvised expressions ($M = 41.46$, $SD = 22.46$), $t(88) = 4.37$, $p < 0.001$, 95% CI [1.64, 3.11], $d = 0.46$. The difference in accuracy between rehearsed and improvised expressions was not significant, $t(88) = 2.23$, $p = 0.028$, 95% CI [0.05, 0.87], $d = 0.24$. The interaction term was not significant, $F(1.23,106.94) = 1.72$, $p = 0.193$ (**Figure 3**). When comparing the accuracy scores to chance performance (33.3%), genuine expressions were classified significantly above chance level, $t(88) = 11.14$, $p < 0.001$, 95% CI [2.63, 3.77], $d = 1.18$, as were improvised expressions, $t(88) = 3.42$, $p = 0.001$, 95% CI [0.44, 1.68], $d = 0.36$. However, rehearsed expressions were no different from chance (Bonferroni corrected), $t(88) = 2.01$, $p = 0.048$, 95% CI [0.01, 1.19].

Intensity

There was a main effect of Expression, $F(2,174) = 15.72$, $p < 0.001$, $\eta_p^2 = 0.153$, but no effect of Format, $F(1,87) = 1.22$, $p = 272$, on perceived intensity. The interaction between the two factors was significant, $F(2,174) = 19.98$, $p < 0.001$, $\eta_p^2 = 0.187$ (**Figure 4**).
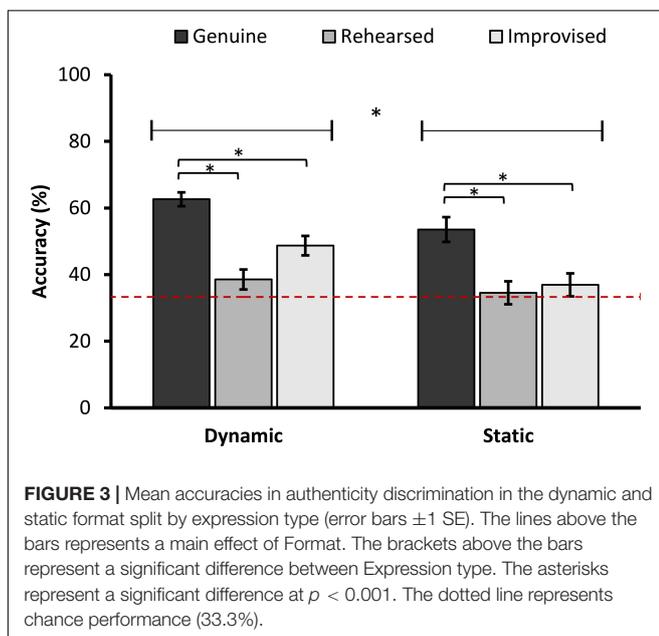
When decomposing the interaction, simple effects analyses revealed a significant main effect of Expression in the dynamic condition, $F(2,86) = 25.38$ $p < 0.001$, $\eta_p^2 = 0.371$. Pairwise comparisons with Bonferroni correction showed that genuine expressions ($M = 43.00$, $SD = 5.07$) were rated as more intense than rehearsed ($M = 37.81$, $SD = 6.27$), $t(42) = 6.63$, $p < 0.001$, 95% CI [3.61, 6.77], $d = 0.70$, and improvised expressions ($M = 38.05$, $SD = 7.34$), $t(42) = 5.57$, $p < 0.001$, 95% CI [3.16, 6.75], $d = 0.59$. Both types of deliberate expressions did not, however, significantly differ from each other, $t < 1$, $p > 0.99$.

Additionally, a significant simple main effect of Expression in the static condition, $F(2,86) = 8.59$, $p < 0.001$, $\eta_p^2 = 0.166$, showed that genuine expressions ($M = 41.02$, $SD = 9.24$) were rated as less intense than improvised expressions ($M = 43.09$, $SD = 8.39$), $t(45) = -2.84$, $p = 0.007$, 95% CI [−3.53, −0.60], $d = 0.30$, but not rehearsed expressions, $t(45) = 1.35$, $p = 0.183$, 95% CI [−0.53, 2.65]. Improvised expressions were perceived as more intense than rehearsed expressions, $t(45) = 3.21$, $p = 0.002$, 95% CI [1.17, 5.10], $d = 0.34$.

When considering differences in intensity ratings between formats, simple effects analyses showed that improvised expressions were judged as significantly more intense when they were presented in dynamic than static format, $F(1,87) = 9.05$, $p = 0.003$, $\eta_p^2 = 0.094$. This difference did not occur in the context of genuine, $F(1,87) = 1.54$, $p = 0.218$, $\eta_p^2 = 0.017$, or rehearsed expressions, $F(1,87) = 1.39$, $p = 0.241$, $\eta_p^2 = 0.016$.

Confidence

The ANOVA revealed a main effect of Expression, $F(2,174) = 6.14$, $p = 0.003$, $\eta_p^2 = 0.066$, and a marginal significant effect of Format, $F(1,87) = 3.66$, $p = 0.059$, $\eta_p^2 = 0.040$, on confidence ratings. These effects were qualified by a significant interaction between the two factors, $F(2,174) = 8.78$, $p < 0.001$, $\eta_p^2 = 0.092$ (**Figure 5**).

When decomposing the interaction, the simple main effect of Expressions was significant in the dynamic condition,

FIGURE 3 | Mean accuracies in authenticity discrimination in the dynamic and static format split by expression type (error bars ±1 SE). The lines above the bars represents a main effect of Format. The brackets above the bars represent a significant difference between Expression type. The asterisks represent a significant difference at $p < 0.001$. The dotted line represents chance performance (33.3%).

FIGURE 4 | Mean ratings for perceived intensity of facial expressions (error bars ±1 SE). The asterisks represent a significant difference at *$p < 0.01$ and **$p < 0.001$.

FIGURE 5 | Confidence scores in the dynamic and static format split by expression type (error bars ±1 SE). The asterisks represent a significant difference at *$p < 0.01$ and **$p \leq 0.001$.

$F(2,86) = 14.29$, $p < 0.001$, $\eta_p^2 = 0.249$. Pairwise comparisons with Bonferroni correction showed that participants were less confident in their ratings of rehearsed ($M = 46.67$, $SD = 6.74$) and improvised expressions ($M = 47.53$, $SD = 6.83$), compared to genuine expressions ($M = 50.00$, $SD = 7.48$), $t(42) = 4.13$, $p < 0.001$, 95% CI [1.70, 4.95], $d = 0.44$, $t(42) = 3.76$, $p = 0.001$, 95% CI [1.14, 3.79], $d = 0.40$. The two deliberate expressions did not significantly differ from each other, $t(42) = 1.11$, $p = 0.27$, 95% CI [−0.70, 2.42].

The simple main effect of Expression was not significant in the static condition, $F < 1$, $p > 0.75$.

When considering differences in confidence ratings between formats, simple effects analyses showed that genuine expressions were more confidently judged in the dynamic than static condition, $F(1,87) = 8.59$, $p = 0.004$, $\eta_p^2 = 0.090$. Neither ratings of improvised, $F(1,87) = 2.11$, $p = 0.150$, $\eta_p^2 = 0.024$, nor rehearsed expressions, $F(1,87) = 1.15$, $p = 0.287$, $\eta_p^2 = 0.013$, were affected by presentation format.

DISCUSSION

Emotions are a central aspect of social interactions, however, not all expressions of emotion are genuine. Knowing the authenticity of an expression can be a crucial factor in determining our perception of and interaction with others (Johnston et al., 2010). Here, we investigated decoders' ability to discriminate genuine expressions of surprise from deliberate expressions produced after a recent experience with actual surprise or in its absence, presented both in dynamic and static format. Our results support our predictions, finding significant effects due to both presentation format and expression type. We extend past emotion perception research by considering how different methods of producing an expression can affect perception and authenticity discrimination.

Genuine expressions, when presented dynamically, were perceived both genuine-looking and intense, echoing past findings (Sato and Yoshikawa, 2004; Krumhuber et al., 2013). These were also the most accurately discriminated as having occurred in the presence of an affective event (i.e., seeing the jack-in-the-box) and the most confidently judged by decoders, compared to the two deliberate expression types. In static presentation, genuine expressions were still the most accurately discriminated, but markedly lower than when presented dynamically. Conversely to the alternative presentation, in static format, these were rated as more genuine than rehearsed expressions, but equal to improvised expressions on genuineness. Decoders' judgmental confidence did not differ between expression types, and was significantly lower than in dynamic presentation.

For the deliberate conditions, in line with our predictions, rehearsed expressions presented dynamically were rated as appearing more genuine than improvised expressions, but still lower than genuine expressions. They were also perceived as less intense than genuine expression, but equal to improvised expressions. Decoders were poor at detecting rehearsed expressions as being deliberate, showing the lowest overall accuracy. Confidence was equal to that of improvised expressions, but still lower than genuine. When presented statically, however, rehearsed expressions were rated lower than improvised expressions in terms of genuineness, but equally on intensity and judgment confidence to genuine expression. Lastly, improvised expressions, in dynamic format, were rated the least genuine-looking of all expressions (rated negatively), but rated equally intense and confidently to rehearsed expressions. These expressions were also poorly discriminated as being deliberate. When presented statically, their intensity ratings were significantly higher than those of all other expressions, confirming our predictions; they also were perceived equally genuine-looking and judged as confidently as genuine expressions.

These findings have important methodological implications for the emotion field. To understand human emotion perception, we argue, considerations must be given to (1) the ability to separate genuine from deliberate expressions of emotions, and (2) differences in how the emotion stimuli are produced, as it is clear that these can significantly impact decoder perception. Presentation format was also an important factor in emotion perception (Hess and Kleck, 1994; Ambadar et al., 2005). Expressions presented dynamically were more accurately discriminated, were judged more confidently, and differences in their perceived intensity and genuineness were more pronounced; static presentation limited such perceptual differences between expressions.

Past inconsistencies reported for decoders' ability to discriminate expression authenticity (e.g., McLellan et al., 2010; Porter et al., 2012), we suggest, may be resolved by considering the type of expressions used and the presentation format. Here, decoders displayed some perceptual ability in recognizing genuine surprise (static and dynamic), but accuracy was not perfect. While for the deliberate expressions, their ability to discriminate these as not being genuine was poor, in

both formats, and varied by expression type (marginally); these performances were even poorer when presenting the expressions as static faces. Decoders, also, showed no self-awareness relating to their accuracy; while they perceived differences in expression intensity, genuineness, and even judgment confidence (predominantly in dynamic presentation), these did not aid authenticity discrimination. Given these performances, it would suggest that decoders do not possess a finely tuned perceptual mechanism to discriminate facial expression authenticity, as they do for emotion categorization.

In the current study, decoders evaluated the expression in the absence of external or contextual information. Eliciting the expressions in a controlled environment permitted a clear comparison between different expression types. However, decoders are unlikely to see such isolated expressions in everyday scenarios with the sole task of detecting authenticity (Reisenzein et al., 2006). This may partly explain why using emotional cues as markers for deception does not produce improvements in accuracy (Porter et al., 2012). Relying on such "cues" will not be beneficial unless decoders can discriminate if these are genuine or deceptive (see Zloteanu, 2015). An interpretation of the current findings is that senders are capable of producing expressions that look sufficiently genuine to fool decoders (Krumhuber and Manstead, 2009; Gunnery et al., 2013). Emotional expressions, thus, can be a strategic tool in communication, used to instill a specific affective belief in the decoder, which benefits the sender. It is not difficult to extend this logic to other deceptive scenarios, such as high-stakes criminal lies, where producing a deceptive expression might help escape suspicion (e.g., Porter et al., 2012). Our findings cast doubt that in a real-world setting where people are not instructed to classify the authenticity of emotional displays, and where emotions tend to more ambiguous, observers could accurately distinguish genuine from deceptive emotional signals. Alternatively, context can, in certain scenarios, aid authenticity judgments (Blair et al., 2010). Removing context from the judgment task may in turn have affected decoders, as the information which may hint that an expression is genuine/fake was absent.

The current consideration for expression type can also aid our understanding of emotion recognition. Intensity is considered an important component in the perception and accurate classification of emotions (Hess et al., 1997). It has been argued that deliberate expressions may appear either *less* intense in presentation, as they are absent of the underlying affect (Levenson, 2014), or *more* intense, as they are attempts by the sender to communicate information successfully (Calder et al., 2000). Given the current results, this may be resolved by considering how the expressions are produced. Namely, rehearsed expression were perceived as less intense than genuine expressions (in dynamic format), while improvised expressions were perceived as more intense (in static format). For this reason, differences on emotion perception tasks may occur based on the authenticity of the stimuli (i.e., genuine or deliberate), the type of production method used (e.g., rehearsed or improvised), the presentation used (i.e., dynamic or static), or a combination of these factors. For instance, using static improvised expressions in a recognition task, due to their perceived high intensity,

may result in overinflated recognition rates for surprise. Regarding authenticity discrimination, intensity did not show any relationship with accuracy, in either dynamic or static presentation. Thus, facial intensity seems not to be diagnostic of authenticity, but more related to the method of production used to elicit the expression.

Finally, dynamic presentation of facial displays offers clear benefits to emotion research. Given the current data, it is clear that using ecologically valid stimuli that reflect genuine expressions allow for subtle differences between expression types to be perceived by decoders, and offer a more realistic approximation of human emotion perception (Trautmann et al., 2009; Sauter and Fischer, 2017). Future research should expand the current findings to explore how decoders perceive other emotions, given the variation in perception and accuracy based on valence and category (see Barrett, 1998), and extended to more social emotions, such as shame and embarrassment (e.g., Tracy and Matsumoto, 2008), to better understand emotion production and perception. Expansions may also consider individual differences in expressive control (Berenbaum and Rotter, 1992) and emotion regulation (Gross, 2002) as factors for the successful production of deliberate expressions. Such work may examine how expressive variability relates to perceptual accuracy, by considering an inter-item analysis of the current stimuli or by directly measuring expressive behavior in the task (e.g., using automated facial expression analysis; Valstar et al., 2006). Also, different emotions could have different effects in terms of senders' ability to display genuine-looking expressions and decoders' ability to discriminate authenticity. For instance, the current approach did not consider the role of the gender of the sender, which some research suggests may affect perception (e.g., Krumhuber et al., 2006); future research should test for gender differences in production and perception.

CONCLUSION

The ability to accurately discriminate and perceive differences in expressions of surprise was affected by both the type of deliberate expressions seen and the way they were presented. Even when asked to specifically judge authenticity, decoders were not adept at separating genuine from deliberate expressions of surprise. While they showed some ability to accurately detect genuine surprise, they also tended to misclassify deliberate expressions as genuine, regardless of expression type. The way in which the deliberate expressions were produced also affected how they were perceived. Rehearsed expressions, in a dynamic format, were perceived as more genuine in appearance than their improvised counterparts and were slightly more difficult to detect as non-genuine. In comparison, improvised expressions were rated as more intense and genuine in appearance in a static format. This supports our predictions of perceptual differences between genuine and deliberate expressions occurring as a result of the method used to produce and present the stimuli. For measuring differences in human emotion perception and accurate authenticity discrimination a dynamic presentation was found to be superior, allowing for nuanced perceptions

of intensity, genuineness, and judgment confidence between expressions. Together, the findings illustrate the complexity of human emotion production and perception, the need for ecologically valid stimuli, and the importance of considering expression type in emotion research.

AUTHOR CONTRIBUTIONS

MZ and DR conceived and designed the experiments. MZ performed the experiments. MZ and EK analyzed the data. MZ, EK, and DR contributed to writing the paper.

REFERENCES

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Bänziger, T., and Scherer, K. R. (2007). "Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus," in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, (Berlin: Springer), 476–487. doi: 10.1007/978-3-540-74889-2_42

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cogn. Emot.* 12, 579–599. doi: 10.1080/026999398379574

Berenbaum, H., and Rotter, A. (1992). The relationship between spontaneous facial expressions of emotion and voluntary control of facial muscles. *J. Nonverbal Behav.* 16, 179–190. doi: 10.1007/bf00988033

Biele, C., and Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Exp. Brain Res.* 171, 1–6. doi: 10.1007/s00221-005-0254-0

Blair, J. P., Levine, T. R., and Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Hum. Commun. Res.* 36, 423–442. doi: 10.1111/j.1468-2958.2010.01382.x

Bond, C. F., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234. doi: 10.1207/s15327957pspr1003_2

Calder, A. J., Rowland, D., Young, A. W., Nimmo-Smith, I., Keane, J., and Perrett, D. I. (2000). Caricaturing facial expressions. *Cognition* 76, 105–146. doi: 10.1016/S0010-0277(00)00074-3

Calvo, M. G., and Nummenmaa, L. (2015). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Casler, K., Bickel, L., and Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Comput. Hum. Behav.* 29, 2156–2160. doi: 10.1016/j.chb.2013.05.009

Cohn, J. F., and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multiresolut. Inf. Process.* 2, 121–132. doi: 10.1142/s021969130400041x

Dawel, A., Wright, L., Irons, J., Dumbleton, R., Palermo, R., O'Kearney, R., et al. (2016). Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behav. Res. Methods* 49, 1539–1562. doi: 10.3758/s13428-016-0813-2

DePaulo, P. J., and DePaulo, B. M. (1989). Can deception by salespersons and customers be detected through nonverbal behavioral cues?. *J. Appl. Soc. Psychol.* 19, 1552–1577. doi: 10.1111/j.1559-1816.1989.tb01463.x

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life.* New York, NY: Owl Books.

Ekman, P. (2004). "Emotional and conversational nonverbal signals," in *Language, Knowledge, and Representation*, eds J. M. Larrazabal, and L. A. P. Miranda (Dordrecht: Springer), 39–50. doi: 10.1007/978-1-4020-2783-3_3

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/bf00987191

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science* 221, 1208–1210. doi: 10.1126/science.6612338

Ekman, P., and Rosenberg, E. (2005). *What the Face Reveals*, 2nd Edn. New York, NY: Oxford.

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146

Field, T. M., and Walden, T. A. (1982). Production and discrimination of facial expressions by preschool children. *Child Dev.* 53, 1299–1311. doi: 10.2307/1129020

Forgas, J. P., and East, R. (2008). How real is that smile? Mood effects on accepting or rejecting the veracity of emotional facial expressions. *J. Nonverbal Behav.* 32, 157–170. doi: 10.1007/s10919-008-0050-1

Frank, M. G., and Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *J. Pers. Soc. Psychol.* 72, 1429–1439. doi: 10.1037/0022-3514.72.6.1429

Gross, J. J. (2002). Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology* 39, 281–291. doi: 10.1017/s0048577201393198

Gunnery, S. D., Hall, J. A., and Ruben, M. A. (2013). The deliberate duchenne smile: individual differences in expressive control. *J. Nonverbal Behav.* 37, 29–41. doi: 10.1007/s10919-012-0139-4

Hess, U., Blairy, S., and Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *J. Nonverbal Behav.* 21, 241–257. doi: 10.1023/A:1024952730333

Hess, U., and Kleck, R. E. (1994). The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions. *Eur. J. Soc. Psychol.* 24, 367–381. doi: 10.1002/ejsp.2420240306

Johnston, L., Miles, L., and Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *Br. J. Soc. Psychol.* 49, 107–127. doi: 10.1348/014466609x412476

Krumhuber, E., and Kappas, A. (2005). Moving smiles: the role of dynamic components for the perception of the genuineness of smiles. *J. Nonverbal Behav.* 29, 3–24. doi: 10.1007/s10919-004-0887-x

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., and Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion* 9, 807–820. doi: 10.1037/a0017844

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emot. Rev.* 6, 100–112. doi: 10.1177/1754073913512003

Levine, T. R., Park, H. S., and McCornack, S. A. (1999). Accuracy in detecting truths and lies: documenting the "veracity effect". *Commun. Monogr.* 66, 125–144. doi: 10.1080/03637759909376468

Lewis, M., Sullivan, M. W., and Vasen, A. (1987). Making faces: age and emotion differences in the posing of emotional expressions. *Dev. Psychol.* 23, 690–697. doi: 10.1037/0012-1649.23.5.690

Maringer, M., Krumhuber, E. G., Fischer, A. H., and Niedenthal, P. M. (2011). Beyond smile dynamics: Mimicry and beliefs in judgments of smiles. *Emotion* 11, 181–187. doi: 10.1037/a0022596

McLellan, T., Johnston, L., Dalrymple-Alford, J., and Porter, R. J. (2010). Sensitivity to genuine versus posed emotion specified in facial displays. *Cogn. Emot.* 24, 1277–1292. doi: 10.1080/02699930903306181

Morecraft, R. J., Louie, J. L., Herrick, J. L., and Stilwell-Morecraft, K. S. (2001). Cortical innervation of the facial nucleus in the non-human primate: a new interpretation of the effects of stroke and related subtotal brain trauma on the muscles of facial expression. *Brain* 124, 176–208. doi: 10.1093/brain/124.1.176

Namba, S., Makihara, S., Kabir, R. S., Miyatani, M., and Nakao, T. (2016). Spontaneous facial expressions are different from posed facial expressions: morphological properties and dynamic sequences. *Curr. Psychol.* 36, 593–605. doi: 10.1007/s12144-016-9448-9

Nelson, N. L., and Russell, J. A. (2013). Universality revisited. *Emot. Rev.* 5, 8–15. doi: 10.1177/1754073912457227

Porter, S., ten Brinke, L., and Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. . *J. Nonverbal Behav.* 36, 23–37. doi: 10.1007/s10919-011-0120-7

Reisenzein, R., Bördgen, S., Holtbernd, T., and Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: the case of surprise. *J. Pers. Soc. Psychol.* 91, 295–315. doi: 10.1037/0022-3514.91.2.295

Reissland, N., Shepherd, J., and Cowie, L. (2002). The melody of surprise: maternal surprise vocalizations during play with her infant. *Infant Child Dev.* 11, 271–278. doi: 10.1002/icd.258

Rinn, W. E. (1984). The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. *Psychol. Bull.* 95, 52–77. doi: 10.1037/0033-2909.95.1.52

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol. Bull.* 115, 102–141. doi: 10.1037/0033-2909.115.1.102

Sato, W., Fujimura, T., and Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *Int. J. Psychophysiol.* 70, 70–74. doi: 10.1016/j.ijpsycho.2008.06.001

Sato, W., and Yoshikawa, S. (2004). BRIEF REPORT: the dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Sauter, D. A., and Fischer, A. H. (2017). Can perceivers recognise emotions from spontaneous expressions? *Cogn. Emot.* 32, 504–515. doi: 10.1080/02699931.2017.1320978

Scherer, K. R., and Bänziger, T. (2010). "On the use of actor portrayals in research on emotional expression," in *A Blueprint for Affective Computing: A Sourcebook and Manual*, eds K. R. Scherer, T. Bänziger, and E. Roesch (Oxford: Oxford University Press), 166–176.

Tracy, J. L., and Matsumoto, D. (2008). The spontaneous expression of pride and shame: evidence for biologically innate nonverbal displays. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11655–11660. doi: 10.1073/pnas.0802686105

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Valstar, M. F., Pantic, M., Ambadar, Z., and Cohn, J. F. (2006). "Spontaneous vs. posed facial behavior: automatic analysis of brow actions," in *Proceedings of the 8th International Conference on Multimodal Interfaces ICMI '06*, (New York, NY: ACM), 162–170. doi: 10.1145/1180995.1181031

Vazire, S., Naumann, L. P., Rentfrow, P. J., and Gosling, S. D. (2009). Smiling reflects different emotions in men and women. *Behav. Brain Sci.* 32, 403–405. doi: 10.1017/S0140525X09991026

Vrij, A., and Mann, S. (2001). Who killed my relative? Police officers' ability to detect real-life high-stake lies. *Psychol. Crime Law* 7, 119–132. doi: 10.1080/10683160108401791

Zloteanu, M. (2015). "The role of emotions in detecting deception," in *Deception: An Interdisciplinary Exploration*, eds E. Williams and I. Sheeha (Oxford: Inter-Disciplinary Press), 203–217.

Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1986). "Humans as deceivers and lie detectors," in *Nonverbal Communication in the Clinical Context*, eds P. D. Blanck, R. Buck, and R. Rosenthal (University Park: Pennsylvania State University Press), 13–35.

Temporal Dynamics of Natural Static Emotional Facial Expressions Decoding: A Study Using Event- and Eye Fixation-Related Potentials

Anne Guérin-Dugué[1]*, Raphaëlle N. Roy[2], Emmanuelle Kristensen[1,3], Bertrand Rivet[1], Laurent Vercueil[4,5] and Anna Tcherkassof[3]

[1] GIPSA-lab, Institute of Engineering, Université Grenoble Alpes, Centre National de la Recherche Scientifique, Grenoble INP, Grenoble, France, [2] Department of Conception and Control of Aeronautical and Spatial Vehicles, Institut Supérieur de l'Aéronautique et de l'Espace, Université Fédérale de Toulouse, Toulouse, France, [3] Laboratoire InterUniversitaire de Psychologie – Personnalité, Cognition, Changement Social, Université Grenoble Alpes, Université Savoie Mont Blanc, Grenoble, France, [4] Exploration Fonctionnelle du Système Nerveux, Pôle Psychiatrie, Neurologie et Rééducation Neurologique, CHU Grenoble Alpes, Grenoble, France, [5] Université Grenoble Alpes, Inserm, CHU Grenoble Alpes, Grenoble Institut des Neurosciences, Grenoble, France

This study aims at examining the precise temporal dynamics of the emotional facial decoding as it unfolds in the brain, according to the emotions displayed. To characterize this processing as it occurs in ecological settings, we focused on unconstrained visual explorations of natural emotional faces (i.e., free eye movements). The General Linear Model (GLM; Smith and Kutas, 2015a,b; Kristensen et al., 2017a) enables such a depiction. It allows deconvolving adjacent overlapping responses of the eye fixation-related potentials (EFRPs) elicited by the subsequent fixations and the event-related potentials (ERPs) elicited at the stimuli onset. Nineteen participants were displayed with spontaneous static facial expressions of emotions (Neutral, Disgust, Surprise, and Happiness) from the DynEmo database (Tcherkassof et al., 2013). Behavioral results on participants' eye movements show that the usual diagnostic features in emotional decoding (eyes for negative facial displays and mouth for positive ones) are consistent with the literature. The impact of emotional category on both the ERPs and the EFRPs elicited by the free exploration of the emotional faces is observed upon the temporal dynamics of the emotional facial expression processing. Regarding the ERP at stimulus onset, there is a significant emotion-dependent modulation of the P2–P3 complex and LPP components' amplitude at the left frontal site for the ERPs computed by averaging. Yet, the GLM reveals the impact of subsequent fixations on the ERPs time-locked on stimulus onset. Results are also in line with the valence hypothesis. The observed differences between the two estimation methods (Average vs. GLM) suggest the predominance of the right hemisphere at the stimulus onset and the implication of the left hemisphere in the processing of the information encoded by subsequent fixations. Concerning the first EFRP, the Lambda response and the P2 component are modulated by the emotion of surprise compared to the neutral emotion, suggesting

an impact of high-level factors, in parieto-occipital sites. Moreover, no difference is observed on the second and subsequent EFRP. Taken together, the results stress the significant gain obtained in analyzing the EFRPs using the GLM method and pave the way toward efficient ecological emotional dynamic stimuli analyses.

Keywords: emotional facial expression, natural faces, event-related potential, eye fixation-related potential, temporal dynamics, General Linear Model

INTRODUCTION

The investigation of the electrocerebral responses to emotional facial expressions (EFEs) is a privileged mean to understand how people process the emotions they see in others' faces (Ahern and Scharwtz, 1985). To evaluate brain responses to EFE processing, most studies use the same experimental protocol. Pictures of EFE are presented during a short time and participants are asked to fixate at the center point of the image while the electroencephalographic signals are recorded. The brain response at the EFE presentation is estimated by averaging the EEG signal time-locked at this stimulus onset. Only synchronous activities elicited at the stimulus presentation contribute to this evoked potential (event-related potential, ERP) when averaging. A main assumption underlies this methodology, that of a unique potential elicited by the event of interest. If the presentation duration is very short and if there is only one ocular fixation at the image center and no eye movement afterward, this estimation for the evoked potential at the image onset is a good solution. Research based on this protocol shows two main stages in the time course of EFE processing. The first stage is a perceptual processing occurring early and stemming from the activity of occipital and temporal regions. The second stage is a conscious recognition one involving a more complex set of activations from frontal and subcortical structures (Adolphs, 2002). Some researchers have posited that the first stage is not impacted by valence and would merely reflect raw structural processing. This view is supported by some ERPs based studies on EFE processing. For instance, Eimer et al. (2003) found that emotional faces elicited higher amplitudes than neutral ones for late components but not for early ones. In the same vein, Almeida et al. (2016) found that arousal, but not valence of the EFE, modulates the amplitude of the N170. However, much more studies have shown that valence does in fact impact early EFE processing, unveiling a very rapid and early top-down modulation during this perceptual stage or at least "rapid emotion processing based on crude visual cues in faces" (Vuillemier and Pourtois, 2007). Indeed, differences in latency and amplitudes of ERP components can occur as early as the first 100 ms post-stimulation, e.g., P1 component (Batty and Taylor, 2003; Neath and Itier, 2015; Itier and Neath-Tavares, 2017), as well as modulations of both latency and amplitude of the face-specific N170 component at posterior temporal-occipital sites (Pizzagalli et al., 1999; Campanella et al., 2002; Blau et al., 2007; Hinojosa et al., 2015; Itier and Neath-Tavares, 2017) and of an anterior negative component around 230 ms (Balconi and Pozzoli, 2003). Moreover, a valence-dependent modulation of a component called early posterior negativity (EPN) between 150 and 300 ms at occipito-parietal sites has

been found with a higher amplitude for EFEs than for neutral faces (Recio et al., 2011; Neath-Tavares and Itier, 2016; Itier and Neath-Tavares, 2017). This component can be computed by subtracting the ERP elicited by neutral faces to that of the emotional ones. If no subtraction is performed, the component is akin to a P2 component at posterior sites and a N2 one at anterior sites. In this article, the subsequent occurrence of a P2 and a P3 components at posterior sites will be referred to as a P2–P3 complex in order to avoid any confusion. Additionally, a modulation of late ERP components has been consistently reported and would reflect a conscious recognition process of EFEs. Hence, a valence-dependent amplitude modulation of a positive component around 350 ms at fronto-central sites and of the late positive potential (LPP) at all sites has been reported (Krolak-Salmon et al., 2001; Batty and Taylor, 2003; Trautmann-Lengsfeld et al., 2013). Moreover, Recio et al. (2011) reported an emotion-dependent modulation for EFE processing of a component akin to the LPP, called the late positive complex (LPC). This long lasting positivity component peaks at 500 ms over centro-parietal sites and is computed by subtracting the neutral ERP from the emotional ones. All in all, the time-course of EFE processing is now precisely documented by studies using the same experimental protocol. This said, the question remains as to whether results obtained with such a protocol can be transposed to everyday occurring EFE processing.

In recent years, there has been a growing interest for the analysis of ecological human behaviors during daily life interactions. It is especially the case for researchers concerned with realism, notably for pragmatic matters (Calvo and D'Mello, 2011). Unfortunately, the generalizability to ordinary emotional behaviors of mostly all results on EFE processing is unlikely because experimental methodologies lack ecological validity. Two key criticisms can be made. The first one concerns the stimuli. Research on the time course of EFE processing is undertaken with EFEs most often coming from the Pictures of Facial Affect database (POFA; Ekman and Friesen, 1976). The use of this dataset promotes comparison between studies and optimizes experimental conditions. However, these non-natural stimuli (EFEs of actors/actresses produced in non-natural contexts) are subjected to many criticisms (Tcherkassof et al., 2007). There are radical differences between non-natural behavioral stimuli (i.e., deliberate emotional displays) used in laboratory studies and those exhibited in everyday life (i.e., spontaneous emotional displays). Research on facial expression has highlighted how a crucial expressive feature of natural displays spontaneity is. Spontaneously occurring behavior differs in various aspects from deliberate behavior (Kanade et al., 2000), including timing and visual appearance (Hess and Kleck, 1990;

Cohn and Schmidt, 2004). From an ecological perspective, affective analyses based on deliberate EFEs have a poor generalization capacity, which is why they need examples of naturally expressed EFEs and not prototypical patterns of facial behaviors such as POFA's ones. Consequently, because they lack ecological validity, one questions the generalizability of experimental results that rely on unnatural stimuli. The second key criticism concerns the stimulus presentation. As mentioned above, participants are asked to fixate the center point of the stimulus which is displayed during a short time. However, this is far from an ordinary kind of activity. In an effort to overcome this issue, a free exploration paradigm using eye-tracking and EEG co-registration has been developed to evaluate the cognitive processing of stimuli in an ecological way. Eye-tracking methods are privileged means to examine the allocation of observers' attention to different facial regions and to examine the relationship between gaze patterns and EFE processing (Schurgin et al., 2014). For joint EEG and eye movements recording, Kaunitz et al. (2014) extracted the eye fixation-related potentials (EFRPs) elicited by faces in a crowd. Simola et al. (2013, 2015) analyzed the eye movement-related brain responses to emotional scenes. Regarding EFEs processing more specifically, Neath-Tavares and Itier (2016) studied the co-registration of eye-tracking and EEG. However, they used a gaze-contingent procedure in order to test the diagnostic impact of different facial regions of interest (i.e., mouth, nose, left eye, right eye) in EFE processing. Hence, they did not study EFE processing in an ecological way since participants could not freely explore the stimuli but rather had specific regions of interest presented directly at their first and only fixation point. In any case, for joint EEG and eye movements recording, the use of both free eye-tracking and EEG co-registration in a free exploration paradigm raises questions as to estimating the evoked potential. When the stimulus is presented for a long duration and if the eye positions are not controlled to be stable, the usual estimation methodology by averaging is debatable: the potential estimated at the image onset not only reveals the potential directly elicited at the image presentation, but also the successive contributions of the visual information processing at each fixation rank. For example, in the case of the window's latency of the LPP component (around 600–800 ms after the stimulus onset), it is reasonable to expect that one ocular fixation *before* this latency already occurred, even one fixation *during* this latency window also. In the study of Trautmann-Lengsfeld et al. (2013), the static stimulus was presented during 1,500 ms and the eye movements were not controlled. As a matter of fact, the neural activity observed during the latency window of the LPP component revealed not only the activity elicited by the stimulus presentation but also the activity provided by the early visual exploration of the stimulus with eye movements. This was in line with the objective of the Trautmann-Lengsfeld's study which was the comparison between the perception of static and dynamic EFE. However, the stimuli in the Trautmann-Lengsfeld's study weren't spontaneous EFE but posed ones thus reducing the ecological significance of the results. Therefore, up to now, no study has yet examined the precise temporal dynamics of the EFE processing using eye-tracking and EEG co-registration with both ecological stimuli

(i.e., freely expressed by ordinary people) and paradigm (i.e., free visual exploration). The present study aims at filling these data gaps.

The goal of our study is to study the EFE processing in ecological settings. We focused on unconstrained visual explorations (i.e., free eye movements) of natural emotional faces (spontaneous EFEs) contrary to what is usually done (i.e., fixed eye gaze and unnatural stimuli). As the Trautmann-Lengsfeld's study, the stimuli in the present study were presented during a long time and the participants were free to visually explore them. This protocol is ecological from a visual exploration point of view both for dynamic and static stimuli, but used here for static stimuli. As a consequence, the methodology to estimate the evoked potentials has to be adapted to this experimental design. At the stimulus presentation and during the subsequent visual exploration by the ocular fixations, several cognitive processes are engaged. The estimation of the evoked potentials by averaging therefore fails to provide a reliable estimation of each of them because these brain responses overlap which each other. Consequently, in order to analyze the temporal dynamics of the EFE processing, a methodology is applied to distinguish between what is due to the stimulus presentation and what is due to its exploration. The precise objectives and hypotheses of the present study are as follow. Recent methodological studies on evoked potentials estimation have shown how promising are linear models decomposing the effects of different neural activities during a same temporal window (Burns et al., 2013; Bardy et al., 2014; Smith and Kutas, 2015a,b; Congedo et al., 2016; Kristensen et al., 2017a). This methodology is based on the General Linear Model (GLM). It is particularly suitable to estimate EFRPs and is more flexible (Kristensen et al., 2017b) than the ADJAR algorithm (Woldorff, 1993). This method has been recently implemented with success in EFRP/ESRPs (Eye Saccade-Related Potentials) estimation (Dandekar et al., 2012; Kristensen et al., 2017b). However, it has never been applied to the emotional field. Our aim is thus to exploit it for EEG activity, in order to examine the time course of the EFE processing during the very beginning of its visual exploration. Using the GLM, we hypothesize that it should be possible to deconvolve adjacent overlapping responses of the EFRPs elicited by the subsequent fixations and the ERPs elicited at the stimulus onset. If our hypothesis is correct, the time course of the early emotional processing could be analyzed through these estimated potentials. For this purpose, a joint EEG-eye tracking experiment was set up to take benefit on both synchronized experimental modalities (Dimigen et al., 2011; Nikolaev et al., 2016). Based on the valence hypothesis, the expected results are differences in ERPs' amplitude depending on both valence and hemisphere (e.g., higher amplitude on left hemisphere for ERP components of positive emotions). According to the literature about the valence hypothesis (see for instance Graham and Cabeza, 2001; Wager et al., 2003; Alves et al., 2008), the effects should be located in the frontal regions. Moreover, and more importantly, based on articles by Noordewier and Breugelmans (2013), Noordewier et al. (2016), we expect that by analyzing the EFRPs of the first fixation, specific processes that might be valence-dependent will be differentiated.

MATERIALS AND METHODS

Database

The DynEmo database (Tcherkassof et al., 2013) is a comprehensive resource of filmed affective facial behaviors which provides a substantial publicly available corpus of validated dynamic and natural facial expressions of pervasive affective states (i.e., representative of daily life affective expression). It supplies 358 EFE recordings performed by a wide range of ordinary people, from young to older adults of both genders (ages 25–65, 182 females and 176 males) filmed in natural but standardized conditions. DynEmo provides genuine facial expressions—or first-order displays—exhibited in the course of a given eliciting episode. The conditions that influence or cause affective behavior, whether internal or external to the expresser (her/his affective state, current situation, etc.), are known to the user. One-third of the EFE recordings have been displayed to observers who have rated (continuous annotations) the emotions displayed throughout the recordings. The dynamic aspects of these EFE recordings and their relationship to the observers' interpretations are displayed in timelines. Such synchronized measures of expressing and decoding activities allow for a moment-by-moment analysis that simultaneously considers the expresser's facial changes and the observer's answers (independently of what is experienced by the expresser). Indeed, for each video, emotional expression timelines instantly signal when a given affective state is considered to be displayed on the face. Therefore, segmentation of the expressions into small emotion excerpts is easily achieved. For our study, the most expressive videos (maximum observers' rates) were selected for four emotions (happiness, surprise, fear, and disgust). Then, inside each short EFE clip, the image corresponding to the maximum expression intensity ("apex") was extracted. The final stimuli dataset was composed of 118 static EFE stimuli: neutral (24), happiness (12), surprise (12), fear (10), disgust (12), distractors (48), i.e., seventy target EFE stimuli with a controlled high recognition rate, and forty eight EFE diversion stimuli.

Stimuli

Stimuli consist of 118 color EFE images, all equalized in luminance. The images' resolution is 768 × 1024 pixels, subtended 30 × 40° of visual angle. **Figures 1A–F** illustrates one EFE stimulus for each emotional category. They were displayed onto a 20-inch ViewSonic CRT monitor located 57 cm in front of the participants of 768 × 1024 pixels and a 75 Hz refresh rate.

Participants

Thirty-one healthy adults participated in the experiment, but data from only nineteen subjects (7 women and 12 men aged from 20 to 32 years – mean age 25 years 7 months (SD = 3 years 2 months, SE = 9 months) – were used for all the analyses. Data from six participants were discarded due to technical problems during acquisition (poor eye-tracking calibration, noisy EEG signals). Data from two participants were discarded due to high energy in the alpha band [8–12 Hz] in the occipital which is a criterion of a loss of attention.

Finally, the data from two other participants were discarded because of a too low number of trials (see the Section "Brain Activity"). The remaining 19 participants were right-handed, except one male left-hander and one female left-hander. All participants had a normal or corrected-to-normal vision. They were free of any medical treatment at the time of the experiment, and had no history of neurological or psychiatric disorder. None of them had prior experience with the experimental task. All gave their written and informed consent prior to the experiment and were recompensed with 15€ in vouchers for their participation. The whole experiment was reviewed and approved by the ethics committee of Grenoble CHU ("Centre Hospitalier Universitaire") (RCB: n° 2011-A00845-36). The co-registration EEG/Eye tracker was performed at the IRMaGe Neurophysiology facility (Grenoble, France).

Experimental Protocol

Each run consisted of two separate and consecutive sessions. The eye movements and the EEG activity were recorded during both sessions. Note that only the recordings of the first session are presented here (the recordings of the second session being out of the scope of this article). In the first session, participants freely explored each static stimulus. To this end, they were asked to attentively watch the stimuli and to "empathize" with the displayed facial expressions. The 118 stimuli were randomly presented. The use of distractors in the first session aimed at preventing a memory effect on the emotional rating task carried out during the second session. Two short breaks were managed to avoid fatigue in participants. In a second session, participants had to rate each target EFE stimulus. The 70 target stimuli were presented in the same order than in the first session. Participants assessed the stimuli according to two scales: arousal on five levels, from −2 to 2 [not (−−) to highly (++) arousing], and the emotional stimulus category (happiness, surprise, fear, disgust, and neutral). As this session was shorter, only one short break was introduced.

The timeline of the trials (**Figures 2A,B**) were similar between the two sessions (except for the two stimuli emotional ratings during the second session). Each trial started with the display of the fixation cross at the center of the screen, followed by the EFE stimulus displayed during 2 s, then the emotional ratings for the second session only (**Figure 2B**), and ended with a gray screen (4 s) before the next trial. The fixation cross was displayed on the center of the screen to initialize the exploration, during a random duration from 700 to 1,200 ms, to avoid the development of saccade anticipation before the visual stimulus presentation. The stimulus was displayed after the stabilization of the participant's gaze on the fixation cross (during 500 ms before the end of its presentation, in a rectangle of 3° × 2° pixels around the fixation cross). During the second session, the display of the scales (arousal followed by emotion category) ended with the participant's answer (key press). The trial terminated with a 4 s gray screen during which the participant could relax and blink.

For eye-tracking purpose, a 9-point calibration routine was carried out at the beginning of each session. It was repeated every 20 trials or when the drift correction, performed every 10 trials, reported an error above 1°. The complete experiment was

FIGURE 1 | Example of EFE stimuli for each emotional category. **(A)** Happiness. **(B)** Surprise. **(C)** Fear. **(D)** Disgust. **(E)** Neutral. **(F)** Distractor.
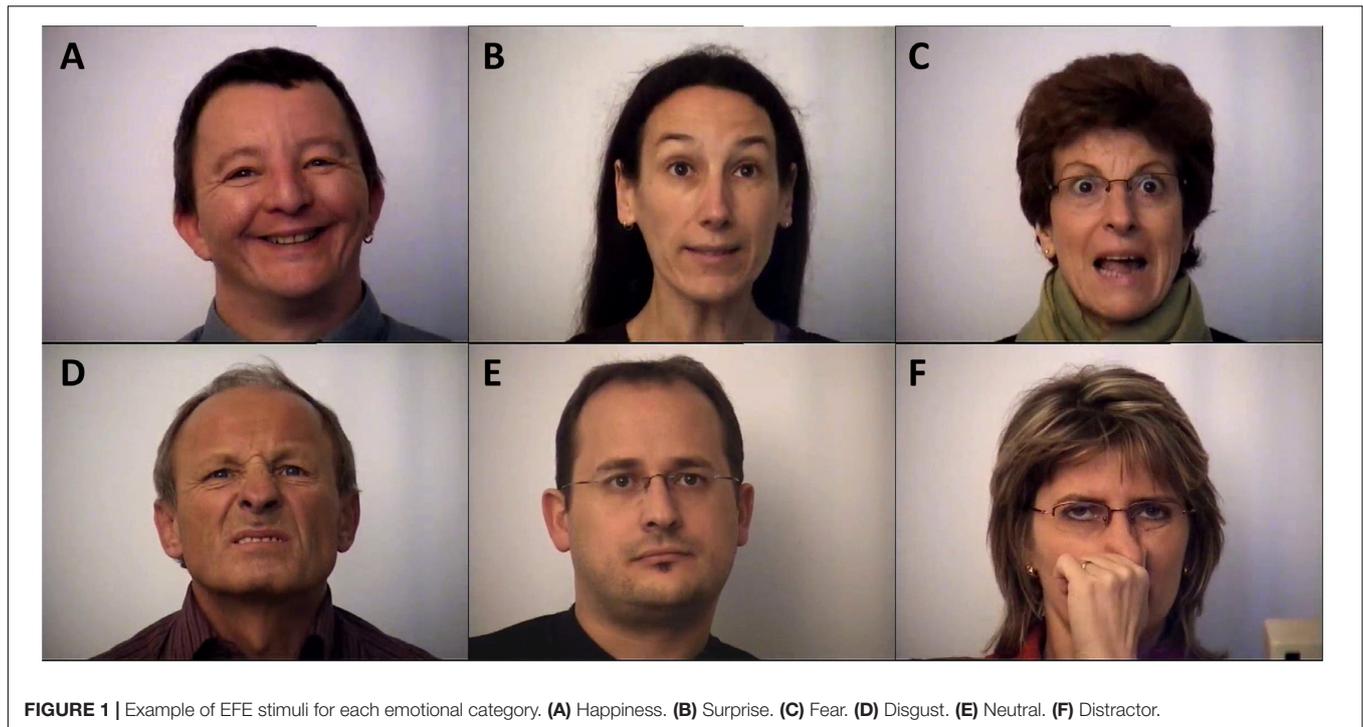
designed thanks to the SoftEye software (Ionescu et al., 2009) to control (1) the timescale for the displays, (2) the eye-tracker and (3) the sending of synchronization triggers to both devices.

Data Acquisition

Behavioral Measures

The behavioral data (EFEs' arousal level and emotion categorization) were analyzed to determine how participants rated the pictures of the database. For the emotion categorization data, we used the index computation method of Wagner (1997). This author has recommended using an unbiased hit rate (Hu) when studying the accuracy of facial expression recognition to take into account possible stimulus and response biases. Wagner's computation method combines the conditional probability that a stimulus will be recognized (given that it is presented) and the conditional probability that a response will be correct (given that it is used) into an estimate of the joint probability of both outcomes. This is done by multiplying together the two conditional probabilities divided by the appropriate marginal total (p. 50). Thus, the accuracy is a proportion of both responses and stimuli frequencies. Confusion matrixes are elaborated so that an unbiased hit rate (Hu) computed for each participant can be used as a dependent variable. The Hu ranges from 0 (no recognition at all) to 1 (complete recognition). Because the fear emotion was badly categorized (28% of the fear stimuli were recognized as surprise, and 16% of the fear stimuli were recognized as neutral), all data on the fear emotion were removed and analyses were conducted on four categories (Neutral, Disgust, Surprise, and Happiness). Moreover, the participants' emotional categorizations during the second session were used as a ground truth to analyze data recorded during the

first session. In other words, for a given participant, each EFE stimulus was re-categorized *post hoc* according to the emotion category the participant had assigned to the EFE. Thus, each participant had decoded a same emotion on slightly different subsets of stimuli. After decoding a given emotion, the associated subsets of stimuli (one subset per participant and per emotion) had large overlaps across participants, such as at least 50% of participants, in average, categorized 75% of same EFE stimuli into the same emotion. More precisely, this percentage of stimuli was distributed as follows across emotion: 75% for neutral, 75% for disgust, 50% for surprise and 100% for happiness.

The main argument supporting this re-categorization procedure is that encoding and decoding processes must not be confused, as stressed by Wagner (1997). Indeed, an encoder can express a given emotion when the decoder interprets this facial expression as displaying another emotion. As we are concerned by the decoding process, it justifies that we rely on the observer's judgment rather than on the encoder's emotion. This is especially relevant because the cerebral signals investigated are the ones corresponding to the observer's own judgment. Let us recall that only EEG data from the first session are analyzed in this study. In this protocol, the neutral condition was the control condition, compared to the three other EFE (disgust, surprise, and happiness).

Ocular Activity

For the sake of compatibility with this EEG acquisition, the remote binocular infrared eye-tracker EyeLink 1000 (SR Research) was used to track the gaze of the guiding eye of each participant while he/she was looking at the screen. The EyeLink system was used in the Pupil-Corneal Reflection tracking mode
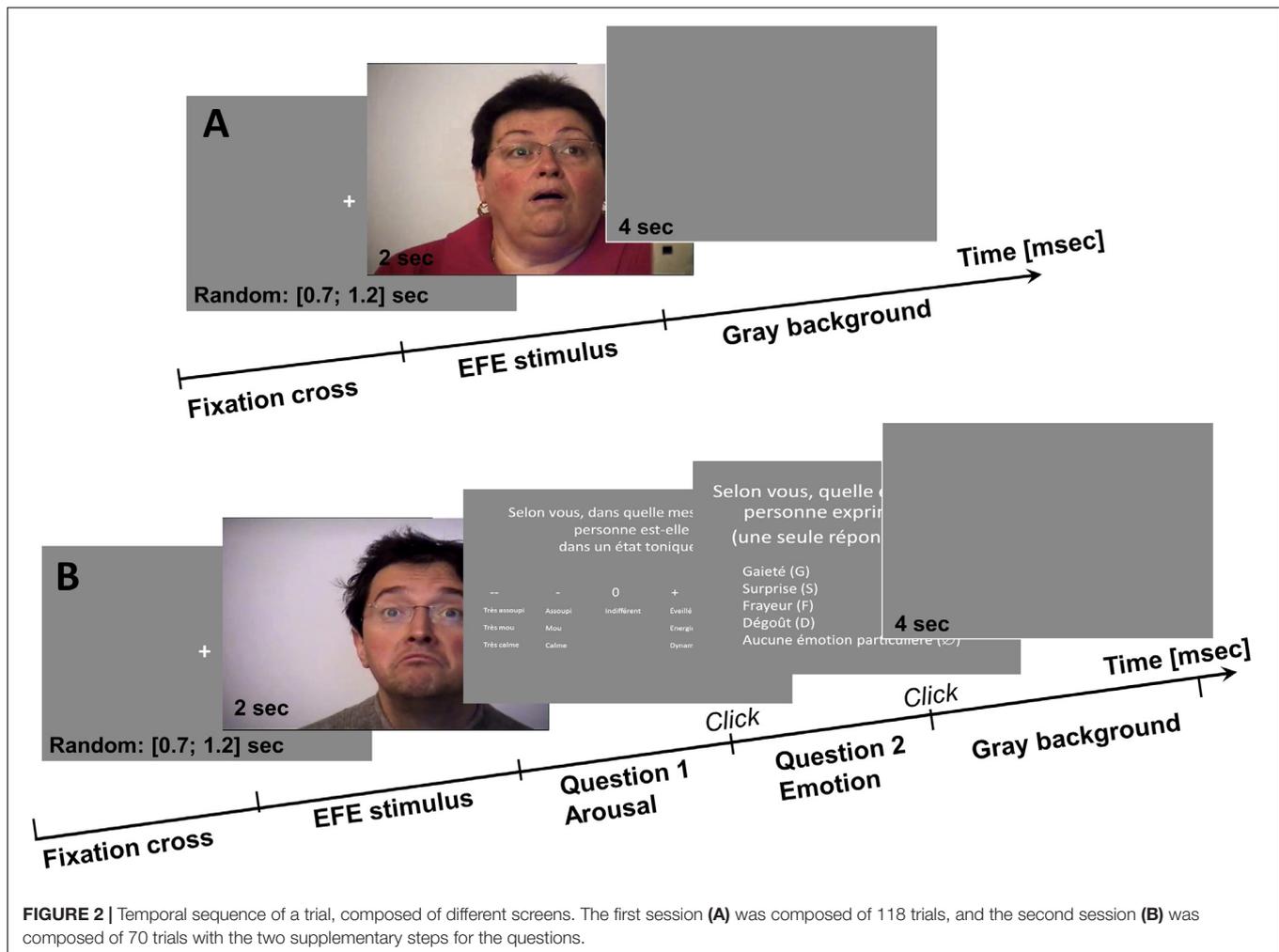
FIGURE 2 | Temporal sequence of a trial, composed of different screens. The first session **(A)** was composed of 118 trials, and the second session **(B)** was composed of 70 trials with the two supplementary steps for the questions.

sampling at 1,000 Hz. For eye-tracking acquisition purposes, the position of the head was stabilized with a chin rest.

Eye gaze and EEG signals were synchronized offline on the basis of triggers sent simultaneously on both signals at each step of the trials, using the SoftEye software (Ionescu et al., 2009). Saccades and fixations were automatically detected by the EyeLink software. The thresholds for saccade detection were a minimum velocity at 30°/s, a minimum acceleration at 8000°/s^2 and a minimum motion at 0.1°/s. In addition, specific triggers were added offline to each eye movement and EEG signals to indicate the beginning of the fixations depending on their localization in the EFE stimuli. Then, in order to select the fixations according to their spatial position, all EFE stimuli had been manually segmented into seven regions of interest (ROI), as illustrated in **Figure 3A**. The seven regions were the forehead, the left and right brows, the corrugator, the left and right eyes, the nose, the mouth, and the chin. An eighth region was added for fixations outside these regions.

Brain Activity

Participants' electroencephalographic (EEG) activity was continuously recorded using an Acticap® (Brain Products, Inc.)

equipped with 64 Ag-AgCl unipolar active electrodes that were positioned according to the extended 10–20 system (Jasper, 1958; Oostenveld and Praamstra, 2001). The reference and ground electrodes used for acquisition were those of the Acticap, i.e., FCz for the reference, and AFz for the ground. The electro-oculographic (EOG) activity was also recorded using two electrodes positioned at the eyes outer canthi, and 2 respectively, above and below the left eye. Participants were free for their eye movements to explore the visual stimulus but they were instructed to limit blinking during the experimental session (see **Figures 3B,C**, for two examples of scanpath). Impedance was kept below 10 kΩ for all electrodes. The signal was amplified using a BrainAmp^{TM} system (Brain Products, Inc.) and sampled at 1,000 Hz with a 0.1 Hz high-pass filter and a 0.1 μV resolution. Data acquisition was performed using Grenoble EEG facility "IRMaGE."

As regards EEG data preprocessing, the raw signal was first band-pass filtered between 1 and 70 Hz and a notch filter was added (50 Hz). The signals were visually inspected for bad channels. The rejected channels were interpolated. The signals were re-referenced offline to the average of all channels. Artifacts related to ocular movements (saccades and blinks)

FIGURE 3 | (A) Example of the masks to define the regions of interest in the EFE stimuli. **(B,C)** Two examples of scanpath for disgust and for happiness. **(D)** Mean percentage for the position of the first fixation depending on ROI and on emotion, based on individual means.
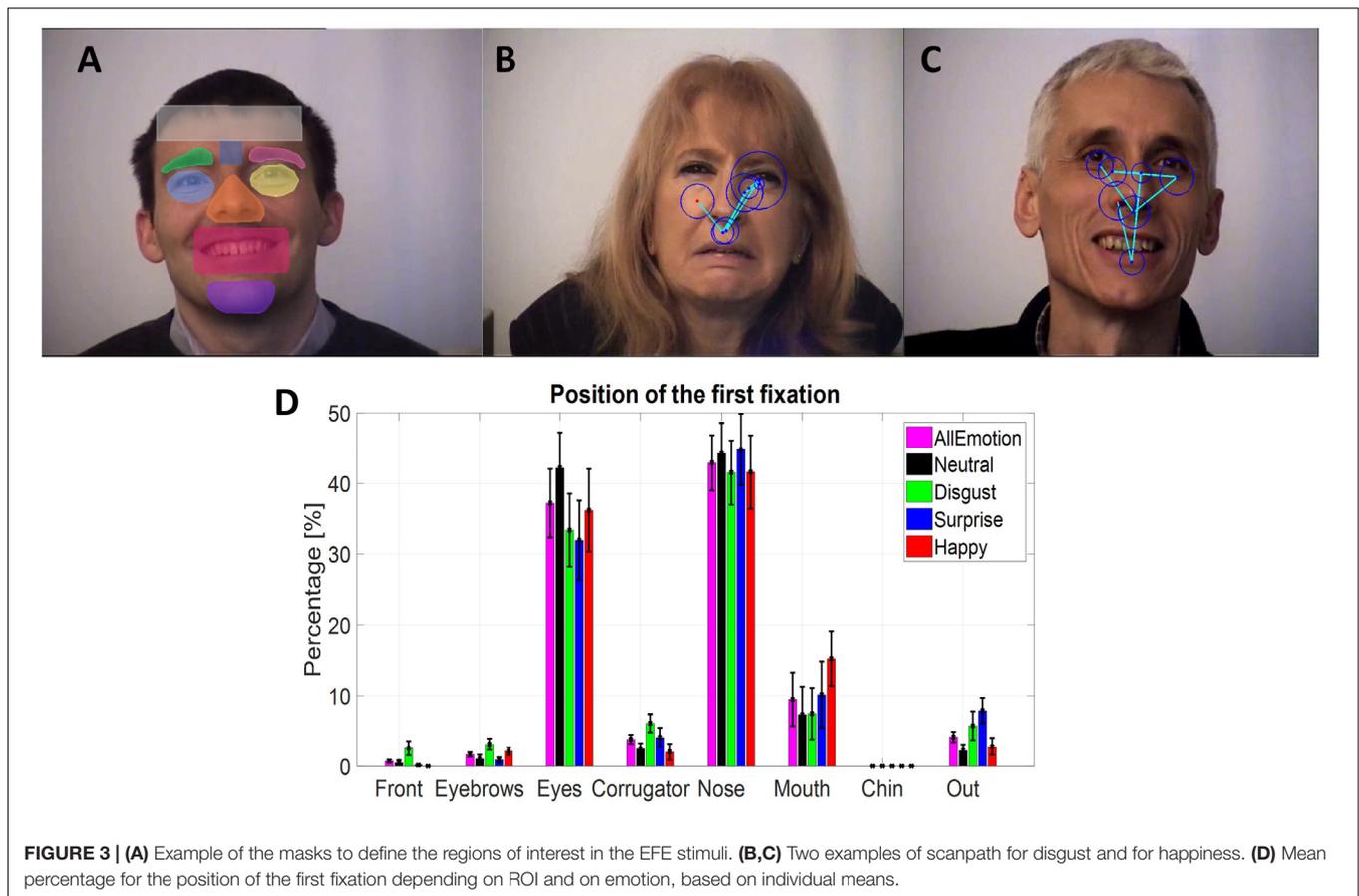
TABLE 1 | Number of epochs per emotion based on individuals measures.

Number of trials	Neutral	Disgust	Surprise	Happiness
mean (SE)	19.3 (1.59)	12.4 (1.03)	12.9 (0.80)	12.2 (0.49)
[min, max]	[8, 32]	[5, 22]	[8, 20]	[8, 16]

were corrected in a semi-automatic fashion using the signal recorded from the EOG electrodes and the SOBI algorithm (Belouchrani et al., 1997). The signal was then segmented into epochs that started 200 ms before and ended 2,000 ms after the image onset. Epochs were rejected when their variance exceeded a restrictive threshold of the mean variance across the epochs plus three standard deviations. Moreover, epochs were also rejected if there were less than two fixations during the 2-s trial. Data from participants without a minimum of five epochs per emotion were excluded. **Table 1** summarizes the number of epochs which were analyzed. EEGlab software (Delorme and Makeig, 2004) was used for all processing steps except the implementation of the GLM for the evoked potential estimation.

Data were then baseline corrected according to the average EEG amplitude over the window from −200 ms to 0 ms before the image onset. Lastly, the signal from seven scalp regions (4/5 electrodes per region) was averaged to create seven virtual electrodes. These regions were evenly distributed across the

scalp ranging from the frontal regions to the parieto-occipital ones, and from left to right, with the median occipital site also. These regions were defined as follows: left frontal (F3, F5, F7, FC5, FC3), right frontal (F4, F6, F8, FC6, FC4), left centro-parietal (C3, C5, T7, CP3, CP5), right centro-parietal (C4, C6, T8, CP4, CP6), left parieto-occipital (P3, P5, P7, PO3, PO7), right parieto-occipital (P4, P6, P8, PO4, PO8), and median occipital (POz, O1, Oz, O2).

Estimation Methods

The two methods (Average and GLM) were applied on the same set of trials (**Table 1**), providing two estimations of the ERP at the stimulus onset, by averaging and by regression, and one estimation of EFRP by regression.

Estimation by Averaging

The estimation of evoked potentials by averaging time-locked EEG signals is the classical method. Let us note the signal $x_i(t)$ time-locked at the image onset during the ith epoch such as:

$$x_i(t) = s(t) + n_i(t)$$

with $s(t)$ the potential evoked at the image onset and $n_i(t)$ the background cortical activity, considered as noise. Assuming that all stimuli elicit the same potential and that the ongoing activity is not synchronized to the fixation onset during the ith epoch,

this potential is estimated by averaging on a given number of epochs as:

$$\widehat{s_{Avg}}(t) = \frac{1}{E}\sum_{i=1}^{E} x_i(t).$$

It is well-known that this estimator is unbiased only if a unique potential is elicited per epoch (Ruchkin, 1965).

In our case, the EFE stimuli were categorized according to each participant's own categorization. The estimation by average was done for each emotion for a given participant. Moreover during the latency of interest (from the image onset up to 600 ms), one or more fixations/saccades occurred (see the Section "Positions on the First Fixations"). Consequently the estimate $\widehat{s_{Avg}}(t)$ is a biased estimation of the evoked potential at the image onset, but it is still an acceptable estimation for the global time-locked activity from the image onset. This global activity includes the activity elicited by the stimulus onset and the activity due to the visual exploration. The statistical results on $\widehat{s_{Avg}}(t)$ are presented in the Section "Event Related Potential at the Image Onset Estimated by Averaging." To separate these two neural activities, supplementary estimations were performed using the GLM, as explained further.

Estimation by Regression With the "General Linear Model"

The evoked potential at the image onset and the potentials elicited at each fixation and saccade rank overlapped one another. To take into account these response overlaps on the observed time-locked neural activity, a more accurate model can be designed such as:

$$x_i(t) = s(t) + fp^{(1)}\left(t - \tau_i^{(1)}\right) + \sum_{l=2}^{L(i)} fp^{(2+)}\left(t - \tau_i^{(l)}\right) +$$

$$\sum_{l'=1}^{L'(i)} sp\left(t - {\tau_i'}^{(l')}\right) + n_i(t)$$

where $s(t)$ is the evoked potential at the image onset, $fp^{(1)}(t)$ is the potential evoked at the first fixation rank, $fp^{(2+)}(t)$ the potential evoked at the second and following ranks, $sp(t)$ the saccadic potential evoked at each saccade rank and $n_i(t)$ the noise of the ongoing activity. In this equation, for a given epoch i, $\tau_i^{(l)}$ is the timestamp of the fixation onset at rank l, and ${\tau_i'}^{(l')}$ is the timestamp at the saccade onset at rank l'. The justification of this model is the following:

— The potential elicited at the first fixation rank is *a priori* different from the one elicited at the following ranks. The rationale for this justification firstly comes from the oculomotor features which can be different at the very first fixation as compared to the followings when the exploration has already begun. Secondly, the categorization of the EFE depends on its recognition which is a fast process with a high contribution of the visual information processed at the

first fixation rank (Batty and Taylor, 2003; Vuillemier and Pourtois, 2007).

— The saccadic activity is taken into account as this activity interacts with the early components of the EFRP at the posterior sites and also at the anterior sites (Nikolaev et al., 2016). Integrating these activities in the linear regression is a good solution (Dandekar et al., 2012). But, contrary to Dandekar et al.'s (2012) study, the saccadic potentials are not here the potentials of interest to analyze, but they are integrated into the model to provide unbiased estimations of the potentials of interest for this study which are mainly $s(t)$ and $fp^{(1)}(t)$, and to a lesser extent $fp^{(2+)}(t)$.

By concatenating all trials, $s(t)$ and $fp^{(1)}(t)$ are estimated by ordinary least square regression to obtain $\widehat{s_{GLM}}(t)$ and $\widehat{fp_{GLM}^{(1)}}(t)$ namely. The statistical results on $\widehat{s_{GLM}}(t)$ are presented in the Section "Event Related Potential at the Image Onset Estimated by Regression" and ones on $\widehat{fp_{GLM}^{(1)}}(t)$ and $\widehat{fp_{GLM}^{(2+)}}(t)$ in the Section "Eye Fixation Related Potentials Estimated by Regression." Mathematical details for the GLM implementation concerning the selected configuration, as well as all configuration parameters for these estimations are given as Supplementary Material in Appendix 1.

Statistical Analysis

For each participant, the averaged and regressed ERPs $[\widehat{s_{Avg}}(t), \widehat{s_{GLM}}(t)]$ were separately computed per emotion condition and virtual electrode. On these evoked potentials, the mean amplitude of four components of interest, namely the P1, the N170, the P2–P3 complex that encompasses both the P2 and the P3 components (or the EPN which is the differential version with the neutral emotion) and the LPP were extracted. Using grand-average inspections, the windows used for the extraction of these amplitude data were adapted from that of Trautmann-Lengsfeld et al. (2013) to fit our data. The latency window for the P1 component was 90–130 ms post-stimulation. The latency window for the N170 component was 140–180 ms post-stimulation. The latency window for the P2–P3 complex was 200–350 ms post-stimulation. The latency window for the LPP component was 400–600 ms post-stimulation. Two components were extracted for the regressed EFRP: the Lambda response and the P2 component within a latency window of 20–100 ms, and 180–300 ms, respectively. In a first step, all these ANOVAs were performed using Statistica, had a 0.05 significance level, used Greenhouse–Geisser adjusted degrees of freedom when sphericity was violated (Significativity of the Mauchly's test of sphericity) and were followed for each significant effect of a given factor by Tukey *post hoc* tests that corrected for multiple comparisons. Regarding the EFRPs, since there is no literature to which these results could be compared, we started by using *t*-tests against zero for the difference in component's amplitude between each EFE and the neutral ones. In a second step, the statistical validity of the results was assessed to determine how the number of both participants and trials interact on each result (Boudewyn et al., 2017). This supplementary verification is undertaken

because the number of participants and the number of trials per participant are in the lower range of the usual values. To do so, 1,000 experiments were simulated for each configuration given by a number of participants (N) and by a number of trials per participant and per emotion. The probability of observing the result is computed on average on all the simulated experiments (1,000) as a function of a given number of participants and a given number of trials per participant. Results are presented as Supplementary Material in Appendix 3.

RESULTS

Behavioral Data

Arousal ratings for each EFE (**Table 2**) were statistically analyzed using a repeated measure ANOVA with emotion as within-participant factor. The main effect on emotion was significant [$F(3,54) = 80.73$, $p < 0.0001$, $\eta_p^2 = 0.82$]. The neutral EFEs (-0.29, $SE = 0.13$) elicited less arousal than disgust (0.73, $SE = 0.11$), surprise (0.70, $SE = 0.07$) and happiness (1.16, $SE = 0.006$) EFEs, which in return elicited more arousal than disgust (0.82, $SE = 0.08$) and surprise (0.75, $SE = 0.07$) EFEs.

The unbiased hit rate was computed (**Table 2**), based on the stimuli emotional categorization provided by the participant, and was statically analyzed using a repeated measure ANOVA with emotion as within-participant factor. The main effect on emotion was significant [$F(3,54) = 67.38$, $p < 0.0001$, $\eta_p^2 = 0.79$]. The unbiased hit rate was the lowest for disgust EFE (0.32, $SE = 0.03$), and was the highest for happiness EFE (0.88, $SE = 0.03$).

Eye Movements' Data

In this section, we first detail the global features of the eye movements' data (the number of fixations and the average fixation duration for a complete trial) and then, more importantly, the specific features for the two first fixations. Results of the repartition of the fixation positions over the ROIs for the first fixation are presented. These results provide an external validation of the experimental data as they reproduce regular results on ocular positions associated to the EFE decoding. The results on the fixation duration, the fixation latency, the incoming saccade amplitude and orientation that are necessary elements for the configuration of the GLM, are detailed as Supplementary Material in Appendix 2.

TABLE 2 | Mean arousal, unbiased hit rate, mean fixations number, and fixation duration (standard error in parentheses) depending on emotion, based on individual means.

	Neutral	Disgust	Surprise	Happiness
Arousal	−0.29 (0.13)	0.82 (0.08)	0.75 (0.07)	1.16 (0.06)
Hu rate	0.53 (0.04)	0.32 (0.03)	0.45 (0.04)	0.88 (0.03)
Fixations number	4.74 (0.21)	4.82 (0.23)	4.74 (0.21)	4.92 (0.20)
Mean fixation duration (ms)	301.91 (10.89)	297.54 (12.64)	293.12 (7.82)	287.31 (9.95)

Global Features

Both the number of fixations and the average fixation duration per trial (synthesized in **Table 2**) were statistically analyzed using two separated repeated measure ANOVAs with the emotion as within-participant factor. The fixations numbers were not different across emotion [$F(3,54) = 1.24$, $p = 0.30$, $\eta_p^2 = 0.02$], nor was the fixation duration [$F(3,54) = 1.49$, $p = 0.23$, $\eta_p^2 = 0.08$].

Positions on the First Fixations

The percentage in each spatial ROI regardless of the emotion (**Table 3**) was analyzed using a repeated measure ANOVA with the fixation rank and the ROI as within-participant factors. Only six ROIs were considered because the forehead and the chin ROIs were not enough fixated (respectively, 0.47% and 0.09%). As expected, a main effect on ROI was observed [$F(5,90) = 37.5$, $p < 0.0001$, $\eta_p^2 = 0.68$]: the eyes (41,94%, $SE = 4.52$%) and the nose (37.96%, $SE = 3.76$%) were the two ROIs the most fixated at the two first ranks. The rank by ROI interaction was significant [$F(5,90) = 15.9$, $p < 0.0001$, $\eta_p^2 = 0.47$] showing that the eyes were more fixated at the second fixation (47.7%, $SE = 4.39$%) than at the first fixation (37.19%, $SE = 4.84$%), and conversely the nose was most fixated at the first fixation (42.91,%, $SE = 3.93$%) than at the second fixation (33.02%, $SE = 3.75$%).

For the six most fixated ROIs, six separated ANOVAs (**Table 4**) were run to analyze specifically the position of the first fixation (**Figure 3D**) according to the emotion (within-participant factor).

For the mouth ROI, a significant larger percentage on this ROI was observed for the happiness emotion (15.24%, $SE = 3.84$%) compared to the disgust (7.49%, $SE = 3.63$%) and neutral (7.35%, $SE = 3.93$%) emotions. A significant difference was observed for the eyebrows ROI, with a larger percentage of first fixation on this ROI for the disgust emotion (3.15%, $SE = 0.81$%) than for the surprise emotion (0.89%, $SE = 0.37$%). For the corrugator ROI, a trend was observed with a larger percentage on this ROI for the disgust emotion (6.12%, $SE = 1.29$%) compared to the happiness emotion (2.03%, $SE = 1.16$%). Finally, a trend was observed for the percentage of the first fixation outside the ROIs, larger for the surprise emotion (7.91%, $SE = 1.81$%) than for the neutral one (2.22%, $SE = 0.90$%).

Brain Activity
Event-Related Potential at the Image Onset Estimated by Averaging

The estimate $\widehat{s_{Avg}}(t)$ was obtained for each participant, each emotion and each virtual electrode (**Figure 4**). Four components were extracted. All statistical results are noticed in **Table 5**. Only significant effects are detailed below.

There was a significant main effect of the virtual electrode for the P1 component, the N170 component, the P2–P3 complex and the LPP (**Table 5**). After *post hoc* decomposition and Tukey corrections, no differences were significant for the N170 component. For each of the three other components, significant differences were observed with higher mean amplitudes at posterior sites than at central sites which were higher in return than at anterior sites.

TABLE 3 | Mean percentages (standard error in parentheses) in each ROI for the first and the second fixation, based on individual means.

Fixation	Eyebrows	Eyes	Corrugator	Nose	Mouth	Out
Rank 1	1.66 (0.33)%	37.2 (4.84)%	3.85 (0.66)%	42.9 (3.93)%	9.51 (3.78)%	4.17 (0.73)%
Rank 2	1.67 (0.47)%	46,7 (4.39)%	3.76 (0.94)%	33.0 (3.75)%	12.5 (2.83)%	1.89 (0.36)%

TABLE 4 | Mean percentages (standard error in parentheses) in each ROI, depending on emotion, for the first and the second fixation, based on individual means.

	Neutral	Disgust	Surprise	Happiness	ANOVA
Eye brows	1.06 (0.57)%	3.15 (0.81)%	0.89 (0.37)%	2.14 (0.57)%	$F(3,54) = 3.22, p = 0.029, \eta_p^2 = 0.15;$ *
			Disgust > Surprise (*)		
Eyes	42.2 (5.04)%	33.4 (5.15)%	32.0 (5.58)%	36.2 (5.84)%	$F(3,54) = 2.59, p = 0.061, \eta_p^2 = 0.13$
Corrugator	2.50 (0.79)%	6.12 (1.29)%	4.12 (1.36)%	2.03 (1.16)%	$F(3,54) = 2.60, p = 0.061, \eta_p^2 = 0.13$
			Disgust > Happiness (trend)		
Nose	44.2 (4.36)%	41.5 (4.56)%	44.8 (5.07)%	41.6 (5.22)%	$F(3,54) = 0.30, p = 0.82, \eta_p^2 = 0.016$
Mouth	7.35 (3.93)%	7.49 (3.63)%	10.2(4.68)%	15.2 (3.84)%	$F(3,54) = 4.52, p = 0.006, \eta_p^2 = 0.20;$ **
			Happiness > Neutral ∼ Disgust (*)		
Out	2.22 (0.90)%	5.77 (2.03)%	7.91 (1.81)%	2.82 (1.22)%	$F(3,54) = 2.84, p = 0.046, \eta_p^2 = 0.14;$ *
			Surprise > Neutral (trend)		

*$p < 0.05$, **$p < 0.01$, bold: significant effect.

For both the P2–P3 complex and the LPP, the interaction between emotion and virtual electrode was significant (**Table 5**). For the P2–P3 complex, the happiness condition lead to a higher mean amplitude ($-2.27 \, \mu V$, $SE = 0.59 \, \mu V$) than for the surprise condition ($-3.67 \, \mu V$, $SE = 0.59 \, \mu V$) at the left frontal site. For the LPP, its mean amplitude was higher in the happiness condition ($-1.56 \, \mu V$, $SE = 0.39 \, \mu V$) than in the surprise condition ($-3.05 \, \mu V$, $SE = 0.62 \, \mu V$) at the left frontal site. Monte Carlo simulations (see Supplementary Material in Appendix 3) were performed on the LPP extracted from the ERP at the stimulus onset estimated by averaging. It confirmed that this difference on the LPP was present at the left frontal site and absent at the right frontal site.

Event-Related Potential at the Image Onset Estimated by Regression

The estimate $\widehat{s_{GLM}}(t)$ was obtained for each participant, each emotion and each virtual electrode (**Figure 5**). Four components were extracted. All statistical results are noticed in **Table 6**.

For all components except N170, there was a significant main effect of virtual electrode (**Table 6**), with a higher mean amplitude of both components at posterior sites than at anterior sites. For the LPP, only a trend difference was observed in the disgust condition with a lower amplitude ($-2.39 \, \mu V$, $SE = 0.79 \, \mu V$) than in the surprise condition ($0.54 \, \mu V$, $SE = 1.69 \, \mu V$) at the right frontal site.

No significant modulation across emotion was observed based on the neural activity estimated on $\widehat{s_{GLM}}(t)$, while a modulation was observed based on the neural activity estimated on $\widehat{s_{Avg}}(t)$. Then, the objective of Monte Carlo simulations realized on the LPP extracted from $\widehat{s_{GLM}}(t)$ (see Supplementary Material in Appendix 3) was to assess the absence of such a modulation (happiness vs. surprise). It confirmed that if this difference between these two EFEs was present on the LPP at the left frontal site on the neural activity $\widehat{s_{Avg}}(t)$, this difference was definitively absent on the LPP at the left and right frontal sites on the neural activity $\widehat{s_{GLM}}(t)$.

Eye Fixation-Related Potentials Estimated by Regression

The difference between the two previous estimations for the evoked potential at the stimulus onset is the inclusion or not of the neural activity linked to fixations. This activity through EFRP estimation is analyzed in this section, and is focused on emotional stimuli compared to neutral, at the occipital sites (VE: Left Parieto-occipital, Right Parieto-occipital and Median Occipital). Two EFRPs were estimated by the GLM: the first at the first fixation onset, namely $\widehat{fp_{GLM}^{(1)}}(t)$ and the next one at the second and following fixation onsets, namely $\widehat{fp_{GLM}^{(2+)}}(t)$ (**Figure 6**). Two components were extracted, the lambda response between 20 and 100 ms, and the P2 component between 180 and 400 ms. The mean amplitude of each component was analyzed using a repeated measure ANOVA with the fixation rank and the emotion as within-participant factors. All statistical results are given in **Table 7**. Only significant effects are detailed below.

Significant differences were observed on the first EFRP and between surprise and neutral. A significant difference was observed on the right parieto-occipital site for the Lambda response, with a higher amplitude for surprise ($4.39 \, \mu V$, $SE = 1.20 \, \mu V$) than for neutral ($1.82 \, \mu V$, $SE = 0.70 \, \mu V$). For the P2 component, there was a significant difference at the right parieto-occipital site with a higher mean amplitude for surprise ($1.40 \, \mu V$, $SE = 0.44 \, \mu V$) than for neutral ($-0.39 \, \mu V$, $SE = 0.40 \, \mu V$). A significant difference was also found for the P2 component between surprise ($0.93 \, \mu V$, $SE = 0.63 \, \mu V$) and neutral ($-0.89 \, \mu V$, $SE = 0.64 \, \mu V$) at the median occipital site. It is known that the local physical features of a visual stimulus influence the amplitude of the Lambda response (Gaarder et al., 1964). The statistical results showed that there was no difference across emotion for
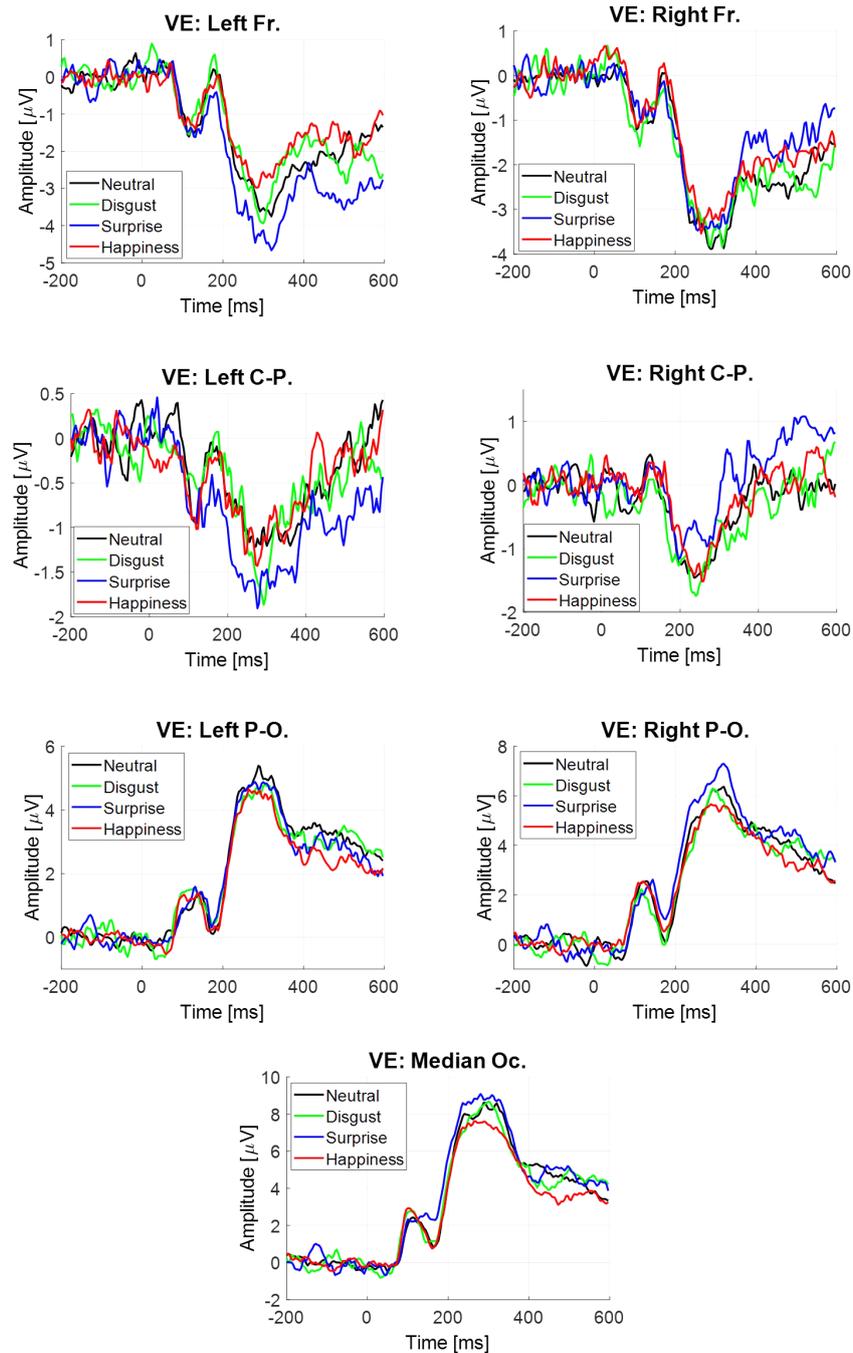
FIGURE 4 | Event-related potentials elicited at the stimulus onset estimated by averaging, depending on emotion and virtual electrode.

the local standard deviation of the luminance on the region gazed at the first fixation, nor any difference across emotion of the local luminance through the first saccade. These results are presented as Supplementary Material in Appendix 2. Moreover, the Monte Carlo simulation (see Supplementary Material in Appendix 3) confirmed that differences were present at the right parieto-occipital site and at the median occipital site for the first EFRP, and were absent for the second and subsequent EFRP.

DISCUSSION

The goal of the present study was to analyze the temporal dynamics of spontaneous and static emotional faces decoding. More precisely, the early visual exploration's temporal dynamics of natural EFEs was explored. Eye movements and EEG activities were jointly recorded and analyzed. Recent studies using such joint recordings have shown the interest of a regression approach

TABLE 5 | Statistical results of the ANOVAs performed on the evoked potential at the image onset, estimated by averaging.

Estimation « Avg »	Evoked potential at the image onset		
	Virtual electrode (VE)	Emotion (EMO)	VE × EMO
P1 [90–130] ms	$F_{(6,108)} = 14.41; p < 0.001; \eta_p^2 = 0.44; ***$ Fr. > CP. > PO, Oc.	$F_{(3,54)} = 1.22; p = 0.31; \eta_p^2 = 0.06$	$F_{(18,324)} = 0.88; p = 0.64; \eta_p^2 = 0.03$
N170 [140–180] ms	$F_{(6,108)} = 3.02; p = 0.009; \eta_p^2 = 0.14; **$ Fr. > C.P. > P.O., Oc.	$F_{(3,54)} = 2.36; p = 0.08; \eta_p^2 = 0.11$	$F_{(18,324)} = 1.78; p = 0.026; \eta_p^2 = 0.09; *$
P2–P3 [200–350] ms	$F_{(6,108)} = 54.86; p < 0.001; \eta_p^2 = 0.75; ***$ Fr. > C.P. > P.O., Oc.	$F_{(3,54)} = 2.08; p = 0.11; \eta_p^2 = 0.10$	$F_{(18,324)} = 2.95; p < 0.001; \eta_p^2 = 0.14; ***$ Left Fr. : H > S
LPP [400–600] ms	$F_{(6,108)} = 38.85; p < 0.001; \eta_p^2 = 0.68; ***$ Fr. > C.P. > P.O., Oc.	$F_{(3,54)} = 0.77; p = 0.51; \eta_p^2 = 0.04$	$F_{(18,324)} = 2,15; p = 0.005; \eta_p^2 = 0.11; **$ Left Fr. : H > S

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, bold: significant effect.

based on the GLM to estimate ERPs as well as eye fixation/saccade related potentials (Dandekar et al., 2012; Kristensen et al., 2017b). The overlapping evoked potentials can be separately estimated by deconvolution when using this method. Consequently, these methodological tools take the temporal dimension into account. This is particularly interesting for the study of the dynamics of EFE processing. For instance, Noordewier et al. (2016) have stressed the importance of taking the temporal dimension into account to understand the nature of surprise. Thus, the contribution of this study is twofold. First, it addresses naturally occurring human affective behavior. Second, it offers a solution to the methodological issue regarding the estimation of overlapping evoked potentials.

Based on an ecological approach, this study used natural EFEs as static stimuli and a free exploration task. Natural EFEs are spontaneous expressions encountered in everyday life and free exploration is an ecological paradigm requiring the consideration of time for analysis. Behavioral results on eye movements are consistent with what is usually observed when studying emotional facial features processing. Since the 1920s, there is evidence that specific facial features, such as the eyes and mouth, are relevant for the decoding of EFEs (Buchan et al., 2007; Schurgin et al., 2014). Present results showed that the eyes and the nose were the two most gazed ROIs at the first fixation, irrespective of the emotion displayed. The former ROI is in accordance with usual results as the eyes are very important for social interaction to decode the emotional state of the other person. As regard to the nose, results are interpreted as an exposition bias. The fixation cross allowing the gaze stabilization before the EFE presentation was at the center of the image, thus close to the nose position in the face. The third most gazed ROI was the mouth which is also an important region for emotion decoding. This region was gazed more for the happiness emotion than for the disgust and neutral emotions at the first fixation. This is also in line with previous research. When looking at happy facial expressions, participants usually fixate the mouth region for a longer time (e.g., Eisenbarth and Alpers, 2011). Eyebrows are likely to be diagnostic features as well. Observers gazed significantly more at this area when looking at EFEs of disgust as compared to EFEs of surprise. They also tended to gaze more at the corrugator area for EFEs of disgust than EFEs of happiness for instance. On the whole, areas of the face attracting attention more

than other areas were quite in line with what is usually observed (Beaudry et al., 2014; Vaidya et al., 2014). Finally, when facing EFEs of surprise, observers tended to collect information out of the face as if they were trying to find in the environment what could have caused such an emotion.

The other key contribution of this study concerns the methodological issue to estimate overlapping evoked potentials. This is a main concern in synchronized EEG and eye movement analysis (Dimigen et al., 2011), and more specifically here as the time was an important issue for the free exploration task. It has been well-established that the estimation of evoked potentials by averaging time-locked EEG signals is biased in the case of overlapping responses. Woldorff (1993) proposed an iterative procedure in the context of ERP experiments where the EEG signal is time-locked on external events. It was called the ADJAR algorithm, and was designed to estimate overlap responses from immediately adjacent events, to converge toward the evoked potential of interest. Moreover, regression techniques, especially the GLM (Kiebel and Holmes, 2003), have proved their efficiency in the estimation of evoked overlapping potentials (Dale, 1999; Dandekar et al., 2012; Burns et al., 2013; Bardy et al., 2014; Kristensen et al., 2017a). Besides, the ADJAR algorithm appears to be poorly suited to EFRP estimation (Kristensen et al., 2017a). In this respect, a regression-based estimation of evoked potentials was done (Smith and Kutas, 2015a,b). Usually, the linear model is designed either with the saccade onset timestamps as regressors (Dandekar et al., 2012), either with the fixation onset timestamps as regressors (Kristensen et al., 2017b). In our study, both types of regressors were integrated into the same model. A third type was added, namely the timestamp of the stimuli onset. The rationale for such a model, with both the timestamps of saccade and fixation onsets, was the observation of different distributions for incoming saccade amplitudes and orientations depending on emotions. It is well known that the saccadic activities just before the fixation onsets modulate the early component (specially the Lambda response) of each EFRP. Thus, to provide an unbiased estimator of the EFRPs from these confounding factors, the timestamps of saccade onsets were added to the model. Moreover, the timestamps of the fixation onsets were split into two different classes, those for the first fixation, and those for the following fixations. This way, the EFRP for the first fixation was discriminated from the EFRP for the second and subsequent
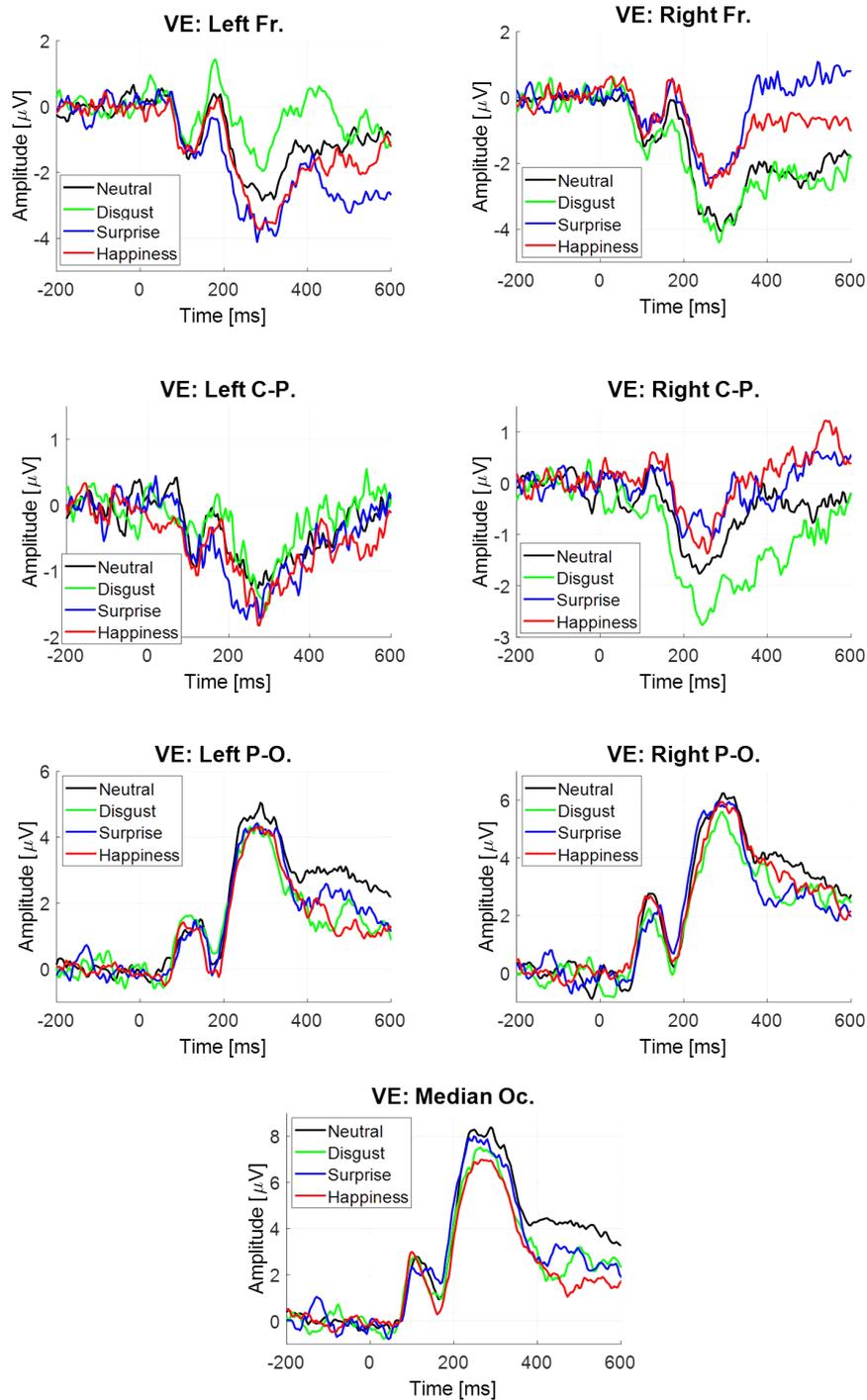
FIGURE 5 | Event-related potentials elicited at the stimulus onset estimated by regression, depending on emotion and virtual electrode

fixations. The timestamps of the stimulus onset were finally also added because the main objective of this study was to distinguish, from the whole neural activity, the one specifically elicited by the stimulus during a given latency window. Altogether, this study shows how the GLM can be adapted to a specific issue and how its configuration plays a central role in the methodological approach.

We also focused on the neural activity during the latency of the P2–P3 complex and of the LPP. It was analyzed with regards to the estimations comparison: Average vs. GLM. The common estimation by averaging takes into account all neural activities time-locked at the stimulus onset. It is commonly accepted that the potential evoked at a visual stimulus presentation lasts about

TABLE 6 | Statistical results of the ANOVAs performed on the evoked potential at the image onset, estimated by regression.

Estimation « GLM »	Evoked potential at the image onset		
	Virtual electrode (VE)	Emotion (EMO)	VE × EMO
P1 [90–130] ms	$F(6,108) = 13.28$; $p < 0.001$; $\eta_p^2 = 0.42$; *** Fr. > C.P. > P.O., Oc.	$F(3,54) = 1.17$; $p = 0.33$; $\eta_p^2 = 0.06$	$F(18,324) = 0.85$; $p = 0.63$; $\eta_p^2 = 0.04$
N170 [140–180] ms	$F(6,108) = 1.98$; $p = 0.07$; $\eta_p^2 = 0.10$	$F(3,54) = 0.64$; $p = 0.59$; $\eta_p^2 = 0.03$	$F(18,324) = 1.07$; $p = 0.39$; $\eta_p^2 = 0.06$
P2–P3 [200–350] ms	$F(6,108) = 27.30$; $p < 0.001$; $\eta_p^2 = 0.60$; *** Fr. > C.P. > P.O., Oc.	$F(3,54) = 1.19$; $p = 0.32$; $\eta_p^2 = 0.06$	$F(18,324) = 0.99$; $p = 0.47$; $\eta_p^2 = 0.05$
LPP [400–600] ms	$F(6,108) = 8.83$; $p < 0.001$; $\eta_p^2 = 0.33$; *** Fr. > C.P. > P.O., Oc.	$F(3,54) = 1.06$; $p = 0.37$; $\eta_p^2 = 0.06$	$F(18,324) = 1.09$; $p = 0.057$; $\eta_p^2 = 0.06$ Right Fr. : D < S (trend)

***$p < 0.001$; bold, significant effect; Fr. for left and right frontal sites; C.P. for left and right centro-parietal sites; P.O., Oc. for left and right parieto-occipital sites and median occipital site; Left Fr. for left frontal site; H for Happiness, S for Surprise.
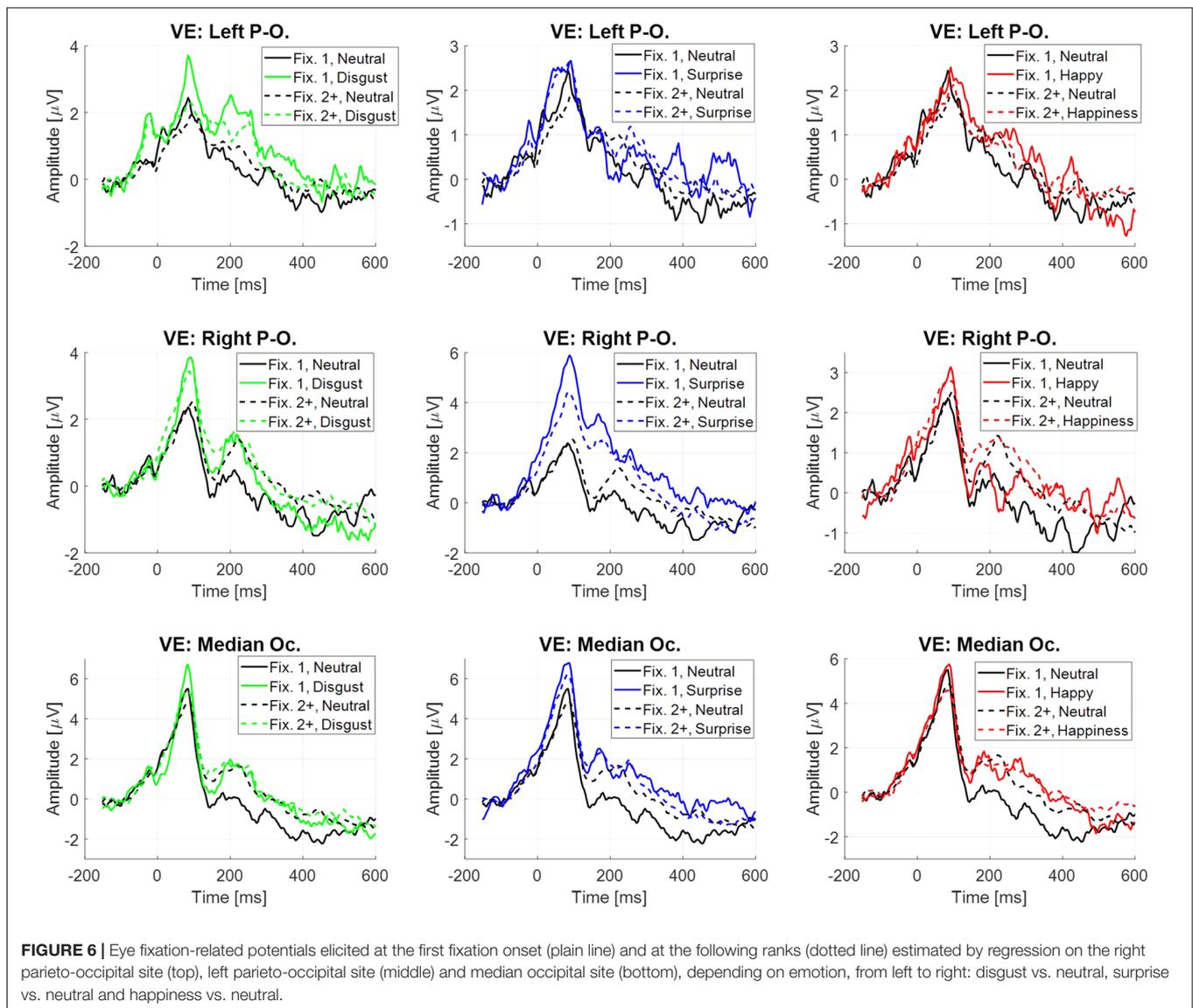
FIGURE 6 | Eye fixation-related potentials elicited at the first fixation onset (plain line) and at the following ranks (dotted line) estimated by regression on the right parieto-occipital site (top), left parieto-occipital site (middle) and median occipital site (bottom), depending on emotion, from left to right: disgust vs. neutral, surprise vs. neutral and happiness vs. neutral.

700 ms, corresponding to the time needed for the stimulus-evoked activity to fade (Dimigen et al., 2011; Nikolaev et al., 2016). This potential had the largest contribution in the neural activity during the latency of the P2–P3 complex ([200; 350] ms). Moreover, its contribution was larger for the P2–P3 latency than for the LPP latency ([400; 600] ms). This evoked potential

TABLE 7 | Statistical results of the Student's tests performed on the EFRPs, estimated by regression.

Estimation « GLM »	Evoked potential at the first fixation onset		
	Disgust – Neutral	**Surprise – Neutral**	**Happiness – Neutral**
Left parieto-occipital site			
Lambda [20–110] ms	$t(18) = 0.49$; $p = 0.63$; $\eta_p^2 = 0.01$	$t(18) = 1.25$; $p = 0.23$; $\eta_p^2 = 0.08$	$t(18) = 0.02$; $p = 0.98$; $\eta_p^2 = 0$
P2 [180–400] ms	$t(18) = 1.28$; $p = 0.22$; $\eta_p^2 = 0.09$	$t(18) = 0.58$; $p = 0.57$; $\eta_p^2 = 0.02$	$t(18) = 0,66$; $p = 0.52$; $\eta_p^2 = 0.02$
Right parieto-occipital site			
Lambda [20–110] ms	$t(18) = 0.60$; $p = 0.55$; $\eta_p^2 = 0.02$	**$t(18) = 2.67$; $p = 0.016$; $\eta_p^2 = 0.30$; ***	$t(18) = 0.32$; $p = 0.76$; $\eta_p^2 = 0.006$
P2 [180–400] ms	$t(18) = 0.73$; $p = 0.48$; $\eta_p^2 = 0.03$	**$t(18) = 3.07$; $p = 0.007$; $\eta_p^2 = 0.36$; ****	$t(18) = 0.62$; $p = 0.54$; $\eta_p^2 = 0.02$
Median occipital site			
Lambda [20–110] ms	$t(18) = 0.17$; $p = 0.87$; $\eta_p^2 = 0.002$	$t(18) = 1.53$; $p = 0.14$; $\eta_p^2 = 0.12$	$t(18) = 0.26$; $p = 0.79$; $\eta_p^2 = 0.004$
P2 [180–400] ms	$t(18) = 1.08$; $p = 0.29$; $\eta_p^2 = 0.06$	**$t(18) = 2.25$; $p = 0.037$; $\eta_p^2 = 0.23$; ***	$t(18) = 1.32$; $p = 0.20$; $\eta_p^2 = 0.09$
Estimation « GLM »	Evoked potential at the second and following fixation onsets		
Left parieto-occipital site			
Lambda [20–110] ms	$t(18) = 0.76$; $p = 0.46$; $\eta_p^2 = 0.03$	$t(18) = 1.22$; $p = 0.24$; $\eta_p^2 = 0.08$	$t(18) = 0.26$; $p = 0.80$; $\eta_p^2 = 0.004$
P2 [180–400] ms	$t(18) = 0.74$; $p = 0.47$; $\eta_p^2 = 0.03$	$t(18) = 0.57$; $p = 0.58$; $\eta_p^2 = 0.02$	$t(18) = -0.01$; $p = 0.99$; $\eta_p^2 = 0$
Right parieto-occipital site			
Lambda [20–110] ms	$t(18) = 0.91$; $p = 0.38$; $\eta_p^2 = 0.05$	$t(18) = 1.72$; $p = 0.10$; $\eta_p^2 = 0.15$	$t(18) = 0.48$; $p = 0.64$; $\eta_p^2 = 0.01$
P2 [180–400] ms	$t(18) = 0.18$; $p = 0.86$; $\eta_p^2 = 0.002$	$t(18) = 1.24$; $p = 0.23$; $\eta_p^2 = 0.08$	$t(18) = 0.71$; $p = 0.49$; $\eta_p^2 = 0.03$
Median occipital site			
Lambda [20–110] ms	$t(18) = 0.36$; $p = 0.72$; $\eta_p^2 = 0.01$	$t(18) = 1.03$; $p = 0.32$; $\eta_p^2 = 0.06$	$t(18) = 0.03$; $p = 0.98$; $\eta_p^2 = 0$
P2 [180–400] ms	$t(18) = 0.48$; $p = 0.64$; $\eta_p^2 = 0.01$	$t(18) = 0.79$; $p = 0.44$; $\eta_p^2 = 0.03$	$t(18) = 0.87$; $p = 0.40$; $\eta_p^2 = 0.04$

*$p < 0.05$, **$p < 0.01$; bold, significant effect; Fr. for left and right frontal sites; C.P. for left and right centro-parietal sites; P.O., Oc. for left and right parieto-occipital sites and median occipital site.

was also estimated by the GLM. The neural activity provided by the free exploration of the stimulus explained the difference between these two estimates. Indeed, the average fixation latency was about 250 ms for the first one and about 500 ms for the second one. We will first discuss these differences on the potential elicited at the stimulus onset (ERPs), before focusing on the evoked potentials at the first fixations (EFRPs).

Regarding the cerebral responses to the natural EFEs, two different activation patterns were observed for the ERPs computed by the averaging on the one hand and by the GLM on the other hand. For the former (Average), the estimated evoked potential includes the potential at the stimulus onset and the activation provided by the visual exploration of the ocular fixations. For the later (Regression), the estimated evoked potential takes only into account the potential elicited at the stimulus onset. As expected, the amplitude of both the P2–P3 complex and the LPP was higher at posterior sites than at anterior sites for both methods, in accordance with the classical topographical distribution of these components. However, for the averaging method there was a low activation pattern (negative) at the left frontal site with a higher amplitude for both components for the happiness than the surprise condition. Rather for the GLM method, no significant modulation across emotions was observed (only a trend at the right frontal site with the amplitude of the LPP, lower for the disgust than the surprise condition which will be discussed below). The discrepancy between the two methods' results is easily explained. Indeed, for the averaging method, the activation amplitude included both the potential at the stimulus onset and the activation provided by the visual exploration on the ocular fixations. And for the GLM method, the

estimated activation included only the potential at the stimulus onset. Therefore, the frontal left negative pattern observed using the averaging method might in fact only rely upon the activations linked to the subsequent fixations and not on the activation from the stimulus onset.

Concerning the hemispheric prevalence, for the time window of the P2–P3 complex and of the LPP, a higher amplitude of the neural activity was observed for the happiness emotion compared to the surprise emotion when the neural activity was estimated by averaging on time-locked signals at the stimulus presentation. The prevalence of the left electrode site is in accordance with the valence hypothesis, with the involvement of the subsequent fixations for discriminating the happiness emotion as compared to others. The left hemisphere would be preferentially dedicated to the analysis of positive emotions such as happiness (Reuter-Lorenz and Davidson, 1981; Adolphs et al., 2001). When using the GLM to analyze the "common" ERPs (i.e., without the involvement of the subsequent fixations), the only trend differences ($p = 0.057$) between EFEs were found at the right frontal site: the cerebral response to the disgust emotion was enhanced compared to the other ones. This is also in line with the valence hypothesis which posits a right hemispheric specialization for negative affects, such as disgust (Reuter-Lorenz et al., 1983). Yet with common ERP, no difference was found on the left frontal site, but only a trend difference on the right frontal site. That means that the perception of EFEs at the stimulus onset might possibly be firstly mostly undertaken by the right hemisphere (Indersmitten and Gur, 2003; Davidson et al., 2004; Tamietto et al., 2006; Torro Alves et al., 2008). And then, the impact of the subsequent fixations that reveal

the involvement of the left hemisphere might therefore reveal the bilateral gain advocated by Tamietto et al. (2006), and in a more general manner, the predominance of the right hemisphere at the stimulus onset and, afterward, the implication of the left hemisphere for the subsequent fixations. It would illustrate the "complex and distributed emotion processing system" detailed by Killgore and Yurgelun-Todd (2007). Hence, it seems that the recruitment of the left hemisphere needs to be primed by a first analysis performed by the right hemisphere. The communication that would take place to ensure such a bilateral recruitment as soon as the first fixation occurs, as well as any causal link, still need to be further explored using spectral and connectivity analyses. Yet, as reported by Tamietto et al. (2007), neuroimaging studies have already shown that the structures involved in EFE processing are various homologous regions of both hemispheres, such as the early sensory cortices, the middle prefrontal cortex and subcortical areas like the amygdala. They also detail that interhemispheric communication might occur at the early stages through connections at the level of the limbic system, while later processing steps allow for an interhemispheric communication through the corpus callosum.

As to the time course of EFE processing, when computing the ERPs by averaging and as expected, we found emotion-dependent modulations of the amplitude of the P2–P3 complex as well as the LPP component. Yet, no difference was found between EFEs for the N170, which might be in favor of the part of the literature that views the first stage as a raw structural processing one (Eimer et al., 2003), which might also be linked to stimuli of low arousal (Almeida et al., 2016). Considering the ERPs computed using the regression method, no significant impact of emotion on the brain response elicited exclusively by the presentation of the stimuli was found. This difference with the literature might be explained by the stimuli and paradigm we used. Indeed, Neath-Tavares and Itier (2016), like most authors interested in this research topic, use prototypical stimuli (whether from POFA or from the MacBrain Face Stimulus Set, Batty and Taylor, 2003). This might explain at least in part why we do not have the same impact of valence on EFEs decoding as revealed by ERPs. In fact, with prototypical stimuli, the displayed emotions are overstated and amplified, whereas in the present study, EFE are natural and spontaneous, thus weaken (Tcherkassof et al., 2013; cf. also Wagner et al., 1986; Valstar et al., 2006). For prototypical stimuli, the actors exaggerate the EFE. For instance, some past studies showed that posed smiles are larger in amplitude and are longer in duration than spontaneous smiles (Ekman and Friesen, 1982; Cohn and Schmidt, 2004; Schmidt et al., 2006). Valstar et al. (2006) also showed that characteristics of brow actions (as such as intensity, speed and trajectory) are different between spontaneous and posed EFE. In our case, the filmed persons expressed spontaneously and naturally the EFE. Consequently, we used less "intensified" or "aroused" EFE than other studies based on prototypical EFE (Tcherkassof et al., 2013). Another explanation is that, when the participants freely explore a stimulus, the brain responses to the presentation of the stimulus can be polluted by the subsequent responses to saccades that can occur after only 200 ms post-stimulation. In our case, in addition to using natural stimuli, we analyzed separately the

brain responses elicited by the stimulus presentation only and the subsequent fixations. Hence, since with natural stimuli and our unconstrained paradigm we found no significant modulation of components' amplitude when using the ERPs computed using the regression method, it might be that the arousal of the used EFEs was too low.

Finally, with respect to EFRPs, the early potential called Lambda response was impacted by the EFE presentation: the amplitude of the Lambda response was significantly higher for the surprise EFE than for the neutral EFE over the right parieto-occipital site. The Lambda response reflects the visual change in the image retina due to the saccade (Yagi, 1979). This response is modulated by low-level visual features as luminance and contrast across the saccade but also by the by the amplitude and orientation of the saccade (Gaarder et al., 1964; Hopfinger and Ries, 2005; Ossandón et al., 2010). High level factors such as task demand and information processing load also modulate the lambda amplitude (Yagi, 1981; Ries et al., 2016). Since low-level factors have entirely been taken into account in this experiment (i.e., global luminance equalization for the stimuli, local luminance verification at the first fixation and saccadic response estimated by the GLM), high-level factors might indeed explain this impact on the Lambda response.

Furthermore, a difference on the P2 component evaluated on the first EFRP was observed between surprise and neutral emotions over the right parieto-occipital and the median occipital site. The visual P2 component is known to be involved in many different cognitive tasks (Key et al., 2005), such as visual feature detection (Luck and Hillyard, 1994). It is modulated by numerous factors like attention allocation, target repetition, task difficulty, but also by the emotional content of faces (Stekelenburg and de Gelder, 2004) and an interaction between valence and arousal was found on EFRP at the visual exploration of emotional scenes (Simola et al., 2015). In our study, the fact that this effect was observed only for the surprise emotion is interesting. The valence of the surprise may be positive or negative depending on the context (Noordewier and Breugelmans, 2013). This effect may be linked to the areas the participant gazed at the first fixation when displayed surprise EFE: a higher number of fixations tended to land out of the selected face ROIs (forehead, eye brows, corrugator, eyes, nose, mouth, and chin) for surprise EFE as compared to neutral EFE. It is as if participants needed to extract information out of the faces to decode the displayed EFE in order to find cues what could have caused such an emotion. This interpretation has to be studied deeper with dedicated experiments. However, this modulation of the first EFRP with the surprise emotion compared to the neutral emotion contributes to the activation pattern of the LPP on the evoked potential at the stimulus onset estimating by averaging, as mentioned above. Lastly, the fact that such a difference only occurs between the surprise and the neutral conditions for the EFRPs cannot rule out completely an impact of arousal for this particular physiological marker. In line with Simola et al. (2015), such an interaction between valence and arousal for fixation-related potentials would be particularly interesting to study in the EFE processing context.

The present investigation is a promising initial work for the study of emotional decoding's time course. More participants

and more trials need to be run to strengthen this exploratory work. Yet, it appears that the visual exploration of emotional faces is a critical ingredient of EFE processing. It is especially the case when stimuli are not prototypical displays, as in ordinary life. For an accurate comprehension of the displayed emotion, observers need to look through the face, and even outside the face. That is why research on facial behavior urgently requires a dynamic approach (Fernández-Dols, 2013). Moreover, the dynamic propriety of EFE is a key feature of facial behavior since it consists of facial features dynamically shifting. The method presented here is an auspicious tool to treat the decoding of this dynamic information. Such work is currently undertaken by the authors.

AUTHOR CONTRIBUTIONS

AT was responsible for the emotion part of this research. AG-D was responsible for the engineering part of this research. LV was responsible for the experimental procedure. RR, AG-D, and EK collected the data. AG-D, RR, EK, and BR are in charge of the data analysis.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg. 2018.01190/full#supplementary-material

REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21–62. doi: 10.1177/1534582302001001003

Adolphs, R., Jansari, A., and Tranel, D. (2001). Hemispheric perception of emotional valence from facial expressions. *Neuropsychology* 15, 516–524. doi: 10.1037/0894-4105.15.4.516

Ahern, G. L., and Scharwtz, G. E. (1985). Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. *Neuropsychologia* 23, 745–755. doi: 10.1016/0028-3932(85)90081-8

Almeida, P. R., Ferreira-Santos, F., Chaves, P. L., Paiva, T. O., Barbosa, F., and Marquez-Teixera, J. (2016). Perceived arousal of facial expressions of emotion modulates the N170, regardless of emotional category: time domain and time–frequency dynamics. *Int. J. Psychophysiol.* 99, 48–56. doi: 10.1016/j.ijpsycho.2015.11.017

Alves, N. T., Fukusima, S. S., and Aznar-Casanova, J. A. (2008). Models of brain asymmetry in emotional processing. *Psychol. Neurosci.* 1, 63–66.

Balconi, M., and Pozzoli, U. (2003). Face-selective processing and the effect of pleasant and unpleasant emotional expressions on ERP correlates. *Int. J. Psychophysiol.* 49, 67–74. doi: 10.1016/S0167-8760(03)00081-3

Bardy, F., Dillon, H., and Dun, B. V. (2014). Least-squares deconvolution of evoked potentials and sequence optimization for multiple stimuli under low-jitter conditions. *Clin. Neurophysiol.* 125, 727–737. doi: 10.1016/j.clinph.2013.09.030

Batty, M., and Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Brain Res. Cogn. Brain Res.* 17, 613–620. doi: 10.1016/S0926-6410(03)00174-5

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., and Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cogn. Emot.* 28, 416–432. doi: 10.1080/02699931.2013.833500

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* 45, 434–444. doi: 10.1109/78.554307

Blau, V. C., Maurer, U., Tottenham, N., and McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behav. Brain Funct.* 23, 3–7.

Boudewyn, M. A., Luck, S. J., Farrens, J. L., and Kappenman, E. S. (2017). How many trials does it take to get a significant ERP effect? *Psychophysiology* 55:e13049. doi: 10.1111/psyp.13049

Buchan, J. N., Paré, M., and Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Soc. Neurosci.* 2, 1–13. doi: 10.1080/17470910601043644

Burns, M. D., Bigdely-Shamlo, N., Smith, N. J., Kreutz- Delgado, K., & Makeig, S. (2013). "Comparison of averaging and regression techniques for estimating event related potentials," in *Proceedings of the IEEE Engineering in Biology and Medicine Conference*, Osaka, 1680–1683. doi: 10.1109/EMBC.2013.6609841

Calvo, R. A., and D'Mello, S. K. (2011). *New Perspectives on Affect and Learning Technologies.* New York, NY: Springer. doi: 10.1007/978-1-4419-9625-1

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., and Guerit, J. M. (2002). Categorical perception of happiness and fear facial expressions: an ERP study. *J. Cogn. Neurosci.* 14, 210–227. doi: 10.1162/089892902317236858

Cohn, J. F., and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multiresolut. Inf. Process.* 2, 121–132. doi: 10.1142/S021969130400041X

Congedo, M., Korczowski, L., Delorme, A., and Lopes da Silva, F. (2016). Spatio-temporal common pattern: a companion method for ERP analysis in the time domain. *J. Neurosci. Methods* 267, 74–88. doi: 10.1016/j.jneumeth.2016.04.008

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114. doi: 10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W

Dandekar, S., Privitera, C., Carney, T., and Klein, S. A. (2012). Neural saccadic response estimation during natural viewing. *J. Neurophysiol.* 107, 1776–1790. doi: 10.1152/jn.00237.2011

Davidson, R. J., Shackman, A. J., and Maxwell, J. S. (2004). Asymmetries in face and brain related to emotion. *Trends Cogn. Sci.* 8, 389–391. doi: 10.1016/j.tics.2004.07.006

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., and Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: analyses and review. *J. Exp. Psychol. Gen.* 140, 552–572. doi: 10.1037/a0023885

Eisenbarth, E., and Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 860–860. doi: 10.1037/a0022758

Eimer, M., Holmes, A., and McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. *Cogn. Affect. Behav. Neurosci.* 3, 97–110. doi: 10.3758/CABN.3.2.97

Ekman, P., and Friesen, W. V. (1976). *Pictures of Facial Affect.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. F. (1982). Felt, false and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/BF00987191

Fernández-Dols, J. M. (2013). Advances in the study of facial expression: an introduction to the special section. *Emot. Rev.* 5, 3–7. doi: 10.1177/1754073912457209

Gaarder, K., Krauskopf, J., Graf, V., Kropfl, W., and Armington, J. C. (1964). Averaged brain activity following saccadic eye movement. *Science* 146, 1481–1483. doi: 10.1126/science.146.3650.1481

Graham, R., and Cabeza, R. (2001). Event-related potentials of recognizing happy and neutral faces. *Neuroreport* 12, 245–248. doi: 10.1097/00001756-200102120-00013

Hess, U., and Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *Eur. J. Soc. Psychol.* 20, 369–385. doi: 10.1002/ejsp.2420200502

Hinojosa, J. A., Mercado, F., and Carretié, L. (2015). N170 sensitivity to facial expression: a meta-analysis. *Neurosci. Biobehav. Rev.* 55, 498–509. doi: 10.1016/j.neubiorev.2015.06.002

Hopfinger, J. B., and Ries, A. J. (2005). Automatic versus contingent mechanisms of sensory-driven neural biasing and reflexive attention. *J. Cogn. Neurosci.* 17, 1341–1352. doi: 10.1162/0898929055002445

Indersmitten, T., and Gur, R. C. (2003). Emotion processing in chimeric faces: hemispheric asymmetries in expression and recognition of emotions. *J. Neurosci.* 23, 3820–3825. doi: 10.1523/JNEUROSCI.23-09-03820.2003

Ionescu, G., Guyader, N., and Guérin-Dugué, A. (2009). SoftEye software. *Registration number:IDDN.FR.001.200017.000.S.P.2010.003.31235*

Itier, R. J., and Neath-Tavares, K. N. (2017). Effects of task demands on the early neural processing of fearful and happy facial expressions. *Brain Res.* 1663, 38–50. doi: 10.1016/j.brainres.2017.03.013

Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.

Kanade, T., Cohn, J. F., and Tian, Y. (2000). "Comprehensive database for facial expression analysis," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble. doi: 10.1109/AFGR.2000.840611

Kaunitz, L. N., Kamienkowski, J. E., Varatharajah, A., Sigman, M., Quiroga, R. Q., and Ison, M. J. (2014). Looking for a face in the crowd: fixation-related potentials in an eye-movement visual search task. *Neuroimage* 89, 297–305. doi: 10.1016/j.neuroimage.2013.12.006

Key, A. P., Dove, G. O., and Maguire, M. J. (2005). Linking brainwaves to the brain: an ERP primer. *Dev. Neuropsychol.* 27, 183–215. doi: 10.1207/s15326942dn2702_1

Kiebel, S. J., and Holmes, À. P. (2003). *The General Linear Model.* Cambridge, MA: Academic Press. doi: 10.1016/b978-012372560-8/50008-5

Killgore, W. D. S., and Yurgelun-Todd, D. A. (2007). The right-hemisphere and valence hypotheses: could they both be right (and sometimes left)? *Soc. Cogn. Affect. Neurosci.* 2, 240–250. doi: 10.1093/scan/nsm020

Kristensen, E., Guerin-Dugué, A., and Rivet, B. (2017a). Regularization and a general linear model for event-related potential estimation. *Behav. Res. Methods* 49, 2255–2274. doi: 10.3758/s13428-017-0856-z

Kristensen, E., Rivet, B., and Guérin-Dugué, A. (2017b). Estimation of overlapped eye fixation related potentials: the general linear model, a more flexible framework than the adjar algorithm. *J. Eye Mov. Res.* 10, 1–27.

Krolak-Salmon, P., Fischer, C., Vighetto, A., and Mauguière, F. (2001). Processing of facial emotional expression: spatio-temporal data as assessed by scalp event-related potentials. *Eur. J. Neurosci.* 13, 987–994. doi: 10.1046/j.0953-816x.2001.01454.x

Luck, S. J., and Hillyard, S. A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1000–1014. doi: 10.1037/0096-1523.20.5.1000

Neath, K. N., and Itier, R. J. (2015). Fixation to features and neural processing of facial expressions in a gender discrimination task. *Brain Cogn.* 99, 97–111. doi: 10.1016/j.bandc.2015.05.007

Neath-Tavares, K. N., and Itier, R. J. (2016). Neural processing of fearful and happy facial expressions during emotion-relevant and emotion-irrelevant tasks: a fixation-to-feature approach. *Biol. Psychol.* 119, 122–140. doi: 10.1016/j.biopsycho.2016.07.013

Nikolaev, A. R., Meghanathan, R. N., and van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: pitfalls and possibilities. *Brain Cogn.* 107, 55–83. doi: 10.1016/j.bandc.2016.06.004

Noordewier, M. K., and Breugelmans, S. M. (2013). On the valence of surprise. *Cogn. Emot.* 27, 1326–1334. doi: 10.1080/02699931.2013.777660

Noordewier, M. K., Topolinski, S., and Van Dijk, E. (2016). The temporal dynamics of surprise. *Soc. Personal. Psychol. Compass* 10, 136–149. doi: 10.1111/spc3.12242

Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/S1388-2457(00)00527-7

Ossandón, J. P., Helo, A. V., Montefusco-Siegmund, R., and Maldonado, P. E. (2010). Superposition model predicts EEG occipital activity during free viewing of natural scenes. *J. Neurosci.* 30, 4787–4795. doi: 10.1523/JNEUROSCI.5769-09.2010

Pizzagalli, D., Regard, M., and Lehman, D. (1999). Rapid emotional face processing in the human right and left brain hemispheres: an ERP study. *Neuroreport* 10, 2691–2698. doi: 10.1097/00001756-199909090-00001

Recio, G., Sommer, W., and Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Res.* 1376, 66–75. doi: 10.1016/j.brainres.2010.12.041

Reuter-Lorenz, P., and Davidson, R. J. (1981). Differential contributions of the two cerebral hemispheres to the perception of happy and sad faces. *Neuropsychologia* 19, 609–613. doi: 10.1016/0028-3932(81)90030-0

Reuter-Lorenz, P. A., Givis, R. P., and Moscovitch, M. (1983). Hemispheric specialization and the perception of emotion: evidence from right-handers and from inverted and non-inverted left-handers. *Neuropsychologia* 21, 687–692. doi: 10.1016/0028-3932(83)90068-4

Ries, A. J., Touryan, J., Ahrens, B., and Connolly, P. (2016). The impact of task demands on fixation-related brain potentials during guided search. *PLoS One* 11:e0157260. doi: 10.1371/journal.pone.0157260 doi: 10.1371/journal.pone.0157260

Ruchkin, D. S. (1965). An analysis of average response computations based upon aperiodic stimuli. *IEEE Trans. Biomed. Eng.* 12, 87–94. doi: 10.1109/TBME.1965.4502353

Schmidt, K. L., Ambadar, Z., Cohn, J. F., and Reed, I. (2006). Movement differences between deliberate and spontaneous facial expressions: zygomaticus major action in smiling. *J. Nonverbal Behav.* 30, 37–52. doi: 10.1007/s10919-005-0003-x

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., and Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *J. Vis.* 14:14. doi: 10.1167/14.13.14 doi: 10.1167/14.13.14

Simola, J., Torniainen, J., Moisala, M., Kivikangas, M., and Krause, C. M. (2013). Eye movement related brain responses to emotional scenes during free viewing. *Front. Syst. Neurosci.* 7:41. doi: 10.3389/fnsys.2013.00041 doi: 10.3389/fnsys.2013.00041

Simola, J., Le Fevre, K., Torniainen, J., and Baccino, T. (2015). Affective processing in natural scene viewing: valence and arousal interactions in eye-fixation-related potentials. *Neuroimage* 106, 21–33. doi: 10.1016/j.neuroimage.2014.11.030

Smith, N. J., and Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. the rerp framework. *Psychophysiology* 52, 157–168. doi: 10.1111/psyp.12317

Smith, N. J., and Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. non-linear effects, overlap correction, and practical considerations. *Psychophysiology* 52, 169–181. doi: 10.1111/psyp.12320

Stekelenburg, J. J., and de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuroreport* 15, 777–780. doi: 10.1097/00001756-200404090-00007

Tamietto, M., Corazzini, L. L., de Gelder, B., and Geminiani, G. (2006). Functional asymmetry and interhemispheric cooperation in the perception of emotions from facial expressions. *Exp. Brain Res.* 171, 389–404. doi: 10.1007/s00221-005-0279-4

Tamietto, M., Adenzato, A., Geminiani, G., and de Gelder, B. (2007). Fast recognition of social emotions takes the whole brain: interhemispheric cooperation in the absence of cerebral asymmetry. *Neuropsychologia* 45, 836–843. doi: 10.1016/j.neuropsychologia.2006.08.012

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., and Adam, J. M. (2007). Facial expressions of emotions: a methodological contribution to the study of spontaneous and dynamic emotional faces. *Eur. J. Soc. Psychol.* 37, 1325–1345. doi: 10.1002/ejsp.427

Tcherkassof, A., Dupré, D., Meillon, B., Mandran, N., Dubois, M., and Adam, J.-M. (2013). DynEmo: a video database of natural facial expressions of emotions. *Int. J. Multimed. Appl.* 5, 61–80. doi: 10.5121/ijma.2013.5505

Torro Alves, N., Fukushima, S. S., and Aznar-Casaova, J. A. (2008). Models of brain asymmetry in emotional processing. *Psychol. Neurosci.* 1, 63–66.

Trautmann-Lengsfeld, S. A., Domìnguez-Borràs, J., Escera, C., Herrmann, M., and Fehr, T. (2013). The perception of dynamic and static facial expressions of happiness and disgust investigated by erps and fmri constrained source analysis. *PLoS One* 8:e66997. doi: 10.1371/journal.pone.0066997 doi: 10.1371/journal.pone.0066997

Vaidya, A. R., Jin, C., and Fellows, L. K. (2014). Eye spy: the predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition* 133, 443–456. doi: 10.1016/j.cognition.2014.07.004

Valstar, M. F., Pantic, M., Ambadar, Z., & Cohn, J. F. (2006). "Spontaneous vs. posed facial behavior: automatic analysis of brow actions," in *Proceedings of the 8th International Conference on Multimodal interfaces, ICMI 2006*, Banff, AB, 162–170. doi: 10.1145/1180995.1181031

Vuillemier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45, 174–194. doi: 10.1016/j.neuropsychologia.2006.06.003

Wager, T. D., Phan, K. L., Liberzon, I., and Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* 19, 513–531. doi: 10.1016/S1053-8119(03)00078-8

Wagner, H. L. (1997). "Methods for the study of facial behavior," in *The Psychology of Facial Expression*, eds J. A. Russell and J. M. Fernández-Dols (Cambridge: Cambridge University Press), 31–54. doi: 10.1017/CBO9780511659911.004

Wagner, H. L., MacDonald, C. J., and Manstead, A. S. R. (1986). Communication of individual emotions by spontaneous facial expressions, *J. Pers. Soc. Psychol.* 50, 737–743. doi: 10.1037/0022-3514.50.4.737

Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction. *Psychophysiology* 30, 98–119. doi: 10.1111/j.1469-8986.1993.tb03209.x

Yagi, A. (1979). Saccade size and lambda complex in man. *Physiol. Psychol.* 7, 370–376. doi: 10.3758/BF03326658

Yagi, A. (1981). Visual signal detection and lambda responses. *Electroencephalogr. Clin. Neurophysiol.* 52, 604–610. doi: 10.1016/0013-4694(81)91434-6

Use and Usefulness of Dynamic Face Stimuli for Face Perception Studies—a Review of Behavioral Findings and Methodology

Katharina Dobs [1,2]*, Isabelle Bülthoff [2] and Johannes Schultz [2,3]

[1] *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States,*
[2] *Department Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Tübingen, Germany,*
[3] *Division of Medical Psychology and Department of Psychiatry, University of Bonn, Bonn, Germany*

Faces that move contain rich information about facial form, such as facial features and their configuration, alongside the motion of those features. During social interactions, humans constantly decode and integrate these cues. To fully understand human face perception, it is important to investigate what information dynamic faces convey and how the human visual system extracts and processes information from this visual input. However, partly due to the difficulty of designing well-controlled dynamic face stimuli, many face perception studies still rely on static faces as stimuli. Here, we focus on evidence demonstrating the usefulness of dynamic faces as stimuli, and evaluate different types of dynamic face stimuli to study face perception. Studies based on dynamic face stimuli revealed a high sensitivity of the human visual system to natural facial motion and consistently reported dynamic advantages when static face information is insufficient for the task. These findings support the hypothesis that the human perceptual system integrates sensory cues for robust perception. In the present paper, we review the different types of dynamic face stimuli used in these studies, and assess their usefulness for several research questions. Natural videos of faces are ecological stimuli but provide limited control of facial form and motion. Point-light faces allow for good control of facial motion but are highly unnatural. Image-based morphing is a way to achieve control over facial motion while preserving the natural facial form. Synthetic facial animations allow separation of facial form and motion to study aspects such as identity-from-motion. While synthetic faces are less natural than videos of faces, recent advances in photo-realistic rendering may close this gap and provide naturalistic stimuli with full control over facial motion. We believe that many open questions, such as what dynamic advantages exist beyond emotion and identity recognition and which dynamic aspects drive these advantages, can be addressed adequately with different types of stimuli and will improve our understanding of face perception in more ecological settings.

Keywords: dynamic faces, facial animation, facial motion, dynamic face stimuli, face perception, social perception, identity-from-motion, facial expressions

INTRODUCTION

Most faces we encounter and interact with move - when we meet a friend, we display continuous facial movements such as nodding, smiling and speaking. From the information conveyed by dynamic faces, we can extract cues about a person's state of mind (e.g., subtle or conversational facial expressions; Ambadar et al., 2005; Kaulard et al., 2012), about their focus of attention (e.g., gaze motion: Emery, 2000; Nummenmaa and Calder, 2009), and about what they are saying (e.g., lip movements; Rosenblum et al., 1996; Ross et al., 2007). Despite this extensive information conveyed by dynamic faces, much of it is already contained in their static counterpart, including sex, age or basic emotions (Ekman and Friesen, 1976; Russell, 1994). Therefore, and for ease of use, most face perception studies rely on static stimuli. When do dynamic faces provide additional information to static faces, and what is this information? What kind of stimuli is appropriate to study different aspects of dynamic face perception? In this review, we will discuss findings on the usefulness of dynamic faces to study face perception, followed by an overview of methodological aspects of this work. We conclude with a brief discussion, future directions and open questions.

Human Sensitivity to Spatio-Temporal Information in Dynamic Faces

Before designing any study using dynamic faces, it seems relevant to ask how sensitive the human visual system is to facial motion. Are simple approximations sufficient, or is the face perception system finely attuned to natural motion? Recent evidence supports the latter: In a recent study, we systematically manipulated the spatio-temporal information contained in animations based on natural facial motion (Dobs et al., 2014). Subjects chose in a delayed matching-to-sample task which of two manipulated animations was more similar to natural motion. Subjects consistently selected the animations closer to natural motion, demonstrating high sensitivity to deviations from natural motion. In line with these results, face stimuli based on motion created by linear morphing techniques (e.g., linear morphing between two frames) can lead to less accurate emotion recognition (Wallraven et al., 2008; Cosker et al., 2010; Korolkova, 2018) and are often perceived as less natural (Cosker et al., 2010) than natural motion. Moreover, humans are sensitive to specific properties of natural motion (e.g., velocity; Pollick et al., 2003; Hill et al., 2005; Bould et al., 2008), to temporal sequencing (e.g., temporal asymmetries in the unfolding of facial expressions; Cunningham and Wallraven, 2009; Reinl and Bartels, 2015; Delis et al., 2016; Korolkova, 2018) and even to perceptual interactions between dynamic facial features (e.g., eye and mouth moving together during yawning; Cook et al., 2015). Given this high sensitivity, what is the additional value of facial motion?

Is There an Added Value of Dynamic Compared to Static Faces?

It seems intuitive to assume that dynamic information (e.g., a video) would facilitate the identification of facial expressions compared to static images (dynamic advantage), because

expressions develop over time. However, this assumption is subject to some controversy (Krumhuber et al., 2013). Most studies report a dynamic advantage for expression recognition (Harwood et al., 1999; Ambadar et al., 2005; Bould et al., 2008; Kätsyri and Sams, 2008; Cunningham and Wallraven, 2009; Horstmann and Ansorge, 2009; Calvo et al., 2016 (for synthetic faces); Wehrle et al., 2000), while others do not (Jiang et al., 2014 (under time pressure); (Widen and Russell, 2015) (for children); (Kätsyri and Sams, 2008) (for real faces); Fiorentini and Viviani, 2011; Gold et al., 2013; Hoffmann et al., 2013).

This controversy might have arisen from differences in stimuli and paradigms or from the methods used to equalize the stimuli (Fiorentini and Viviani, 2011). For example, most studies reporting a lack of a dynamic advantage have tested basic emotions and compared the expression's peak frame as static stimulus against the video sequence (e.g., Kätsyri and Sams, 2008; Fiorentini and Viviani, 2011; Gold et al., 2013; Hoffmann et al., 2013). In contrast, in studies reporting a dynamic advantage, either the authors presented degraded or attenuated basic emotion stimuli (Bassili, 1978; see also Bruce and Valentine, 1988; but see Gold et al., 2013) or observers had difficulty extracting information from the stimuli (for example, autistic children and adults: Gepner et al., 2001; Tardif et al., 2006; but see Back et al., 2007); individuals with prosopagnosia: (Richoz et al., 2015), or more complex and subtle facial expressions were tested (Cunningham et al., 2004; Cunningham and Wallraven, 2009; Yitzhak et al., 2018). These findings suggest that the dynamic advantage is stronger for subtle than basic expressions, while a dynamic advantage for basic emotions can be best observed under suboptimal conditions (Kätsyri and Sams, 2008).

Perception of Dynamic Face Information Beyond Emotional Expressions

Facial motion does not only enhance facial expression understanding, but can also improve the perception of other face aspects. For example, one robust finding is that facial motion enhances speech comprehension when hearing is impaired (Bernstein et al., 2000; Rosenblum et al., 2002). Facial motion also conveys cues about a person's gender (Hill and Johnston, 2001) and identity (Hill and Johnston, 2001; O'Toole et al., 2002; Knappmeyer et al., 2003; Lander and Bruce, 2003; Lander and Chuang, 2005; Girges et al., 2015). Interestingly, the amount of identity information contained in facial movements depends on the type of facial movement: In a recent study (Dobs et al., 2016), we recorded from several actors three types of facial movements: emotional expressions (e.g., happiness), emotional expressions in social interaction (e.g., laughing with a friend), and conversational expressions (e.g., introducing oneself). Using a single avatar head animated with these facial movements, we found that subjects could better match actor identities based on conversational compared to emotional facial movements. Importantly, ideal observer analyses revealed that conversational movements contained more identity information, suggesting that humans move their face more idiosyncratically when in a conversation. Similar to the dynamic advantage for facial expressions, these findings show that the visual system can

use identity cues in facial motion when form information is degraded or absent. However, whether this phenomenon occurs in real life in the presence of identity cues carried by facial form was still unclear (O'Toole et al., 2002). In a recent study (Dobs et al., 2017), we systematically modified the amount of identity information contained in facial form versus motion while subjects performed an identity categorization task. Based on optimal integration models, we showed that subjects integrated facial form and motion using each cue's respective reliability, suggesting that in the presence of naturally moving faces, we combine static and dynamic cues in a near-optimal fashion. However, which dynamic aspects exactly contain useful and additional information compared to static faces is still under debate.

Which Dynamic Aspects Contain Information Beyond Static Face Information?

An obvious first hypothesis is that the dynamic face advantage is due to a dynamic stimulus providing more samples of the information contained in snapshots of static faces. This was tested using dynamic stimuli in which visual noise masks were inserted between the images making up the stimulus, maintaining the information content of the sequence but eliminating the experience of motion (Ambadar et al., 2005). This manipulation reduced recognition to the level observed with single static frames, thus falsifying this hypothesis. The authors further found that motion enhanced the perception of subtle changes occurring during facial expressions. In a series of experiments, Cunningham and Wallraven (2009) used a similar approach by presenting displays with several static faces as an array or dynamic stimuli with partially or fully randomized frame order. Results again confirmed that dynamic information was coded in the natural deformation of the face over time. Other studies revealed that motion induces a representational momentum during perception of facial expressions which facilitates the detection of changes in the emotion expressed by a face (Yoshikawa and Sato, 2008), that face movement draws attention and increases perception of emotions (Horstmann and Ansorge, 2009) and evokes stronger emotional reactions (Sato and Yoshikawa, 2007). Importantly, most studies investigating the mechanisms underlying the dynamic advantage focused on emotional expressions, ignoring other aspects in which motion contributes less information than form yet still increases performance, such as recognition of facial identity or speech. Therefore, the full picture of what drives the dynamic advantage during face processing is still incomplete.

Advantages and Disadvantages of Different Kinds of Dynamic Face Stimuli

In this section, we give an overview of different types of stimuli that can be used to investigate dynamic face perception. **Figure 1** compares five types of stimuli based on the following characteristics: level of naturalness and level of control for

form and motion, possibility of manipulating form and motion separately and level of technical demand.

The simplest way to investigate dynamic face perception is to use video recordings of faces (row "Videos" in **Figure 1**). This has several advantages. First, these stimuli are intuitively more ecologically valid than other types of stimuli since both form and motion are kept natural. Second, videos avoid discrepancies between form and motion naturalness which can reduce perceptual acceptability (e.g., uncanny valley; Mori, 1970). Third, the technical demand is low. Fourth, videos convey spontaneous facial expressions occurring in real-life well, compared to posed facial expressions which tend to be more stereotyped and artificial (Cohn and Schmidt, 2004; Kaulard et al., 2012). Videos have been used to investigate neural representations of emotional valence that generalize across different types of stimuli (Skerry and Saxe, 2014; Kliemann et al., 2018). Other studies have manipulated the order of video frames to investigate the importance of the temporal unfolding of facial expressions (Cunningham and Wallraven, 2009; Reinl and Bartels, 2015; Korolkova, 2018), or the neural sensitivity to natural facial motion dynamics (Schultz and Pilz, 2009; Schultz et al., 2013). While for these research questions, videos of faces achieved a good balance between ecological validity and experimental control, the content of information in such videos is technically challenging to assess (compare "photo-realistic face rendering" below), let alone to parametrically control it.

This control can be achieved using point-light face stimuli (row "Point-light faces" in **Figure 1**), in which only reflective markers attached to the surface of a moving face are visible. In these stimuli, static form information is typically reduced, while motion information is preserved and fully controllable (i.e., the time courses of marker positions). Studies showed that point-light faces enhance speech comprehension (Rosenblum et al., 1996), that facial expressions can be recognized from such displays (Atkinson et al., 2012) and that subjects are sensitive to the modulation of different properties of point-light faces (Pollick et al., 2003). Despite these valuable findings, one obvious disadvantage of these stimuli is that pure motion and form-from-motion information can hardly be disentangled. For example, what appears like a random point cloud as static display is clearly perceived as a face when in motion. Therefore, the information in facial point-light displays contains both facial motion properties and static face information derived from motion. Taken together, despite their usefulness to investigate perception, point-light stimuli have large drawbacks because they are highly degraded and unnatural and because motion and form-from-motion cues are intermingled.

To address the trade-off between naturalness (e.g., videos of faces) and high degree of control (e.g., point-light faces), an increasing number of studies use image-based morphing techniques (row "image-based morphing" in **Figure 1**; e.g., by linearly morphing between neutral and peak expression) to create dynamic stimuli. These stimuli represent a compromise between naturalness and experimental control since they allow controlling for motion properties such as intensity or velocity, while the face appears natural. Such stimuli have been used to

	Naturalness		Control		Separate motion from form	Technical demand	Example
	facial form	facial motion	facial form	facial motion			
Videos	high	high	low (e.g., blurring)	low (e.g., velocity)	no	low (e.g., image / video processing)	
Point-light faces	low	low	intermediate (e.g., shape-from-motion)	high (e.g., velocity, intensity)	yes	intermediate (e.g., motion capture / processing)	
Image-based morphing	high	intermediate (e.g., linearly approximated)	low (e.g., blurring)	intermediate (e.g., velocity, intensity)	no	low (e.g., image / video processing)	
Synthetic facial animation	intermediate (varies with avatar quality)	intermediate (varies with motion synthesis)	high (avatar parameter-limited)	high (varies with animation technique)	yes	high (e.g., motion retargeting, avatar design)	
Photo-realistic face rendering	high (may contain artifacts)	high (may contain artifacts)	intermediate (varies with animation technique)	high (varies with animation technique)	yes	high (e.g., motion retargeting, machine learning)	

FIGURE 1 | Schematic overview of five different kinds of face stimuli used to investigate dynamic face perception with their respective characteristics. Characteristics include (from left to right): Naturalness of facial form and motion varying between high (e.g., videos), intermediate (e.g., synthetic facial animation), and low (e.g., point-light faces); control of form and motion varying between high (e.g., synthetic facial animation), intermediate (e.g., photo-realistic rendering for form and image-based morphing for motion) or low (e.g., videos); potential for separating motion from form information (e.g., synthetic facial animation); and technical demand varying from low (e.g., videos), to high (e.g., photo-realistic rendering). For ease of comparison, advantages are colored green, intermediate in yellow and disadvantages in orange. Stimuli are listed in no particular order. While the first four kinds of stimuli are commonly used in face perception research, photo-realistic rendering is the most recent advancement and has not yet entered face perception research. [Sources of example stimuli: Videos: (Skerry and Saxe, 2014); Point-light faces: recorded with Optitrack (NaturalPoint, Inc., Corvallis, OR, USA); Image-based morphing: (Ekman and Friesen, 1978); Facial animation: designed in Poser 2012 (SmithMicro, Inc., Watsonville, CA, USA); Photo-realistic rendering: (Suwajanakorn et al., 2017)].

compare the recognition thresholds for static and dynamic faces (Calvo et al., 2016) or the perception of the intensity of facial expressions (Recio et al., 2014). Despite these useful findings, such stimuli represent solely a coarse linear approximation of natural face motion, which might lead to less accurate emotion recognition than their natural counterparts (Wallraven et al., 2008; Cosker et al., 2010; Korolkova, 2018). Moreover, these stimuli do not allow separating form and motion information, which is necessary to investigate identity-from-motion for example.

To gain full control over form and motion of faces, many studies use synthetic faces animated with facial motion properties (Hill and Johnston, 2001; Knappmeyer et al., 2003; Ku et al., 2005). While such stimuli appear more natural than stimuli based on linear morphing between images (Cosker et al.,

2010), perceived naturalness of form and motion varies with the quality of the synthetic faces and the motion used for animation (Wallraven et al., 2008). One way to generate such stimuli is to use recorded marker-based motion data (see "Point-light faces" above) from actors performing facial actions, and to map these to synthetic faces (e.g., Hill and Johnston, 2001; Knappmeyer et al., 2003). Drawbacks are the difficulty to map specific markers to face regions, and artifacts resulting from shape differences between recorded and target faces. Further, while the resulting animations can closely approximate natural expressions, systematically manipulating and interpreting the underlying motion properties remains complex. To address this challenge, complex and detailed movements can be created using a common coding scheme for facial motion called Facial Action Coding System (FACS;

Ekman and Friesen, 1978). This system uses a number of discrete 'face movements' - termed Action Units - to describe the basic components of most facial actions. Importantly, the motion properties of each Action Unit can be semantically described (e.g., eyebrow raising) and modified separately to induce systematic local changes in facial motion (Jack et al., 2012; Yu et al., 2012). Synthetic faces can be animated based on Action Unit time courses extracted from real motion-capture data (Curio et al., 2006) or synthesized in the absence of actor data (Roesch et al., 2010; Yu et al., 2012). Overall, such animations allow meaningful interpretation, quantification as well as systematic manipulation of motion properties, with full control over form. The main shortcomings are the high technical demands to create these stimuli, and the fact that the faces are synthetic.

Major advancements in the development of face tracking and animations have recently been made. In particular, it is now possible to animate faces in a photo-realistic fashion (see row "Photo-realistic rendering" in **Figure 1**). These recent developments bear potential for face perception research. First, new developments reduce the technical demands of recording facial movements allowing markerless tracking by using for example depth sensors (e.g., Walder et al., 2009; Girges et al., 2015), automated landmark detection (Korolkova, 2018), or simply RGB channels in videos (Thies et al., 2016). Second, recent facial animation and machine learning advancements (e.g., deep learning) now allow creating naturalistic dynamic face stimuli indistinguishable from real videos (e.g., Thies et al., 2016; Suwajanakorn et al., 2017). While these technologies have hardly entered face perception research to date, we believe that a novel and promising approach will consist in collaborating with computer vision labs to address open questions in face perception.

CONCLUSION AND FUTURE DIRECTIONS

In this review, we discuss the usefulness of dynamic faces for face perception studies, review the conditions under which dynamic advantages arise, and compare different kinds of stimuli used to investigate dynamic face processing. The finding that the dynamic advantage was less pronounced when other cues convey similar or more reliable information fits the view that the brain constantly integrates sensory cues (e.g., dynamic and static) based on their respective reliabilities to achieve robust perception. While such an integration mechanism was shown for identity recognition (Dobs et al., 2017), the mechanisms underlying the perception of other facial aspects (e.g., gender, age or health) still need to be unraveled. Moreover, most studies investigated faces presented alone; yet when interpreting the mood or intention of a vis-à-vis in daily life, humans do not take solely facial form and movements into account, but also gaze motion, voice, speech, so as motion of the head or even the whole body (e.g., Van den Stock et al., 2007; Dukes et al., 2017).

To better understand these aspects of face perception, future face perception studies would benefit from the use of models of cue integration as well as dynamic and multisensory face stimuli (e.g., gaze, voice).

What kind of dynamic stimulus is appropriate to study which aspect of face perception? Each of the dynamic stimuli reviewed here has specific advantages and disadvantages; it is thus difficult to make general suggestions. Findings showed that the face perception system is highly sensitive to natural facial motion, which supports the use of dynamic face stimuli based on real face motion. However, to our knowledge, a systematic investigation of differences in processing faces across different types of stimuli (e.g., synthetic faces vs. videos) is still lacking, and thus the generalizability of findings from studies using synthetic or point-light faces is still unclear and should be addressed in future studies.

Furthermore, it is still unclear which motion properties are used by the face perception system. Advances were made in the realm of dynamic expressions of emotions, but more controlled studies and paradigms are needed. Synthetic facial animations or even photo-realistic face rendering providing high control over form and motion are promising candidate stimuli to investigate these questions. For example, using synthetic facial animations and a reverse correlation technique, Jack et al. (2012) revealed cultural differences in perception of emotions from dynamic stimuli and identified the motion properties contributing to these differences. Similar techniques might help to characterize which properties convey idiosyncratic facial movements for example, and the dynamic advantage in general.

Finally, a major remaining question addresses the representation of facial motion in the human face perception system. How many dimensions are used to encode the full space of facial motions, and what are these dimensions? Recent evidence suggests that a small number of dimensions are sufficient (Dobs et al., 2014; Chiovetto et al., 2018) but more studies based on larger data sets are needed. If a set of basic components can be characterized, can we identify behavioral and neural correlates of a facial motion space, similar to what is known as face space for static faces (Valentine, 2001; Leopold et al., 2006; Chang and Tsao, 2017)?

AUTHOR CONTRIBUTIONS

KD, IB, and JS designed the concept of the article, reviewed the literature and wrote the article.

ACKNOWLEDGMENTS

REFERENCES

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Atkinson, A. P., Vuong, Q. C., and Smithson, H. E. (2012). Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. *Neuroimage* 59, 1700–1712. doi: 10.1016/j.neuroimage.2011.08.073

Back, E., Ropar, D., and Mitchell, P. (2007). Do the eyes have it? Inferring mental states from animated faces in autism. *Child Dev.* 78, 397–411. doi: 10.1111/j.1467-8624.2007.01005.x

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 373–379. doi: 10.1037/0096-1523.4.3.373

Bernstein, L. E., Tucker, P. E., and Demorest, M. E. (2000). Speech perception without hearing. *Percept. Psychophys.* 62, 233–252. doi: 10.3758/BF03205546

Bould, E., Morris, N., and Wink, B. (2008). Recognising subtle emotional expressions: the role of facial movements. *Cogn. Emotion* 22, 1569–1587. doi: 10.1080/02699930801921156

Bruce, V., and Valentine, T. (1988). "When a nod's as good as a wink: the role of dynamic information in face recognition," in *Practical Aspects of Memory: Current Research and Issues, Vol. 1,* eds M. M. Gruneberg, P. E. Morris, and R. N. Sykes (Oxford: Wiley), 169–174.

Calvo, M. G., Avero, P., Fernández-Martín, A., and Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion* 16, 1186–1200. doi: 10.1037/emo0000192

Chang, L., and Tsao, D. Y. (2017). The code for facial identity in the primate brain. *Cell* 169, 1013.e4–1020.e14. doi: 10.1016/j.cell.2017.05.011

Chiovetto, E., Curio, C., Endres, D., and Giese, M. (2018). Perceptual integration of kinematic components in the recognition of emotional facial expressions. *J. Vis.* 18, 13–19. doi: 10.1167/18.4.13

Cohn, J. F., and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multiresolution Inf. Process.* 2, 121–132. doi: 10.1142/S021969130400041X

Cook, R., Aichelburg, C., and Johnston, A. (2015). Illusory feature slowing: evidence for perceptual models of global facial change. *Psychol. Sci.* 26, 512–517. doi: 10.1177/0956797614567340

Cosker, D., Krumhuber, E., and Hilton, A. (2010). "Perception of linear and nonlinear motion properties using a FACS validated 3D facial model," in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization.* (Los Angeles, CA), 101–108.

Cunningham, D. W., Nusseck, M., Wallraven, C., and Bülthoff, H. H. (2004). The role of image size in the recognition of conversational facial expressions. *Comp. Anim. Virtual Worlds* 15, 305–310. doi: 10.1002/cav.33

Cunningham, D. W., and Wallraven, C. (2009). Dynamic information for the recognition of conversational expressions. *J. Vis.* 9, 1–17. doi: 10.1167/9.13.7

Curio, C., Breidt, M., Kleiner, M., Vuong, Q. C., Giese, M. A., and Bülthoff, H. H. (2006). "Semantic 3d motion retargeting for facial animation," in *Proceedings of the 3rd symposium on Applied Perception in Graphics and Visualization* (Boston, MA), 77–84.

Delis, I., Chen, C., Jack, R. E., Garrod, O. G. B., Panzeri, S., and Schyns, P. G. (2016). Space-by-time manifold representation of dynamic facial expressions for emotion categorization. *J. Vis.* 16, 14–20. doi: 10.1167/16.8.14

Dobs, K., Bülthoff, I., Breidt, M., Vuong, Q. C., Curio, C., and Schultz, J. (2014). Quantifying human sensitivity to spatio-temporal information in dynamic faces. *Vis. Res.* 100, 78–87. doi: 10.1016/j.visres.2014.04.009

Dobs, K., Bülthoff, I., and Schultz, J. (2016). Identity information content depends on the type of facial movement. *Sci. Rep.* 6, 1–9. doi: 10.1038/srep34301

Dobs, K., Ma, W. J., and Reddy, L. (2017). Near-optimal integration of facial form and motion. *Sci. Rep.* 7, 1–9. doi: 10.1038/s41598-017-10885-y

Dukes, D., Clément, F., Audrin, C., and Mortillaro, M. (2017). Looking beyond the static face in emotion recognition: the informative case of interest. *Vis. Cogn.* 25, 575–588. doi: 10.1080/13506285.2017.1341441

Ekman, P., and Friesen, W. V. (1976). Measuring facial movement. *J. Nonverbal Behav.* 1, 56–75. doi: 10.1007/BF01115465

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurements of Facial Movement.* Palo Alto, CA.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-7

Fiorentini, C., and Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *J. Vis.* 11, 1–15. doi: 10.1167/11.3.17

Gepner, B., Deruelle, C., and Grynfeltt, S. (2001). Motion and emotion: a novel approach to the study of face processing by young autistic children. *J. Autism Dev. Disord.* 31, 37–45. doi: 10.1023/A:1005609629218

Girges, C., Spencer, J., and O'Brien, J. (2015). Categorizing identity from facial motion. *Q. J. Exp. Psychol.* 68, 1832–1843. doi: 10.1080/17470218.2014.993664

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., et al. (2013). The efficiency of dynamic and static facial expression recognition. *J. Vis.* 13, 1–12. doi: 10.1167/13.5.23

Harwood, N. K., Hall, L. J., and Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *Am. J. Ment. Retard.* 104, 270–278. doi: 10.1352/0895-8017(1999)104%253C0270:ROFEEF%253E2.0.CO;2

Hill, H. C. H., Troje, N. F., and Johnston, A. (2005). Range- and domain-specific exaggeration of facial speech. *J. Vis.* 5, 4–15. doi: 10.1167/5.10.4

Hill, H., and Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Curr. Biol.* 11, 880–885. doi: 10.1016/S0960-9822(01)00243-3

Hoffmann, H., Traue, H. C., Limbrecht-Ecklundt, K., Walter, S., and Kessler, H. (2013). Perception of dynamic facial expressions of emotion. *Psychology* 4, 663–668. doi: 10.1007/11768029_17

Horstmann, G., and Ansorge, U. (2009). Visual search for facial expressions of emotions: a comparison of dynamic and static faces. *Emotion* 9, 29–38. doi: 10.1037/a0014147

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., and Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proc. Nat. Acad. Sci. U.S.A.* 109, 7241–7244. doi: 10.1073/pnas.1200155109/-/DCSupplemental/sm01.avi

Jiang, Z., Li, W., Recio, G., Liu, Y., Luo, W., Zhang, D., et al. (2014). Time pressure inhibits dynamic advantage in the classification of facial expressions of emotion. *PLoS ONE* 9:e100162. doi: 10.1371/journal.pone.0100162

Kätsyri, J., and Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *Int. J. Hum. Comp. Stud.* 66, 233–242. doi: 10.1016/j.ijhcs.2007.10.001

Kaulard, K., Cunningham, D. W., Bülthoff, H. H., and Wallraven, C. (2012). The MPI facial expression database — a validated database of emotional and conversational facial expressions. *PLoS ONE* 7:e32321. doi: 10.1371/journal.pone.0032321.s002

Kliemann, D., Richardson, H., Anzellotti, S., Ayyash, D., Haskins, A. J., Gabrieli, J. D. E., et al. (2018). Cortical responses to dynamic emotional facial expressions generalize across stimuli, and are sensitive to task-relevance, in adults with and without Autism. *Cortex* 103, 24–43. doi: 10.1016/j.cortex.2018.02.006

Knappmeyer, B., Thornton, I. M., and Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vis. Res.* 43, 1921–1936. doi: 10.1016/S0042-6989(03)00236-0

Korolkova, O. A. (2018). The role of temporal inversion in the perception of realistic and morphed dynamic transitions between facial expressions. *Vis. Res.* 143, 42–51. doi: 10.1016/j.visres.2017.10.007

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Ku, J., Jang, H. J., Kim, K. U., Kim, J. H., Park, S. H., Lee, J. H., et al. (2005). Experimental results of affective valence and arousal to avatar's facial expressions. *Cyberpsychol. Behav.* 8, 493–503. doi: 10.1089/cpb.2005.8.493

Lander, K., and Bruce, V. (2003). The role of motion in learning new faces. *Vis. Cogn.* 10, 897–912. doi: 10.1080/13506280344000149

Lander, K., and Chuang, L. (2005). Why are moving faces easier to recognize? *Vis. Cogn.* 12, 429–442. doi: 10.1080/13506280444000382

Leopold, D. A., Bondar, I. V., and Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442, 572–575. doi: 10.1038/nature04951

Mori, M. (1970). The uncanny valley. *Energy* 7, 33–35.

Nummenmaa, L., and Calder, A. J. (2009). Neural mechanisms of social attention. *Trends Cogn. Sci.* 13, 135–143. doi: 10.1016/j.tics.2008.12.006

O'Toole, A. J., Roark, D. A., and Abdi, H. (2002). Recognizing moving faces: a psychological and neural synthesis. *Trends Cogn. Sci.* 6, 261–266. doi: 10.1016/S1364-6613(02)01908-3

Pollick, F. E., Hill, H., Calder, A., and Paterson, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception* 32, 813–826. doi: 10.1068/p3319

Recio, G., Schacht, A., and Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: specificity and intensity effects in event-related brain potentials. *Biol. Psychol.* 96, 111–125. doi: 10.1016/j.biopsycho.2013.12.003

Reinl, M., and Bartels, A. (2015). Perception of temporal asymmetries in dynamic facial expressions. *Front. Psychol.* 6:403. doi: 10.3389/fpsyg.2015.01107

Richoz, A. R., Jack, R. E., Garrod, O. G. B., Schyns, P. G., and Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex* 65, 50–64. doi: 10.1016/j.cortex.2014.11.015

Roesch, E. B., Tamarit, L., Reveret, L., Grandjean, D., Sander, D., and Scherer, K. R. (2010). FACSGen: a tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *J. Nonverbal Behav.* 35, 1–16. doi: 10.1007/s10919-010-0095-9

Rosenblum, L. D., Johnson, J. A., and Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *J. Speech Hear. Res.* 39, 1159–1170. doi: 10.1044/jshr.3906.1159

Rosenblum, L. D., Yakel, D. A., Baseer, N., Panchal, A., Nodarse, B. C., and Niehus, R. P. (2002). Visual speech information for face recognition. *Percept. Psychophys.* 64, 220–229. doi: 10.3758/BF03195788

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what i am saying? exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153. doi: 10.1093/cercor/bhl024

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychol. Bull.* 115, 102–141. doi: 10.1037/0033-2909.115.1.102

Sato, W., and Yoshikawa, S. (2007). Enhanced experience of emotional arousal in response to dynamic facial expressions. *J. Nonverbal Behav.* 31, 119–135. doi: 10.1007/s10919-007-0025-7

Schultz, J., Brockhaus, M., Bülthoff, H. H., and Pilz, K. S. (2013). What the human brain likes about facial motion. *Cereb. Cortex* 23, 1167–1178. doi: 10.1093/cercor/bhs106

Schultz, J., and Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Exp. Brain Res.* 194, 465–475. doi: 10.1007/s00221-009-1721-9

Skerry, A. E., and Saxe, R. (2014). A common neural code for perceived and inferred emotion. *J. Neurosci.* 34, 15997–16008. doi: 10.1523/JNEUROSCI.1676-14.2014

Suwajanakorn, S., Seitz, S. M., and Kemelmacher-Shlizerman, I. (2017). Synthesizing obama. *ACM Trans. Graph.* 36, 1–13. doi: 10.1145/3072959.3073640

Tardif, C., Lainé, F., Rodriguez, M., and Gepner, B. (2006). Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition and induces facial–vocal imitation in children with autism. *J. Autism Dev. Disord.* 37,1469–1484. doi: 10.1007/s10803-006-0223-x

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., and Nießner, M. (2016). "Face2Face: real-time face capture and reenactment of RGB videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas), 2387–2395

Valentine, T. (2001). "Face-space models of face recognition," in *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges*, eds M. J. Wenger and J. T. Townsend (Hillsdale, NJ: Lawrence Erlbaum Associates Inc), 83–113.

Van den Stock, J., Righart, R., and de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion* 7, 487–494. doi: 10.1037/1528-3542.7.3.487

Walder, C., Breidt, M., Bülthoff, H., Schölkopf B., and Curio, C. (2009). "Markerless 3d face tracking," in *Joint Pattern Recognition Symposium* (Berlin; Heidelberg), 41–50.

Wallraven, C., Breidt, M., Cunningham, D. W., and Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Trans. Appl. Percep.* 4, 1–20. doi: 10.1145/1278760.1278764

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Pers. Soc. Psychol.* 78, 105–119. doi: 10.1037//0022-3514.78.1.105

Widen, S. C., and Russell, J. A. (2015). Do dynamic facial expressions convey emotions to children better than do static ones? *J. Cogn. Dev.* 16, 802–811. doi: 10.1080/15248372.2014.916295

Yitzhak, N., Gilaie-Dotan, S., and Aviezer, H. (2018). The contribution of facial dynamics to subtle expression recognition in typical viewers and developmental visual agnosia. *Neuropsychologia* 117, 265–35. doi: 10.1016/j.neuropsychologia.2018.04.035

Yoshikawa, S., and Sato, W. (2008). Dynamic facial expressions of emotion induce representational momentum. *Cogn. Affect. Behav. Neurosci.* 8, 25–31. doi: 10.3758/CABN.8.1.25

Yu, H., Garrod, O. G. B., and Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Comput. Graph.* 36, 152–162. doi: 10.1016/j.cag.2011.12.002

Metaphorical Action Retrospectively but Not Prospectively Alters Emotional Judgment

Tatsuya Kato[1]*[†], Shu Imaizumi[1,2]*[†] and Yoshihiko Tanno[1]

[1] Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan, [2] Japan Society for the Promotion of Science, Tokyo, Japan

Metaphorical association between vertical space and emotional valence is activated by bodily movement toward the corresponding space. Upward or downward manual movement "following" observation of emotional images is reported to alter the perceived valence as more positive or negative. This study aimed to clarify this retrospective emotional modulation. Experiment 1 investigated the effects of temporal order of emotional stimuli and manual movements. Participants performed upward, downward, or horizontal manual movements immediately before or after observation of emotional images; they then rated the valence of the image. The images were rated as more negative in downward- than in horizontal-movement conditions only when the movements followed the image observation. Upward movement showed no effect. Experiment 2 examined the effects of temporal proximity between images, movements, and ratings. The results showed that a 2-s interval either between image and movement or movement and rating nullified the retrospective effect. Bodily movement that corresponds to space–valence metaphor retrospectively, but not prospectively, alters the perceived valence of emotional stimuli. This effect requires temporal proximity between emotional stimulus, the subsequent movement, and rating of the stimulus. With respect to the lack of effect of upward–positive correspondence, anisotropy in effects of movement direction is discussed.

Keywords: human cognition, action, emotion, space–valence metaphor, embodiment, postdiction

INTRODUCTION

Human cognition (e.g., thought, emotion) drives bodily action and can also be affected by the action and its entailed somatosensory input. Such aspects of cognition formed by the body are called as embodied cognition (Niedenthal, 2007; Barsalou, 2008; Landau et al., 2010). For example, after filling out a questionnaire attached to a clipboard, people who had a heavy clipboard estimated social problems to be more serious compared with those who had a light clipboard (Jostmann et al., 2009). In another scenario, people who held a hot beverage felt more social proximity to a known other compared with people who held a cold beverage (Ijzerman and Semin, 2009). As such, somatosensory input representing physical weight and warmth may affect the importance of a problem and the psychological warmth of others, respectively. An underlying mechanism of embodied cognition is a metaphorical relationship between concrete and abstract concepts. In the above examples, the concrete concepts of physical weight and warmth are metaphorically

associated with the abstract concepts of importance and psychological warmth. Humans are able to understand various abstract concepts in the mental and social worlds by associating them with corresponding concrete concepts through somatosensory information from bodily action and external stimuli (Barsalou, 2008; Lee, 2016).

The concept of space, such as up or down, can represent emotional valence and power as a metaphor. Up represents goodness, strength, and joyful, whereas down represents the opposite as in the examples "He moved up the rank," "My friend has been feeling down." These metaphorical expressions are seen in various languages besides English (Marmolejo-Ramos et al., 2013). Indeed, such metaphorical association can influence cognitive performance. For instance, upward visual attention activates a concept of "up" associated with positive valence and consequently engenders a faster response to positive stimuli (e.g., word) (Meier and Robinson, 2004; Santiago et al., 2012). On the contrary, after being presented positive words on the center of display, the reaction time to a cue at the top of the display becomes faster (Xie et al., 2015). This "metaphor congruency effect" promotes cognitive processing that occurs when two concepts are in a corresponding metaphorical relationship (e.g., upward–positive). Furthermore, bodily movements can serve as a trigger of space–valence metaphor congruency effect and change the ongoing and subsequent processing of emotional stimuli. For instance, moving objects upward or downward can concurrently promote recollection of positive or negative autobiographic memory (Casasanto and Dijkstra, 2010), and sensation of upward self-motion (i.e., upward vection) induced by moving gratings can promote recollection of positive memories (Seno et al., 2013). Hence, vertical bodily movements and their related sensory input may affect the simultaneous and/or subsequent emotional processing.

Humans do not only predict future events from present and past stimuli but also retrogradely reorganize perceptions and interpretations of past stimuli by later stimuli, in a process called "postdiction" (Shimojo, 2014). For example, when a dot is flashed once at a position vertically aligned with another smoothly and horizontally moving dot, the flashed dot is perceived at a lagged position relative to the moving dot position, despite the two dots being in the same vertical position at the flashing moment (Eagleman and Sejnowski, 2000). In this "flash-lag illusion," the moving dot's motion signals, within a time window of ∼80 ms *after* the flashed dot, are used to generate the percept of the relative position of the moving dot when flashing (Eagleman and Sejnowski, 2007). In addition to the postdictive perception in a short time scale, athletes who won a match tend to reconstruct their prediction of performance reported before the match as more positive, and vice versa (Shimojo, 2014). Thus, postdiction can be observed even in a relatively long-time scale.

Based on theory of embodied cognition and metaphor congruency effect, Sasaki et al. (2015) hypothesized that if postdiction can also occur in emotional processing, the emotional valence of visual stimuli would be reconstructed by the subsequent "vertical" information activated by bodily movements. In their experiments, participants were instructed to move a dot on a touch panel (virtually, participants' hand)

upward, downward, leftward, or rightward after the presentation of visual stimuli representing positive, negative, and neutral emotions. Finally, the participants rated the valence of the stimuli. Their results showed that, when moving the dot upward, the stimuli were rated as more positive than in those conditions where there were horizontal movements, regardless of the valence of the stimuli (i.e., valence rating scores for positive, negative, and neutral images were biased to be more positive). Conversely, in the moving down condition, the stimuli were rated as more negative compared with those in the horizontal conditions. Therefore, the perceived valence of emotional visual stimuli can be postdictively or retrospectively reorganized by the vertical bodily movements that metaphorically corresponded to emotional valence.

Nevertheless, the underlying mechanisms of the metaphorical, postdictive modulation of emotional valence by bodily movements (Sasaki et al., 2015) have yet to be fully understood. Specifically, it remains unclear whether this effect is limited to be postdictive or can be generalized to the predictive or prospective effect. To our knowledge, no study has investigated the effect of motor action on subsequent emotional processing of visual stimuli. Furthermore, the prerequisites for this postdictive effect have not been determined. Sasaki et al. (2015) showed that a substantial temporal discrepancy (i.e., 2-s delay) between emotional stimuli and the following vertical action nullifies the emotional modulation effect, suggesting that temporal proximity between stimuli and movement is a prerequisite. However, the crucial temporal relationship, among visual stimuli, movements, and the following retrospective evaluation, has not been identified.

Therefore, the present study conducted two experiments according to the experimental paradigm in Sasaki et al. (2015), to extend their findings. In Experiment 1, we investigated the relationship between vertical manual movements and perceived emotional valence of visual stimuli not only in the condition with action *following* visual stimuli but also in the condition with action *preceding* visual stimuli. If the action corresponding to space–valence metaphor affects the perceived valence of stimuli regardless of the temporal order of stimuli and action, it will be perceived as more positive and negative by upward and downward manual movements, respectively, in both conditions. Additionally, as upward and downward arm movements can alter the perceived valence of emotional images, regardless of their actual valence (Sasaki et al., 2015), we expected that this image valence-independent effect would also be observed in the present study. In Experiment 2, we tested the influence of temporal proximity between stimuli, action, and evaluation on metaphorical emotional modulation, by inserting 2-s intervals between stimuli and action, or between action and evaluation.

EXPERIMENT 1

Materials and Methods
Participants
Thirty-nine healthy Japanese undergraduates participated for monetary compensation of 500 Japanese yen (∼4.5 US dollars).

FIGURE 1 | Schematic of experimental setup. In the actual experiment, the participant's chin was placed on a chin rest.
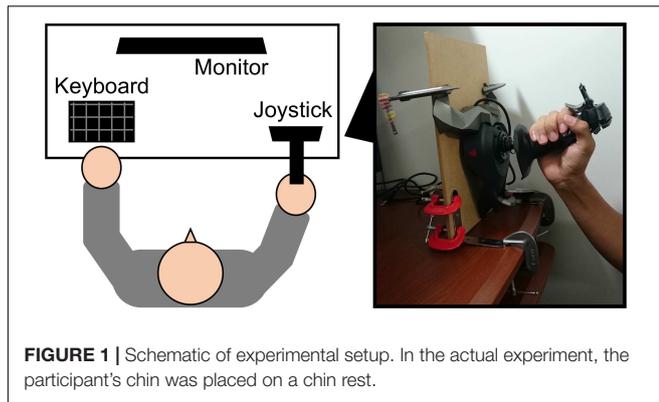
Four participants were excluded from the analysis because their number of error trials (see "Procedures") exceeded 2 SD from the mean. Finally, data from 18 participants in the retrospect condition (13 females; mean age 19.7 years, $SD = 1.25$) and 17 in the prospect condition (9 females; mean age 20.3 years, $SD = 1.57$) were analyzed. All reported that they were right-handed and had normal or corrected-to-normal visual acuity. The sample size was determined based on *a priori* power analysis using G*Power (Faul et al., 2007) version 3.1.9.3 for a one-sample, two-tailed *t*-test to check the effect of upward and downward manual movements on emotional valence rating. The power analysis indicated that at least 16 participants were required for a statistical power of 0.90, assuming an effect size Cohen's $|d|$ of 0.88 and 0.90, reported by Sasaki et al. (2015), and Type I error probability of 0.05. This study was carried out in accordance with the recommendations of the ethical committee of the Graduate School of Arts and Sciences, The University of Tokyo. The protocol was approved by the ethical committee of the Graduate School of Arts and Sciences, The University of Tokyo (approval number: 468). All participants gave written informed consent in accordance with the Declaration of Helsinki.

Apparatus

Visual stimuli were presented on a 24-inch liquid crystal display monitor (V242, Hewlett Packard, Palo Alto, CA, United States) with resolution of 1920 × 1080 pixels and refresh rate of 60 Hz. Participants viewed the monitor at a distance of approximately 57 cm with a chin rest. A joystick (Cyborg V1, Mad Catz, Hong Kong) was installed on a board along with a coronal plane parallel to the participants' coronal plane. Participants could move the joystick with their right hand in all orientations on the coronal plane. The joystick was placed on the right side of the participants' visual periphery (i.e., without direct obstacle to the visual stimuli). The setup (**Figure 1**) followed the one used in a previous study on the relationship between space–valence metaphor and manual action (Sasaki et al., 2016). Participants responded using a standard QWERTY keyboard with their left hand. Stimulus presentation and response collection were controlled by MATLAB R2016a (MathWorks, Natick, MA, United States) with Psychophysics Toolbox 3 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) running on a Windows 10 computer.

Stimuli

Visual stimuli included a fixation dot, action cues, emotional images, and a rating scale, and were presented on a gray background (**Figure 2**). The chromatic and luminance parameters of stimuli followed those in a previous study (Sasaki et al., 2015). The fixation dot was a solid white circle (0.3° diameter) and presented at the center of the monitor. The action cues consisted of the fixation dot, a solid black dot (0.3° diameter), and rectangles. The black dot was superimposed onto the fixation dot and could be moved by the joystick. Each of blue- and red-colored solid rectangles were placed on the top and bottom ends or the left and right ends on the monitor. The rectangles subtended by 7.2° × 51.9° when displayed on the top and bottom ends, whereas they subtended by 32.4° × 17.8° when displayed on the left and right. The rectangles were presented at a distance of 11.4° from the center of the monitor.

Twenty images from each of positive, neutral, and negative affective categories in the International Affective Picture System (IAPS) (Lang et al., 2008) were derived (**Table 1**). Each image subtended by 12.8° × 16.8°. The IAPS images used by Sasaki et al. (2015) varied in size; we chose images with a fixed size to eliminate potential confounding factor. The fixation dot was superimposed at the center of the image. To confirm that three image categories varied in the emotional valence rating scores but were comparable in the arousal rating scores, we performed an analysis of variance (ANOVA) with a factor of *Image category* on the valence and arousal scores. The results showed a significant effect of *Image category* [$F(2, 57) = 634.1$, $p < 0.01$, $\eta_p^2 = 0.96$]. Comparison between image categories with Bonferroni correction revealed that the valence score of positive stimuli was higher compared with neutral [$t(57) = 17.77$, $p < 0.01$, $d = 5.62$] and negative stimuli [$t(57) = 36.46$, $p < 0.01$, $d = 11.53$], and that the score of neutral stimuli was higher compared with negative stimuli [$t(57) = 17.86$, $p < 0.01$, $d = 5.65$]. There was no difference in arousal scores between image categories [$F(2, 57) = 1.37$, $p = 0.26$, $\eta_p^2 = 0.05$].

The rating scale, from −3 to +3, was written with white lines (vertical lines, 1.2°; horizontal line, 11.4°), also presented at the center of the monitor. When participants chose a number, a solid white dot (0.3° diameter) moved to the intersection of the vertical and horizontal lines under the selected number.

Procedures

The experiment was individually conducted in a quiet darkroom. Participants sat at the designated seat and then manipulated the joystick with their right hand and the keyboard with their left. Before the experiment, the participants controlled a black dot on the screen freely, using the joystick for 10 s, to get accustomed to the apparatus. A trial (**Figure 2**) began by pressing the space key during the presentation of "start" on the screen. At first, the fixation dot was presented for 500 ms. Then, in the retrospect condition, the emotional image was displayed for 500 ms followed by the action cue; in the prospect condition, the action cue was followed by the emotional image. The action cue was presented for 1,500 ms or until the participants moved the black dot to either target or non-target area. At the end of the

FIGURE 2 | Schematic of trials in Experiment 1. In **(A)** the retrospect and **(B)** prospect conditions, action cue followed or preceded the presentation of emotional image, respectively.
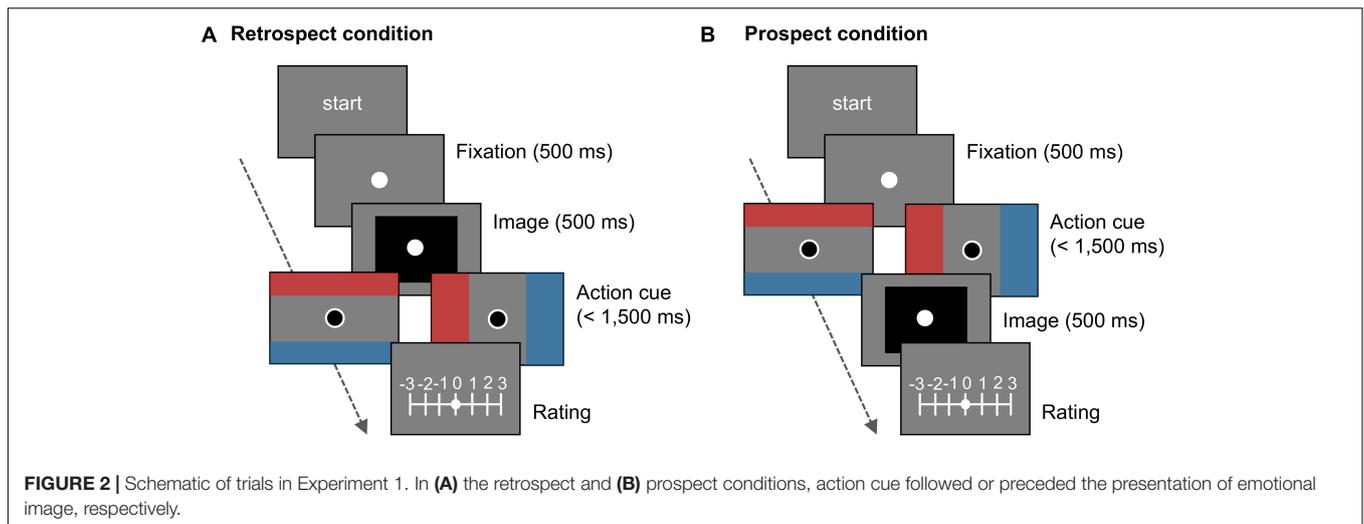
TABLE 1 | Images from the International Affective Picture System (IAPS) used in Experiments 1 and 2.

	Image category		
	Positive	**Neutral**	**Negative**
IAPS number	1440, 1604, 1750, 1920, 1999, 2150, 2398, 5200, 5700, 5760, 5830, 5831, 5982, 7280, 7330, 7508, 8420, 8497, 8501, 8540	1390, 1560, 1670, 1947, 2020, 2025, 2220, 2690, 2850, 5395, 5532, 5661, 5920, 7182, 7188, 7211, 7351, 7484, 7503, 7620	2455, 2490, 2750, 2800, 2900, 6242, 9000, 9041, 9110, 9220, 9280, 9290, 9330, 9342, 9390, 9435, 9471, 9830, 9902, 9925
Mean valence (SD)	7.63 (0.34)	5.41 (0.44)	2.77 (0.48)
Mean arousal (SD)	4.63 (0.82)	4.33 (0.76)	4.70 (0.66)

trial, the participants were asked to rate the emotional valence of the image using a seven-point Likert scale ranging from −3 (strongly negative) to +3 (strongly positive) with the keyboard. Negative values were displayed on the left side of the screen and positive values were on the other side for all participants.

The experiment consisted of a vertical and a horizontal session. Each participant in the retrospect- and prospect-condition groups completed both sessions. The session order was counterbalanced across participants. In the vertical session, the target area was displayed on either the top or bottom of the screen (i.e., upward or downward condition, respectively), and the non-target area was displayed on the other side. As such, the participants were required to move their right arm up or down to move the black dot upward or downward on the screen. In the horizontal session, the target area was displayed on the left or right of the screen (i.e., leftward or rightward condition), and the non-target area was displayed on the other side. The horizontal session was considered to provide a baseline measure by collapsing responses under leftward and rightward conditions. The color of the target area (i.e., blue or red) was fixed per participant but counterbalanced across participants.

Each session included 20 practice trials and 60 main trials. In the practice trials, a neutral image, which was not used in the main trials, was presented. In the main trials, 30 images (i.e., 10 each of positive, neutral, and negative images) were randomly chosen from the set of 60 images and then presented in a randomized order according to one condition; the other 30 images were used in the other condition. The order of conditions was also randomized within a session.

Data Availability
All datasets analyzed for this study are included in the **Data Sheet S1** of the **Supplementary Material**.

Results
We excluded from the analyses error trials where the black dot did not reach the target area within 1,500 ms or reached the non-target area (1.3% of trials in total). We performed an ANOVA with *Direction* (i.e., upward, downward, leftward, and rightward arm movements) as a within-participant factor and *Order* (i.e., retrospect and prospect) as a between-participant factor on the averaged valence rating for emotional images. There was a significant main effect of *Direction* [Greenhouse–Geisser corrected, $F(2.34, 77.14) = 3.35$, $p = 0.03$, $\eta_p^2 = 0.09$]; however, we did not find the main effect of *Order* [$F(1, 33) = 3.13$, $p = 0.09$, $\eta_p^2 = 0.09$] and their interaction [$F(2.34, 77.14) = 1.65$, $p = 0.19$, $\eta_p^2 = 0.05$]. *Post hoc* planned comparisons using Tukey's test revealed no differences in valence ratings between leftward and rightward movements in the retrospect and prospect conditions [$t(99) = 0.56$, $p = 0.99$, $d = 0.11$; $t(99) = -0.91$, $p = 0.98$, $d = -0.18$, respectively]. Thus, in the following analyses, averaged data of the leftward and rightward conditions (hereafter, "horizontal condition") served as a baseline measure.

To investigate whether the manual action of moving the dot upward and downward biased the valence ratings, we calculated the valence bias score by subtracting the averaged score in the horizontal condition from that in the upward condition (i.e., upward bias) and downward condition (i.e., downward bias)

(Sasaki et al., 2015). The positive and negative values of the valence bias score indicated that the perceived valence of the emotional images was modified as more positive and negative owing to vertical manual movements, respectively.

Average and individual data for valence bias scores are summarized in **Figure 3**. To test for significant upward or downward bias, we performed one-sample, two-tailed t-tests against zero. In the retrospect condition, there was no significant upward bias [$t(17) = 1.52$, $p = 0.15$, $d = 0.51$], although we found a significant downward bias [$t(17) = -2.69$, $p = 0.02$, $d = -0.90$]. The results suggested that downward movement made the perceived emotional valence of the image more negative. In the prospect condition, upward and downward biases were comparable to zero [upward: $t(16) = 0.87$, $p = 0.40$, $d = 0.30$; downward: $t(16) = -0.09$, $p = 0.93$, $d = -0.03$]. Furthermore, ANOVA with the factors of *Direction* (upward, downward) and *Order* (retrospect, prospect) on valence bias scores revealed a main effect of *Direction* [$F(1, 33) = 5.88$, $p = 0.02$, $\eta_p^2 = 0.15$] but not effect of *Order* [$F(1, 33) = 0.55$, $p = 0.46$, $\eta_p^2 = 0.02$] and their interaction [$F(1, 33) = 2.05$, $p = 0.16$, $\eta_p^2 = 0.06$]. *Post hoc* planned comparisons using Bonferroni correction revealed a significant difference between upward and downward movements in the retrospect condition [$F(1, 33) = 7.66$, $p < 0.01$, $\eta_p^2 = 0.19$] but not in the prospect condition [$F(1, 33) = 0.48$, $p = 0.49$, $\eta_p^2 = 0.01$]. Finally, to further ensure the null effects of vertical movements in the prospect condition, we performed the Bayesian one-sample two-tailed t-test (i.e., null hypothesis: bias score = 0) with the Cauchy prior width of 0.707 using JASP 0.8.6 (JASP JASP Team, 2018). Results of the Bayesian analysis provided the Bayes factor (BF$_{01}$; for detailed results, see **Supplementary Figures S1–S8**

in the **Data Sheet S2** of the **Supplementary Material**. For example, BF$_{01}$ of 3 indicates that the observed data are three times more likely to occur under the null hypothesis than the alternative hypothesis. We interpreted > 3.00 BF$_{01}$ value as substantial evidence of null hypothesis, 1.00–3.00 BF$_{01}$ value as weak evidence of null hypothesis, 0.33–1.00 BF$_{01}$ value as weak evidence of alternative hypothesis, and 0.10–0.33 BF$_{01}$ value as substantial evidence of alternative hypothesis (Jeffreys, 1961). The null effects of vertical movements in the prospect condition were supported by weak and substantial evidence for the null hypothesis; upward movement: BF$_{01} = 2.88$; downward movement: BF$_{01} = 4.00$. In contrast, the effect of downward movement in the retrospect condition was suggested by substantial evidence of the alternative hypothesis (BF$_{01} = 0.27$) while we obtained weak evidence of the null hypothesis for the upward movement (BF$_{01} = 1.52$). In sum, these results suggest that vertical arm movements following but not preceding observation of emotional images modulated the perceived valence of the images.

Based on visual inspections of **Figure 3**, one might notice potential outliers (e.g., a very low score in the upward, prospect condition), which would cause doubt concerning any confounding effects that could result in a null effect of the vertical arm movements. However, we have confirmed that statistically comparable results were obtained from the analyses with and without four outliers (for details, see **Supplementary Figures S9–S13** in the **Data Sheet S2** of the **Supplementary Material**).

Discussion

Our results indicated that vertical manual movements could affect the perceived valence of emotional images when the action was performed after, but not before, the observation of the emotional images. As such, bodily movements corresponding to space–valence metaphorical association may retrospectively, but not prospectively, modulate our visual experience of emotional valence. Our findings support and extend those in Sasaki et al. (2015), while also contradicting them. That is, we found only the biasing effect of downward movement, whereas Sasaki et al. (2015) showed both upward and downward biases. We speculated that a methodological difference might have caused the different results. In their experiment, visual stimuli were presented on a touch panel; participants reached their hand forward and moved it on the surface of the panel. In our experiment, participants held the joystick at a space near their shoulder. One potential explanation for the null effect of upward movement is that the difficulty of upward arm movement owing to arm posture and/or the weight and stiffness of joystick may have interfered with the metaphorical and emotional modulation by upward movement, although the upward movement itself has been accomplished in all analyzed trials.

Our *post hoc* analysis revealed that downward arm movements also had a specific effect by which the perceived negative valence of negative images was enhanced retrospectively. As space–valence metaphor postulates specific associations, such as down–negative (Meier and Robinson, 2004; Casasanto and Dijkstra, 2010; Santiago et al., 2012; Seno et al., 2013; Xie et al., 2015), it may be reasonable that the
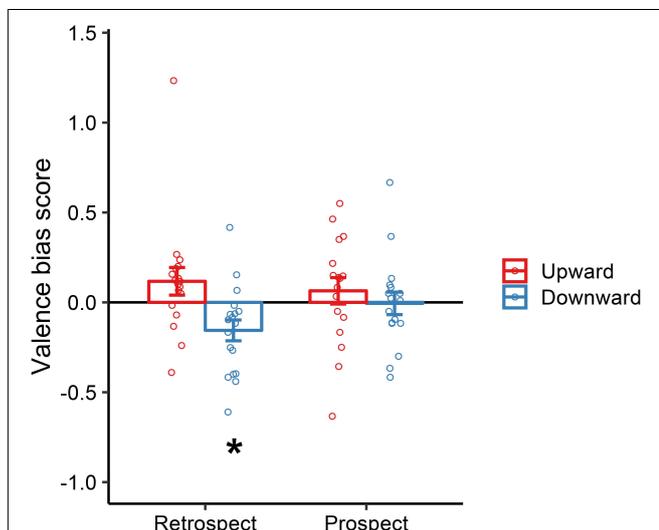
FIGURE 3 | Valence bias score by upward and downward movements in the retrospect and prospect conditions of Experiment 1. Error bars show the standard error of the mean across participants. Open circles represent each participant's data. The asterisk represents significant difference between the mean score and zero (*$p < 0.05$).

space–valence metaphor activated by movement with a certain direction influences only the stimuli with corresponding emotional valence. Nevertheless, this downward-specific effect may not be powerful such that the positive stimuli are rated as less positive.

Why do vertical movements performed "after" the visual experience of emotional stimuli modulate the perceived valence of the stimuli? The null effect found in the prospect condition suggests that space–valence metaphor activated by a preceding action does not affect the following visual experience of emotional valence. Thus, the visual emotional experience might be modified by the activated space–valence metaphor on a retrospective stage of recalling and evaluating past perceptions and impressions. If so, this retrospection may be deteriorated by a substantial temporal discrepancy between emotional stimuli, metaphorical bodily movements, and retrospection (e.g., rating), consequently nullifying the effect of the vertical movements on the perceived emotional valence. Specifically, we hypothesized three potential underlying mechanisms. First, temporally proximate visual information (i.e., emotional images) and motor information (i.e., vertical movements activating space–valence metaphor) would be bound at the following stage of evaluation (i.e., valence rating), resulting in biased recollection of the visual information. Second, temporal proximity between vertical manual movements and subsequent evaluation would be necessary so that the movement could bias the immediately subsequent evaluation. Third, temporal proximity between visual information, manual movements, and the subsequent evaluation would be necessary. Indeed, Sasaki et al. (2015) already reported that vertical manual movements do not influence the perceived valence of emotional images in the condition with temporal interval of 2 s between the images and movements (valence rating immediately followed the movements). As such, the first and/or third hypothetical mechanisms may be plausible, while the second may not. Therefore, it remains still unclear whether temporal proximity between emotional stimuli and manual movements itself is sufficient for the effect, or whether proximity between stimuli, movements, and evaluation is required.

To this end, in Experiment 2, we examined how temporal proximity between emotional images, vertical manual movements, and valence rating influences the retrospective metaphorical modulation effect on emotional experience by vertical manual movements corresponding to space–valence metaphor. The methods were identical to the retrospect condition in Experiment 1, except that we inserted 2-s temporal intervals between emotional stimuli and movements [i.e., image–action condition; similar to Sasaki et al. (2015)], and between movements and valence rating (i.e., action–rating condition). If proximity between emotional stimuli and manual movements is crucial, metaphorical modulation effect would be observed in the action–rating condition but not in the image–action condition. Meanwhile, if proximity between emotional stimuli, movements, and evaluation is required, the effect would not be observed in both conditions.

EXPERIMENT 2

Materials and Methods

Participants

Thirty-two healthy right-handed Japanese undergraduates participated for monetary compensation. None of them participated in Experiment 1. Three participants were excluded from the analysis because their number of error trials exceeded 2 SD from the mean. Finally, data from 15 participants in the image–action condition (six females; mean age 19.5 years, $SD = 0.99$) and 14 in the action–rating condition (one female; mean age 19.6 years, $SD = 0.76$) were analyzed.

Apparatus and Stimuli

Identical to those in Experiment 1.

Procedures

The task and procedure were identical to the retrospect condition in Experiment 1, except that 2-s intervals were inserted either between the presentation of emotional images and action cues (i.e., image–action condition) or between action cues and valence rating (i.e., action–rating condition), as illustrated in **Figure 4**. A gray screen and a white fixation dot were displayed during the blank interval. The participants were assigned to either the image–action or action–rating condition. The duration of the blank interval was in accordance with that in a previous study (Sasaki et al., 2015).

Results

Trials in which the black dot did not reach the target area within 1,500 ms or reached the non-target area were excluded (2.2% of trials in total). We performed ANOVA with *Direction* (i.e., upward, downward, leftward, and rightward arm movements) as a within-participant factor and *Interval* (i.e., image–action and action–rating) as a between-participant factor on the averaged valence rating for emotional images. There was a significant main effect of *Direction* [$F(3, 81) = 3.03$, $p = 0.03$, $\eta_p^2 = 0.10$] but no main effect of *Interval* [$F(1, 27) = 4.17$, $p = 0.05$, $\eta_p^2 = 0.13$] and their interaction [$F(3, 81) = 1.24$, $p = 0.30$, $\eta_p^2 = 0.04$]. *Post hoc* planned comparisons using Tukey's test revealed no differences in valence ratings between leftward and rightward movements in the image–action and action–rating conditions [$t(81) = 1.52$, $p = 0.80$, $d = 0.34$; $t(99) = -0.93$, $p = 0.98$, $d = -0.21$, respectively]. Hence, leftward and rightward conditions were collapsed into the horizontal condition as a baseline index.

Average and individual data for the bias scores are summarized in **Figure 5**. In the image–action condition, upward and downward bias scores did not significantly differ from zero [upward: $t(14) = 1.53$, $p = 0.15$, $d = 0.56$; downward: $t(14) = -1.72$, $p = 0.11$, $d = -0.63$]. In the action–rating condition, there were also no such biases [upward: $t(13) = 1.17$, $p = 0.27$, $d = 0.44$; downward: $t(13) = -0.61$, $p = 0.56$, $d = -0.23$]. To ensure the null effects of vertical movements, we performed the Bayesian one-sample two-tailed t-test (null hypothesis: valence bias score = 0) as in Experiment 1. The null effects in both tasks were supported by weak and substantial evidence for the

FIGURE 4 | Schematic of trials in Experiment 2. In **(A)** the image–action and **(B)** action–rating conditions, an interval of 2 s was inserted before or after the action cue, respectively.
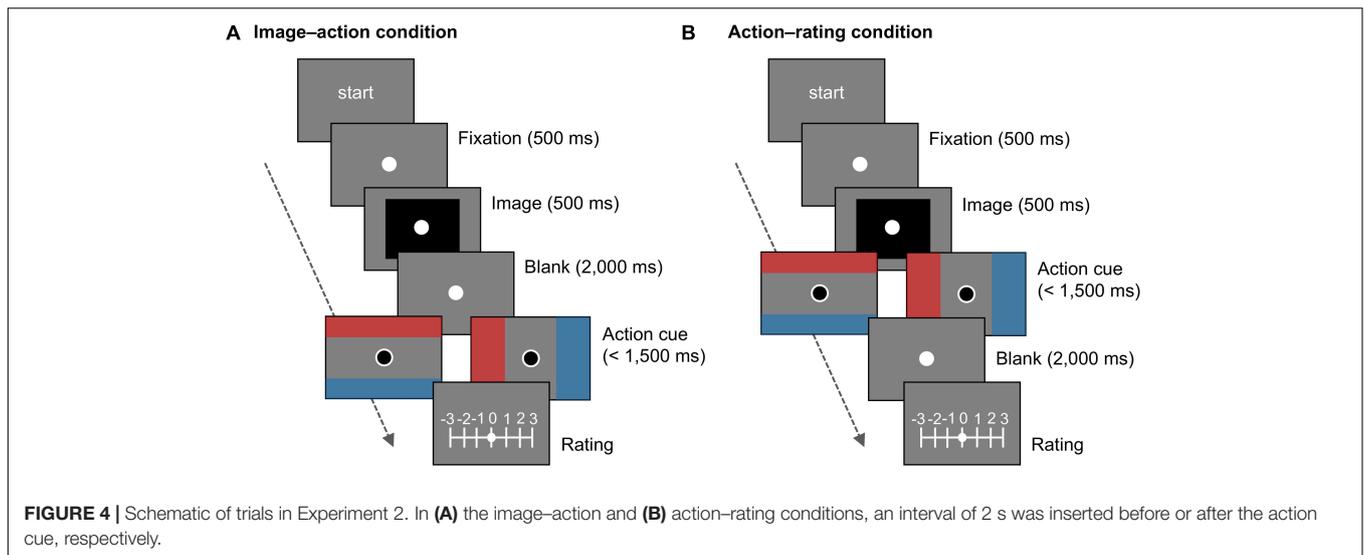
null hypothesis; upward in image–action condition: $BF_{01} = 1.47$; downward in image–action condition: $BF_{01} = 1.15$; upward in action–rating condition: $BF_{01} = 2.08$; downward in action–rating condition: $BF_{01} = 3.14$. Furthermore, ANOVA with the factors of *Direction* (upward, downward) and *Interval* on valence bias scores revealed a main effect of *Direction* $[F(1, 27) = 5.47, p = 0.03, \eta_p^2 = 0.17]$ but no effect of *Interval* $[F(1, 27) = 0.87, p = 0.36, \eta_p^2 = 0.03]$ and their interaction $[F(1, 27) < 0.01, p = 0.95, \eta_p^2 < 0.01]$. *Post hoc* planned comparisons using Bonferroni correction revealed no significant difference between upward and downward movements in the image–action and action–rating conditions $[F(1, 27) = 3.00, p = 0.10, \eta_p^2 = 0.10; F(1, 27) = 2.49, p = 0.13, \eta_p^2 = 0.08$, respectively]. In sum, space–valence metaphorical effect did not emerge in both conditions.

As in Experiment 1, one might doubt the confounding effects of potential outliers resulting in null effects of the vertical movements. We have confirmed that comparable results were obtained from analyses with and without four outliers, leading to the same conclusions (see **Supplementary Figures S9**, **S14–S17** in the **Data Sheet S2** of the **Supplementary Material**).

Discussion

The retrospective effect of space–valence metaphor activated by arm movements did not appear when a 2-s interval was inserted between the emotional image and action and between the action and valence rating. These results are consistent with previous findings (Sasaki et al., 2015) and also extend them by demonstrating that temporal contiguity between emotional image, action, and recollection/evaluation of the image is essential for the retrospective emotional modulation by metaphorical movements.

GENERAL DISCUSSION

The two experiments in this study aimed to extend the findings in Sasaki et al. (2015); the experiments showed results partially

FIGURE 5 | Valence bias score by upward and downward movements in the image–action and action–rating conditions in Experiment 2. Error bars show the standard error of the mean across participants. Open circles represent each participant's data.
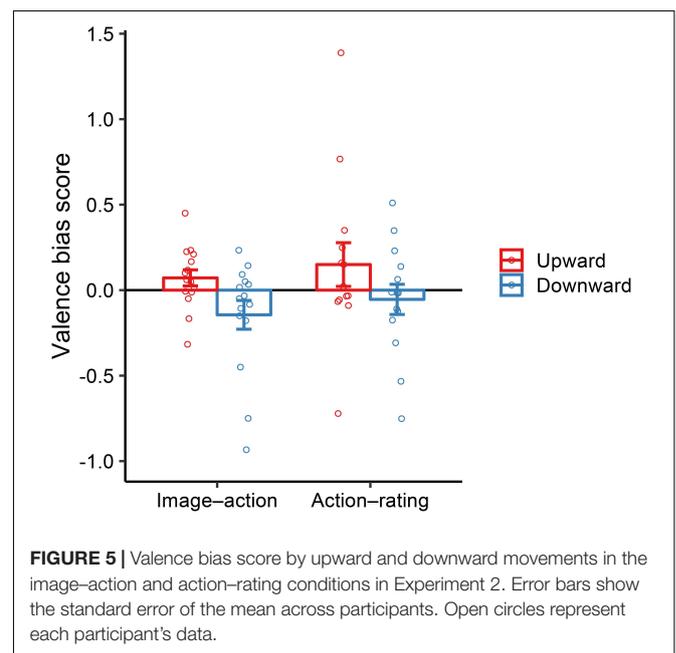
similar to theirs. Experiment 1 suggested that vertical manual movements corresponding to the space–valence metaphor (e.g., down–negative) had retrospective influence on perceived valence of emotional visual stimuli: downward manual movements following visual stimuli modified the perceived emotional valence of the stimuli more negatively. Nevertheless, the influence of the manual movements was observed only for downward movements but not in upward movements, inconsistent with Sasaki et al. (2015). Importantly, we showed that manual movements *preceding* visual stimuli did not modify the perceived emotional valence, suggesting that metaphorical action retrospectively, but not prospectively, alters emotional experience. In Experiment 2, when time intervals of 2 s were inserted between the stimuli and manual movement or between the manual movement

and valence rating, the influence of the vertical manual movements was nullified, suggesting that retrospective emotional modulation requires temporal proximity between emotional stimuli, metaphorical movements, and *post hoc* valence rating.

Retrospective but Not Prospective Effect of Metaphorical Action

Our findings, consistent with Sasaki et al. (2015), showed that vertical manual action corresponding to space–valence metaphor, which was performed after emotional stimulus, affected valence rating. In addition, we showed that this effect was limited to retrospective situation; that is, manual action performed before the stimulus did not affect valence rating. Hence, the manual action corresponding to and activating space–valence metaphor may modulate emotional visual experience retrospectively.

As instances of prospective influences of bodily movements on later perceptual experience, previous studies have shown that visual temporal resolution increases during motor preparation periods (Hagura et al., 2012) and that voluntary movement changes the timing and duration perceptions of later stimulus (Haggard et al., 2002; Park et al., 2003; Imaizumi and Asai, 2017). Furthermore, studies have demonstrated that words meaning vertical space (Ansorge et al., 2013) and/or vertical attentional cueing (Meier and Robinson, 2004; Santiago et al., 2012) prospectively facilitated classification of emotional words with valence metaphorically corresponding to space primed by the preceding words/cues. Thus, one might hypothesize that vertical bodily movements may be able to prospectively modulate later emotional processing. However, our results reject this hypothesis. Methodological differences between the previous and present studies might explain the lack of prospective effect of the space–valence metaphorical correspondence. Manual movements themselves seem to be able to activate the representation of space–valence metaphorical correspondence, according to Sasaki et al. (2015) and the present study, although the effect may be limited to be retrospective. Thus, the difference between metaphorical priming by arm movements, words (Ansorge et al., 2013), and attentional cueing (e.g., Santiago et al., 2012) cannot solely explain the lack of prospective effect in the present study. However, previous experiments have employed speeded discrimination of emotional valence of words (Meier and Robinson, 2004; Santiago et al., 2012; Ansorge et al., 2013), whereas we employed non-speeded rating of valence of images. Therefore, a longer time for non-speeded rating than for speeded discrimination might have decayed the effect of previous manual movements and/or the metaphorical representation activated by them. This could be indirectly supported by the results of Experiment 2, indicating the requirement of temporal proximity between action, stimulus, and rating for the *retrospective* effect.

Retrospective or postdictive (Shimojo, 2014) phenomena have been characterized by low-level visual and tactile processing, such as flash lag (Eagleman and Sejnowski, 2000) and cutaneous rabbit effects (Goldreich, 2007), in which subsequent sensory information overwrites past sensory, perceptual experience. Sasaki et al. (2015) added a new postdictive effect regarding emotional modulation by metaphorical bodily movements, and the present study supported this effect. The mechanisms

underlying this retrospective emotional modulation remain unclear but may be different from those of the above perceptual illusions. At the inferential evaluation stage (e.g., valence rating), metaphorical information activated by bodily movements might be implicitly used for causal inference for past experience (Wegner, 2003) and consequently modulate valence rating.

Temporal Proximity Among Visual Experience, Action, and Evaluation

Experiment 2 examined the conditions necessary to modulate retrospectively past visual emotional experiences by bodily movement corresponding to the space–valence metaphor. Given the absence of prospective effects in Experiment 1, we speculated that, when recalling and evaluating the perceived emotional valence of visual stimulus, manual movement temporally close to the recollection and evaluation might have effects on them but not on the preceding visual experience itself. Indeed, manual movement corresponding to space–valence metaphor, performed simultaneously with recollection, enhances retrieval of emotional memories (Casasanto and Dijkstra, 2010). However, in a condition with temporal interval of 2 s between emotional images and the subsequent vertical manual movements, there was no effect on the perceived valence of the images (Sasaki et al., 2015), suggesting that temporal proximity between manual movements and the subsequent evaluation *per se* is not necessary for the retrospective effect. Thus, Experiment 2 tested the other two possibilities. First, temporally proximate visual and motor information (i.e., stimuli and manual movements) would be bound at the following stage of evaluation (i.e., valence rating), resulting in a biased recollection of the perceived valence of visual stimuli. Second, temporal proximity between stimuli, movements, and evaluation is essential. To investigate these possibilities, temporal proximity between visual stimulus and manual movement or between manual movement and evaluation was manipulated by inserting a temporal interval of 2 s. The results showed that, in both conditions, the influence of vertical manual movements was nullified, supporting the second possibility: metaphorical manual movement retrospectively affects the perceived valence of visual stimuli only when all stimuli, movements, and evaluation are temporally proximate. Nevertheless, it remains unclear which of the temporal proximities, whether that between stimulus and movements or between movements and evaluation, were more crucial. To answer this, future investigation may need to manipulate separately various amounts of temporal delays between visual stimuli, manual movements, and evaluation.

Anisotropy of the Effects of Vertical Movements

Different effects of upward and downward manual movements were suggested. In the retrospect condition of Experiment 1, the effect of manual movements corresponding to space–valence metaphor was induced only by the downward movement (i.e., images were rated as more negative), potentially suggesting a negativity bias (Rozin and Royzman, 2001). Negative events tend to elicit more causal attribution and reasoning in individuals

compared with positive events (Bohner et al., 1988), and negative feedback of one's voluntary action retrospectively distorts time perception more than positive feedback does (Takahata et al., 2012; Yoshie and Haggard, 2013). Such negativity bias may potentially explain our results: only downward movement metaphorically activating negative valence modulates the perceived emotional valence. However, as such negativity bias was not observed in Sasaki et al. (2015), care should be taken when interpreting our results. As the other possible explanation, in upward conditions, the participants moved the joystick in the direction opposite to gravity by raising their hands from the height of their shoulder. Hence, this may have caused difference in mobility between upward and downward movements. If so, difficulty to move upward, not negativity bias, might have canceled out the positive effect of the upward movement. Several studies have reported positivity but not negativity biases, suggesting that the effect of metaphorical correspondences between positive emotional valence and upward location and movement can be stronger than that of negative–downward correspondences (Crawford et al., 2006; Lakens, 2012; Gozli et al., 2013; Lynott and Coventry, 2014; Xie et al., 2015; Damjanovic and Santiago, 2016; Sasaki et al., 2016). For example, positive face presented at the top of a screen can be detected faster than when presented at the bottom, but there was no such metaphor congruency effect for negative face (Damjanovic and Santiago, 2016). In addition, the subsequent manual movement with a joystick is more strongly biased upward by a positive image than downward by a negative image (Sasaki et al., 2016). Further, horizontal saccadic trajectory deviates upward after the observation of a positive word; however, a negative word does not affect the saccade (Gozli et al., 2013). Based on these studies and our results, the effect of space–valence metaphorical correspondence may be task-independent (i.e., perceptual processing, bodily and eye movements), but potentially dependent on movement parameters. We speculate that kinematic characteristics of vertical manual movements and their entailing physical and/or cognitive loads might affect the metaphor congruency effect; consequently, a positivity bias may decay and change to negativity bias in our Experiment 1. Further investigations are needed to explore requirements for the emergence and switching of the two biases.

In our experiments, as in a previous study (Sasaki et al., 2015), the leftward and rightward conditions were regarded as the baseline horizontal condition, in which the effect of space–valence metaphor does not appear. However, the space corresponding to one's dominant hand (e.g., right for right-handers) and the stimulus presented there are felt and considered as more positive than the opposite side (Casasanto, 2009; de la Vega et al., 2013; Marmolejo-Ramos et al., 2013). Hence, our participants (all right-handed) may have rated the rightward condition more positively compared with the leftward condition. Furthermore, as positive values were displayed on the right side of the valence rating scale, the rightward manual movement might have primed the participants to attend rightward (Corbetta and Shulman, 2002), consequently causing bias to the participants' responses toward positive (right-sided) values, and vice versa. However, our results indicated no difference in valence rating

between the rightward and leftward conditions, suggesting that biases attributable to hand-dominance and priming by the movement–scale correspondence were not strong enough to alter the valence rating, and this was consistent with the previous study (Sasaki et al., 2015). Another recent study has also shown no effect of the horizontal location of a visual stimulus on emotional processing (Xie et al., 2015). Nevertheless, we cannot rule out the potential, selective effect of horizontal movements on stimuli with corresponding emotional valence (e.g., rightward movement on positive stimuli), although our experimental design with its relatively small number of trials may be insufficient to statistically test this possibility by making comparisons between emotional image categories. Moreover, a few participants in our study reported having slight difficulty moving the joystick rightward. This difficulty might also have canceled out the potential effects of the rightward movements. Therefore, detailed future studies are required to elucidate not only the "anisotropy" of the metaphorical effects of vertical and horizontal bodily movements on emotional processing but also the potential effects of mobility, gravity which affects visuomanual processing (Scotto Di Cesare et al., 2014), and their accompanying physical loads.

CONCLUSION

This study suggests that vertical bodily movement corresponding to space–valence metaphor (e.g., down–negative) retrospectively, but not prospectively, alters the perceived emotional valence of visual stimuli. This effect requires temporal proximity between the stimuli, bodily movement, and evaluation. Given the modulation only by downward movement found in Experiment 1, mechanisms underlying the potential anisotropy in movement direction and/or space–valence metaphor should be investigated in future studies. Finally, examining the modulation of emotional processing by bodily movement in affective disorders, such as alexithymia (Taylor, 2000), might be a fruitful research direction for clinical application.

AUTHOR CONTRIBUTIONS

TK, SI, and YT conceived the study and wrote the manuscript. TK and SI performed the experiments and analyzed the data. All authors approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01927/full#supplementary-material

REFERENCES

Ansorge, U., Khalid, S., and Konig, P. (2013). Space-valence priming with subliminal and supraliminal words. *Front. Psychol.* 4:81. doi: 10.3389/fpsyg.2013.00081

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Bohner, G., Bless, H., Schwarz, N., and Strack, F. (1988). What triggers causal attributions? The impact of valence and subjective probability. *Eur. J. Soc. Psychol.* 18, 335–345. doi: 10.1002/ejsp.2420180404

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right- and left-handers. *J. Exp. Psychol. Gen.* 138, 351–367. doi: 10.1037/a0015854

Casasanto, D., and Dijkstra, K. (2010). Motor action and emotional memory. *Cognition* 115, 179–185. doi: 10.1016/j.cognition.2009.11.002

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Crawford, L. E., Margolies, S. M., Drake, J. T., and Murphy, M. E. (2006). Affect biases memory of location: evidence for the spatial representation of affect. *Cogn. Emot.* 20, 1153–1169. doi: 10.1080/02699930500347794

Damjanovic, L., and Santiago, J. (2016). Contrasting vertical and horizontal representations of affect in emotional visual search. *Psychon. Bull. Rev.* 23, 62–73. doi: 10.3758/s13423-015-0884-6

de la Vega, I., Dudschig, C., De Filippis, M., Lachmair, M., and Kaup, B. (2013). Keep your hands crossed: the valence-by-left/right interaction is related to hand, not side, in an incongruent hand-response key assignment. *Acta Psychol.* 142, 273–277. doi: 10.1016/j.actpsy.2012.12.011

Eagleman, D. M., and Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science* 287, 2036–2038. doi: 10.1126/science.287.5460.2036

Eagleman, D. M., and Sejnowski, T. J. (2007). Motion signals bias localization judgments: a unified explanation for the flash-lag, flash-drag, flash-jump, and Frohlich illusions. *J. Vis.* 7:3. doi: 10.1167/7.4.3

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Goldreich, D. (2007). A Bayesian perceptual model replicates the cutaneous rabbit and other tactile spatiotemporal illusions. *PLoS One* 2:e333. doi: 10.1371/journal.pone.0000333

Gozli, D. G., Chow, A., Chasteen, A. L., and Pratt, J. (2013). Valence and vertical space: saccade trajectory deviations reveal metaphorical spatial activation. *Vis. Cogn.* 21, 628–646. doi: 10.1080/13506285.2013.815680

Haggard, P., Clark, S., and Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nat. Neurosci.* 5, 382–385. doi: 10.1038/nn827

Hagura, N., Kanai, R., Orgs, G., and Haggard, P. (2012). Ready steady slow: action preparation slows the subjective passage of time. *Proc. Biol. Sci.* 279, 4399–4406. doi: 10.1098/rspb.2012.1339

Ijzerman, H., and Semin, G. R. (2009). The thermometer of social relations: mapping social proximity on temperature. *Psychol. Sci.* 20, 1214–1220. doi: 10.1111/j.1467-9280.2009.02434.x

Imaizumi, S., and Asai, T. (2017). My action lasts longer: potential link between subjective time and agency during voluntary action. *Conscious. Cogn.* 51, 243–257. doi: 10.1016/j.concog.2017.04.006

JASP Team (2018). *JASP (Version 0.8.6) [Computer software] [Online]*. Available at: https://jasp-stats.org [accessed August 29, 2018].

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press.

Jostmann, N. B., Lakens, D., and Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychol. Sci.* 20, 1169–1174. doi: 10.1111/j.1467-9280.2009.02426.x

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception* 36, 14–14. doi: 10.1177/03010066070360s101

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: structural overlap predicts categorization times for bipolar concepts presented in vertical space. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 726–736. doi: 10.1037/a0024955

Landau, M. J., Meier, B. P., and Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychol. Bull.* 136, 1045–1067. doi: 10.1037/a0020970

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual.* Technical Report A–8 (Gainesville, FL: University of Florida).

Lee, S. W. S. (2016). Multimodal priming of abstract constructs. *Curr. Opin. Psychol.* 12, 37–44. doi: 10.1016/j.copsyc.2016.04.016

Lynott, D., and Coventry, K. (2014). On the ups and downs of emotion: testing between conceptual-metaphor and polarity accounts of emotional valence-spatial location interactions. *Psychon. Bull. Rev.* 21, 218–226. doi: 10.3758/s13423-013-0481-5

Marmolejo-Ramos, F., Elosua, M. R., Yamada, Y., Hamm, N. F., and Noguchi, K. (2013). Appraisal of space words and allocation of emotion words in bodily space. *PLoS One* 8:e81688. doi: 10.1371/journal.pone.0081688

Meier, B. P., and Robinson, M. D. (2004). Why the sunny side is up: association between affect and vertical position. *Psychol. Sci.* 15, 243–247. doi: 10.1111/j.0956-7976.2004.00659.x

Niedenthal, P. M. (2007). Embodying emotion. *Science* 316, 1002–1005. doi: 10.1126/science.1136930

Park, J., Schlag-Rey, M., and Schlag, J. (2003). Voluntary action expands perceived duration of its sensory consequence. *Exp. Brain Res.* 149, 527–529. doi: 10.1007/s00221-003-1376-x

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897x00366

Rozin, P., and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Pers. Soc. Psychol. Rev.* 5, 296–320. doi: 10.1207/S15327957pspr0504_2

Santiago, J., Ouellet, M., Roman, A., and Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cogn. Sci.* 36, 1051–1077. doi: 10.1111/j.1551-6709.2012.01240.x

Sasaki, K., Yamada, Y., and Miura, K. (2015). Post-determined emotion: motor action retrospectively modulates emotional valence of visual images. *Proc. Biol. Sci.* 282:20140690. doi: 10.1098/rspb.2014.0690

Sasaki, K., Yamada, Y., and Miura, K. (2016). Emotion biases voluntary vertical action only with visible cues. *Acta Psychol.* 163, 97–106. doi: 10.1016/j.actpsy.2015.11.003

Scotto Di Cesare, C., Sarlegna, F. R., Bourdin, C., Mestre, D. R., and Bringoux, L. (2014). Combined influence of visual scene and body tilt on arm pointing movements: gravity matters. *PLoS One* 9:e99866. doi: 10.1371/journal.pone.0099866

Seno, T., Kawabe, T., Ito, H., and Sunaga, S. (2013). Vection modulates emotional valence of autobiographical episodic memories. *Cognition* 126, 115–120. doi: 10.1016/j.cognition.2012.08.009

Shimojo, S. (2014). Postdiction: its implications on visual awareness, hindsight, and sense of agency. *Front. Psychol.* 5:196. doi: 10.3389/fpsyg.2014.00196

Takahata, K., Takahashi, H., Maeda, T., Umeda, S., Suhara, T., Mimura, M., et al. (2012). It's not my fault: postdictive modulation of intentional binding by monetary gains and losses. *PLoS One* 7:e53421. doi: 10.1371/journal.pone.0053421

Taylor, G. J. (2000). Recent developments in alexithymia theory and research. *Can. J. Psychiatry.* 45, 134–142. doi: 10.1177/070674370004500203

Wegner, D. M. (2003). The mind's best trick: how we experience conscious will. *Trends Cogn. Sci.* 7, 65–69. doi: 10.1016/S1364-6613(03)00002-0

Xie, J., Huang, Y., Wang, R., and Liu, W. (2015). Affective valence facilitates spatial detection on vertical axis: shorter time strengthens effect. *Front. Psychol.* 6:277. doi: 10.3389/fpsyg.2015.00277

Yoshie, M., and Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Curr. Biol.* 23, 2028–2032. doi: 10.1016/j.cub.2013.08.034

The Role of Pattern Extrapolation in the Perception of Dynamic Facial Expressions in Autism Spectrum Disorder

Letizia Palumbo[1]*, Sylwia T. Macinska[2] and Tjeerd Jellema[2]

[1] Department of Psychology, Liverpool Hope University, Liverpool, United Kingdom, [2] Psychology, School of Life Sciences, University of Hull, Hull, United Kingdom

Changes in the intensity and type of facial expressions reflect alterations in the emotional state of the agent. Such "direct" access to the other's affective state might, top-down, influence the perception of the facial expressions that gave rise to the affective state inference. Previously, we described a perceptual bias occurring when the last, neutral, expression of offsets of facial expressions (joy-to-neutral and anger-to-neutral), was evaluated. Individuals with high-functioning autism (HFA) and matched typically developed (TD) individuals rated the neutral expression at the end of the joy-offset videos as slightly angry and the identical neutral expression at the end of the anger-offset videos as slightly happy ("overshoot" bias). That study suggested that the perceptual overshoot response bias in the TD group could be best explained by top-down "emotional anticipation," i.e., the involuntary/automatic anticipation of the agent's next emotional state of mind, generated by the immediately preceding perceptual history (low-level mind reading). The experimental manipulations further indicated that in the HFA group the "overshoot" was better explained by contrast effects between the first and last facial expressions, both presented for a relatively long period of 400 ms. However, in principle, there is another, more parsimonious, explanation, which is pattern extrapolation or representational momentum (RM): the extrapolation of a pattern present in the dynamic sequence. This hypothesis is tested in the current study, in which 18 individuals with HFA and a matched control group took part. In a base-line condition, joy-offset and anger-offset video-clips were presented. In the new experimental condition, the clips were modified so as to create an offset-onset-offset pattern within each sequence (joy-to-anger-to-neutral and anger-to-joy-to-neutral). The final neutral expressions had to be evaluated. The overshoot bias was confirmed in the base-line condition for both TD and HFA groups, while the experimental manipulation removed the bias in both groups. This outcome ruled out pattern extrapolation or RM as explanation for the perceptual "overshoot" bias in the HFA group and suggested a role for facial contrast effects in HFA. This is compatible with the view that ASD individuals tend to lack the spontaneous "tracking" of changes in the others' affective state and hence show no or reduced emotional anticipation.

Keywords: dynamic facial expressions, perceptual distortions, pattern extrapolation, emotional anticipation, embodied simulation

INTRODUCTION

The dynamic expressivity of the face greatly facilitates social communication. Very subtle changes in facial expressivity can be detected and may reflect subtle positive or negative alterations in the affective state of the agent (Krumhuber et al., 2013). The ability to detect such emotional state alterations over time enables us to make predictions about other people's behavior. Typically, we read facial expressions without explicit intention to do so or without inferential efforts. This ability to tacitly understand others' mental states has been referred to as low-level mind reading (Goldman, 2006). Its implicit (automatic, non-volitional) nature can be contrasted to the deliberate, effortful, use of cognitive resources or conceptual and linguistic mediators, involved in explicit Theory of Mind (Baron-Cohen, 1995), which is referred to by Goldman as high-level mind reading (Goldman, 2006).

There is growing evidence that these implicit, involuntary or spontaneous, skills for reading others' emotional or mental states are compromised in Autism Spectrum Disorder (ASD; Hudson et al., 2009, 2012; Jellema et al., 2009; Senju et al., 2009), and possibly also in other disorders like schizophrenia (Van 't Wout et al., 2009) and OCD (Obsessive compulsive disorder; Tumkaya et al., 2014), resulting in inadequate social exchanges. ASD is a pervasive neurodevelopment condition characterized by impaired social development and stereotypical, repetitive behaviors, often associated with obsessive interests and a lack of empathy (Rutter, 1978; World Health Organization [WHO], 2008; DSM-V, 2013). Symptom severity varies hugely in ASD (DSM-V, 2013). High-functioning autism (HFA) is a relatively mild form of ASD with normal intelligence quotient (IQ) distribution, but with a delayed development of language skills and difficulties in social and emotional domains (McPartland and Klin, 2006).

The perceptual processing stage of social cues, such as facial expressions, gaze direction, bodily postures, and action sequences, provides a mechanistic description of these cues, grounded on physical or geometrical features and dynamics of the stimuli, possibly in relation to contextual cues and objects in the environment (Jellema and Perrett, 2003, 2006, 2012). These mechanistic descriptions next trigger inferences about the emotional/mental state of the agent (e.g., Blakemore and Decety, 2001). Besides this bottom-up route (from perception of bodily cues to attribution of social meaning), there is also a top-down route where attributions of other people's mental states (such as intentions) can in turn influence the low-level perception of bodily cues (Hudson et al., 2009; Teufel et al., 2009). These top-down processes can be highly inferential or reflective, but they can also be quite reflexive and automatic (Satpute and Lieberman, 2006; Lieberman, 2007). The bi-directional interaction between bottom-up and top-down streams has been captured under the term "perceptual mentalizing" (Teufel et al., 2010). This model, however, only considers explicit mental attributions, neglecting the possibility that implicit attributions might also influence social perception. Hudson et al. (2009)

reported a study where participants' estimations of how far an agent's head had rotated were influenced by the agent's gaze direction. With gaze direction ahead of head rotation the head rotation was overestimated as compared to when the gaze was lagging behind head rotation. Importantly, participants were not aware of the eye gaze manipulation. The bias thus seemed to be induced by implicit attributions of the intention to continue/discontinue to move in the direction of the head rotation. This study therefore supports the idea that early stages of the visual processing of social stimuli can be influenced by implicit attributions made by the observer about the agent's mental state.

In individuals with ASD an impaired top-down route would result in relatively unbiased perception of bodily cues, not "contaminated" by attributions of mental/emotional states (cf. Wang and Hamilton, 2012). However, at the same time it might make them more susceptible to perceptual illusions driven by low-level perceptual features, such as geometries or patterns present in the stimuli. In this respect, it is interesting to note that the perceptual bias reported in TD individuals in Hudson et al. (2009) was also found in individuals with ASD (Hudson and Jellema, 2011). However, in contrast to the TD group, the ASD group continued to show this bias in response to a non-social stimulus designed to match the low-level physical characteristics (eye gaze) of the agent stimulus. These results suggest that individuals with ASD fail to grasp the mental states in an involuntary manner and instead rely on the mechanistic descriptions of the physical features of the social cues (such as directional cues, input-output relations, or statistical regularities).

Emotional Anticipation in TD and HFA

In a series of studies, we presented a phenomenon occurring when evaluations have to be made of dynamic offsets of facial expressions (Jellema et al., 2011; Palumbo and Jellema, 2013; Palumbo et al., 2015). TD participants observed dynamic presentations of an intense facial expression of joy or anger, which gradually weakened until the actor posed a neutral expression. Participants' task was to evaluate the last neutral frame on a 5-point Likert scale, ranging from slightly angry (1), via neutral (3), to slightly happy (5). Results showed a perceptual bias (which we call "overshoot" bias), such that the neutral expression at the end of the joy-to-neutral videos was evaluated as slightly angry, and the identical neutral expression at the end of anger-to-neutral videos as slightly happy (in the remainder of the text we refer to this condition as the Offset condition). We proposed that the perceptual history led the observer to automatically anticipate what the emotional state of the agent would be *after* the sequence stopped. The "emotional anticipation" is thought to drive, in a top-down fashion, the perceptual bias. This interpretation fits with the "perceptual mentalizing" model proposed by Teufel et al. (2010). However, as emotional anticipation acted involuntarily', it highlights the role of implicit attributions in social perception.

In subsequent studies, we found that participants with HFA also reported the perceptual overshoot bias

(Palumbo et al., 2015). However, when in an additional condition we changed the identity of the agent in the last frame of the video-clips (the new identity was unfamiliar to the observer), the influence of the perceptual history was nullified in the TD group (the overshoot bias was removed), while the HFA group continued to report an overshoot bias. This suggested that the perceptual distortion found in the TD group was not due to sequential contrast effects (Tanaka-Matsumi et al., 1995) as the degree of expressive contrast between the first and last frames of the videos remained unaffected by the identity change manipulation. The removal of the overshoot bias was compatible with the emotional anticipation hypothesis. Actor B in the last frame was someone for whom no perceptual history was available, so the observer did not know anything about B's emotional state other than that B had a neutral expression, and therefore rated B as neutral. The finding that the HFA group continued to show an overshoot bias suggested that they had not used an anticipation mechanism linked to the agent (we established that they did detect the change in identity). We hypothesized that the persistence of the overshoot bias in the HFA group might have resulted from susceptibility to low-level stimulus features, most probably the contrast between the first (happy or angry) and the last (neutral) expression (both presented for a relatively long period of 400 ms), which is not affected by the identity change (Palumbo et al., 2015).

For the TD group, the results of the identity-change condition also seemed to rule out an explanation in terms of representational momentum (RM; Freyd and Finke, 1984; Yoshikawa and Sato, 2008), or at least suggest that in TD individuals RM can be modulated, or even overruled, by top-down information (such as information referring to the agent's identity). RM is the phenomenon that an observer's memory for the final position of a moving target is displaced further along the observed trajectory (Freyd and Finke, 1984), which also applies to the gradual changes in dynamic facial expressions (Yoshikawa and Sato, 2008). However, also for the HFA group, it is in principle possible that RM, rather than sequential contrast effects, could explain the response bias, as they continued to report the bias in the identity-change condition. These experiments therefore did not allow to discriminate between these two competing low-level explanations in the HFA group. Another condition (Palumbo et al., 2015) in which video-clips started with a neutral expression that morphed via happy (or angry) back to neutral (forming a "loop") did not produce a bias in the evaluations of the HFA group. However, as the extrapolation direction in this condition is ambiguous (in which direction does the pattern continue?), it cannot be used to exclude RM as the underpinning mechanism.

Representational momentum at work in the base-line condition in HFA individuals would mean that the negative going trend (happy offset) would be extrapolated into a slightly angry expression, and that the positive going trend (angry-offset) would be extrapolated into a slightly happy expression. Individuals with HFA tend to be adept at detecting regularities, input-output relations or statistical regularities, which typically govern the physical world (Baron-Cohen, 2002; Baron-Cohen et al., 2003).

In individuals with HFA, this tendency for low-level pattern detection and extrapolation may be quite prominent and may not easily get overruled by top-down information relating to the object, such as information that the agent's identity had changed (cf. Vivanti et al., 2011).

The Current Study

The current study aimed to clarify what drove the perceptual bias in the HFA group in our previous experiments, specifically targeting the role of extrapolation (or RM) of patterns present in the dynamic facial expressions. To this end video-clips were created in which an intense facial expression (happy or angry) gradually morphed via a neutral expression into its "opposite" expression (angry or happy), after which it morphed back to neutral (joy-to-anger-to-neutral and anger-to-joy-to-neutral sequences). The final neutral expressions of the videos were again evaluated using the 5-point Likert scale. The rationale was that if the overshoot effect is driven by pattern extrapolation (or RM) then the last, neutral, expression should be evaluated as slightly happy in the joy-to-anger-to-neutral videos, and as slightly angry in the anger-to-joy-to-neutral videos. In other words, observers would implicitly expect the pattern to continue. Pattern extrapolation and RM predict the same outcome in this paradigm: a slightly happy overshoot for the joy-to-anger-to-neutral videos, and a slightly angry overshoot for the anger-to-joy-to-neutral videos. Further, if the evaluations in the TD group would be driven by emotional anticipation (as suggested by Palumbo and Jellema, 2013), then we would predict the absence of a response bias in this new condition in the TD group, as in terms of "tracking and anticipating" the agent's emotional state of mind, the videos would, if anything, suggest the agents to remain emotionally neutral after the clip stopped.

MATERIALS AND METHODS

Participants
HFA Group
Twenty-one individuals with HFA participated in the experiment. All were recruited through disability services from universities in the North-East of England (United Kingdom).

They all had previously received a diagnosis of HFA or Asperger's syndrome from a clinical psychologist or psychiatrist based on DSM-IV-TR (American Psychiatric Association [APA], 2013) or ICD-10 (World Health Organization [WHO], 2008) criteria. Diagnosis of HFA was confirmed using the ADOS (Autism Diagnostic Observation Schedule, module 4), administered by a qualified experimenter (SM). The ADOS is a semi-structured, standardized assessment of communication, social interaction, and imagination, designed for use with children and adults suspected of having ASD. They also completed the Autism Spectrum Quotient questionnaire (AQ; Baron-Cohen et al., 2001), which is a fifty-statement, self-administered questionnaire, designed to measure the degree to which an adult with normal intelligence possesses autistic-like traits. IQ scores were determined using the

Wechsler Adult Intelligence Scale, WAIS-III (Wechsler, 1997). From the 21 students with HFA participating in the study, three were removed following the application of exclusion criteria to the data set (see Data reduction below for details). The remaining 18 students (6 females, 12 males; mean age = 19.9 years, SD = 1.1) had a mean total ADOS score of 9.3 (SD = 2.6) and a mean AQ score of 32.3 (SD = 9.5). Their mean total IQ score was 117.7 (SD = 8.1).

TD Group

All TD participants were undergraduate Psychology students from Hull University. All were asked if they had previously obtained a head injury or had received a diagnosis of ASD or another mental health or developmental disorder. No participants disclosed this. Twenty-two TD individuals took part in the study; applying the exclusion criteria to the data set (see below) removed four individuals. The remaining 18 participants (5 females, 13 males; mean age = 20.5 years, SD = 1.4) had a mean AQ score of 17.6 (SD = 3.9) and a mean total IQ score of 114 (SD = 6.7). The TD group did not differ from the HFA group in terms of age [$t(34) = 1.43$, $p = 0.163$], gender ratio [$X^2(1,35) = 0.13$, $p = 0.72$], or IQ [$t(34) = -1.56$, $p = 0.13$]. As expected, AQ scores were significantly higher in the HFA group [$t(34) = 7.59$, $p < 0.001$]. Importantly, the HFA group matched very closely to the control TD group, as both groups consisted of university students with fairly similar daily routines, resulting in a good approximation of the influence of the factor "HFA." All HFA and TD participants had normal or corrected-to-normal vision, and provided written informed consent prior to the experiment. Participants received course credits or a fee for taking part. The study was approved by the Ethics committee of the Department of Psychology of Hull University.

Stimuli

The stimuli used in the current experiment were similar to those used in Palumbo and Jellema (2013) and in Palumbo et al. (2015). Pictures of facial expressions of joy and anger were selected from the Pictures of Facial Affect (eight actors, four males: EM, JJ, PE, WF, and four females: C, MO, PF, SW) (Ekman and Friesen, 1976; Young et al., 2002). All faces were frontally oriented with their eye gaze directed straight ahead. The photographs were in grayscale. The pictures were digitally adjusted to match in contrast and brightness. The eyes of all actors were positioned at approximately the same screen location. Faces measured about 13 × 20 cm when displayed on the screen, subtending approximately 8° vertically. Nine interpolated images, in between the full-blown expression of joy or anger (which is called 100%) and the neutral expression (0%) were created at equal steps of 10% intensity change, using computer morphing procedures (Perrett et al., 1994). In the Offset condition, the morph sequences depicted a maximally happy or angry expression of which the intensity gradually decreased until a neutral expression was reached (joy-to-neutral or anger-to-neutral). The first and last frames of the sequences were displayed for 400 ms. The duration of the morph sequence

was 270 ms (9 frames × 30 ms), the total duration of the stimulus presentation was 1070 ms. In the new condition, the initial full-blown facial expression (happy or angry) gradually morphed smoothly via a neutral expression into its "opposite" expression, after which it morphed back to neutral (joy-to-anger-to-neutral and anger-to-joy-to-neutral sequences). We will refer to this manipulation as the offset-onset-offset condition. The first and the last frames again both lasted 400 ms. The duration of these morph sequences was 870 ms (29 frames × 30 ms), the total duration of the stimulus presentation was 1940 ms. An illustration of the morph sequences in both conditions is presented in **Figure 1**.

FIGURE 1 | Illustration of stimulus presentations. Shown are the Offset condition **(A)** and the Offset-onset-offset condition **(B)**, for joy and anger initial emotions. Face pictures are shown in Palumbo and Jellema (2013).
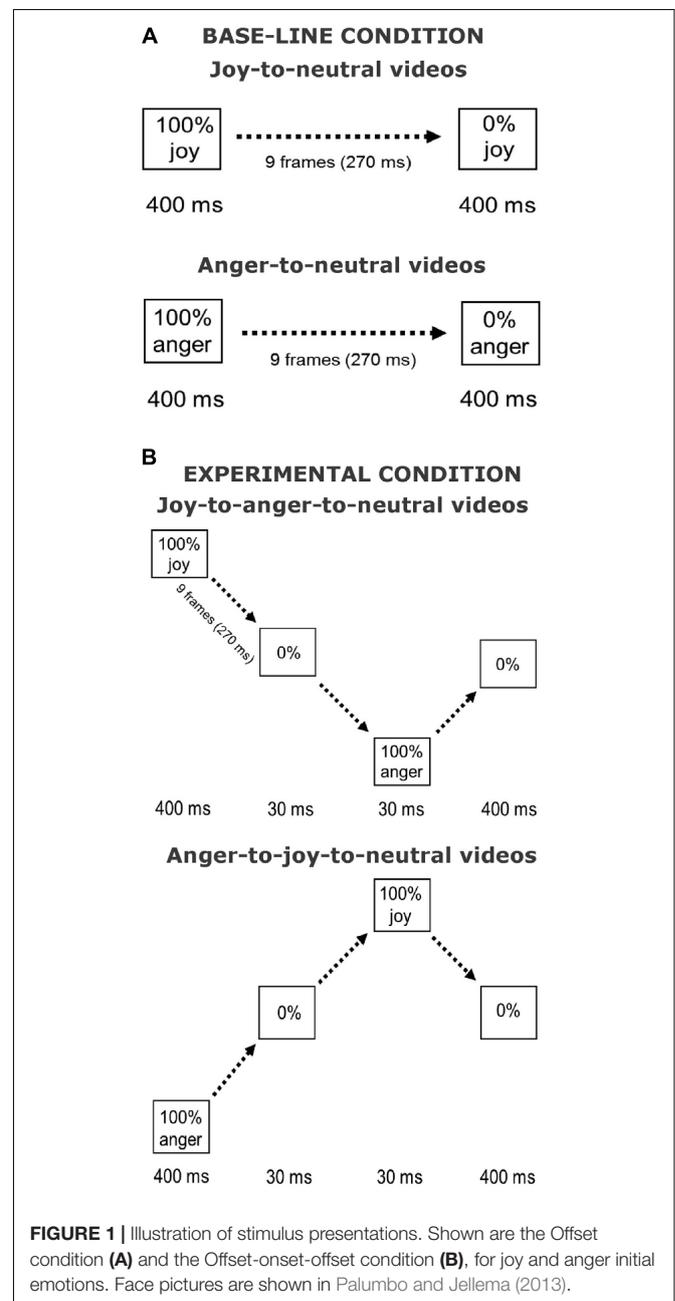
Experimental Procedure

Participants were seated at a viewing distance of 80 cm from a PC screen (17-inch monitor, 1024 × 768 pixels, 100 Hz). The stimuli were presented using E-Prime (v. 1.2; Psychology Software Tools, Inc.). The software uploaded each single frame at specific durations as illustrated in **Figure 1**. This generated smooth morph sequences that resembled short video clips. First participants completed a calibration phase in which they rated the static neutral expressions of the eight actors (i.e., neutral expressions according to the ratings from Ekman and Friesen, 1976). Each calibration trial started with a fixation cross displayed in the center of the screen for 500 ms, followed by the static neutral face displayed for 600 ms. Sixteen calibration trials were presented (eight actors, two repetitions each) in randomized order. Participants were prompted to rate these "neutral" expressions using a 5-point scale, ranging from slightly angry (1) via neutral (3) to slightly happy (5), by pressing one of the five labeled keys on a button box (SR-Box, Psychology Software Tools, Inc., United States). Directly following the calibration phase, the experimental session started. First, 6 practice trials were completed (displaying two actors not used in the experiment), followed by 64 randomized experimental trials (8 actors × 2 expressions × 2 conditions × 2 repetitions). Each trial started with a fixation cross displayed for 500 ms, followed by the video-clips. Participants were prompted to rate the last neutral expression of the sequence using the same 5-point scale, and were instructed to respond within 3 s.

RESULTS

Calibration

The mean calibration scores for the neutral expression for each of the eight stimulus actors, obtained at the start of the experiment of each TD and HFA participant, are shown in **Figure 2**. The TD and HFA groups reported very similar scores, with the neutral expression of actors C and WF consistently rated as slightly angry. These calibration scores were used to adjust the scores in the subsequent experimental trials on an individual participant basis for each actor: a calibration factor (equal to 3.00 minus the calibration score) was added to the experimental scores. All statistical analyses were performed on the calibrated scores. The finding that the HFA and TD groups produced very similar evaluations of the "neutral" expressions of the eight actors, and in particular that all individuals of both groups consistently rated actors C and WF as slightly angry (**Figure 2**), indicates that HFA individuals did not show anomalies in processing subtle differences in these facial expressions. These results also mirror those in our previous studies (Palumbo et al., 2015).

Data Reduction and Analysis

Trials in which RTs were below 250 ms or above 3000 ms were considered outliers and were removed (HFA, 10.4%; TD, 3.5%). Participants were excluded if more than 25% of their RT values fell outside the above range (HFA, $n = 2$; TD, $n = 0$) and when they pressed the same key for more than 90% of trials (HFA, $n = 0$; TD, $n = 2$). A ± 2.5 SD rule was applied to the mean
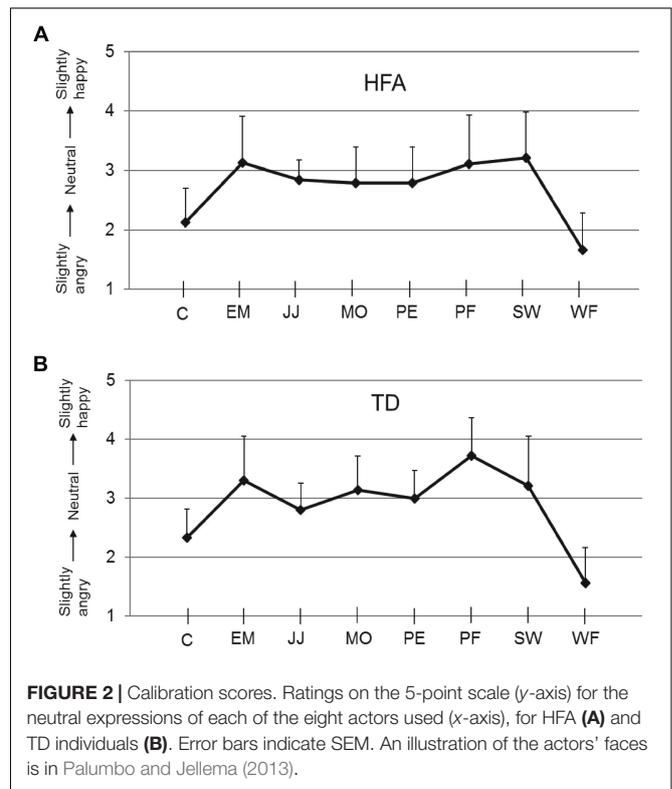
FIGURE 2 | Calibration scores. Ratings on the 5-point scale (y-axis) for the neutral expressions of each of the eight actors used (x-axis), for HFA **(A)** and TD individuals **(B)**. Error bars indicate SEM. An illustration of the actors' faces is in Palumbo and Jellema (2013).

difference of the ratings per participant, i.e., rating in the Anger-to-neutral condition minus rating in the Joy-to-neutral condition (HFA, $n = 1$; TD, $n = 2$).

Following application of these exclusion criteria, the data of 18 TD individuals and 18 HFA individuals was analyzed with a 2 × 2 × 2 repeated measures ANOVA, with Offset history (Offset vs. Offset-onset-offset) and Initial emotion (Joy vs. Anger) as within-subject factors, and Group (HFA vs. TD) as between-subjects factor. The main effects of Offset history [$F(1,34) = 2.14$, $p = 0.15$, $\eta_p^2 = 0.06$] and Group [$F(1,34) = 0.04$, $p = 0.85$, $\eta_p^2 = 0.00$] were not significant, while the main effect of the factor Initial emotion was highly significant [$F(1,34) = 40.44$, $p < 0.001$, $\eta_p^2 = 0.54$], reflecting that the evaluations of the final neutral expressions were significantly different when the initial emotion was anger as compared to joy. Importantly, the interaction of Offset history by Initial Emotion was significant [$F(1,34) = 6.96$, $p = 0.01$, $\eta_p^2 = 0.17$]. *Post hoc* analyses showed the overshoot bias to be more pronounced in the Offset condition (Joy-to-neutral: $M = 2.83$, SD = 0.05; Anger-to-neutral: $M = 3.19$, SD = 0.04; paired samples t-test: $t(35) = -5.97$, $p < 0.001$) than in the Offset-onset-offset condition (Joy-to-anger-to-neutral: $M = 3.01$, SD = 0.06; Anger-to-joy-to-neutral: $M = 3.10$, SD = 0.05; paired samples t-test: $t(35) = -1.31$, $p = 0.20$). The interactions Offset history by Group [$F(1,34) = 0.05$, $p = 0.82$, $\eta_p^2 = 0.00$] and Initial emotion by Group [$F(1,34) = 2.26$, $p = 0.14$, $\eta_p^2 = 0.06$] were not significant, nor was the 3-way interaction [$F(1,34) = 1.12$, $p = 0.30$, $\eta_p^2 = 0.03$]. Thus, the TD and HFA groups responded in a very similar fashion in both conditions, with a significant overshoot response bias in the Offset condition and an absence of

a response bias in the Offset-onset-offset condition. The results are shown in **Figure 3** (to illustrate consistency in these effects across the eight different actors, group means separated per actor can be found in **Figure A1**).

As in the calibration phase the "neutral" expressions of actors C and WF were consistently evaluated as slightly angry, whereas the other six actors were consistently evaluated as fairly neutral (see **Figure 2**), we conducted the same analyses on the data from just these six actors (i.e., excluding C and WF). This, however,

resulted in the same outcome as was obtained for all eight actors.

DISCUSSION

The current study examined whether pattern extrapolation might give rise to distortions in the perception of dynamic facial expressions in individuals with HFA. Pattern extrapolation refers
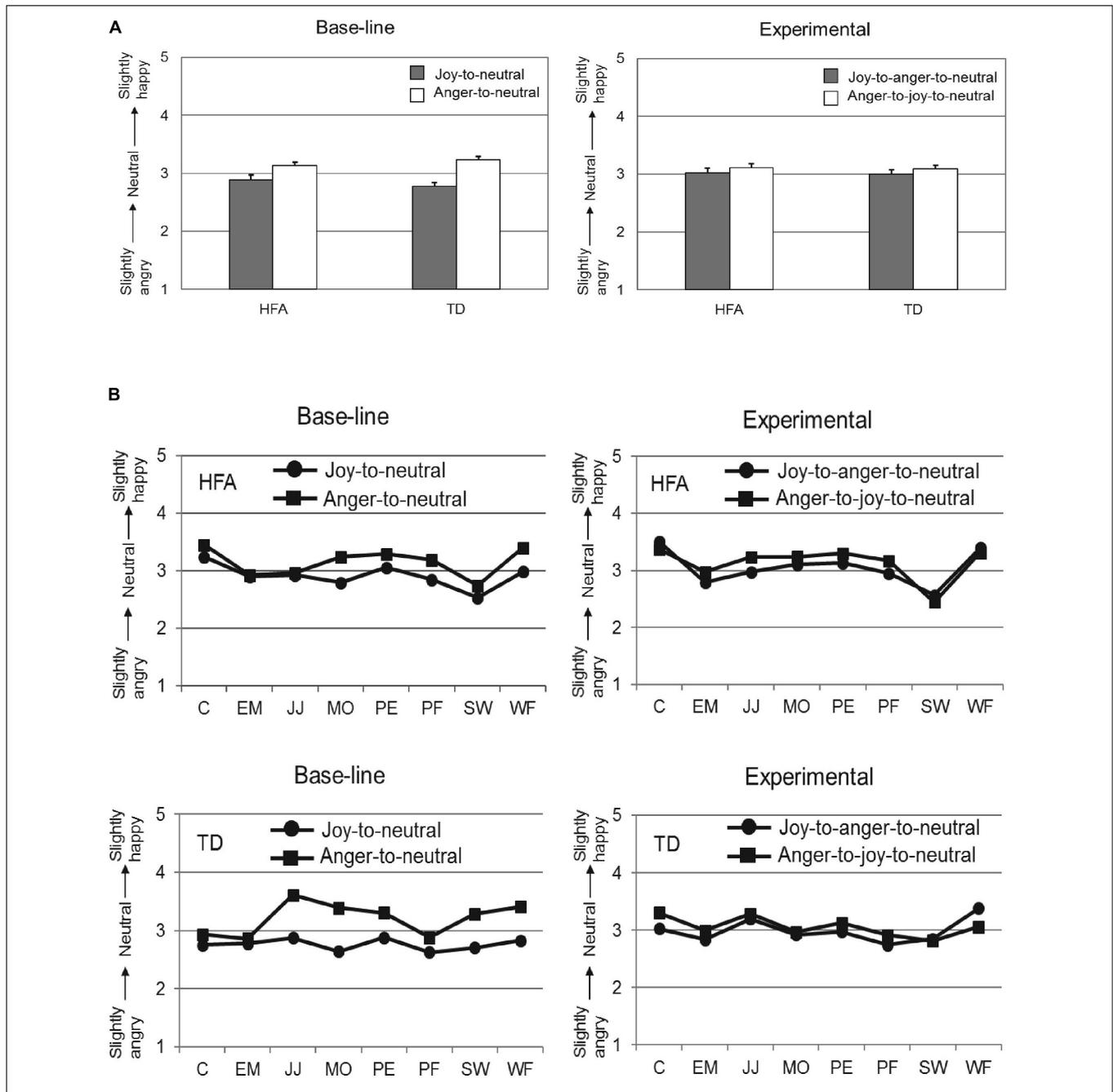
FIGURE 3 | Mean ratings on the 5-point scale (y-axis) for the neutral expressions at the end of the video-sequences for the HFA and TD groups (x-axis) in the Offset condition **(A)** and in the Offset-onset-offset condition **(B)**. Error bars indicate SEM. Significant results are indicated with ***$p < 0.001$.

to the human tendency to detect patterns in presented stimuli and to extrapolate them. We argued that an "overshoot" response in the new experimental condition (happy-to-angry-to-neutral and angry-to-happy-to neutral) would support the notion that pattern extrapolation could underpin perceptual distortion in the HFA. Individuals with HFA are adept at detecting regularities and cause-effect relations, which typically rule object dynamics. Importantly, they may apply this vision of a rigid, rule based, environment to the social world to make sense of social signals (Vivanti et al., 2011; Hudson et al., 2012).

We found that in the Offset condition, offsets of happy and angry facial expressions reproduced the robust overshoot bias that was first reported in Palumbo and Jellema (2013): the last neutral expressions of the Joy-to-neutral and Anger-to-neutral videos were misjudged as slightly angry and slightly happy, respectively, in both groups. However, in the Offset-onset-offset condition we found an absence of the perceptual bias in both HFA and TD groups. These latter results suggest that pattern extrapolation did not play a major role in bringing about the overshoot bias in the Offset condition in the HFA group. Extrapolation of the facial expression dynamics would have resulted in a slightly happy evaluation of the neutral expression at the end of the joy-to-anger-to-neutral clips, and in a slightly angry evaluation of the neutral expression at the end of the anger-to-joy-to-neutral clips, whereas the results showed no response biases.

We previously suggested that the most likely mechanism underpinning the overshoot bias in the Offset condition in HFA was a sequential contrast effect, as the HFA group continued to show the perceptual bias after the agent's identity had changed at the end of the clip (Palumbo and Jellema, 2013; Palumbo et al., 2015). Although the change of the agent's identity suggested that the HFA group could have relied on sequential contrast effects, in these previous studies pattern extrapolation could not be excluded as explanatory mechanism. The results of the current experiment make this explanation very unlikely, as we found no evidence for an extrapolation of the observed pattern in the HFA group in the Offset-onset-offset condition. Therefore, the original suggestion that sequential contrast effects are the best candidate for explaining the overshoot bias in the HFA group still stands. This interpretation is also supported by the "Loop" condition (neutral-to-happy-to-neutral, and neutral-to-angry-to-neutral; Palumbo et al., 2015), where the contrast hypothesis would predict the absence of a perceptual bias (because the contrast is between the first and last frames, each presented for 400 ms, which were both neutral), which was exactly what was found. However, it should be stressed that the results from the current study on themselves do not allow to make any inference about the specific mechanism that underpinned the response bias in the HFA group. It merely allows to conclude that it was not pattern extrapolation or RM that caused the bias.

We previously argued that for the TD group the perceptual bias could not be explained by contrast effects, as the agent's identity-change does not interfere with the contrast between the first emotional expression and the last neutral frame, and no bias was reported by the TD individuals in the identity-change condition. We therefore proposed an emotional anticipation mechanism (i.e., a low-level mind reading mechanism; Goldman, 2006) for the TD individuals, which would be susceptible to top-down information, such as identity information. The emotional anticipation hypothesis would predict the absence of a response bias in the new manipulation presented in the current study, which is what we found. The rationale is that because the perceptual history is equally divided over the two "opposite" emotions (joy and anger), the final neutral expression adequately sums up the agent's (final) emotional state of mind.

Thus, the current study ruled out an explanation of the overshoot bias based on pattern extrapolation in both HFA and TD, while the findings are compatible with the notion that the perceptual bias in the Offset condition was caused by sequential contrast effects in HFA and by emotional anticipation in TD, but does not itself provide any new evidence for the latter.

Emotional Anticipation: An Implicit Mechanism of Social Understanding

On the basis of our previous studies (Palumbo and Jellema, 2013; Palumbo et al., 2015) in conjunction with the current study, we propose to extend the notion of "perceptual mentalizing" (Teufel et al., 2010) by suggesting that the perceptual processing of social actions also interacts with implicit attributions made on the basis of the immediate perceptual history (Palumbo, 2012). These latter attributions are thought to reflect the operations of an anticipation mechanism, which operates automatically and involuntarily, not involving any deliberate reasoning, and which could be considered part of the perceptual system (Palumbo, 2012). In Teufel et al. (2010) model the observer is fully aware of the attributions, as they reflect explicit knowledge provided by the experimenter to the observer prior to the task. In our model, the processing of the dynamic facial expressions generates, in an automatic/involuntary fashion, an anticipation in the observer about what the actor's most likely next mental/emotional state of mind will be. This happens "on line" during the task, whereby the most likely next state of mind is continuously updated on the basis of the immediately preceding events (Palumbo, 2012). These ideas blur the distinction between perception and mentalizing, as the latter is embedded within the perceptual process. It is as if the mere perception of the social stimulus automatically induces "mentalizing" activities, which then in turn modulate the perception (cf. Hudson et al., 2009; Hudson and Jellema, 2011).

We postulate that in ASD there may be an impairment in the ability to generate anticipations about the other's immediate future action, or future state of mind, on the basis of the immediately preceding perceptual history, which could explain at least part of the communication difficulties they experience during social interaction. Taken together it suggest that individuals with HFA use an alternative route, which may rely more on physical characteristic rather than social meaning. The proposed mechanism of emotional anticipation matches recent theories of embodiment of facial expressions, which proposed that the categorization of facial expressions could be determined, or facilitated, by the experiential understanding of the agent's emotional/mental state (Wicker et al., 2003;

Botvinick et al., 2005). As such, emotional anticipation fits in well with embodied simulation models (Gallese, 2007; Niedenthal et al., 2010), which emphasize that the recognition of facial expressions is not purely the result of visual processing, but also relies on motor simulation (Palumbo, 2012). Recently substantial evidence has accumulated that the observation of dynamic facial expressions activates mirror neuron mechanisms (Dapretto et al., 2006; Pitcher et al., 2008; Likowski et al., 2012). Mirror neuron mechanisms have been argued to provide the observer with a notion of the upcoming action before it is executed (Fogassi et al., 2005; Cattaneo et al., 2007). As such, mirror mechanisms may underpin emotional anticipation (Palumbo, 2012). However, at this stage, direct evidence for this interpretation is not yet available and future research should shed light on the possible contribution of a simulation account.

ETHICS STATEMENT

The study was approved by the Ethics Committee of the University of Hull and was conducted in accordance with the Declaration of Helsinki (2008). All participants signed written informed consent before taking part.

REFERENCES

American Psychiatric Association [APA] (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: Author. doi: 10.1176/appi.books.9780890425596

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cognit. Sci.* 6, 248–254. doi: 10.1016/S1364-6613(02)01904-6

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., and Weelwright, S. (2003). The systemising quotient (SQ): an investigation of adults with Asperger syndrome or high functioning autism and normal sex differences. *Philos. Trans. R. Soc. Lond. B* 358, 361–374. doi: 10.1098/rstb.2002.1206

Baron-Cohen, S., Wheelwright, S., Skinner, S., Martin, J., and Clubley, E. (2001). The autism spectrum quotient (AQ): evidence from asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17. doi: 10.1023/A:100565341

Blakemore, S. J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567. doi: 10.1038/35086023

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., and Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage* 25, 312–319. doi: 10.1016/j.neuroimage.2004.11.043

Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., et al. (2007). Impairment of actions chains in autism and its possible role in intention understanding. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17825–17830. doi: 10.1073/pnas.0706273104

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., et al. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat. Neurosci.* 9, 28–30. doi: 10.1038/nn1611

Ekman, P., and Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667. doi: 10.1126/science.1106138

Freyd, J. J., and Finke, R. A. (1984). Representational momentum. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 126–132. doi: 10.1037/0278-7393.10.1.126

Gallese, V. (2007). London Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. B* 362, 659–669. doi: 10.1098/rstb.2006.2002

Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford: Oxford University Press. doi: 10.1093/0195138929.001.0001

Hudson, M., Burnett, H. G., and Jellema, T. (2012). Anticipation of action intentions in autism spectrum disorder. *J. Autism. Dev. Disord.* 42, 1684–1693. doi: 10.1007/s10803-011-1410-y

Hudson, M., and Jellema, T. (2011). Resolving ambiguous behavioural intentions by means of involuntary prioritization of gaze processing. *Emotion* 11, 681–686. doi: 10.1037/a0023264

Hudson, M., Liu, C. H., and Jellema, T. (2009). Anticipating intentional actions: the effect of eye gaze direction on the judgement of head rotation. *Cognition* 112, 423–434. doi: 10.1016/j.cognition.2009.06.011

Jellema, T., Lorteije, J. A. M., van Rijn, S., van't Wout, M., De Haan, E. H. F., Van Engeland, H. et al. (2009). Involuntary interpretation of social cues is compromised in Autism Spectrum Disorders. *Autism Res.* 2, 192–204. doi: 10.1002/aur.83

Jellema, T., Pecchinenda, A., Palumbo, L., and Tan, E. G. (2011). Biases in the perception and affective valence of neutral facial expressions induced by the immediate perceptual history. *Vis. Cogn.* 19, 616–634. doi: 10.1080/13506285.2011.569775

Jellema, T., and Perrett, D. I. (2003). Perceptual history influences neural responses to face and body postures. *J. Cogn. Neurosci.* 15, 961–971. doi: 10.1162/089892903770007353

Jellema, T., and Perrett, D. I. (2006). Neural representations of perceived bodily actions using a categorical frame of reference. *Neuropsychologia* 44, 1535–1546. doi: 10.1016/j.neuropsychologia.2006.01.020

Jellema, T., and Perrett, D. I. (2012). "Neural pathways of social cognition," in *The Oxford Handbook of Evolutionary Psychology*, eds R. I. M. Dunbar and L. Barrett (Oxford: Oxford University Press), 163–177. doi: 10.1093/oxfordhb/9780198568308.013.0013

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58, 259–289. doi: 10.1146/annurev.psych.58.110405.085654

Likowski, K. U., Mühlberger, A., Gerdes, A. B. M., Wieser, M. J., Pauli, P., and Weyers, P. (2012). Facial mimicry and the mirror neuron system: simultaneous

acquisition of facial electromyography and functional magnetic resonance imaging. *Front. Hum. Neurosci.* 6:214. doi: 10.3389/fnhum.2012.00214

McPartland, J., and Klin, A. (2006). Asperger syndrome. *Adolesc. Med. Clin. N. Am.* 17, 771–778.

Niedenthal, P. M., Mermillod, M., Maringer, M., and Hess, U. (2010). The simulation of smiles (sims) model: embodied simulation and the meaning of facial expression. *Brain Behav. Sci.* 33, 417–480. doi: 10.1017/S0140525X10000865

Palumbo, L. (2012). *Beyond Face Value: Involuntary Emotional Anticipation in Typical Development and Asperger's Syndrome.* Ph.D. thesis,, University of Hull, Hull.

Palumbo, L., Burnett, H. G., and Jellema, T. (2015). Atypical emotional anticipation in high-functioning autism. *Mol. Autism* 6:47. doi: 10.1186/s13229-015-0039-7

Palumbo, L., and Jellema, T. (2013). Beyond face value: does involuntary emotional anticipation shape the perception of dynamic facial expressions? *PLoS One* 8:e56003. doi: 10.1371/journal.pone.0056003

Perrett, D. I., May, K. A., and Yoshikawa, S. (1994). Facial shape and judgements of female actractiveness. *Nature* 368, 239–242. doi: 10.1038/368239a0

Pitcher, D., Garrido, L., Walsh, V., and Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J. Neurosci.* 28, 8929–8933. doi: 10.1523/JNEUROSCI.1450-08.2008

Rutter, M. (1978). Diagnosis and definition of childhood autism. *J. Autism Child. Schizophr.* 8, 139–161. doi: 10.1007/BF01537863

Satpute, A. B., and Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Res.* 1079, 86–97. doi: 10.1016/j.brainres.2006.01.005

Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in asperger syndrome. *Science* 325, 883–885. doi: 10.1126/science.1176170

Tanaka-Matsumi, J., Attivissimo, D., Nelson, S., and D'urso, T. (1995). Context effects on the judgment of basic emotions in the face. *Motiv. Emot.* 19, 139–155. doi: 10.1007/BF02250567

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., and Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Curr. Biol.* 19, 1274–1277. doi: 10.1016/j.cub.2009.05.069

Teufel, C., Fletcher, P. C., and Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends Cognit. Sci.* 14, 376–382. doi: 10.1016/j.tics.2010.05.005

Tumkaya, S., Karadag, F., Jellema, T., Oguzhanoglu, N. K., Ozdel, O., Atesci, F. C., et al. (2014). Involuntary social cue integration in patients with obsessive compulsive disorder. *Compr. Psychiatry* 55, 137–144. doi: 10.1016/j.comppsych.2013.08.007

Van 't Wout, M., van Rijn, S., Jellema, T., Kahn, R. S., and Aleman, A. (2009). Deficits in implicit attention to social signals in schizophrenia and high risk groups: behavioural evidence from a new illusion. *PLoS One* 4:e5581. doi: 10.1371/journal.pone.0005581

Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, A., et al. (2011). Intact and impaired mechanisms of action understanding in autism. *Dev. Psychol.* 47:841. doi: 10.1037/a0023105

Wang, Y., and Hamilton, A. F. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Front. Hum. Neurosci.* 6:153. doi: 10.3389/fnhum.2012.00153

Wechsler, D. (1997). *WAIS-III Administration and Scoring Manual.* San Antonio, TX: The Psychological Corporation.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2

World Health Organization [WHO] (2008). *ICD-10: International Statistical Classification of Diseases and Related Health Problems*, 10th Rev. Edn. New York, NY: World Health Organization.

Yoshikawa, S., and Sato, W. (2008). Dynamic expressions of emotion induce representational momentum. *Cognit. Affect. Behav. Neurosci.* 8, 25–31. doi: 10.3758/CABN.8.1.25

Young, A. W., Perrett, D., Calder, A., Sprengelmeyer, R., and Ekman, P. (2002). *Facial Expressions of Emotion: Stimuli and Tests (FEEST).* Bury St. Edmunds: Thames Valley Test Company.

APPENDIX

**FIGURE-A1 |** Mean ratings on the 5-point scale (*y*-axis) for the neutral expressions at the end of the video sequences for each of the eight actors (*x*-axis) in the Offset condition (left panels) and in the Offset-onset-offset condition (right panels) for the HFA (top row) and TD (bottom row) groups.

Check for
updates

Dyadic Dynamics: The Impact of Emotional Responses to Facial Expressions on the Perception of Power

Shlomo Hareli[1,2]*, Mano Halhal[1,2] and Ursula Hess[3]

[1] *The Laboratory for the Study of Social Perception of Emotions, University of Haifa, Haifa, Israel,* [2] *Department of Business Administration, University of Haifa, Haifa, Israel,* [3] *Department of Psychology, Humboldt University, Berlin, Germany*

Emotion expressions play a central role in social communication, which, by definition is a dynamic process. Social communication involves the exchange of signals with temporal dynamic properties between two or more individuals. Nonetheless, emotion perception research has strongly focused on the study of single, static, unidirectional images. The goal of this research is to illustrate the dynamic nature of emotion communication by showing how the back and forth of a dyadic emotional interaction affects its perception by uninvolved observers. To that aim, we conducted three studies that investigated how observer's inferences of social power are influenced by an exchange of emotions between members of a dyad. In Study 1, participants saw one person showing either anger or sadness to which the second member of the dyad reacted by showing either anger, fear or neutrality. In Study 1, only still photos were shown in sequence. In Studies 2 and 3, more dynamic stimuli and other emotions were included. Even though an angry expresser was always perceived as more powerful than a sad expresser, the emotional reactions of the interaction partner modulated perceived power. Across all three studies and different levels of dynamic stimuli, fear reactions always increased perceived power. Happiness, contempt and neutrality affected perceived power more selectively. This effect was mediated by the extent to which participants felt that the reaction of the second interaction partner suggested that the second interaction partner agreed with regard to the power differential between the two. Taken together, these experiments show that the social signal value of emotion expressions changes meaningfully as a function of the emotional response of the expressions' target. Thus, the social signal value of emotions does not stand alone but has to be understood in the fuller context of the interaction.

Keywords: dynamic expression of emotions, emotional interaction, reactive emotions, social power, anger, sadness

OPEN ACCESS

INTRODUCTION

Emotion expressions serve a social communicative function (Darwin, 1872/1965; Eibl-Ebesfeldt, 1989; Ekman, 1992; Fridlund, 1994; Hess et al., 1995; Fischer and Manstead, 2008; Shariff and Tracy, 2011; Scarantino, 2017) and most social interactions include exchanges of emotional expressions between the people involved (Frijda and Mesquita, 1994; Keltner and Haidt, 1999;

Hareli and Rafaeli, 2008; Van Kleef, 2009). The study of the communication of emotions aims to understand how emotional signals are perceived and used by observers. This research focuses both on observers' recognition of such expressions and the inferences about the expressers and the situation that they draw based on these expressions (Ekman et al., 1972; Hess et al., 2008; Van Kleef, 2010; Hareli, 2014).

However, the extant research is limited in two important respects. First, much of this research is restricted to the study of how a single, static, unidirectional expression of emotion is perceived (Krumhuber et al., 2013). Yet, social communication is by definition a dynamic process that involves an exchange of expressions between interaction partners (Hareli and Rafaeli, 2008). That is, the expressions shown by one interaction partner elicit expressions by the receiver. These can take different forms. Thus, the receiver may mimic the expression shown (Hess et al., 1999; Hess and Blairy, 2001; Hess and Fischer, 2013). Alternatively, the emotion shown by one interaction partner may elicit a reactive emotion in the other, which is then expressed by the addressee of the first expression (Hess and Fischer, 2013; Fischer and Hess, 2017). This latter response by the addressee of an emotion is an integral, but so far neglected, aspect of the emotion communication process.

Another limitation is the use of static images that are often bereft of context. This approach neglects informative aspects of expressive signals (see e.g., Ambadar et al., 2005; Krumhuber et al., 2013). Temporal characteristics only evident in dynamic displays impact both on the labeling of expressions and on the inferences about the expresser drawn from them (Krumhuber et al., 2013; Hess et al., 2016a).

The present research focuses on both of these points. Specifically, three studies explored how expressions of anger and sadness affect the attribution of social power as a function of the emotional response of the addressee of these emotions (i.e., reactive emotions). For this, participants saw not only the emotional expression of the person to be judged but also the emotional response of the addressee of this person's expression. In addition, we assessed the impact of the temporal dynamics of the expressions by both parties. This goal was achieved by employing a strategy in which the complexity of the dynamic context was gradually increased across three studies. Study 1 used the frozen dynamism approach (Hareli et al., 2016), in which a timed sequence of still photos is shown to simulate an exchange of emotions between members of a dyad. This approach focusses participants' attention on the different stages of the interaction. This enabled us to study, first, the effect of a mere exchange of expressive signals in an interaction. Study 2 went one step further by replacing the still photos of emotion expressions with videos. This allowed us to study whether dynamic expressions, which more closely resemble real life expressions, lead to the same effects in a social interaction. Finally, Study 3 used a video depicting the unfolding of an interaction involving the exchange of emotions between two persons appearing together. In all studies, a control condition in which only the expression of the person who is the focus of the judgment was included without any interactive context. Also, we tested a possible mechanism responsible for the combined effect of the emotions exchanged

between the parties to the interaction. Overall, this research contributes to the understanding of how the dynamics of the social communication of emotions affects attributions of social power. In addition, it offers a research strategy allowing for a controlled examination of the social perception of dynamic interactions involving emotional exchanges between interaction partners.

In what follows, we will discuss how emotion expressions lead to inferences regarding an expresser's social power and how reactive emotions are expected to affect such perceptions. This will serve as the basis for the specific hypotheses tested in this research.

Social power is a fundamentally important social factor (Russell, 1938), because it reflects a person's ability to control others (Keltner et al., 2003). One cue to social power are emotion expressions (Keltner et al., 2003; Hareli and Rafaeli, 2008). Anger and sadness are among the most studied emotions in this context. Specifically, anger signals high social power and related constructs such as dominance (Keltner, 1995; Knutson, 1996; Averill, 1997; Hareli et al., 2009; Tiedens et al., 2016). Based on appraisal theory, anger signals high social power because it is associated with an appraisal that the expresser is able to control the environment (Keltner et al., 2003; Lerner and Tiedens, 2006). By contrast, expressions of sadness reflect low levels of social power as they are associated with appraisals of lack of control (Tiedens, 2001). Accordingly, anger expressions can be considered to be signals of high social power and expressions of sadness to be signals of a lack of power. This notion is in line with the assumption that emotion expressions communicate the expressers' viewpoint in the situation (Dawkins and Krebs, 1978; Hess et al., 1995; Hareli and David, 2017; Scarantino, 2017).

Targets of such signals may respond with an expression of their own. Such responses are termed *reactive emotions* (Hess and Fischer, 2013). Reactive emotions are a direct response to the expression that elicited them. For example, if someone laughs in amusement and another person laughs as well, this can be seen as agreement that something funny happened. By contrast, if someone laughs and the other person looks irritated, this may suggest a fax pas.

In the present context, we focus on facial expressions that regulate the relationship between interaction partners. That is, we create a situation where the expressions of the interaction partners refer to each other. In that context, an anger expression, for example, signals that the expresser has control over the situation, and more specifically, control over the interaction partner (Keltner et al., 2003). In fact, the expression suggests that the interaction partner should conform to the angry person's wishes. The reactive emotion of the addressed interaction partner then signals their perception of the power differential. Thus, if the other person shows a submissive emotion such as fear or sadness, they signal that the first person has more power than they have. By contrast, a dominant expression such as anger, contempt but also happiness (Knutson, 1996) should signal that they do not agree that the other person has more power than they do. The same rationale works for emotions signaling lack of social power such as sadness. If the second person shows fear in response to sadness, they signal that in their view, the other person, even though s/he

does not signal much power, still has more power than they do. And, conversely, if the second person shows a dominant emotion, they signal that they also think that they have more power than the other person. That is, both emotion expressions "comment" on the power relationship within the dyad. These comments may agree or disagree with each other.

Importantly, for the observer, the second interaction partner is a second source of information. It makes sense for the observer to assume that this interaction partner has additional information about the sender and the situation and therefore can evaluate the relative power of the sender. As such, it makes only sense to prevail oneself of this additional information.

This implies that in a social interaction, anger or sadness expressions are not an absolute signal of power or the lack thereof simply because power is not an absolute attribute. The social power of any person depends on who else is present and therefore on the reactions of the addressee of these expressions. Also, we do not suggest that reactive emotions can completely change the perception of the initial emotions. As regards anger, since it is a signal of high social power, ignoring such a signal involves risks since even if a second person may think that they are at least equal in power, the first person may still have more power than the observer. Expressions of sadness, by contrast, reflect the admission of low social power. Since low power is socially undesirable, it is less likely to be attributed to ulterior motives and hence is likely to be trusted (Robinson et al., 1995). As such, there are (different) reasons for both emotions to be taken seriously. This is why reactive emotions are expected to modulate but not fundamentally change the meaning of anger and sadness for attribution of social power. This does, however, not mean that in real life interactions, where relative power and status are more relevant than absolute power and status, reactive emotions may not play a decisive role.

Hareli and David (2017) provided first evidence for such modulation. Specifically, they found that a person showing anger was perceived as having more social power when this anger was responded to with fear or sadness than when it was responded to with neutrality or anger. Further, this research also showed that the degree to which the expression of the second person was perceived by participants as congruent with the notion that the first person has more power than the second person mediated the effect of reactive emotions on perceived social power. Overall, they concluded that the perceived social power of the expresser is determined by the emotion shown and modified but not reversed by the reactive emotions of the interaction partner. While this research underscored the important role that reactive emotions play in social perception of emotions, several questions were left open.

First, Hareli and David (2017) did not include a no interaction control condition. That is, they could only compare the effect of different types of reactive emotions but could not assess the relevance of the absence or presence of a reaction. Second, as noted above, emotion expressions themselves are dynamic. Accordingly, it is important to understand whether the dynamics of the expressions exchanged between the interaction partners affect the attribution of power. The present research therefore addressed three questions. First, are attributions of social power

to a person whose initial expression was reacted to by someone else, different from attributions of the same initial expression when shown alone? The latter situation reflects the typical paradigm used in this line of research. Second, we assessed the impact of expressive dynamics on this process.

Finally, we assessed the specific effect of different emotions. In particular, we compared the effect of fear reactions – which had previously been shown to increase perceptions of power (Hareli and David, 2017) to expressions of happiness (Study 1–3), contempt (Study 2) and anger (Study 3), as well as neutrality (Study 1 and 2). We predicted that emotions that signal high dominance (anger and happiness, Knutson, 1996) and emotions that suggest a devaluation of the expresser (contempt, Fischer and Roseman, 2007) would reduce perceptions of power. Also, specifically, they would reduce the degree to which the second person's reaction is seen as a sign of acceptance of the high power signaled by the first person's anger or conversely of the low power signaled by the first person's sadness. Happiness also is an emotion that signals that the expresser considers that all is well (Scherer, 1987) thus in the present context this emotion may also be seen as mocking or denying especially expressions of anger signaling high power.

STUDY 1

In Study 1 participants first saw one person expressing anger or sadness and then another person responding to this reaction with either fear, happiness or neutrality. In addition, in a control condition, participants saw only the first expresser. This enabled us to study the impact of the same expressions witnessed in isolation when they are not part of an unfolding social interaction.

Methods

Participants

A total of 915 (477 men, 2 other) participants with a mean age of 38 years (SD = 11.5) who were recruited through Amazon MTurk, completed the study. Data collection continued with random assignment until a minimum of 25 participants per experimental cell was reached.

Materials and Procedure

Participants were randomly assigned to the social interaction or no social interaction condition. Participants in the social interaction condition were informed that they will see a series of photos taken from videos of an interaction between two persons, depicting a sequence of events in the interaction. The first photo was described as showing an expression by one interaction partner and the second the interaction partner's response to this expression. No information about the nature of relationship between the two was provided. We assumed that in many situations this information is unknown to observers, although they may have guesses.

Participants in the no social interaction condition were informed that they will see a photo of a person. All participants

were told that they will have to rate different things about what they saw. Each participant completed only one trial.

As posers we randomly chose 8 men and 8 women from the Radboud Faces Database (Langner et al., 2010). Of these, four posers from each gender showing either anger or sadness, served as the first expresser in the social interaction condition or the only expresser in the no social interaction condition. The remaining 4 posers of each gender expressing fear, happiness and neutrality, served as the second expresser. Dyads were formed by randomly selecting one poser from the set of first expressers and one from the set of second expressers. To increase the impression that reactions were taken from actual interactions, we used the 45° left and right orientations versions of the photographs, so that the expressers appeared to orient their reactions toward one another. To control for the effect of orientation and side of presentation, half of the participants saw the sets with first expresser person appearing on the right hand side of the screen, orienting the expression toward the left, and the person reacting to this expression appearing on the left and orienting the reaction toward the right. The rest of the participants saw the sets with the reversed position of expressers and orientations. To further establish the impression that the stimuli represent a sequence of reactions, the photograph depicting the person expressing the emotion first appeared for 1,500 ms after which it disappeared, and the person reacting to this expression then appeared on the other side for 1,500 ms. Below the photographs was written: "The reaction of the first person" and "The response of the second person," for the first and second photos, respectively (for an example of a stimulus and the sequence of events, see **Figure 1**). Next, both photographs were presented in their original position and rating scales appeared below.

In the no social interaction condition, a poser from the first expresser set was selected. This poser appeared either in the right gaze or left gaze orientation in the respective position as the first poser in the social interaction position. The photo appeared first for 1,500 ms and then disappeared. Then the photo reappeared together with the rating scales. No inscription appeared under the photo in this condition.

Dependent Measures

Participants were asked to rate their perception of the first expresser's dominance, submissiveness and competence as well the expresser's control over the situation. Since these measures correlated highly ($\alpha = 0.76$; $\omega = 0.85$[1]), they were combined into one social power scale by computing the average of these ratings with submissiveness being reverse scored. Then participants were asked to rate the intensity of anger and sadness of the person who was shown first (or the only person shown, for the no social interaction condition). We further assessed to what degree participants considered the expression of the second person to signal that they accepted the first expresser's dominance, submitted to the first expresser and confirmed the first expresser's standing in the interaction. These measures correlated ($\alpha = 0.82$,

---

[1]Since Cronbach's alpha, as a measure of reliability of a composite measure is considered to rely on assumptions that are often violated, we also report Omega as a measure of Composite Reliability (McNeish, 2017). This was done for all constructs across the studies.

$\omega = 0.89$) and hence were combined by averaging the ratings into one scale, which we labeled "acceptance of power." All ratings were made on 7-point Likert scales anchored with 1 = not at all and 7 = to a large extent.

RESULTS

Emotion Perception

Emotions of first expresser

Initial analyses did not reveal any significant main effects nor interactions involving sex of either interaction partner for anger ratings for the first expression. A significant effect on sadness ratings for the first expression did not yield any significant *post hoc* effects. The two gender factors were therefore dropped from further analyses. A 2 (emotion shown by the first person: sadness, anger) × 4 (emotion shown by the second person: no emotion, neutral, fear, happiness) analysis of variance on the emotion ratings yielded for anger, $F(1,907) = 829.83$, $p < 0.001$, $\eta_p^2 = 0.48$, a main effect of first emotion such that anger expressions were rated as showing more anger ($M = 5.43$, $SD = 1.56$, CI: 5.29, 5.58) than sadness expressions ($M = 2.46$, $SD = 1.55$, CI: 2.32, 2.61). For sadness, a main effect of first emotion emerged, $F(1,907) = 992.85$, $p < 0.001$, $\eta_p^2 = 0.52$, as well, such that sadness expressions were rated as sadder ($M = 5.96$, $SD = 1.30$, CI: 5.82, 6.10) than anger expressions ($M = 2.78$, $SD = 1.73$, CI: 2.64, 2.92). In addition, for sadness only, a main effect of second emotion emerged, $F(3,907) = 3.08$, $p = 0.027$, $\eta_p^2 = 0.01$, such that overall, across both emotion conditions, expressions that were reacted to with fear were rated as less sad ($M = 4.17$, $SD = 2.24$, CI: 3.93, 4.33) than those that were shown alone ($M = 4.48$, $SD = 2.18$, CI: 4.37, 4.76). Expressions reacted to with happiness ($M = 4.40$, $SD = 2.23$, CI: 4.21, 4.61) and with a neutral expression ($M = 4.38$, $SD = 2.18$, CI: 4.17, 4.57) were not rated differently from one another. The interaction effect was not significant, $F(3,907) = 0.30$, $p = 0.826$, $\eta_p^2 = 0.00$. Thus, overall, the emotions were interpreted as intended. It is interesting to note that a fear reaction by the addressee of either an anger or sad expression makes this expression appear sadder. The absence of an interaction effect suggests that this may be more of a halo effect.

Perceived social power of the first expresser

Initial analyses did not reveal any significant main effects nor interactions involving sex of either interaction partner for perceived power or perceived acceptance of power by the second person. The two gender factors were therefore dropped from further analyses. A 2 (emotion shown by the first person: sadness, anger) × 4 (emotion shown by the second person: no emotion, neutral, fear, happiness) analysis of variance was conducted on the attribution of social power. A significant main effect of first expression, $F(1,907) = 358.11$, $p < 0.001$, $\eta_p^2 = 0.28$, emerged, such that individuals who showed anger were rated as higher in social power ($M = 4.60$, $SD = 1.14$, CI: 4.51, 4.72) than those who showed sadness ($M = 3.22$, $SD = 1.16$, CI: 3.12, 3.32). Further, a significant main effect of second expression, $F(3,907) = 21.30$, $p < 0.001$, $\eta_p^2 = 0.07$, emerged. *Post hoc* analyses revealed that

FIGURE 1 | Sequence of stimulus presentation and set-up with example of stimuli used in Study 1.

any expression reacted to with fear resulted in higher attributions of social power ($M = 4.41$, $SD = 1.34$, CI: 4.28, 4.57) than expressions reacted to with happiness ($M = 3.70$, $SD = 1.32$, CI: 3.55, 3.84) or neutrality ($M = 3.78$, $SD = 1.34$, CI: 3.64, 3.93), or not responded to at all ($M = 3.79$, $SD = 1.26$, CI: 3.62, 3.90) which did not differ. The interaction was not significant, $F(3,907) = 0.26$, $p = 0.857$, $\eta_p^2 = 0.00$. That is, contrary to expectations, the effect of the reactive emotion did not depend on the first emotion shown. Thus, being responded to with fear increased perceived social power regardless of whether high or low social power were signaled.

Perceived acceptance of power by the second expresser

A 2 (emotion shown by the first person: sadness, anger) × 3 (emotion shown by the second person: neutral, fear, happiness) analysis of variance was conducted on the degree to which participants considered that the second person accepted that the first person has more power. A significant main effect of first expression, $F(1,671) = 44.39$, $p < 0.001$, $\eta_p^2 = 0.06$, and of second expression, $F(2,671) = 72.84$, $p < 0.001$, $\eta_p^2 = 0.18$, emerged. This main effect was qualified by an interaction $F(2,671) = 9.11$, $p < 0.001$, $\eta_p^2 = 0.03$, such that when anger was shown first, acceptance of power was perceived as strongest when the reactive emotion was fear ($M = 4.78$, $SD = 1.40$, CI: 4.50, 5.05), followed by neutrality ($M = 3.61$, $SD = 1.59$, CI: 3.34, 3.88), and least for happiness ($M = 2.62$, $SD = 1.52$, CI: 2.35, 2.89). The same pattern was found for sadness: fear ($M = 3.64$, $SD = 1.45$, CI: 3.37, 3.91), then neutrality ($M = 2.58$, $SD = 1.37$, CI: 2.31, 2.85) and happiness ($M = 2.54$, $SD = 1.43$, CI: 2.27, 2.81), yet, neutrality and happiness did not differ significantly. Thus, independent of whether the first expression was anger or sadness, participants saw fear as a sign that the second expresser considered the first to be high(er) in power, and neutrality and happiness as doing so to a much lesser

degree. This is congruent with the finding reported above that fear reactions always increased the perceived power of the first expresser. We therefore conducted a mediation analysis to assess whether this increase in perceived power is due to the fact that the expression was seen as supportive of the notion that the first expresser is high(er) in power.

Mediation Analysis

To analyze the proposed mediation, we calculated a mediation model (Hayes model 4) with reactive emotion as a multicategorial index coded variable comparing fear and happiness to neutral. The analysis used Process 3.0 (Hayes, 2017).

A significant positive indirect effect on perceived social power for reactive fear expressions ($b = 0.41$, $SE = 0.06$, CI: 0.29, 0.54) and a significant indirect effect for happiness ($b = -0.19$, $SE = 0.06$, CI: -0.30, -0.08) compared to neutral emerged. Specifically, reactive fear expressions were rated as signaling acceptance of the first person's power by the second person and this acceptance in turn increased attributions of social power to the first person by the participants. The converse effect was found for happiness reactions (even though this effect did not yield a significant effect in the ANOVA). Thus, as predicted, the emotional expression of the addressee of an expression impacts on the inferences that observers draw about the sender of that expression because these expressions themselves speak meaningfully toward the social power of the first person.

DISCUSSION

Overall, the present findings replicate and extend findings by Hareli and David (2017). We found again that a fear reaction by the addressee of an expression leads to attributions of higher social power to the person sending the initial expression. In

Study 1, this was independent of whether the initial expression was anger or sadness.

We further found that this increase in attributed social power was mediated by the fact that anyone who is reacted to with fear is seen as more powerful than someone who is reacted to with neutrality. Interestingly, the converse was found for happiness in the mediation analysis. That is, anyone who was reacted to with happiness was rated as lower in social power to the degree that this expression seemed to dispute claims of social power. This finding is suggestive of the notion that reactions of happiness may contradict signals of high social power.

STUDY 2

Even though the findings of Study 1 support our basic hypotheses that the emotional reactions of both partners in an interaction are relevant for observers' social judgments, the setting we used was somewhat artificial. Participants saw two still photos of individuals supposedly interacting rather than actual dynamic expressions. Thus, in Study 2, using the same methodology as in Study 1, the still photos were replaced by videos of expressions of emotions with the goal of examining to what degree the findings of Study 1 replicate in such conditions.

In addition, we added expressions of contempt as an additional reactive emotion as one goal of the present research was to examine if and under what conditions a reactive emotion can decrease the perceived social power of the first expresser. Contempt is considered a response that devalues its objects to the point of nullifying them and their capabilities (Fischer and Roseman, 2007). Thus, a contempt reaction by the addressee of a "power claim" by the first expresser should undermine this claim.

In Study 2 we also measured the perceived intensity of reactive emotions. We did this because ratings of perceived emotions more accurately reflect the participants' perception of these expressions than do the categorical condition codes. Finally, since we did not find significant effects for gender composition in Study 1, we simplified the design by using same-sex dyads only.

Methods
Participants
A total of 593 (343 women, 1 other) participants with a mean age of 40 years (SD = 12.6) who were recruited through Amazon MTurk completed the study. Data collection continued with random assignment until a minimum of 25 participants per experimental cell was reached.

Materials and Procedure
The procedure was the same as in Study 1 except for the fact that videos were used as the primary stimuli. As posers we randomly chose 4 men and 4 women from the Amsterdam Dynamic Facial expressions Set (Van der Schalk et al., 2011). To increase the impression that reactions were taken from actual interactions, we used the 45° turning right versions from the set. To control for the effect of orientation and side of presentation, videos were rotated 180° using video editing software (Camtasia Studio 8,

Techsmith[2]). Thus, as in Study 1, the orientation of the first expresser was counterbalanced. Videos were edited to start with the expresser showing a neutral expression. Emotion expressions started after 500 ms. and the reaction unfolded and lasted for an additional 5000 ms. Combination of expressers was random with the restrictions that the two posers were different actors of the same sex. As in Study 1, the first expresser appeared first and the video with the person reacting to this expression then appeared on the other side after the end of the first video. Next, photographs created from the apex of the reaction in the video were presented in their original position and rating scales appeared below. Below the videos and photographs it was written: "The reaction of the first person" and "The response of the second person," for the first and second stimuli, respectively. In the no social interaction condition, only one poser appeared either in the right gaze or left gaze orientation. When the video was finished, the video and photo of the apex of the reaction appeared with the rating scales. No inscription appeared under the video and photo in this condition. This resulted in a 2 (Emotion of first expresser: anger or sadness) × 2 (Gender of the expressers) × 5 (Reactive emotion of second expresser: fear, contempt, happiness and neutrality, no reaction) between-subjects design.

Dependent Measures
The same dependent measures as in Study 1 were used. Ratings of perceived dominance, submissiveness, competence and control over the situation were combined into one social power scale (α = 0.73, ω = 0.82). The ratings of the extent to which the person who was second to express an emotion submitted to the first expresser, accepted the first expresser's dominance and confirmed the first expresser's standing in the interaction were combined into one acceptance of social power scale (α = 0.68, ω = 0.64). For self-report questionnaire items, internal consistencies of 0.70 are often considered acceptable if scales consist of very few items (Hahn et al., 2012), as is the case here.

Participants further rated the perceived intensity of anger and sadness of the person who was shown first (or the only person shown, for the no social interaction condition) as well as perceived intensity of the reactive emotions of fear, contempt, happiness and neutrality in the social interaction condition. All ratings were made on 7-point Likert scales anchored with 1 = not at all and 7 = to a large extent.

RESULTS

Emotion Perception
Emotions of first expresser
A 2 (First expression) × 2 (Gender of expressers) × 5 (Reactive emotion) ANOVA was conducted on ratings of anger and sadness intensity. For ratings of anger, a significant main effect of first expression emerged, $F_{(1,573)} = 573.46$, $p < 0.001$, $\eta_p^2 = 0.50$, such that expressions of anger were rated as angrier ($M = 5.78$, $SD = 1.40$, CI: 5.60, 5.95) than expressions of sadness ($M = 2.76$,

---

[2] www.techsmith.com/camtasia.html

$SD = 1.66$, CI: 2.60, 2.95). The main effect of reactive emotion was also significant, $F(4,573) = 2.69$, $p = 0.03$, $\eta_p^2 = 0.02$. *Post hoc* tests revealed that anger intensity was rated somewhat lower when it was responded to by contempt ($M = 4.00$, $SD = 2.22$, CI: 3.75, 4.28) or neutrality ($M = 4.19$, $SD = 2.16$, CI: 3.85, 4.40) compared to when shown alone ($M = 4.65$, $SD = 2.08$, CI: 4.34, 4.90). When anger was responded to by fear ($M = 4.22$, $SD = 2.15$, CI: 4.02, 4.58) or by happiness ($M = 4.26$, $SD = 2.13$, CI: 4.04, 4.59) perceived intensity of anger did not differ from any other condition.

For ratings of sadness, a significant main effect of first expression emerged, $F(1,573) = 502.21$, $p < 0.001$, $\eta_p^2 = 0.47$, such that expressions of sadness were rated as sadder ($M = 5.70$, $SD = 1.69$, CI: 5.52, 5.90) than expressions of anger ($M = 2.64$, $SD = 1.67$, CI: 2.46, 2.84). In addition, a significant gender by first emotion interaction emerged, $F(1,573) = 13.10$, $p < 0.001$, $\eta_p^2 = 0.02$. *Post hoc* tests indicated that women's sadness was perceived as somewhat more intense ($M = 5.99$, $SD = 1.42$, CI: 5.71, 6.25) than men's sadness ($M = 5.44$, $SD = 1.88$, CI: 5.17, 5.70) and men's anger was rated as somewhat sadder ($M = 2.87$, $SD = 1.75$, CI: 2.60, 3.14) than women's anger ($M = 2.43$, $SD = 1.55$, CI: 2.17, 2.70). Overall, these results indicate that the emotions of the first expresser were perceived as planned.

Perceived intensity of reactive emotions

A 2 (First expression) × 2 (Gender of expressers) × 4 (Reactive emotion) ANOVA was conducted on ratings of fear, contempt, happiness and neutrality. A main effect of reactive emotion emerged for all emotions (see **Table 1**). Ratings on each of the four emotion scales were highest for the video with the corresponding focal emotion expression. However, additional effects emerged for secondary emotion ratings. That is, for emotions not actually expressed, for example, perceived fear of a face showing anger. Contempt expressions were rated as more neutral than fear and happiness expressions and fear expressions were rated as less contemptful than happiness and neutral expressions. Contempt expressions were rated as happier than expressions of fear and neutrality.

For fear ratings, a significant main effect of first emotion emerged, $F(1,464) = 21.86$, $p < 0.001$, $\eta_p^2 = 0.05$, such that fear was rated somewhat more intensely when it was expressed in response to anger ($M = 3.30$, $SD = 2.26$, CI: 3.16, 3.54) than in response to sadness ($M = 2.69$, $SD = 2.16$, CI: 2.51, 2.89). For neutrality ratings, a main effect of expresser gender emerged, $F(1,464) = 4.41$, $p = 0.04$, $\eta_p^2 = 0.009$, such that men were rated as somewhat more neutral overall ($M = 3.08$, $SD = 2.18$, CI: 2.92, 3.31) than women ($M = 2.85$, $SD = 2.04$, CI: 2.62, 3.02). Thus, overall, reactive emotions were perceived as planned.

Perceived social power of the first expresser

We first compared the effect of a reactive emotion on the evaluation of the expression alone. For this a 2 (First expression) × 2 (Gender of expressers) × 4 (Reactive emotion) ANOVA was conducted on ratings of social power. A significant main effect of first expression emerged, $F(1,573) = 266.05$, $p < 0.001$, $\eta_p^2 = 0.32$, such that individuals showing anger expressions ($M = 4.53$, $SD = 1.13$, CI: 4.41, 4.65) were rated as higher in social power than those who showed sadness ($M = 3.10$, $SD = 1.10$, CI: 2.98, 3.23). A first expression × gender interaction was significant, $F(1,573) = 4.20$, $p = 0.041$, $\eta_p^2 = 0.01$, but *post hoc* tests did not reveal significant differences as a function of gender. Further, as in Study 1, the main effect of second expression was significant, $F(4,573) = 13.93$, $p < 0.001$, $\eta_p^2 = 0.09$, but in Study 2 also qualified by a first expression by second expression interaction, $F(4,573) = 3.30$, $p = 0.011$, $\eta_p^2 = 0.02$. As shown in **Table 2**, for both sadness and anger expressions, as in Study 1, reactive fear expressions increased attributions of social power relative to the expression shown alone. In addition, for anger expressions, reactive happiness expressions reduced the attribution of social power relative to the expression shown alone. This effect of reactive happiness was hinted at in the mediation analysis for Study 1, but not significant when comparing means. No other significant differences emerged. In sum, for sad expressions only fear and for anger expressions both fear and happiness moderated the perception of social power compared to the expression alone.

TABLE 1 | Ratings of perceived intensity of reactive emotions as a function of expressed reactive emotion – Study 2 and Study 3.

Reactive emotion	Fear			Happiness			Contempt			Neutrality			F (3,464)	p	η_p^2
Study 2	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI	M	SD	95% CI			
Fear	5.74$_a$	1.58	5.51, 6.06	1.67$_b$	1.36	1.40, 1.94	2.04$_c$	1.47	1.78, 2.31	2.63$_d$	1.80	2.33, 2.31	180.86	<0.001	0.54
Happiness	1.74$_a$	1.36	1.50, 1.99	6.17$_b$	1.11	5.94, 6.42	2.38$_c$	1.65	2.14, 2.61	1.74$_a$	1.10	1.52, 2.01	305.79	<0.001	0.66
Contempt	2.60$_a$	1.64	2.27, 2.93	3.41$_b$	2.08	3.07, 3.73	4.46$_c$	1.81	4.14, 4.78	3.79$_b$	1.81	3.46, 4.13	21.82	<0.001	0.12
Neutrality	1.68$_a$	1.27	1.38, 1.95	1.67$_a$	1.10	1.39, 1.95	3.33$_b$	1.86	3.06, 3.60	5.21$_c$	1.79	4.93, 5.50	140.29	<0.001	0.48
Study 3	Fear			Happiness			Anger						(2, 334)		
Fear	4.62$_a$	1.84	4.29, 4.90	1.99$_b$	1.42	1.67, 2.28	3.22$_c$	1.72	4.29, 4.90				71.35	<0.001	0.30
Happiness	1.99$_a$	1.36	1.70, 2.28	5.44$_b$	1.81	5.17, 5.74	2.09$_a$	1.40	1.78, 2.35				185.93	<0.001	0.53
Anger	2.51$_a$	1.64	2.24, 2.83	2.09$_a$	1.57	1.79, 2.39	4.84$_b$	1.68	4.54, 5.13				94.72	<0.001	0.36
Neutrality	2.06	1.55	1.79, 2.36	2.21	1.62	1.91, 2.48	2.32	1.55	2.02, 2.60				0.65	=0.52	0.00

Means with different subscripts differ at $p < 0.05$. In Study 3 there was no condition of a reactive emotion of neutrality but participants were asked to rate each expression on perceived neutrality.

TABLE 2 | Perceived social power as a function of first expresser's emotion and reactive emotion – Study 2 and Study 3.

	Reactive emotion									
Study 2	**No Emotion**		**Fear**		**Happiness**		**Contempt**		**Neutral**	
First emotion	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI
Sadness	3.03_a 0.99	2.75, 3.32	3.58_b 1.05	3.31, 3.84	3.01_a 1.21	2.74, 3.27	2.86_a 0.98	2.60, 3.13	3.04_a 1.15	2.76, 3.32
Anger	4.64_c 0.95	4.38, 4.92	5.25_d 0.85	4.97, 5.53	3.84_b 1.09	3.57, 4.11	4.50_c 1.16	4.24, 4.77	4.44_c 1.11	4.14, 4.68
Study 3	**No Emotion**		**Fear**		**Happiness**		**Anger**			
Sadness	2.84_a 0.90	2.57, 3.08	3.79_c 1.21	3.53, 4.05	3.57_c 1.19	3.27, 3.79	4.17_b 1.18	3.93, 4.43		
Anger	4.75_d 0.72	4.49, 5.00	5.18_e 0.79	4.93, 5.42	4.66_d 0.85	4.41, 4.90	4.86_{de} 0.86	4.61, 5.11		

Means with different subscripts differ at $p < 0.05$.

We then assessed the effects of reactive emotions as perceived by the participants. Specifically, it can be argued that the effect of the reactive emotions depends on the perceived emotion rather than emotion condition. Specifically, even if a face has been validated as showing anger, a given participant may also perceive secondary emotions such as sadness and fear. Secondary emotions have been shown to affect perceptions of interactions in meaningful ways (Hess et al., 2016b). In fact, as can be seen in **Table 1** above, even though the focal emotion was rated as strongest for each of the expressions, participants perceived a mix of expressions as is common in emotion perception (Russell and Fehr, 1987; Russell et al., 1993; Yrizarry et al., 1998; Hess et al., 2016b). We therefore conducted multiple regression analyses with the emotion ratings for the reactive emotion as predictors. Given the first emotion by second emotion interaction reported above, we ran separate analyses for sad and anger first expressions. Given the weakness of the gender × first emotion effect, gender was dropped from this analysis.

For reactions to sadness, the MR model explained 12% of the variance, $F_{(4,238)} = 7.89$, $p < 0.001$. Significant effects emerged for fear ($\beta = 0.28$, $p < 0.001$, CI: 0.15, 0.42) and contempt ($\beta = -0.17$, $p = 0.007$, CI: −0.30, 0.05). Specifically, whereas fear reactions to sadness increased perceptions of social power, contempt reactions to sadness reduced it. That is, contempt reduced the already weak signal of power shown by the first expresser.

For reactions to anger, the MR model explained 29% of the variance, $F_{(4,232)} = 23.80$, $p < 0.001$. Significant effects emerged for fear ($\beta = 0.33$, $p < 0.001$, CI: 0.21, 0.45), contempt ($\beta = -0.13$, $p = 0.027$, CI: −0.24, −0.01) and happiness ($\beta = -0.27$, $p < 0.001$, CI: −0.39, −0.15). Again, whereas fear increased perceptions of social power both contempt and happiness decreased it.

Perceived acceptance of power by the second expresser
We then assessed to what degree the reactive emotions shown by the addressee of the first expressions were seen as accepting that the first person has more power. Congruent with the analyses above, we calculated MR separately for sadness and anger first expressions with reactive emotion ratings as predictors.

For reactions to sadness, the MR model explained 19% of the variance, $F_{(4,238)} = 13.84$, $p < 0.001$. Only fear significantly and positively predicted the degree to which the expression of the second person signaled that they considered the first person to have (more) power ($\beta = 0.42$, $p < 0.001$, CI: 0.30, 0.55).

For reactions to anger, the MR model explained 40% of the variance, $F_{(4,232)} = 39.12$, $p < 0.001$). Significant effects emerged for fear ($\beta = 0.49$, $p < 0.001$, CI: 0.38, 0.60), contempt ($\beta = -0.19$, $p < 0.001$, CI: −0.29, −0.09) and happiness ($\beta = -0.15$, $p = 0.010$, CI: −0.26, −0.04). Specifically, reactions of fear increased, whereas reactions of contempt and happiness decreased the degree to which the response by the addressee of an anger expression was considered supportive of the notion that the anger expresser had high(er) social power.

Mediation Analysis

As for Study 1, we conducted mediation analyses to assess whether the increases and decreases in perceived social power as a function of reactive emotion can be explained by the degree to which these expressions were perceived as accepting that the first person has high(er) power. For this, we defined a saturated model in AMOS (22.0) in which the four emotion rating variables predicted the degree of acceptance of power and this variable in turn predicted perceived social power. We conducted the analyses separately for sadness and anger first expressions. Bootstrap was set to 3000.

For reactions to sadness, only for fear was the indirect effect significant ($\beta = 0.12$, $p < 0.001$, CI: 0.06, 0.20). For reactions to anger, significant indirect effect were found for fear ($\beta = 0.24$, $p < 0.001$, CI: 0.17, 0.33), contempt ($\beta = -0.10$, $p = 0.002$, CI: −0.16, −0.04) and happiness ($\beta = -0.07$, $p = 0.012$, CI: −0.14, −0.02).

DISCUSSION

In sum, the mediation analyses confirmed the notion that the effects of reactive emotions on perceived social power were mediated by the perception that the second expresser considered the first expresser to be high(er) in power. Specifically, fear reactions in response to both anger and sadness expressions increased perceived social power to the degree to which these reactions were seen as accepting the power signaled by the first person. For contempt and happiness expressions shown in reaction to anger, the converse effect was found. The effects for

fear and happiness replicate findings from Study 1. The finding for contempt supports the notion that contempt can invalidate the power signaled by anger expressions. For sadness expressions, contempt also had the effect of eroding the already low level of power signaled by that expression even further. Yet, this was not mediated through the perception that this expression signals that the second perceiver disagrees with the power claim by the first perceiver. One possibility is that contempt shown toward a sad person may devalue the person as such (Fischer and Roseman, 2007) – rather than their "claim" and this may also lead to perceived lack of social power.

Overall, the results of Study 2 further support the notion that not only the expression shown by a person but also the reactions of others to this expression are relevant for the assessment of the social power of the individual. That is, in a dyadic interaction, the emotional expressions of both interaction partners meaningfully inform observers about the expressers. Interestingly, whereas in Study 1, the type of emotion shown by the first expresser did not affect the impact of the reactive emotions, it did so for Study 2. Specifically, as proposed in the introduction, reactions of happiness in response to anger but not in response to sadness had a power eroding effect. This, because showing happiness and signaling that all is well in the face of an aggressive signal such as anger suggests that the happy person does not consider the threat display threatening. Someone who smiles at a sad person, by contrast, might be seen as callous more than anything else and hence their display is disregarded for evaluations of the social power of the sad person.

Importantly, the replication of findings from Study 1, showed that the effects were not driven by the artificial nature of the stimulus display in that study. That stronger additional effects of happiness were found may be due to the use of dynamic rather than static images.

Yet, this study too is limited in two important respects. First, the emotion expressions were presented to the participants in sequence and then were shown as stills during the rating task. Even though this enables participants to focus carefully on the sequence of the events, it may also over sensitize them to aspects of the situation that otherwise may be more subtle. That is, when people witness a dyadic social interaction, both partners appear together and the focus of the observers may shift between the two, forcing them to be less aware of each individual expression. In addition, the mere presence of both partners together may provide important information about the interaction that is missing when the stimuli are presented sequentially. Further, the videos we used in Study 2 showed expressions that were quite intense. Real-life expressions of emotions are often considerably less intense (Motley and Camden, 1988).

STUDY 3

Given the limitations of Study 2, as described above, the goal of Study 3 was to test our hypotheses using a more ecological valid design in which both interaction partners are showing more subtle facial expressions concurrently. Finally, the emotion ratings showed that contempt was notably less well recognized than fear and happiness. Since the recognition of contempt would likely be even further reduced for expressions with lower intensity, we replaced contempt in Study 3 with expressions of anger. Anger was also expected to serve as a signal that the addressee, especially of an anger expression, does not agree that the other is (more) powerful (Hareli and David, 2017).

Methods

Participants

A total of 457 (274 women) participants with a mean age of 40 years (SD = 13.3) who were recruited through Amazon MTurk. Data collection continued with random assignment until a minimum of 25 participants per experimental cell was reached.

Materials and Procedure

The still photos that were created from the videos and were used in Study 2 as the stimuli for the second phase of the study, were used to create morphed videos with an expression changing from neutral to one of the expressions (anger, sadness, fear, or happiness) using Fantamorph 5.0 (Abrosoft)[3]. Morphed videos were saved as AVI video files. Videos ended when the expression reached 80% of their peak intensity along the continuum from a neutral expression to the apex of the emotion. For the conditions involving an interaction, videos of two posers of the same sex were placed, one next to the other, each orienting toward the other. The video of the first expresser was edited so that the expression started after 500 ms. The expression in the video of the responder started 1000 ms later. Both reactions reached their respective apex (80% of the original apex) after 1000 ms, respectively and the entire sequence lasted for 3000 ms. We provided the participants with the explanation that the video shows an expression by one interaction partner and how the other interaction partner responded to this expression and that each partner was filmed with a different camera. To clarify this fictitious set up supposedly creating the presented stimuli, a figure depicting how the scene was created was shown (in the social interaction condition only, see **Figure 2**). Participants were further told that the video would be shown twice so that they can have a better sense of what went on. As in Study 2, combination of expressers was random with the restrictions that both posers were different individuals and that they were of the same sex. Presentation orientation of the posers was counterbalanced, as in Studies 1 and 2. In the no social interaction condition, only one poser appeared either in the right gaze or left gaze orientation in the respective position as the first poser in the social interaction condition. Unlike in Studies 1 and 2, no inscription appeared under the videos in any condition. This resulted in a 2 (Emotion of first expresser: anger or sadness) × 2 (Gender of the expressers) × 4 (Reactive emotion of second expresser: no reactive emotion, fear, happiness, and anger) between-subjects design.

Dependent Measures

The same ratings and scales as in Study 2 were used, except that the contempt rating was replaced with an anger rating.

---

[3] www.fantamorph.com

FIGURE 2 | Diagram of supposed set-up of creating the videos of social interactions for Study 3 presented to participants in the social interaction condition as part of the instructions. The individual videos of each person were supposedly merged into one video as presented to participants.
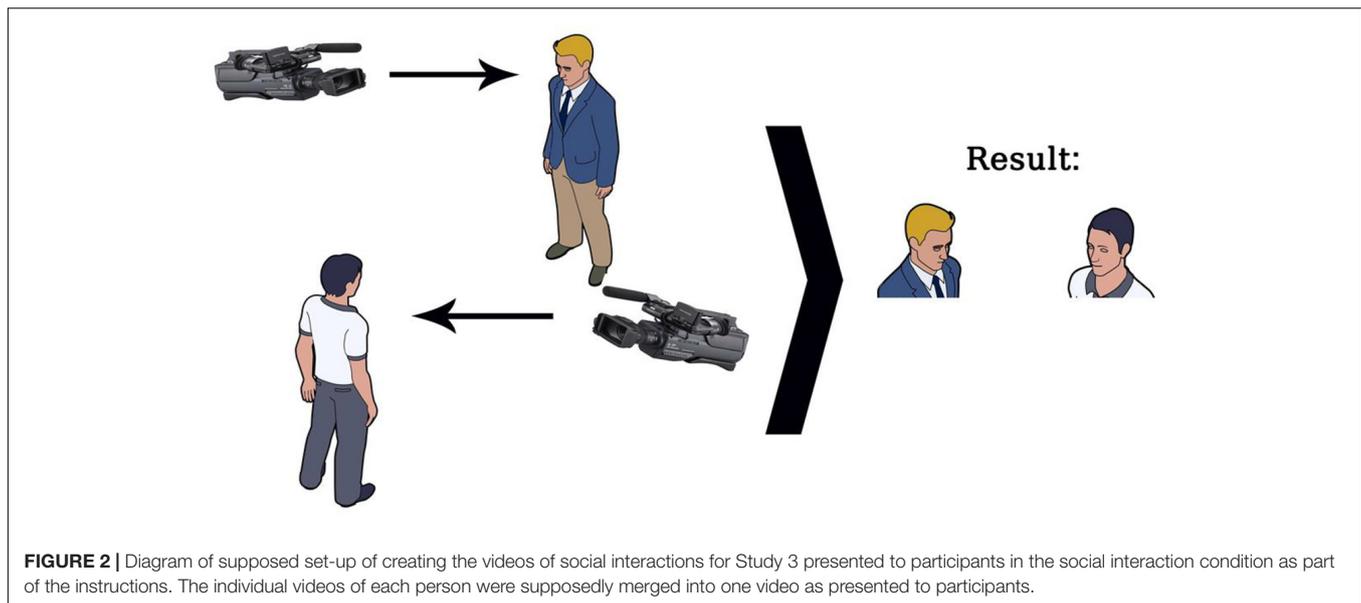
As in the previous studies, measures of perceived dominance, submissiveness, competence and control over the situation were combined into one social power scale ($\alpha = 0.71$, $\omega = 0.63$) and measures of submission to the first expresser, acceptance of the first expresser's dominance and confirmation of the first express's standing in the interaction were combined into the acceptance of power scale ($\alpha = 0.61$, $\omega = 0.79$).

RESULTS

Emotion Perception

Emotions of first expresser

A 2 (First expression) × 2 (Gender of expressers) × 4 (Reactive emotion) ANOVA was conducted on ratings of anger and sadness intensity. For ratings of anger, a significant main effect of first expression emerged, $F(1,441) = 188.98$, $p < 0.001$, $\eta_p^2 = 0.30$, such that expressions of anger were rated as angrier ($M = 5.15$, $SD = 1.64$, CI: 4.93, 5.35) than expressions of sadness ($M = 3.07$, $SD = 1.65$, CI: 2.84, 3.26). In addition, a significant interaction between first emotion and reactive emotion emerged, $F(3,441) = 4.67$, $p = 0.003$, $\eta_p^2 = 0.03$. *Post hoc* tests revealed that perceived anger intensity was always higher for anger expressions than for sadness expressions and all anger expressions were rated similarly irrespective of reactive emotion ($M = 5.43$, $SD = 1.26$, CI: 4.99, 5.84; $M = 5.10$, $SD = 1.50$, CI: 4.67, 5.52; $M = 4.72$, $SD = 2.00$, CI: 4.28, 5.10; and, $M = 5.36$, $SD = 1.61$, CI: 4.94, 5.76, for anger with no reaction, anger reacted to with anger, happiness, and fear, respectively). However, anger ratings of sadness expressions varied with reactive emotions. Specifically, sadness responded to with fear was rated as less angry ($M = 2.42$, $SD = 1.49$, CI: 1.98, 2.86) than sadness responded to with anger ($M = 3.57$, $SD = 1.55$, CI: 4.67, 5.52). No difference emerged between sadness shown alone ($M = 3.13$, $SD = 1.74$, CI: 3.14, 3.98) and sadness reacted to with happiness ($M = 3.11$, $SD = 1.63$, CI: 2.66, 3.52).

For ratings of sadness, a significant main effect of first expression, $F(1,441) = 147.64$, $p < 0.001$, $\eta_p^2 = 0.25$, emerged, such that expressions of sadness were rated as sadder ($M = 4.73$, $SD = 2.00$, CI: 4.51, 4.97) than expressions of anger ($M = 2.78$, $SD = 1.57$, CI: 2.55, 2.99). In addition, there was a significant main effect of reactive emotion, $F(3,441) = 10.35$, $p < 0.001$, $\eta_p^2 = 0.07$. *Post hoc* tests indicated that sadness was perceived as somewhat more intense when participants saw it alone ($M = 4.50$, $SD = 1.97$, CI: 4.19, 4.84) than in any other condition which did not differ ($M = 3.59$, $SD = 1.80$, CI: 3.28, 3.91; $M = 3.59$, $SD = 2.18$, CI: 3.31, 3.95; and $M = 3.22$, $SD = 1.98$, CI: 2.97, 3.61, for sadness – anger, sadness – happiness and, sadness – fear, respectively). Overall, these results indicate that the emotions of the first expresser were perceived as planned.

Perceived intensity of reactive emotions

A 2 (First expression) × 2 (Gender of expressers) × 3 (Reactive emotion) ANOVA was conducted on ratings of fear, happiness, anger and neutrality intensity. A main effect of reactive emotion emerged for all emotions, except for ratings of neutrality (see lower part of **Table 1**). For each emotion, as expected, ratings were highest on the scale that corresponded to the focal emotion for that expression.

For ratings of neutrality, the only effect that emerged was a main effect of gender, $F(1,334) = 8.08$, $p = 0.005$, $\eta_p^2 = 0.02$, indicating that men were rated as somewhat more neutral ($M = 2.43$, $SD = 1.68$, CI: 2.20, 2.67) than women ($M = 1.95$, $SD = 1.42$, CI: 1.72, 2.19).

For ratings of anger, an interaction between first emotion and reactive emotion emerged, $F(2,334) = 4.04$, $p = 0.018$, $\eta_p^2 = 0.02$. As can be seen in **Table 3**, anger expressions in response to anger, were perceived angrier than in any other condition. The next most intense rating was for anger expressions in response to sadness, which was higher than in

TABLE 3 | Perceived intensity of reactive emotions of anger and fear as a function of first expresser's emotion and reactive emotion – Study 3.

	First emotion											
	Anger						Sadness					
	Reactive emotion											
	Fear		Anger		Happiness		Fear		Anger		Happiness	
Perceived emotion	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI	*M SD*	95% CI
Anger	2.25_{cd} 1.58	1.84, 2.66	5.16_a 1.36	4.74, 5.58	2.13_c 1.67	1.71, 2.53	2.81_d 1.67	2.38, 3.26	4.52_b 1.90	4.09, 4.93	2.04_c 1.47	1.63, 2.49
Fear	5.03_a 1.80	4.61, 5.45	3.12_c 1.68	2.67, 3.53	2.15_d 1.48	1.70, 2.54	4.15_b 1.79	3.71, 4.60	3.33_c 1.76	2.90, 3.76	1.82_d 1.34	1.38, 2.27

Means with different subscripts differ at p < 0.05.

any remaining condition. When sadness was responded to with fear, fear expressions were rated as angrier than when sadness was responded to with happiness. No other differences between conditions emerged.

A significant interaction between first emotion and reactive emotion also emerged for fear ratings, $F(2,334) = 3.15$, $p = 0.04$, $\eta_p^2 = 0.02$, as the lower part of **Table 3** indicates, in response to anger, fear expressions were rated as more fearful than in any other condition, followed by the condition were fear was a response to sadness, which was still higher than in any other condition. Anger expressions in response to sadness or to anger, which did not differ, were rated as more fearful than happiness expressions in response to sadness or to anger which did not differ. Thus, overall, the focal emotion for each reactive emotion were perceived as planned; yet, as expected, expressions were rated as less intense and more mixed than the more intense expressions used in Study 1 and 2.

Perceived social power of first expresser

First, a 2 (First expression) × 2 (Gender of expressers) × 4 (Reactive emotion) ANOVA was conducted on ratings of social power. A significant main effect of first emotion emerged, $F(1,441) = 197.44$, $p < 0.001$, $\eta_p^2 = 0.31$, such that individuals who showed anger were rated as higher in power ($M = 4.87$, $SD = 0.83$, CI: 4.73, 4.98) than those who showed sadness ($M = 3.60$, $SD = 1.22$, CI: 3.45, 3.71). A significant main effect of second emotion, $F(3,441) = 14.43$, $p < 0.001$, $\eta_p^2 = 0.09$, was qualified by a first emotion by second emotion interaction $F(3,441) = 8.05$, $p < 0.001$, $\eta_p^2 = 0.05$. As shown in **Table 2**, compared to sadness shown alone, all reactive emotions increased perceptions of social power of the sad person. The increase was highest for anger followed by fear and significantly lower for happiness.

Compared to anger shown alone, anger reacted to with fear lead to increased attributions of social power. No other reactive emotion led to significantly different attributions when compared to anger alone.

Only the effect of fear responses to both angry and sad expressers replicated previous findings. It is curious that both happiness and anger when shown in response to sadness increased social power. In fact, we had expected that these expressions would either not impact on the power attributed to a sad person or reduce it. Also we had expected that anger

and happiness responses to angry expressions would reduce attributions of social power to an angry expresser, which was not found.

The MR model for attributions of power to sad expressions explained 13% of the variance, $F(4,161) = 6.07$, $p < 0.001$. Only reactive emotion ratings of fear significantly and positively predicted attributions of social power ($\beta = 0.25$, $p = 0.002$, CI: 0.09, 0.40). The MR model for attributions of power to anger expressions explained 12% of the variance, $F(4,175) = 5.88$, $p < 0.001$. Significant effects emerged for anger ($\beta = -0.23$, $p = 0.005$, CI: −0.41, −0.07) and happiness ($\beta = -0.23$, $p = 0.016$, CI: −0.41, −0.04) which both decreased attributed power. Thus, the regressions based on the actual ratings of the expressions yielded a different picture than the ANOVA comparing reactive emotion conditions with the ratings for the expression shown alone. The findings from the MR are also more congruent with findings from Study 2. The emotion ratings reported above, might give an insight into the reason for this. As intended, the emotions were more subtle and hence rated less intensely, but also less distinctly. Further, a stronger interaction between first and second emotion was observed. Hence, the categorical emotion conditions may not have reflected the actual perceived emotions as closely as was the case for Study 2. Thus, while we can say that reactive emotions did make a difference for the attribution of power when compared to judgments of the expression alone, the direction and intensity of the impact depend strongly on the actual emotion perceived rather than on categorical emotion conditions.

Perceived acceptance of power by the second person

We then assessed to what degree the reactive emotions were perceived as supporting the notion that the first person has high(er) power. The MR for sadness expressions explained 33% of the variance, $F(4,161) = 20.07$, $p < 0.001$. Both fear reactions ($\beta = 0.52$, $p < 0.001$, CI: 0.38, 0.66) and perceived neutrality ($\beta = 0.16$, $p = 0.015$, CI: 0.03, 0.29) predicted less acceptance of social power. The MR for anger expressions explained 24% of the variance, $F(4,175) = 13.84$, $p < 0.001$). Only reactions of fear ($\beta = 0.51$, $p < 0.001$, CI: 0.36, 0.65) were perceived as supporting the notion that the first person has high(er) power. The finding for fear replicates findings from Study 1 and 2, however,

neutrality did not contribute to the acceptance of power for either study. Also, the previously found effect for happiness as eroding a claim of high social power did not emerge.

Mediation Analysis

As for Study 2, we calculated two saturated path-models, one for each first emotion. For sadness expressions, significant indirect effects emerged for fear (β = 0.14, p = 0.005, CI: 0.07, 0.40) and neutrality (β = 0.04, p = 0.020, CI: 0.02, 0.33) such that to the degree that these emotions were seen as signaling that the first person has high(er) power, the sad expresser was rated as higher in social power. For anger expressions, only a significant positive indirect effect for fear emerged (β = 0.20, p < 0.001, CI: 0.10, 0.34).

DISCUSSION

Based on the mediation analysis we were able to replicate the finding that the fear reactions of the addressee of sadness or anger are interpreted as acceptance of the notion that the first person has (high)er social power and in turn this support leads to attributions of higher social power by the participants. The effect for happiness reactions, which decreased such attributions for anger in Study 2 and to some degree also in Study 1, could not be replicated. In addition, ratings of the neutrality of the emotional reaction of the addressee of a sad expression also positively predicted social power as mediated by acceptance of power.

The findings overall suggest that participants paid attention to the expressions of both interaction partners and based their judgment of the social power of the first person on the expressions of both. This even when the expressions were subtle and dynamically evolved and overlapped.

Yet, the incongruence between ANOVA results, which were based on the categorical label of the focal emotion expression and the regression analyses, which were based on the actual emotion ratings effectuated by the participants, suggest that the effects of subtle emotions, which are perceived as more mixed and less distinctive, can not necessarily be predicted by the focal emotion alone. This in turn points to the importance of secondary emotion ratings. This is also evident from the effect of neutrality observed here. In Studies 1 and 2 only neutral expressions (which were not included in Study 3) were rated as neutral. But the more subtle expressions used here were rated as somewhat neutral. One speculation could be that an expressive reaction that is seen as emotional but somewhat controlled or constrained and thus somewhat neutral is perceived as indicative of the social power of the person it is addressed to. This is an interesting question for future research as it suggests that efforts at emotion regulation could have social signal value in their own right.

GENERAL DISCUSSION

The congruent finding of all three studies suggests that reactive fear is a strong signal of the social power of another person. It also supports the notion that observers base judgments of social power not only on the expression of the person whose social power they judge but also on the reactions of their interaction partner. However, the effect of reactive emotions was clearest when prototypical expressions were shown as dynamic videos one after the other. This allowed participants to clearly see the expressions and facilitated their labeling. Once emotion expressions were more subtle and shown concurrently, only the effect of reactive fear remained stable. More importantly in Study 3 it became evident that not only the social signal value of the focal emotion (i.e., fear for a fear expression) but also the secondary emotions that can be perceived in such expressions (i.e., neutrality in a fear expression or fear in a happy expression) are relevant. This is an interesting finding as previous research on the attribution of social power based on facial expressions has not only focused exclusively on the expression of the person whose power is to be judged but also exclusively focused on focal emotions, completely neglecting secondary emotions.

That secondary emotions are of importance in social interactions has been shown in recent research that links the perception of secondary emotions to the perception of social interaction quality such that to the degree that people perceive more intense secondary emotions they report less satisfying social interactions (Hess et al., 2016b). If secondary emotions also interfere with the perception of social attributes such as power or affiliation, this could be one path to explain this reduced social interaction quality.

In sum, the results of three studies suggest that the emotional reactions of the addressee of emotion expressions are meaningful signals which are used to infer the social power of the sender of the first expression. As discussed above, social power can be best conceived of as a person's ability to influence others (Keltner et al., 2003). Emotion's expressions can serve as cues to this ability (Knutson, 1996; Tiedens, 2001; Hareli et al., 2009) but also as signals of power (Hareli and David, 2017; Scarantino, 2017). Accordingly, the way an interaction partner reacts to such expressions is important for the degree to which such a signal should be believed. Hence, observers should prevail themselves of this information and they do.

The results also suggest that this basic finding is not depended on the somewhat artificial approach chosen by Hareli and David (2017) and by us in Study 1 and to a lesser degree in Study 2. Overall, the use of a set of studies that gradually and in a controlled manner add to the complexity involved in the perception of emotions in a social interaction enabled us to carefully assess the factors that influence how reactive emotions contribute to social judgments of power. Taken together, our research shows that the social signal value of emotion expressions depends in part on the emotional reaction of the interaction partner. Thus, the social signal value of emotions does not stand alone but has to be understood in the fuller context of the interaction. The present research highlights the importance of studying the social signal value of emotions in an interactional context and to acknowledge that observers do not necessarily perceive emotions as "pure" instantiations of a single emotional state, but more often as mixed. This is especially the case for the

more subtle dynamic emotion displays that are typical for real life interactions.

ETHICS STATEMENT

This study was carried out in accordance with the ethical standards of the American Psychological Association and the ethics committee of the faculty of management of the University of Haifa. The protocol was approved by the ethics committee of the faculty of management of the University of Haifa. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

SH planned the studies and supervised its running and contributed in writing the paper, and conducing the analyses. MH was running the studies, created the experiments, and helped with the writing and analysis. UH helped plan the studies, contributed in writing the paper, and was leading the analyses.

FUNDING

REFERENCES

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Averill, J. R. (1997). "The emotions: an integrative approach," in *Handbook of Personality Psychology*, eds R. Hogan, J. A. Johnson, and S. R. Briggs (San Diego, CA: Academic Press), 513–541. doi: 10.1016/B978-012134645-4/50022-6

Darwin, C. (1872/1965). *The Expression of the Emotions in Man and Animals.* Chicago, IL: The University of Chicago Press.

Dawkins, R., and Krebs, J. R. (1978). Animal signals: information or manipulation. *Behav. Ecol. Evol. Approach* 2, 282–309.

Eibl-Ebesfeldt, I. (1989). *Human Ethology.* New York, NY: Aldine de Gruyter.

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068

Ekman, P., Friesen, W. V., and Ellsworth, P. (1972). *Emotion in the Human Face: Guidelines for Research and an Integration of Findings.* New York, NY: Pergamon Press.

Fischer, A. H., and Manstead, A. S. R. (2008). "Social functions of emotion," in *Handbook of Emotions*, 3 Edn, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York: Guilford Press), 456–468.

Fischer, A. H., and Hess, U. (2017). Mimicking emotions. *Curr. Opin. Psychol.* 17, 151–155. doi: 10.1016/j.copsyc.2017.07.008

Fischer, A. H., and Roseman, I. J. (2007). Beat them or ban them: the characteristics and social functions of anger and contempt. *J. Pers. Soc. Psychol.* 93, 103–115. doi: 10.1037/0022-3514.93.1.103

Fridlund, A. J. (1994). *Human Facial Expression: An Evolutionary View.* New York, NY: Academic Press.

Frijda, N. H., and Mesquita, B. (1994). "The social roles and functions of emotions," in *Emotion and Culture: Empirical Studies of Mutual Influence*, eds S. Kitayama and H. R. Markus (Washington, DC: American Psychological Association), 51–87. doi: 10.1037/10152-002

Hahn, E., Gottschling, J., and Spinath, F. M. (2012). Short measurements of personality–validity and reliability of the GSOEP big five inventory (BFI-S). *J. Res. Pers.* 46, 355–359. doi: 10.1016/j.jrp.2012.03.008

Hareli, S. (2014). Making sense of the social world and influencing it by using a naïve attribution theory of emotions. *Emot. Rev.* 6, 336–343. doi: 10.1177/1754073914534501

Hareli, S., and David, S. (2017). The effect of reactive emotions expressed in response to another's anger on inferences of social power. *Emotion* 17, 717–727. doi: 10.1037/emo0000262

Hareli, S., David, S., and Hess, U. (2016). The role of emotion transition for the perception of social dominance and affiliation. *Cogn. Emot.* 30, 1260–1270. doi: 10.1080/02699931.2015.1056107

Hareli, S., and Rafaeli, A. (2008). Emotion cycles: on the social influence of emotions in organizations. *Res. Organ. Behav.* 28, 35–59. doi: 10.1016/j.riob.2008.04.007

Hareli, S., Shomrat, N., and Hess, U. (2009). Emotional versus neutral expressions and perceptions of social dominance and submissiveness. *Emotion* 9, 378–384. doi: 10.1037/a0015958

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach.* New York, NY: Guilford Publications.

Hess, U., and Fischer, A. H. (2013). Emotional mimicry as social regulation. *Pers. Soc. Psychol. Rev.* 17, 142–157. doi: 10.1177/1088868312472607

Hess, U., Adams, R. B. Jr., and Kleck, R. E. (2008). "The role of perceived emotion in first impressions," in *First Impressions*, eds N. Ambady and J. J. Skowronski (New York, NY: Guilford Press), 234–254.

Hess, U., and Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *Int. J. Psychophysiol.* 40, 129–141. doi: 10.1016/S0167-8760(00)00161-6

Hess, U., David, S., and Hareli, S. (2016a). Emotional restraint is good for men only: the influence of emotional retraint on the perception of competence. *Emotion* 16, 208–213. doi: 10.1037/emo0000125

Hess, U., Kafetsios, K., Mauersberger, H., Blaison, C., and Kessler, C. L. (2016b). Signal and noise in the perception of facial emotion expressions: from labs to life. *Pers. Soc. Psychol. Bull.* 42, 1092–1110. doi: 10.1177/0146167216651851

Hess, U., Kappas, A., and Banse, R. (1995). The intensity of facial expressions is determined by underlying affective state and social situation. *J. Pers. Soc. Psychol.* 69, 280–288. doi: 10.1037/0022-3514.69.2.280

Hess, U., Philippot, P., and Blairy, S. (1999). "Mimicry: facts and fiction," in *The Social Context of Nonverbal Behavior. Studies in Emotion and Social Interaction*, eds P. Philippot and R. S. Feldman (Cambridge: Cambridge University Press), 213–241.

Keltner, D. (1995). Signs of appeasement: evidence for the distinct displays of embarrassment, amusement, and shame. *J. Pers. Soc. Psychol.* 68, 441–454. doi: 10.1037/0022-3514.68.3.441

Keltner, D., Gruenfeld, D. H., and Anderson, C. (2003). Power, approach, and inhibition. *Psychol. Rev.* 110, 265–284. doi: 10.1037/0033-295X.110.2.265

Keltner, D., and Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cogn. Emot.* 13, 505–521. doi: 10.1080/026999399379168

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *J. Nonverbal Behav.* 20, 165–182. doi: 10.1007/BF02281954

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., and Van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cogn. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076

Lerner, J. S., and Tiedens, L. Z. (2006). Portrait of the angry decision maker: how appraisal tendencies shape anger's influence on cognition. *J. Behav. Decis. Mak.* 19, 115–137. doi: 10.1002/bdm.515

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychol. Methods* 23, 412–433. doi: 10.1037/met0000144

Motley, M. T., and Camden, C. T. (1988). Facial expression of emotion: a comparison of posed expressions versus spontaneous expressions in an interpersonal communications setting. *West. J. Speech Commun.* 52, 1–22. doi: 10.1080/10570318809389622

Robinson, M. D., Johnson, J. T., and Shields, S. A. (1995). On the advantages of modesty: the benefits of a balanced self-presentation. *Commun. Res.* 22, 575–591. doi: 10.1177/009365095022005003

Russell, B. (1938). *Power: A New Social Analysis*. London: Allen and Unwin.

Russell, J. A., and Fehr, B. (1987). Relativity in the perception of emotion in facial expressions. *J. Exp. Psychol. Gen.* 116, 223–237. doi: 10.1037/0096-3445.116.3.223

Russell, J. A., Suzuki, N., and Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expressions. *Motiv. Emot.* 17, 337–351.

Scarantino, A. (2017). How to do things with emotional expressions: the theory of affective pragmatics. *Psychol. Inq.* 28, 165–185. doi: 10.1080/1047840X.2017.1328951

Scherer, K. R. (1987). Towards a dynamic theory of emotion: the component process model of affective states. *Geneva Stud. Emot. Commun.* 1, 1–98.

Shariff, A. F., and Tracy, J. L. (2011). What are emotion expressions for? *Curr. Dir. Psychol. Sci.* 20, 395–399. doi: 10.1177/0963721411424739

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative emotion expressions on social status conferral. *J. Pers. Soc. Psychol.* 80, 86–94. doi: 10.1037/0022-3514.80.1.86

Tiedens, L. Z., Ellsworth, P. C., and Mesquita, B. (2016). Sentimental stereotypes: emotional expectations for high-and low-status group members. *Pers. Soc. Psychol. Bull.* 26, 560–575. doi: 10.1177/0146167200267004

Van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the amsterdam dynamic facial expression set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

Van Kleef, G. A. (2009). How emotions regulate social life. *Curr. Dir. Psychol. Sci.* 18, 184–188. doi: 10.1111/j.1467-8721.2009.01633.x

Van Kleef, G. A. (2010). The emerging view of emotion as social information. *Soc. Pers. Psychol. Compass* 4, 331–343. doi: 10.1111/j.1751-9004.2010.00262.x

Yrizarry, N., Matsumoto, D., and Wilson-Cohn, C. (1998). American-Japanese differences in multiscalar intensity ratings of universal facial expressions of emotion. *Motiv. Emot.* 22, 315–327. doi: 10.1023/A:1021304407227

Human Observers and Automated Assessment of Dynamic Emotional Facial Expressions: KDEF-dyn Database Validation

Manuel G. Calvo[1,2], Andrés Fernández-Martín[3], Guillermo Recio[4*] and Daniel Lundqvist[5]

[1] Department of Cognitive Psychology, Universidad de La Laguna, San Cristóbal de La Laguna, Spain, [2] Instituto Universitario de Neurociencia (IUNE), Universidad de La Laguna, Santa Cruz de Tenerife, Spain, [3] Department of Health Sciences, Universidad Internacional de la Rioja, Logroño, Spain, [4] Institute of Psychology, Universität Hamburg, Hamburg, Germany, [5] Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Most experimental studies of facial expression processing have used static stimuli (photographs), yet facial expressions in daily life are generally dynamic. In its original photographic format, the Karolinska Directed Emotional Faces (KDEF) has been frequently utilized. In the current study, we validate a dynamic version of this database, the KDEF-dyn. To this end, we applied animation between neutral and emotional expressions (happy, sad, angry, fearful, disgusted, and surprised; 1,033-ms unfolding) to 40 KDEF models, with morphing software. Ninety-six human observers categorized the expressions of the resulting 240 video-clip stimuli, and *automated face analysis* assessed the evidence for 6 expressions and 20 facial action units (AUs) at 31 intensities. Low-level image properties (luminance, signal-to-noise ratio, etc.) and other purely perceptual factors (e.g., size, unfolding speed) were controlled. Human recognition performance (accuracy, efficiency, and confusions) patterns were consistent with prior research using static and other dynamic expressions. Automated assessment of expressions and AUs was sensitive to intensity manipulations. Significant correlations emerged between human observers' categorization and automated classification. The KDEF-dyn database aims to provide a balance between experimental control and ecological validity for research on emotional facial expression processing. The stimuli and the validation data are available to the scientific community.

Keywords: facial expression, dynamic, action units, KDEF, FACET

INTRODUCTION

Research on facial expression processing (see reviews in Nelson and Russell, 2013; Calvo and Nummenmaa, 2016) has generally utilized static faces as stimuli, obtained from standardized databases such as the Pictures of Facial Affect (PoFA; Ekman and Friesen, 1976), the Karolinska Directed Emotional Faces (KDEF; Lundqvist et al., 1998), the NimStim Stimulus Set (Tottenham et al., 2002), the Radboud Faces Database (RaFD; Langner et al., 2010), FACES (Ebner et al., 2010) and others (for a review and evaluation, see Cowie et al., 2005; Anitha et al., 2010; Sandbach et al., 2012). Yet, in social encounters and face-to-face communication, facial expressions are generally

dynamic. Further, research has shown that motion benefits affect recognition (see Krumhuber et al., 2013; Calvo et al., 2016; Wingenbach et al., 2016). Accordingly, it is important to use dynamic stimuli for investigating recognition of facial expressions.

A number of *dynamic* expression databases have been developed, generally involving on-line video recordings of facial activity, which represent a valuable advance (e.g., van der Schalk et al., 2011; Banziger et al., 2012; Kaulard et al., 2012; Zhang et al., 2014; O'Reilly et al., 2016; Wingenbach et al., 2016). Krumhuber et al. (2017) have reviewed and discussed the major issues of 22 dynamic expression databases. In the current study, the proposal of a new stimulus set (KDEF-dyn) aims to make a contribution by taking two issues into account. First, the control of possible *perceptual confounds* with non-expressive factors that may affect expression recognition. They involve low-level image properties of the stimuli, such as illumination and light source, size of the face relative to the background, head-face orientation, or changes in facial appearance like hair, make up, eyeglasses, jewelry, etc. They may be difficult to control for in video-recordings of spontaneous expressions. Yet, to unequivocally attribute emotion recognition to facial expression *per se*, all the facial stimuli across types of expressions must be comparable on these non-expressive factors. Further, the control of such factors may be critical for paradigms using neurophysiological (such as event-related potentials, ERPs; see Naples et al., 2015) or eyetracking (e.g., probability of first fixation in a particular face region, or pupillometry; e.g., Calvo and Nummenmaa, 2011) measures, which are particularly sensitive to physical image properties. To this end, all the face stimuli in our KDEF-dyn set are standardized in size, resolution, location, and frontal view, in addition to multiple low-level image properties (luminance, contrast, etc.).

A second issue is concerned with the objective validation of expressions and component facial actions across *multiple intensities*. According to Valstar et al. (2015, 2017), many existing benchmark databases show expressions at fixed intensities (generally, the apex or maximum intensity) or do not support the evaluation of intensity effects. Computational algorithms have been developed to automatically detect Facial Action Coding System (FACS) action units (AUs; Ekman et al., 2002), which are anatomical changes in the facial morphology that can be associated to specific emotions (e.g., AU12 or lip corner puller, to happiness; or AU4, brow lowerer, to anger; etc.). Manual FACS-coding by expert raters (van der Schalk et al., 2011; Banziger et al., 2012), and also automated computation (Lucey et al., 2010; Cosker et al., 2011; Mavadati et al., 2013; Zhang et al., 2014), have been applied to dynamic expression databases only on the apex. The estimation at multiple intensities is, however, required because, in real life, expressions vary in intensity, which is often a critical cue to interpret their meaning. Accordingly, we computed the objective evidence of each of six basic expressions and also the evidence of each of 20 AUs, across 31 intensities from neutral (0% intensity) to emotional (100% intensity) in 3.33% intensity steps. This adds to recent work (Calvo et al., 2016; Wingenbach et al., 2016) regarding the role of intensity on the categorization of dynamic expressions. This approach will be particularly useful for expression discrimination studies, e.g., the lowest intensity

or threshold at which a particular emotion is recognized and differentiated from others and from neutral faces.

With these two issues in mind, in the current study we developed and validated a dynamic version (KDEF-dyn) of the original KDEF database in static format (Lundqvist et al., 1998), to extend research possibilities. The photographic KDEF stimuli have been validated in large norming studies (Calvo and Lundqvist, 2008; Goeleven et al., 2008), and widely used in behavioral (e.g., Calvo et al., 2013; Sanchez et al., 2014; Gupta et al., 2016) and neurophysiological (e.g., Bublatzky et al., 2014; Calvo and Beltrán, 2014; Adamaszek et al., 2015) research. The original KDEF database has been cited in over 1,980 published articles, according to Google Scholar[1] (accessed 18.09.2018). We took advantage of this research on the static KDEF stimuli to produce dynamic expressions of 40 different models, each portraying the six basic emotions.

To *develop* dynamic expressions, we applied morphing animation software (FantaMorph, v. 5.4.2; Abrosoft) to the original KDEF photographs. For each encoder and emotion, we created a 1,033-ms video-clip of 31 frames starting with a neutral face and ending with a full-blown emotional face. Thus, we tried to mimic real-life expressions and approximate the average natural speed of emotional expression development from a neutral face, since apex of facial expression is generally reached within 1 s for basic emotions (Pollick et al., 2003; Hoffmann et al., 2010). Admittedly, dynamic morphing creates linear movement, which can make expressions appear as less natural than on-line video recordings. Nevertheless, although non-linear changes are generally judged as more natural than linear motion, morphing does not necessarily compromise naturalness (Cosker et al., 2010, 2015). In fact, dynamically morphed facial expressions have often been employed in prior research on facial emotion recognition, with behavioral (Hoffmann et al., 2010; Fiorentini and Viviani, 2011; Recio et al., 2013; Calvo et al., 2016) and neurophysiological (Popov et al., 2013; Harris et al., 2014; Recio et al., 2014; Vrticka et al., 2014) measures being sensitive to expression manipulations. The morphing technique involves some advantages, such as fine-grained control and standardization of expressive intensity, unfolding speed, and duration. We chose this approach as a balance between (reduced) ecological validity and (enhanced) experimental control.

To *validate* the KDEF-dyn database, we followed two approaches, each with several measures. First, we collected data from *human observers* in an expression categorization task including measures of (a) correct recognition responses, i.e., the probability that they coincided with the intended KDEF expression, (b) reaction times indicating processing efficiency, and (c) the probability of confusions across different expressions, for each of the six basic emotions. Second, with Emotient FACET software (v. 6.1.2667.3; iMotions), we performed *automated facial expression analyses* (Bartlett and Whitehill, 2011; Olderbak et al., 2014; Cohn and De la Torre, 2015; Girard et al., 2015; Dente et al., 2017) of (a) the probability of each expression to be detected, as a function of spatial maps of

---

[1]https://scholar.google.com/scholar?cites=93971208802805184&as_sdt=2005

facial features, and also (b) the probability of each of 20 AUs to be activated, i.e., muscle movements, according to FACS (Ekman and Friesen, 1978; Ekman et al., 2002). The automated analyses of expressions and AUs were performed for 31 intensities (including the neutral baseline) of each emotional facial expression (including apex), while the human recognition measures were obtained for the maximum expressive intensity only. These measures indicate to what extent each KDEF stimulus is consistently categorized, the objective evidence for each facial expression configuration, and the specific morphological features.

The current KDEF-dyn database contributes to existing databases of dynamic facial expression stimuli in several respects. First, the combined validation approach (with both 'subjective' human categorization data and 'objective' automated assessment data) provides researchers with empirical and theoretical criteria to select stimuli depending on various dimensions (recognition accuracy and efficiency, susceptibility to specific confusions, and automated classification of expressions and AUs). In a dataset file (see **Supplementary Dataset S1**), each stimulus can be ordered according to each of these measures. Second, due to the standardization of expression unfolding speed and duration for all the stimuli, the present database allows for a fine-grained investigation of emotion recognition as a function of expressive intensity. We provide evidence values from automated analysis of expressions and AUs for each frame of each video-clip. In a dataset file (see **Supplementary Dataset S2**), such values are shown for each of 31 intensity levels of each stimulus, from 0 (neutral) to 100% (full-blown emotion). Third, another novel contribution involves the control of multiple non-expressive perceptual factors (e.g., low-level image properties) that might otherwise confound expression recognition differences. In a dataset file (see **Supplementary Dataset S3**), each stimulus has been quantified in terms of such perceptual factors across each of 31 expressive intensity levels. Potential applications and limitations will be considered in the Section "Discussion."

MATERIALS AND METHODS

Participants

Ninety-six university undergraduates (56 females and 40 males; aged 18–30 years; M = 21.2 years) from different courses (Psychology, Medicine, Law, Economics, and Education) participated voluntarily for payment (5 €) or course credit, after signing written informed consent. Four more participants were excluded from the analyses because their mean correct recognition rate was below 50% for three or more expressions. An *a priori* power calculation using G^*Power (v. 3.1.9.2; Faul et al., 2007) showed that 46 participants would be sufficient to detect a medium effect size (Cohen's d = 0.60) at α = 0.05, with power of 0.98. As this was a norming study of stimulus materials, a larger participant sample was used to obtain stable and representative average scores for each stimulus. The study was approved by the Ethics Committee of University of La Laguna (protocol CEIBA2017-0227), and was conducted in accordance with the Declaration of Helsinki 2008.

Stimuli

The color photographs of 40 posers (20 females and 20 males) in frontal view from the KDEF database (Lundqvist et al., 1998) displaying six emotional facial expressions (happiness, sadness, anger, fear, disgust, and surprise) were used. The KDEF identities (see **Supplementary Dataset S1**) were the same as in a previous norming study using photographic stimuli (Calvo and Lundqvist, 2008). For the current study, 240 dynamic video-clip versions (1,033-ms duration) of the original KDEF photographs were constructed. The face stimuli were morphed with FantaMorph (Abrosoft) computer software. For each expression of each poser, we created a 1,033-ms sequence of 31 (33.33-ms) frames smoothly increasing expressive intensity at 30 frames per second (fps), starting with a neutral face as the first frame (frame 0; original KDEF), and ending with an emotional face (happy, sad, etc.) as the final frame (frame 30; original KDEF). Video-clips are shown as supporting information (see **Supplementary Dataset S4**). A very similar or identical procedure and display duration was used previously (Schultz and Pilz, 2009; Johnston et al., 2013; Wingenbach et al., 2016). Each face stimulus subtended a visual angle of 10.6° (height) × 8° (width) at a 70-cm viewing distance (this approximates the size of a real face, i.e., 18.5 × 13.8 cm, from a 1-m distance).

Procedure

The 96 participants were presented with all 240 video-clips (40 posers × 6 expressions) in six blocks of 40 trials each, and a short break after each block. Block order was counterbalanced, and trial order and type of expression were randomized within each block. The stimuli were displayed on a computer screen (12-in TFT LED LCD with a 1,366 × 768 resolution) by means of E-Prime 2.0 software. Participants were told that short videos of faces with different expressions would be presented, and were asked to indicate which expression was shown on each trial, by pressing a key out of six, as soon and as accurately as possible, with their dominant index finger. Between trials, the index finger was placed at a predetermined location in the middle of the spacebar, equidistant from all six response keys (from 4 to 9). During the instructions, the six basic expressions were identified, as well as the location of the keys to be pressed for each category. Twelve video-clips of two additional, non-KDEF encoders displaying six emotional expressions served as practice trials.

The sequence of events on each trial was as follows. After an initial 500-ms central fixation cross on a screen, a video-clip showed a facial expression that unfolded for 1,033 ms. Following face offset, graphical instructions appeared on the screen for responding: Six small boxes were arranged horizontally, numbered from 4 to 9, with each box/number associated to a verbal label (e.g., 4: happy; 5: sad, etc.). The assignment of expressions to numbers was counterbalanced across participants. For categorizing each expression, participants pressed one key (from 4 to 9) in the upper row of a standard computer keyboard. The selected response and reaction times (RTs; from the video-clip offset) were recorded. There was a 1,500-ms intertrial interval.

Design and Measures

We used a within-subjects experimental design, with expressive category (happiness, sadness, anger, fear, disgust, and surprise) as a factor. As dependent variables, we measured hits, i.e., the probability that responses coincided with the displayed expression (e.g., responding "happy" when the face stimulus was intended to convey happiness), and RTs. In addition, we identified the type of confusions, i.e., the probability that each target (the actually displayed expression) was categorized as each of the other five, non-target expressions (e.g., if the target was anger on a trial, the five non-targets were happiness, sadness, disgust, fear, and surprise). These measures, along with those involving automated expression analysis (see below), are provided as supplementary data for each KDEF-dyn stimulus (see **Supplementary Dataset S1**).

Automated Facial Expression Analysis

In addition to the human observers' performance measures, we subjected the video stimuli to automated face analysis by means of Emotient FACET software, which is assumed to detect facial features (e.g., mouth corners) and feature groups, and then to classify the image as belonging to a particular emotional expression category by comparing the resulting output maps with template images. Recently, FACET has been used in psychological and applied research (see Dente et al., 2017). The automated analysis provides two types of measures (see Gordon et al., 2011; Olderbak et al., 2014): (a) *expression* evidence scores for each category: joy, anger, surprise, fear, disgust, sadness, and contempt, in addition to neutral; and (b) AUs evidence scores (for 20 AUs: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 18, 20, 23, 24, 25, 26, 28, and 43), according to FACS (Ekman et al., 2002); see also (Cohn et al., 2007; Cohn and De la Torre, 2015). AUs are anatomically related to the movement of specific face muscles (e.g., AU12 involves the contraction of the zygomaticus major muscle, which draws the angle of the mouth superiorly and posteriorly to allow for smiling).

We obtained expression and AU evidence scores for each of 31 frames across the 1,033-ms unfolding, for each poser and expression (see **Supplementary Dataset S2**). The FACET evidence scores quantify the odds (in decimal logarithmic scale) of each expression or AU to be present in a given face stimulus, and can be transformed into probabilities (p) with the formula $p = 1/(1 + 10^{-\text{evidence score}})$. An evidence score of zero indicates chance level (0.50/0.50). Positive values indicate greater probabilities that a given expression or AU is present, and negative values indicate greater probabilities that an expression or AU is unlikely to be present in the stimulus. All evidence scores above 1 will approach the probability value of 1, and all evidence scores below −1 will approach a 0 probability. This implies that evidence scores (in odds ratios) are more discriminative than probabilities to detect subtle changes, and the former are more suitable for statistical tests because they tend to be normally distributed. The evidence scores ranged in a continuous scale between −12 and 12. We conducted Kolmogorov–Smirnov and Levene's tests to exam the assumptions of ANOVA regarding normality and homoscedasticity, respectively. Results revealed

that most residuals of the evidence scores for expressions and AUs were normally distributed and homoscedastic (for multivariate ANOVA with the evidence scores used as dependent variables and expression category as a fixed factor; see **Supplementary Dataset S2**).

Low-Level Stimulus Image Properties

To examine potential physical and perceptual differences among expression categories across the 1,033-ms unfolding display, we computed (with Matlab 7.0, The Mathworks) the following low-level image statistics of each neutral face and the respective emotional faces for each of 31 frames, at consecutive expressive intensity levels, from 0% intensity (i.e., neutral face) to full-blown emotion (i.e., 100% intensity), in 3.33% steps: mean and variance of luminance, RMS or root mean square contrast, skewness, kurtosis, SNR or signal-to-noise ratio, and entropy. Each low-level property was analyzed by means of a (6: Expression Stimulus) × 31 (Intensity Levels) ANOVA. All the measures were sensitive to the effects of intensity, all $Fs(30,7020) \geq 38.44$, $p < 0.0001$, $\eta_p^2 \geq 0.14$), but, importantly, the main effect of expression was never significant (all $Fs < 1$, except for skewness: $F(5,234) = 1.51$, $p = 0.19$, *ns*; see **Supplementary Dataset S3**). Accordingly, the face stimuli of the different expressions did not significantly differ in such physical properties. This rules out purely perceptual factors as responsible for the differences observed in categorization performance by human observers or automated facial expression classification (see below).

RESULTS

We wanted to relate human observers' performance and automated facial expression analysis, which had to be conducted for each *stimulus*. Further, the study aimed to obtain and provide other researchers with validation measures for each *stimulus* (i.e., KDEF model identity). Accordingly, the statistical analyses were performed on the stimuli as the error term. This means that the recognition performance scores of the 96 participants were averaged for each of the 240 video-clip stimuli, which served as the units of analysis, with an $N = 40$ for each expression category. All the multiple *post hoc* comparisons in the following analyses involved Bonferroni corrections (with a $p < 0.05$ threshold).

Analyses of Recognition Performance and Confusions by Human Observers

For response *accuracy*, a one-way (6: Expression) ANOVA yielded significant effects, $F(5,234) = 32.07$, $p < 0.0001$, $\eta_p^2 = 0.41$. *Post hoc* contrasts revealed significantly better recognition of happiness, surprise, and anger, than sadness and disgust, which were recognized better than fear (see **Table 1**). The correct response *reaction times*, $F(5,234) = 69.91$, $p < 0.0001$, $\eta_p^2 = 0.60$, were faster for happiness than for any other expression, followed by surprise and anger (which did not differ from each other), and by disgust and sadness (which did not differ from each other), with fear being recognized more slowly than the other categories. Pairwise (Pearson) *correlations* between response accuracy and

TABLE 1 | Mean proportion (%) of hits and confusions in human observers' responses, and reaction times (for hits only) for each target (stimulus) expression.

	Expression response					
Expression stimulus	**Happiness**	**Surprise**	**Anger**	**Sadness**	**Disgust**	**Fear**
Happiness	**98.5**[a]	1.0[b]	0.0[b]	0.0[b]	0.3[b]	0.2[b]
Surprise	2.8[b]	**93.7**[a]	0.1[c]	0.0[c]	0.2[c]	3.2[b]
Anger	0.2[c]	0.8[bc]	**91.7**[a]	1.3[bc]	3.6[b]	2.4[b]
Sadness	0.5[c]	0.7[c]	1.8[c]	**80.7**[a]	5.7[b]	10.6[b]
Disgust	0.1[d]	0.5[d]	13.4[b]	4.7[bc]	**77.8**[a]	3.5[c]
Fear	0.8[d]	18.5[b]	1.1[d]	2.5[d]	8.5[c]	**68.6**[a]
Hits	98.5[a]	93.7[a]	91.7[a]	80.7[b]	77.8[b]	68.6[c]
Hit RTs	868[a]	1,061[b]	1,140[b]	1,253[c]	1,229[c]	1,431[d]

Within each expression stimulus category (horizontally), scores with different letters across expression response (i.e., on the same line) are significantly different in post hoc multiple contrasts (p < 0.05, Bonferroni corrected); expressions sharing a letter are equivalent. Boldface for hits in columns.

TABLE 2 | Mean raw evidence scores (odds ratios) of each expression (response) for each target (stimulus) expression.

	Expression response						
Expression stimulus	**Happiness**	**Surprise**	**Anger**	**Sadness**	**Disgust**	**Fear**	**Neutral**
Happiness	**6.4**[a]	−8.6[e]	−7.7[d]	−9.2[e]	−5.4[c]	−3.9[b]	−11.9[f]
Surprise	−5.6[d]	**3.5**[a]	−3.9[c]	−6.1[d]	−4.5[c]	0.8[b]	−4.8[cd]
Anger	−6.3[e]	−4.8[d]	**1.7**[a]	−2.8[c]	−0.4[b]	−2.8[c]	−2.6[bc]
Sadness	−4.6[e]	−4.1[e]	−2.3[d]	**1.7**[a]	−1.3[c]	−0.2[b]	−2.2[cd]
Disgust	−5.4[d]	−8.2[e]	−1.6[b]	−5.7[d]	**4.2**[a]	−3.8[c]	−8.2[e]
Fear	−4.0[d]	−1.3[b]	−2.9[cd]	−3.1[cd]	−2.1[bc]	**1.6**[a]	−4.4[de]
Target	6.4[a]	3.5[b]	1.7[c]	1.7[c]	4.2[b]	1.6[c]	

Automated analysis computed by Emotient FACET software. Within each expression stimulus category (horizontally), scores with different letters across expression response (i.e., on the same line) are significantly different in post hoc multiple contrasts (p < 0.05, Bonferroni corrected); expressions sharing a letter are equivalent. Boldface for correct responses to target (stimulus) expressions. Target: correct classification of each stimulus.

reaction times for all the expressions showed that reaction times decreased as accuracy increased (Happiness: $r = -0.67$; Surprise: $r = -0.72$; Anger: $r = -0.78$; Sadness: $r = -0.64$; Disgust: $r = -0.81$; Fear: $r = -0.71$; all $ps < 0.0001$; $N = 40$).

For the analysis of *confusions*, a 6 (Expression Stimulus) × 6 (Expression Response) ANOVA was conducted. Interactive effects, $F(25,1170) = 836.53$, $p < 0.0001$, $\eta_p^2 = 0.95$, were decomposed by means of separate one-way (6: Expression Response) ANOVAs for each expression stimulus. See the mean scores and multiple contrasts in **Table 1**. Facial *happiness*, $F(5,195) = 11922.15$, $p < 0.0001$, $\eta_p^2 = 1$, was very unlikely to be confused. *Surprise*, $F(5,195) = 2952.68$, $p < 0.0001$, $\eta_p^2 = 0.99$, was slightly confused with fear and happiness. *Anger*, $F(5,195) = 1625.02$, $p < 0.0001$, $\eta_p^2 = 0.98$, was slightly confused with disgust and fear. *Sadness*, $F(5,195) = 427.46$, $p < 0.0001$, $\eta_p^2 = 0.92$, was confused with fear and disgust more than with other expressions. *Disgust*, $F(5,195) = 228.31$, $p < 0.0001$, $\eta_p^2 = 0.85$, was confused with anger and sadness, followed by fear. Finally, *fear*, $F(5,195) = 315.88$, $p < 0.0001$, $\eta_p^2 = 0.89$, was confused with surprise, followed by disgust.

Automated Assessment of Expressions With FACET

The evidence scores for each expression were subjected to a 6 (Expression Stimulus) × 7 (Expression Response,

i.e., the six basic emotions plus neutral) ANOVA. Main effects of expression stimulus, $F(5,234) = 73.25$, $p < 0.0001$, $\eta_p^2 = 0.61$, and response, $F(6,1404) = 142.17$, $p < 0.0001$, $\eta_p^2 = 0.38$, and an interaction, $F(30,1404) = 152.43$, $p < 0.0001$, $\eta_p^2 = 0.77$, emerged. To decompose the interaction, separate one-way (7: Expression Response) ANOVAs were conducted for each expression stimulus. All the expressions were correctly classified (e.g., facial happiness was classified as joy), with target responses being significantly higher (after Bonferroni corrections) than alternative responses (e.g., happiness classified as surprise, etc.), which were assigned negative scores: Facial *happiness*, $F(6,234) = 636.60$, $p < 0.0001$, $\eta_p^2 = 0.94$; *surprise*, $F(6,234) = 150.16$, $p < 0.0001$, $\eta_p^2 = 0.79$; *anger*, $F(6,234) = 66.31$, $p < 0.0001$, $\eta_p^2 = 0.63$; *sadness*, $F(6,234) = 61.98$, $p < 0.0001$, $\eta_p^2 = 0.61$; *disgust*, $F(6,234) = 196.70$, $p < 0.0001$, $\eta_p^2 = 0.86$; and *fear*, $F(6,234) = 31.44$, $p < 0.0001$, $\eta_p^2 = 0.45$. The interaction reflected the fact that the correct response scores were higher for happy expressions, followed by disgust and surprise (which did not differ from each other), followed by anger, sadness, and fear (which did not differ from one another), as indicated by a one-way (6: Expression Stimulus) ANOVA, $F(5,234) = 64.34$, $p < 0.0001$, $\eta_p^2 = 0.58$, and multiple *post hoc* comparisons. See the mean scores and contrasts in **Table 2**.

Automated Assessment of Expressive Intensity With FACET

To examine expression classification by FACET as a function of expressive *intensity*, we conducted a 6 (Stimulus Expression) $\times$ 31 (Intensity Levels: 0% or neutral, 3.3%, 6.7%, etc., and 100% or full-blown emotion) ANOVA on the evidence scores. Effects of expression, $F(5,7254) = 420.79$, $p < 0.0001$, $\eta_p^2 = 0.23$, intensity, $F(30,7254) = 593.43$, $p < 0.0001$, $\eta_p^2 = 0.71$, and an interaction, $F(150,7254) = 23.66$, $p < 0.0001$, $\eta_p^2 = 0.33$, emerged. Separate one-way (Intensity: 31) ANOVAs were performed for each expression to determine the intensity *threshold*, i.e., when significant evidence of each emotion started relative to the neutral face baseline. Facial *happiness*, $F(30,1209) = 232.76$, $p < 0.0001$, $\eta_p^2 = 0.85$, started to be correctly classified as such at 13.3% intensity ($p = 0.003$, after Bonferroni corrections); *disgust*, $F(30,1209) = 146.76$, $p < 0.0001$, $\eta_p^2 = 0.78$, at 20.0% intensity ($p = 0.002$); *surprise*, $F(30,1209) = 109.37$, $p < 0.0001$, $\eta_p^2 = 0.73$, at 23.3% ($p = 0.012$); *anger*, $F(30,1209) = 43.38$, $p < 0.0001$, $\eta_p^2 = 0.52$, at 26.7% ($p = 0.02$); *fear*, $F(30,1209) = 52.47$, $p < 0.0001$, $\eta_p^2 = 0.57$, at 26.7% ($p = 0.039$); and *sadness*, $F(30,1209) = 44.45$, $p < 0.0001$, $\eta_p^2 = 0.53$, at 36.7% intensity ($p = 0.007$). **Figure 1** shows the pattern of automated expression classification as a function of expressive intensity.

Automated Assessment of Action Units (AUs) With FACET

The evidence scores (at 100% intensity of expression) of AUs were subjected to a 6 (Expression Stimulus) $\times$ 20 (AUs) ANOVA. Effects of expression, $F(5,234) = 30.69$, $p < 0.0001$, $\eta_p^2 = 0.40$, AUs, $F(19,4446) = 433.60$, $p < 0.0001$, $\eta_p^2 = 0.65$, and an interaction, $F(95,4446) = 100.63$, $p < 0.0001$, $\eta_p^2 = 0.68$, emerged. For all the AUs, there were significant differences across expressions, all $Fs(5,234) \geq 23.64$, $p < 0.0001$, $\eta_p^2 \geq 0.34$. **Table 3** shows the 100% intensity AU scores.

To interpret the interaction and determine the association of specific AUs to particular expressions, we used two complementary approaches. First, we examined whether, for each AU and emotional expression, the scores were positive and above 0 (thus revealing that an AU was in fact present), by means of t-tests for dependent samples. Significant differences appeared for all the AUs in boldface in **Table 3**, all $ts(39) \geq 5.53$, $p < 0.0001$, $d \geq 0.87$. Second, for each AU, we examined whether scores were higher for each emotional expression (at any intensity level from 3.33 to 100%) relative to those for the neutral face, in one-way (31: Intensity level) ANOVAs, followed by Bonferroni ($p < 0.05$) corrections. Significant differences appeared for all the AUs in boldface in **Table 3**, $Fs(30,1170) \geq 59.62$, $p < 0.0001$, $\eta_p^2 = 0.61$. **Figure 2** shows the variations in the selected AUs (those that fulfilled both criteria, i.e., significantly above 0 and above neutral faces) across expressive intensities. In sum, facial happiness or joy was significantly characterized by AUs 6, 12, and 25; surprise, by AUs 1, 2, 5, 25, and 26; anger, by AUs 4 and 7; sadness, by AUs 1, 4, and 15; disgust, by AUs 4, 6, 7, 9, and 10; and fear, by AUs 1, 5, and 25.

Relationships Between Human Observers' Performance (Responses and RTs) and Automated Assessment With FACET (Evidence Scores of Expressions and AUs)

Intra-class correlation (*ICC*, 2) analyses revealed high classification consistency between the automated evidence
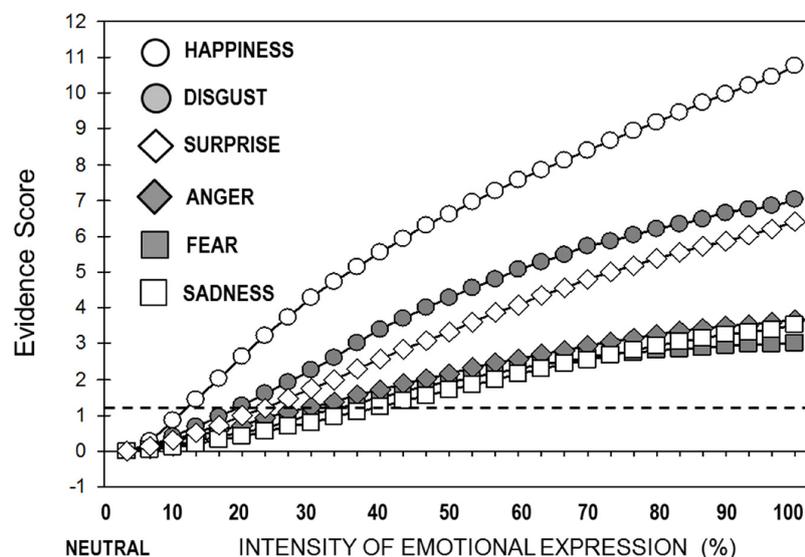
FIGURE 1 | Automated assessment of expressive intensity. Mean automated (FACET) *difference* (emotional minus neutral) evidence scores of each type of expression (response) across intensity levels for each expression stimulus. Above the dotted line: significantly different from the 0% (neutral) baseline (happy: 13.3% of intensity; disgust: 20.0%; surprise: 23.3%; anger and fear: 26.7%; sadness: 36.7%).

TABLE 3 | Mean raw evidence scores (odds ratios) of action units (AUs) for each expression (100% expressive intensity).

		Expression response					
Action Units		Happiness	Surprise	Anger	Sadness	Disgust	Fear
AU1	Inner brow raiser	−1.11	**1.51**	−1.76	**1.38**	−2.23	**1.58**
AU2	Outer brow raiser	−0.71	**1.93**	−1.75	−0.15	−1.60	0.71
AU4	Brow lowerer	−1.65	−1.08	**1.86**	**1.55**	**1.70**	0.92
AU5	Upper lid raiser	−1.41	**1.99**	0.31	−0.19	−0.97	**1.26**
AU6	Cheek raiser	**2.88**	−2.32	−0.24	−0.28	**1.02**	−0.72
AU7	Lid tightener	0.43	−1.10	**0.89**	0.09	**1.28**	−0.15
AU9	Nose wrinkle	−2.49	−5.22	0.19	−2.45	**3.48**	−3.25
AU10	Upper lip raiser	−0.40	−1.79	0.43	−0.17	**3.55**	−0.34
AU12	Lip corner puller	**4.06**	−1.60	−1.80	−1.04	−1.29	−0.76
AU14	Dimpler	−1.73	−2.62	−1.94	−1.37	−3.58	−1.88
AU15	Lip corner depressor	−1.98	−1.87	−0.99	**0.98**	0.16	−1.22
AU17	Chin raiser	−1.79	−2.53	−0.31	0.25	0.49	−2.02
AU18	Lip puckerer	−9.79	−3.04	−1.69	−2.08	−4.97	−3.83
AU20	Lip stretcher	−0.23	−0.74	−1.48	−0.08	−0.07	0.37
AU23	Lip tightener	−1.58	−1.13	−0.23	−0.80	−0.89	−1.09
AU24	Lip pressor	−2.89	−3.38	−1.07	−0.97	−2.54	−2.77
AU25	Lips part	**2.07**	**2.58**	−1.41	−1.31	0.94	**1.48**
AU26	Jaw drop	−0.05	**2.27**	−2.44	−1.76	−1.82	0.02
AU28	Lip suck	−3.42	−4.84	−3.80	−2.99	−6.07	−3.54
AU43	Eyes closed	−3.45	−0.96	−1.20	−1.47	−0.99	−1.33

Automated analysis computed by Emotient FACET Software. Boldface: AU evidence scores for emotional faces significantly higher than those for neutral faces and above 0. They represent the AUs specifically associated with each expression.

scores and hits from human raters, separately for each emotional category ($N = 40$; Happiness: $ICC = 0.93$; Surprise: $ICC = 0.94$; Anger: $ICC = 0.89$; Sadness: $ICC = 0.95$; Disgust: $ICC = 0.76$; Fear: $ICC = 0.65$; all $ps < 0.001$; 95% CI). $ICCs$ were calculated as consistency between the proportion of hits for each KDEF model (averaged across all 96 human observers) and the evidence scores recalculated into probabilities as $p = 1/(1 + 10^{-\text{evidence score}})$. Also, RTs for observers' hits were negatively related to automated evidence of expressions (Happiness: $r = -0.45$; Surprise: $r = -0.51$; Anger: $r = -0.40$; Sadness: $r = -0.41$; Disgust: $r = -0.58$; Fear: $r = -0.47$; all $ps \leq 0.01$; $N = 40$).

In addition, there were positive correlations between specific AUs and the probability of human categorization responses. Most of the significantly related (all $ps < 0.0001$; $N = 240$) AUs were those that typically characterize each expression: The probability that observers categorized expressions (a) as *happy* was related to AU6 ($r = 0.67$) and AU12 ($r = 0.90$); (b) as *surprised*, to AU1 ($r = 0.45$), AU2 ($r = 0.73$), AU5 ($r = 0.68$), AU25 ($r = 0.45$), and AU26 ($r = 0.77$); (c) as *angry*, to AU4 ($r = 0.41$), AU7 ($r = 0.37$), and AU23 ($r = 0.48$); (d) as *sad*, to AU1 ($r = 0.36$), AU4 ($r = 0.34$), AU15 ($r = 0.63$), and AU24 ($r = 0.44$); (e) as *disgusted*, to AU4 ($r = 0.36$), AU7 ($r = 0.50$), AU9 ($r = 0.73$), and AU10 ($r = 0.77$); and (f) as *fearful*, to AU1 ($r = 0.42$) and AU5 ($r = 0.34$).

DISCUSSION

We aimed to provide researchers of emotional facial expression processing with a set of useful and valid dynamic stimuli. To

this end, with agreed time parameters (i.e., unfolding speed to expressive apex within 1 s; Schultz and Pilz, 2009; Hoffmann et al., 2010; Johnston et al., 2013; Wingenbach et al., 2016), we animated static face stimuli of the KDEF database (Lundqvist et al., 1998). The current study examined the resulting KDEF-dyn video-clip stimuli from two complementary approaches: human observer judgments and automated assessment of facial expression. A variety of measures (recognition accuracy, efficiency, and confusions, as well as automated classification of expressions and detection of AUs as a function of intensity, in addition to low-level image properties) were obtained, and are shown on a stimulus level as supplementary data. They will supply researchers with an instrument to select the stimuli as a function of multiple criteria.

Recognition Patterns of Static and Dynamic Expressions

Human observers correctly recognized all the expressions (as they were intended) well-above chance level ($M = 85.2\%$). Happy faces were recognized better and faster—and fearful faces, less accurately and more slowly—than others, with confusions of fear as surprise, disgust as anger, and sadness as fear. The patterns of recognition accuracy, processing efficiency, and confusions across dynamic expressions converge with those found in prior research for *static expressions*, using different stimulus databases. Regarding *recognition accuracy*, Nelson and Russell reviewed 38 sets of data from 17 studies (Nelson and Russell, 2013): Scores were highest for facial happiness (89%), followed by surprise
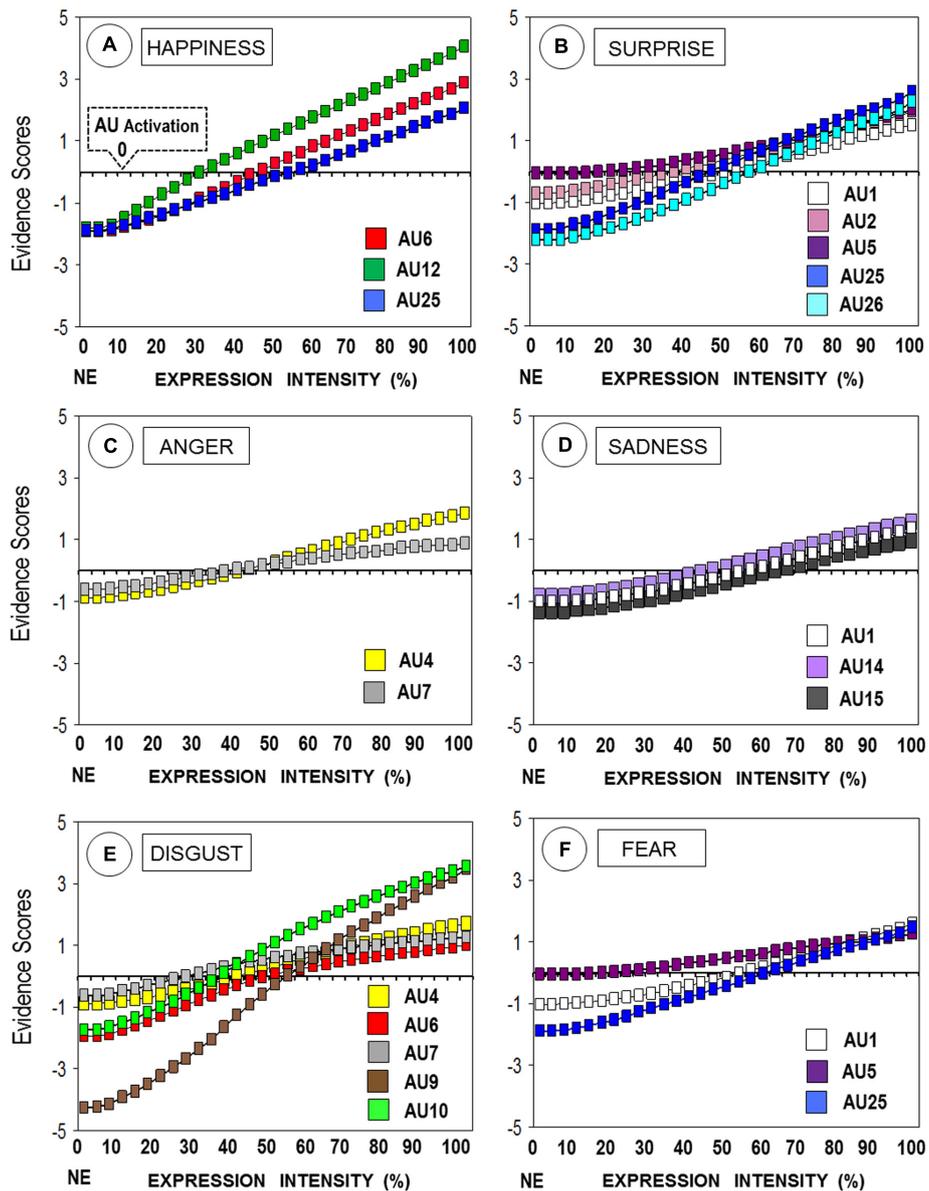
FIGURE 2 | Distribution of AU evidence scores across levels of expressive intensity. For each expression, AUs were selected if scores were higher than for the neutral face (NE, or 0% expression intensity) and were positive and above the 0 AU activation baseline. AU1: inner brow raiser; AU2: outer brow raiser; AU4: brow lowerer; AU5: upper lid raiser; AU6: cheek raiser; AU7: lid tightener; AU9: nose wrinkle; AU10: upper lip raiser; AU12: lip corner puller; AU15: lip corner depressor; AU25: lips part; AU26: jaw drop. **(A)** Happiness; **(B)** Surprise; **(C)** Anger; **(D)** Sadness; **(E)** Disgust; **(F)** Fear.

(83%), which were higher than for sadness and anger (71 and 68%, respectively), followed by disgust and fear (65 and 59%, respectively). This coincides with our own relative differences (see also Tottenham et al., 2009; Recio et al., 2014; Calvo et al., 2016). Such a consistency extends also to *processing efficiency*, as happy faces are typically recognized faster, followed by surprise, while fear is recognized most slowly (Calder et al., 2000; Elfenbein and Ambady, 2003; Palermo and Coltheart, 2004; Calvo and Nummenmaa, 2009). The pattern of *confusions* is also consistent, as they have been found to occur systematically between disgust and anger, and between surprise and fear, and to a lesser extent

between sadness and fear (Palermo and Coltheart, 2004; Calvo and Lundqvist, 2008; Tottenham et al., 2009; Recio et al., 2013).

Further validation comes from prior research using *dynamic expression* stimuli. First, three studies included all six basic expressions in dynamic *morphing format* from three different databases. Calvo et al. (2016) presented real faces (24 models of the KDEF-dyn database) for 1 s. Recio et al. (2014) presented real faces (from the RaFD; Langner et al., 2010) for 600 ms. Recio et al. (2013) displayed computer-generated faces (FACSGen 2.0; Krumhuber et al., 2012) for 900 ms. The pattern of recognition accuracy across expressions was similar in all three studies, with

happy faces being identified most accurately (also including higher *A'* sensitivity; Calvo et al., 2016), and disgusted and fearful faces, least accurately (and lower *A'* sensitivity; Calvo et al., 2016). In addition, in all three studies, fear was likely to be confused with surprise, disgust with anger, and there was some confusion between sadness and fear. Second, regarding the dynamic stimulus sets based on on-line *video recordings* (e.g., van der Schalk et al., 2011; Banziger et al., 2012; Kaulard et al., 2012; Zhang et al., 2014; O'Reilly et al., 2016; Wingenbach et al., 2016; see the 22 databases reviewed by Krumhuber et al., 2017), it is difficult to make comparisons because some studies did not measure recognition performance (accuracy, RTs, or confusions), and due to considerable variations in number of expressive categories and display times (among many other methodological differences). The study conducted by Wingenbach et al. (2016) was methodologically more similar to our own. Their relative recognition accuracies and the pattern of RTs across the six basic expressions were comparable to those in the current study. Altogether, this empirical consistency validates the current database.

Automated Assessment vs. Human Observers

Another major source of validation for the current database involves the use of automated facial expression analysis. First, the automated classification of expressions showed discrimination specificity, with the evidence of each expression being significantly greater for the corresponding stimulus category than for the others. Nevertheless, some expressions, especially, happiness, and also disgust and surprise, were classified better than sadness, anger, and fear (see **Table 2**), which is in total agreement with results obtained with other automated computation algorithms (Lucey et al., 2010). Second, AUs generally discriminated between expressive categories, and this was in accordance with FACS proposals (Ekman et al., 2002; Olderbak et al., 2014). Some AUs characterized expressions more specifically or strongly than others (see **Table 3**), e.g., AU12 for happiness, AU25 for surprise, AUs 9 and 10 for disgust, AU1 for fear, and AU4 for anger and sadness (the AU4 combination with other AUs allowed for a clear discrimination between these two expressions; see **Table 2**). A related pattern has been obtained with different automated AU detection systems (Lucey et al., 2010; Mavadati et al., 2013; Zhang et al., 2014). Third, automated expression classification and also AU evidence scores increased significantly across 3.33% expressive intensity steps between a neutral and an emotional face (see **Figures 1**, **2**). The steepness of such a progressive increase as a function of intensity varied for different expressions and AUs. This approach and results regarding intensity represent a novel contribution and further validate the current video-clip stimuli.

Fourth, importantly, significant correlations emerged between human observers' performance and automated evidence of expressions (large effect sizes: Cohen's $ds \geq 1.71$) and AUs (medium to large effects: $ds \geq 0.72$). This has implications for expression recognition theories concerning the type of information that is processed and the cognitive processes involved. Computational models such as EMPATH (Dailey et al., 2002, 2010) and support vector machine (SVM) based techniques (Susskind et al., 2007)—and, presumably, FACET—simulate face processing and expression recognition in humans. In these models, facial expressions are computed by "emotionless machines" on purely perceptual grounds, i.e., physical image properties (the morphological structure of facial configurations and the visual saliency of distinctive facial cues), in the absence of affective processing. Accordingly, the fact that the automated classifications of expressions converged with human observers' judgments in the current study suggests that human expression recognition also relies to a significant extent on the perceptual (devoid of affect) analysis of facial features. Nevertheless, first, while this may be true for photographs or videos of faces, the role of human affective processing is probably greater in actual face-to-face social encounters, when emotional significance becomes relevant for adaptive purposes. Second, it is likely that the morphological facial features of expressions have become associated (through practice) with their affective significance, and thus both would be processed in tandem, therefore explaining the observed correlations.

Applications and Limitations

The KDEF-dyn database aims to extend the research possibilities of dynamic facial expression stimuli. First, regarding *experimental control*, all the stimuli are equated in multiple image properties that are non-specific of expression—but can act as confounds—(luminance, signal-to-noise ratio, size, orientation, etc.), in addition to standardization of dynamic properties (unfolding speed and duration). Such controls will be particularly useful for neurophysiological and eyetracking research, where the dependent measures are especially sensitive to physical stimulus factors; and also useful for paradigms in which the stimuli must be presented briefly, where display duration needs to be strictly comparable for the different stimuli. A second benefit is related to the role of *expressive intensity*. Instead of considering only the apex, we have established the assessment of expressions and AUs at fine-grained intensities. This is important, as intensity is often critical to interpret the meaning of expressions. By knowing the evidence for each expression and AU at each intensity level, and the time-intensity correspondence in the video-clips (as shown in **Supplementary Dataset S3**), researchers can easily manipulate the display time of the stimuli to investigate the desired intensity (e.g., by cutting, masking, or stopping each video-clip at the respective time point). This approach will be useful for the investigation of visual processing, particularly for studies of expression discrimination thresholds. A third promising application is concerned with the use of these stimuli in the investigation of *cognitive biases* (attentional and interpretative) in *psychopathology*. For example, it has been shown that individuals with clinical levels of social anxiety are especially prone to detect negatively valenced dynamic expressions at low intensities (Gutiérrez-García and Calvo, 2016, 2017; Gutiérrez-García et al., 2018). A reason for the usefulness of this application to psychopathology research is that dynamic information improves identification of facial affect, particularly for lower intensity and subtle stimuli

(Krumhuber et al., 2013), which would increase sensitivity for individuals that are hypervigilant to threat and incongruities in facial expressions.

Researchers should, nonetheless, be aware of potential limitations. First, although standardization of unfolding *speed* is beneficial for experimental control, it can reduce the natural speed variance across expressions. For example, we averaged the 1-s unfolding speed from neutral baseline to emotional apex for all the expressions (see Schultz and Pilz, 2009; Johnston et al., 2013; Wingenbach et al., 2016). However, facial surprise is considered as most natural when it unfolds at a fast pace while sadness is judged as more realistic when the facial expression changes slowly (Sato and Yoshikawa, 2004; Adamaszek et al., 2015). To remedy this potential limitation, it is possible to slow down or speed up the video-clips, by means of video-editing software. Second, we used *posed* instead of spontaneous expressions. The majority of extant dynamic stimulus sets, in fact, include posed expressions, either in response to instructions to perform facial actions or as the enactment of emotional scenarios (van der Schalk et al., 2011; Banziger et al., 2012; Kaulard et al., 2012; O'Reilly et al., 2016; Wingenbach et al., 2016), although some have included spontaneous expressions (Mavadati et al., 2013; Zhang et al., 2014). Posed expressions may lose naturalness and their recognition rates may be inflated, although the former avoid the ambiguity of spontaneous expressions. Third, we used *morphed* expressions. Morphing creates linear movement where all the facial components change at the same time and speed, whereas natural expressions appear to change in a non-linear manner. However, some studies indicate that natural expressions look smooth, uniform, and ballistic (Weiss et al., 1987; Hess et al., 1989), thus actually sharing properties with morphed dynamic expressions. Further, in the current study, automated assessment revealed specificity and sensitivity to expressions and also to AUs in accordance with FACS proposals. This suggests that the possible reduction of naturalness was not critical (see Cosker et al., 2010, 2015).

CONCLUSION

We present a set of dynamic facial expressions (KDEF-dyn) based on a widely used database of static expressions (KDEF). The new stimuli have been validated by means of several measures from two approaches: expression categorization by human observers and automated analysis of facial expressions and AUs with computer software. Results show good convergence with prior research using static and dynamic expression stimuli. Although not devoid of limitations, this convergence reinforces the validation of the current database, while offering additional

advantages: (a) the use of automated facial expression and AU analysis, with significant correlations between human and automated performance; (b) the control of perceptual properties (e.g., size and multiple low-level image statistics) and stimulus dynamic properties (e.g., duration and unfolding speed); and (c) the systematic and fine-grained gradation of expressive intensities of an otherwise relatively large sample of encoders. This will be useful for behavioral, computational, and neurophysiological studies investigating facial expression processing.

AVAILABILITY OF DATA

The KDEF-dyn stimuli and datasets are freely available for scientific purposes, and can be downloaded from http://kdef.se/versions.html (KDEF-dyn I).

AUTHOR CONTRIBUTIONS

MC and DL conceived and designed the experiments. AF-M prepared the materials, performed the experiments, and conducted the statistical analyses. MC wrote the first draft of the manuscript. MC, AF-M, GR, and DL wrote sections and revised the whole manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02052/full#supplementary-material

DATASET S1 | Human and automated expression and action unit categorization. Measures of recognition performance by human observers and automated analysis for each KDEF-dyn stimulus.

DATASET S2 | FACET assessment of intensities. Evidence values for each expression and action units, at each of 31 intensity levels (in 3.33% steps).

DATASET S3 | Low-level image statistics of intensities. Image values at each of 31 intensity levels (in 3.33% steps).

DATASET S4 | Stimuli. Video-clip stimuli_MP4. Two hundred and forty video-clips, separated for each of six emotional expression categories (40 video-clips each).

REFERENCES

Adamaszek, M., Kirkby, K. C., D'Agata, F., Olbrich, S., Langner, S., Steele, C., et al. (2015). Neural correlates of impaired emotional face recognition in cerebellar lesions. *Brain Res.* 1613, 1–12. doi: 10.1016/j.brainres.2015.01.027

Anitha, B., Venkatesha, M. K., and Adiga, B. S. (2010). A survey on facial expression databases. *Int. J. Eng. Sci. Technol.* 2, 5158–5174.

Banziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827

Bartlett, M., and Whitehill, J. (2011). "Automated facial expression measurement: recent applications to basic research in human behavior, learning, and education," in *Handbook of Face Perception*, eds A. Calder, G. Rhodes, M. Johnson, and J. Haxby (Oxford: Oxford University Press), 489–513.

Bublatzky, F., Gerdes, A. B. M., White, A. J., Riemer, M., and Alpers, G. W. (2014). Social and emotional relevance in face processing: happy faces of future interaction partners enhance the late positive potential. *Front. Hum. Neurosci.* 8:493. doi: 10.3389/fnhum.2014.00493

Calder, A. J., Young, A. W., Keane, J., and Dean, M. (2000). Configural information in facial expression perception. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 527–551. doi: 10.1037/0096-1523.26.2.527

Calvo, M. G., Avero, P., Fernandez-Martin, A., and Recio, G. (2016). Recognition thresholds for static and dynamic emotional faces. *Emotion* 16, 1186–1200. doi: 10.1037/emo0000192

Calvo, M. G., and Beltrán, D. (2014). Brain lateralization of holistic versus analytic processing of emotional facial expressions. *Neuroimage* 92, 237–247. doi: 10. 1016/j.neuroimage.2014.01.048

Calvo, M. G., Gutiérrez-García, A., Avero, P., and Lundqvist, D. (2013). Attentional mechanisms in judging genuine and fake smiles: eye-movement patterns. *Emotion* 13, 792–802. doi: 10.1037/a0032317

Calvo, M. G., and Lundqvist, D. (2008). Facial expressions of emotion (KDEF): identification under different display-duration conditions. *Behav. Res. Methods* 40, 109–115. doi: 10.3758/BRM.40.1.109

Calvo, M. G., and Nummenmaa, L. (2009). Eye-movement assessment of the time course in facial expression recognition: neurophysiological implications. *Cogn. Affect. Behav. Neurosci.* 9, 398–411. doi: 10.3758/CABN.9.4.398

Calvo, M. G., and Nummenmaa, L. (2011). Time course of discrimination between emotional facial expressions: the role of visual saliency. *Vis. Res.* 51, 1751–1759. doi: 10.1016/j.visres.2011.06.001

Calvo, M. G., and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). "Observer-based measurement of facial expression with the Facial Action Coding System," in *The Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York, NY: Oxford University Press), 203–221.

Cohn, J. F., and De la Torre, F. (2015). "Automated face analysis for affective computing," in *The Oxford Handbook of Affective Computing*, eds R. A. Calvo, S. DiMello, J. Gratch, and A. Kappas (New York, NY: Oxford University Press), 131–150.

Cosker, D., Krumhuber, E., and Hilton, A. (2010). "Perception of linear and nonlinear motion properties using a FACS validated 3D facial model," in *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV)*, New York, NY. doi: 10.1145/1836248.1836268

Cosker, D., Krumhuber, E., and Hilton, A. (2011). "A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling," in *Proceedings of the 13th IEEE International Conference on Computer Vision (ICCV)*, Barcelona. doi: 10.1109/ICCV.2011.6126510

Cosker, D., Krumhuber, E., and Hilton, A. (2015). "Perceived emotionality of linear and nonlinear AUs synthesised using a 3D dynamic morphable facial model," in *Proceedings of the Conference on Facial Analysis and Animation (FAA)*, New York, NY. doi: 10.1145/2813852.2813859

Cowie, R., Douglas-Cowie, E., and Cox, C. (2005). Beyond emotion archetypes: databases for emotion modelling using neural networks. *Neural Netw.* 18, 371–388. doi: 10.1016/j.neunet.2005.03.002

Dailey, M. N., Cottrell, G. W., Padgett, C., and Adolphs, R. (2002). EMPATH: a neural network that categorizes facial expressions. *J. Cogn. Neurosci.* 14, 1158–1173. doi: 10.1162/089892902760807177

Dailey, M. N., Joyce, C., Lyons, M. J., Kamachi, M., Ishi, H., Gyoba, J., et al. (2010). Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion* 10, 874–893. doi: 10.1037/a0020019

Dente, P., Küster, D., Skora, L., and Krumhuber, E. (2017). Measures and metrics for automatic emotion classification via FACET. *Paper Presented at the Artificial Intelligence and Simulation of Behaviour Annual Convention*, Bath.

Ebner, N. C., Riediger, M., and Lindenberger, U. (2010). FACES—a database of facial expressions in young, middle-aged, and older women and men: development and validation. *Behav. Res. Methods* 42, 351–362. doi: 10.3758/BRM.42.1.351

Ekman, P., and Friesen, W. V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *Facial Action Coding System*. Salt Lake City, UT: A Human Face.

Elfenbein, H. A., and Ambady, N. (2003). When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *J. Pers. Soc. Psychol.* 85, 276–290. doi: 10.1037/0022-3514.85.2.276

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fiorentini, C., and Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *J. Vis.* 11, 1–15. doi: 10.1167/11.3.17

Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., and De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behav. Res. Methods* 47, 1136–1147. doi: 10.3758/s13428-014-0536-1

Goeleven, E., De Raedt, R., Leyman, L., and Verschuere, B. (2008). The Karolinska Directed Emotional Faces: a validation study. *Cogn. Emot.* 22, 1094–1118. doi: 10.1080/02699930701626582

Gordon, I, Tanaka, J. W., Pierce, M., and Bartlett, M. (2011). Facial expression production and training. *J. Vis.* 11, 565–565. doi: 10.1167/11.11.565

Gupta, R., Hur, Y. J., and Lavie, N. (2016). Distracted by pleasure: effects of positive versus negative valence on emotional capture under load. *Emotion* 16, 328–337. doi: 10.1037/emo0000112

Gutiérrez-García, A., and Calvo, M. G. (2016). Social anxiety and trustworthiness judgments of dynamic facial expressions of emotion. *J. Behav. Ther. Exp. Psychiatry* 52, 119–127. doi: 10.1016/j.jbtep.2016.04.003

Gutiérrez-García, A., and Calvo, M. G. (2017). Social anxiety and threat-related interpretation of dynamic facial expressions: sensitivity and response bias. *Pers. Individ. Differ.* 107, 10–16. doi: 10.1016/j.paid.2016.11.025

Gutiérrez-García, A., Calvo, M. G., and Eysenck, M. W. (2018). Social anxiety and detection of facial untrustworthiness: spatio-temporal oculomotor profiles. *Psychiatry Res.* 262, 55–62. doi: 10.1016/j.psychres.2018.01.031

Harris, R. J., Young, A. W., and Andrews, T. J. (2014). Dynamic stimuli demonstrate a categorical representation of facial expression in the amygdala. *Neuropsychologia* 56, 47–52. doi: 10.1016/j.neuropsychologia.2014. 01.005

Hess, U., Kappas, A., McHugo, G. J., Kleck, R. E., and Lanzetta, J. T. (1989). An analysis of the encoding and decoding of spontaneous and posed smiles: the use of facial electromyography. *J. Nonverb. Behav.* 13, 121–137. doi: 10.1007/BF00990794

Hoffmann, H., Traue, H. C., Bachmayr, F., and Kessler, H. (2010). Perceived realism of dynamic facial expressions of emotion: optimal durations for the presentation of emotional onsets and offsets. *Cogn. Emot.* 24, 1369–1376. doi: 10.1080/02699930903417855

Johnston, P., Mayes, A., Hughes, M., and Young, A. W. (2013). Brain networks subserving the evaluation of static and dynamic facial expressions. *Cortex* 49, 2462–2472. doi: 10.1016/j.cortex.2013.01.002

Kaulard, K., Cunningham, D. W., Bulthoff, H. H., and Wallraven, C. (2012). The MPI facial expression database — A validated database of emotional and conversational facial expressions. *PLoS One* 7:e32321. doi: 10.1371/journal. pone.0032321

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Krumhuber, E. G., Tamarit, L., Roesch, E. B., and Scherer, K. R. (2012). FACSGen 2.0 animation software: generating three-dimensional FACS-valid facial expressions for emotion research. *Emotion* 12, 351–363. doi: 10.1037/a0026632

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and Validation of the Radboud Faces Database. *Cogn. Emot.* 24, 1377–1388. doi: 10.1080/02699930903485076

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). "The Extended Cohn–Kanade Dataset (CK +): a complete dataset for action unit and emotion-specified expression," in *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010)*, San Francisco, CA.

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF [CD-ROM]*. Stockholm: Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., and Cohn, J. F. (2013). DISFA: a spontaneous facial action intensity database. *IEEE Trans. Affect. Comput.* 4, 151–160. doi: 10.1109/T-AFFC.2013.4

Naples, A., Nguyen-Phuc, A., Coffman, M., Kresse, A., Faja, S., Bernier, R., et al. (2015). A computer-generated animated face stimulus set for psychophysiological research. *Behav. Res. Methods* 47, 562–570. doi: 10.3758/s13428-014-0491-x

Nelson, N. L., and Russell, J. A. (2013). Universality revisited. *Emot. Rev.* 5, 8–15. doi: 10.1177/1754073912457227

Olderbak, S., Hildebrandt, A., Pinkpank, T., Sommer, W., and Wilhelm, O. (2014). Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behav. Res. Methods* 46, 992–1006. doi: 10.3758/s13428-013-0421-3

O'Reilly, H., Pigat, D., Fridenson, S., Berggren, S., Tal, S., Golan, O., et al. (2016). The EU-Emotion Stimulus Set: a validation study. *Behav. Res. Methods* 48, 567–576. doi: 10.3758/s13428-015-0601-4

Palermo, R., and Coltheart, M. (2004). Photographs of facial expression: accuracy, response times, and ratings of intensity. *Behav. Res. Methods Instrum. Comput.* 36, 634–638. doi: 10.3758/BF03206544

Pollick, F. E., Hill, H., Calder, A., and Paterson, H. (2003). Recognizing facial expression from spatially and temporally modified movements. *Perception* 32, 813–826. doi: 10.1068/p3319

Popov, T., Miller, G. A., Rockstroh, B., and Weisz, N. (2013). Modulation of alpha power and functional connectivity during facial affect recognition. *J. Neurosci.* 33, 6018–6026. doi: 10.1523/JNEUROSCI.2763-12.2013

Recio, G., Schacht, A., and Sommer, W. (2013). Classification of dynamic facial expressions of emotion presented briefly. *Cogn. Emot.* 27, 1486–1494. doi: 10.1080/02699931.2013.794128

Recio, G., Schacht, A., and Sommer, W. (2014). Recognizing dynamic facial expressions of emotion: specificity and intensity effects in event-related brain potentials. *Biol. Psychol.* 96, 111–125. doi: 10.1016/j.biopsycho.2013.12.003

Sanchez, A., Vazquez, C., Gomez, D., and Joormann, J. (2014). Gaze-fixation to happy faces predicts mood repair after a negative mood induction. *Emotion* 14, 85–94. doi: 10.1037/a0034500

Sandbach, G., Zafeiriou, S., Pantic, M., and Yin, L. (2012). Static and dynamic 3D facial expression recognition: a comprehensive survey. *Image Vis. Comput.* 30, 683–697. doi: 10.1016/j.imavis.2012.06.005

Sato, W., and Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Schultz, J., and Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Exp. Brain Res.* 194, 465–475. doi: 10.1007/s00221-009-1721-9

Susskind, J. M., Littlewort, G., Bartlett, M. S., Movellan, J., and Anderson, A. K. (2007). Human and computer recognition of facial expressions of emotion. *Neuropsychologia* 45, 152–162. doi: 10.1016/j.neuropsychologia.2006.05.001

Tottenham, N., Borscheid, A., Ellertsen, K., Marcus, D., and Nelson, C. A. (2002). *The Nimstim Face Set1*. Available at: http://www.macbrain.org/faces/index.htm

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168, 242–249.

Valstar, M. F., Almaev, T., Girard, J. M., McKeown, G., Mehu, M., Yin, L., et al. (2015). "FERA 2015 – Second facial expression recognition and analysis challenge," in *Proceedings of the Eleventh IEEE International Conference on Face and Gesture Recognition*, Ljubljana. doi: 10.1109/FG.2015.7284874

Valstar, M. F., Sanchez-Lozano, T., Cohn, J. F., Jeni, L. A., Girard, J. M., Zhang, Z., et al. (2017). "FERA 2017 - Addressing head pose in the third facial expression recognition and analysis challenge," in *Proceedings of the Twelfth IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC. doi: 10.1109/FG.2017.107

van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

Vrticka, P., Lordier, L., Bediou, B., and Sander, D. (2014). Human amygdala response to dynamic facial expressions of positive and negative surprise. *Emotion* 14, 161–169. doi: 10.1037/a0034619

Weiss, F., Blum, G. S., and Gleberman, L. (1987). Anatomically based measurement of facial expressions in simulated versus hypnotically induced affect. *Motiv. Emot.* 11, 67–81. doi: 10.1007/BF00992214

Wingenbach, T. S., Ashwin, C., and Brosnan, M. (2016). Correction: validation of the Amsterdam Dynamic Facial Expression Set – Bath Intensity variations (ADFES-BIV): a set of videos expressing low, intermediate, and high intensity emotions. *PLoS One* 11:e0168891. doi: 10.1371/journal.pone.0168891

Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., et al. (2014). BP4D-spontaneous: a high-resolution spontaneous 3D dynamic facial expression database. *Image Vis. Comput.* 32, 692–706. doi: 10.1016/j.imavis.2014.06.002

Check for updates

Naturalistic Emotion Decoding From Facial Action Sets

Sylwia Hyniewska [1,2,3*], Wataru Sato [1], Susanne Kaiser [2,3] and Catherine Pelachaud [4]

[1] Kokoro Research Center, Kyoto University, Kyoto, Japan, [2] Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland, [3] Human Behaviour Analysis Laboratory, Department of Psychology, University of Geneva, Geneva, Switzerland, [4] Institut des Systèmes Intelligents et de Robotique (ISIR), Université Pierre et Marie Curie/Centre National de la Recherche Scientifique (CNRS), Paris, France

Researchers have theoretically proposed that humans decode other individuals' emotions or elementary cognitive appraisals from particular sets of facial action units (AUs). However, only a few empirical studies have systematically tested the relationships between the decoding of emotions/appraisals and sets of AUs, and the results are mixed. Furthermore, the previous studies relied on facial expressions of actors and no study used spontaneous and dynamic facial expressions in naturalistic settings. We investigated this issue using video recordings of facial expressions filmed unobtrusively in a real-life emotional situation, specifically loss of luggage at an airport. The AUs observed in the videos were annotated using the Facial Action Coding System. Male participants ($n = 98$) were asked to decode emotions (e.g., anger) and appraisals (e.g., suddenness) from facial expressions. We explored the relationships between the emotion/appraisal decoding and AUs using stepwise multiple regression analyses. The results revealed that all the rated emotions and appraisals were associated with sets of AUs. The profiles of regression equations showed AUs both consistent and inconsistent with those in theoretical proposals. The results suggest that (1) the decoding of emotions and appraisals in facial expressions is implemented by the perception of set of AUs, and (2) the profiles of such AU sets could be different from previous theories.

Keywords: emotional facial expression, spontaneous expressions, naturalistic, cognitive appraisal, nonverbal behavior

INTRODUCTION

Reading emotions of other individuals from their facial expressions is an important skill in managing our social relationships. Researchers have postulated that emotional categories (e.g., anger) (Ekman and Friesen, 1978, 1982) or elementary components of emotions, such as cognitive appraisals (e.g., suddenness) (Scherer, 1984; Smith and Scott, 1997), can be decoded based on the recognition of specific sets of facial movements (**Tables A1, A2** in Supplementary Material). For example, Ekman and Friesen (1978) proposed that specific sets of facial action units (AUs), which could be coded through the Facial Action Coding System (FACS; Ekman et al., 2002), could signal particular emotions and specified the required action unit sets. For instance, in the case of sadness, the facial action set includes inner eyebrows raised (AU 1) and drawn together (AU 4), and lip corners pulled down (AU 15) (Ekman and Friesen, 1975). Scherer (1984), on the other hand, proposed that sets of AUs could signal cognitive appraisals. These researchers developed their theories based on previous theories and findings and their intuitions (Ekman, 2005).

However, only a few previous empirical studies systematically investigated the theoretical predictions on the relationships between the decoding of emotional categories or cognitive appraisals and AU sets and these studies did not provide clear supportive evidence (Galati et al., 1997; Kohler et al., 2004; Fiorentini et al., 2012; Mehu et al., 2012). For example, Kohler et al. (2004) investigated how participants categorize four emotions expressed by actors (Kohler et al., 2004). Based on the results from their decoding study, the authors described the necessary facial AUs for recognizing emotional expressions of high intensity happy, sad, angry, and fearful faces. The analysis showed that the four emotions could be identified with sets of AUs specific to these emotions which are characteristic of the target emotions and distinct from the other three analyzed emotions. However, the profiles of AUs were only partially consistent with theoretical predictions. For example, the brow lowerer (AU4) was associated with the decoding of sadness, however the inner eyebrows raiser (AU 1) and lip corner depressor (AU 15) were not. In short, these studies showed that decoding of emotions and appraisals in facial expressions was associated with sets of AUs, but the profiles of AU sets were only partially consistent with the theoretical predictions.

Furthermore, it must be noted that none of the aforementioned studies evaluated spontaneous emotional expressions in naturalistic settings. Facial displays encountered in everyday life situations show high variability including blends between emotions (Scherer and Ellgring, 2007; Calvo and Nummenmaa, 2015), and spontaneous behavior is more ambiguous (e.g., Yik et al., 1998). This issue is particularly important as behaviors we see in real-life emotional situations are often not the prototypical ones described in literature– they are very varied in terms of co-existing facial movements and sometimes subtle, with rare and low-intensity facial actions (e.g., see Hess and Kleck, 1994; Russell and Fernández-Dols, 1997).

In this study, we investigated whether and how sets of facial actions could be associated with the decoding of emotions and appraisals in spontaneous facial expressions in a naturalistic setting. As stimuli of such spontaneous facial expressions, we used unobtrusive recordings from a hidden camera showing face-to-face interactions of passengers claiming the loss of their luggage at an airport (Scherer and Ceschi, 1997, 2000). All the AUs in the passengers' facial expressions were first coded using FACS (Ekman et al., 2002). We then asked participants to rate six emotions—two positive (Joy, Relief) and four negative (Anger, Sadness, Contempt, Shame)—as well as six appraisals: suddenness, goal obstruction, importance and relevance, coping potential, external norm violation, and internal norm violation. Surprise was not included given that some previous studies (Kohler et al., 2004; Mehu et al., 2012) showed no agreement regarding its valence (Fontaine et al., 2007; Reisenzein and Meyer, 2009; Reisenzein et al., 2012; Topolinski and Strack, 2015), and described its duration as shorter than that of other emotions, making it an affect that could be potentially of a different nature than the other studied emotions (Reisenzein et al., 2012). We explored the relationships between the emotion/appraisal decoding and facial actions using stepwise multiple regression analyses. We expected to observe that sets of facial actions enable the decoding of emotions and appraisals (Ekman, 1992; Scherer and Ellgring, 2007). We did not formulate predictions for the AUs expected in each set given the lack of former decoding studies focusing on data from naturalistic settings.

MATERIALS AND METHODS

Participants

One hundred and twenty-two students from a French technical university took part in the study. A psychologist conducted a short interview with the participants and found that women, a minority in this technical school, were a non-homogenous population (great age distribution, reported psychological history, and intake of substances). Therefore, only data from male students were considered in the analysis ($n = 98$; age = 17–25, means $\pm$ SD = 19.0 $\pm$ 1.5). The interview did not lead to the detection of any neuropsychiatric or psychological history in any of the participants. All participants provided written informed consent prior to participation in the study and were debriefed after the study. The study was approved by the University of Geneva ethics committee and conducted in accordance with the approved guidelines.

Stimuli

Our data relies on unobtrusive recordings from a hidden camera showing face-to-face interactions of passengers claiming the loss of their luggage at an airport (Scherer and Ceschi, 1997, 2000). The aim of such a naturalistic corpus was to obtain dynamic non-acted expressions, including non-typical and subtle facial displays.

Videos from this Lost Luggage corpus focus on the passenger, with a head and torso framing, while showing in the right corner a reduced size video of the face of the hostess (see **Figure 1** for a schematic representation of stimuli).

The original corpus included 1 min long video clips (16-bit colors) that have not been cut in a way to depict only one mental state per segment, therefore the first task was to segment emotional extracts, i.e., define when an emotional state starts and when it ends.

We asked laypersons to watch and to mark in time all mental states and point out state changes. The task was explained through guidelines that were provided in a written format that was additionally read orally to make sure the participants thought carefully about all the provided examples.

The judges were told that their task was to indicate changes between different mental states of one person and that each mental state could be made of several affects happening at the same time, e.g., one mental state composed of 50% joy and 50% guilt. They had to select a period of time (by indicating a starting and an ending time) for each state and to define this mental state. To avoid guiding participants into a particular theoretical framework, guidelines provided examples of action tendencies, motivational changes, appraisal attributions and emotional labels. Judges were told orally that the focus is on "internal states" of passengers that have lost their luggage and that the films come from a hidden camera at an airport. Judges

FIGURE 1 | Schematic illustration of stimuli: frame presenting a passenger claiming the loss of luggage to a hostess. Actual stimuli were real-life dynamic videos.

were told that in one video clip a passenger can display several mental states and moments of neutrality and that they had to indicate them all. They could describe what they see in sentences, through expressions or labels either orally (transcribed by the experimenter) or in a written format on a piece of paper or directly in the provided ANVIL software, with which they were assisted (ANVIL, Video Annotation Research Tool. http://www. dfki.de/kipp/anvil/).

Seven laypeople, administrative staff from the technical university, were invited to act as judges for the task. The two first judges to participate (an account officer and a junior secretary) found the task to be extremely difficult. They gave the following reasons:

- it is impossible to say that a state is changing;
- it could be possible to point that there is an emotion such as anger in a video, but not point to a time;
- defining what the passengers feel or in what state they are, without being guided by specific emotional labels is difficult.

A third judge reported that the observed passengers are talking and not experiencing any affects or changes in mental states and therefore it is impossible to fulfill the task.

Consequently we decided to assign this procedure to individuals who we expected to have some ease to fulfill the task: e.g., individuals who have developed some acuity in the perception of facial expressions. Three individuals were recruited according to their professional activity (virtual character synthesis; facial graphics; FACS coding) and one for his

interest in the non-verbal communication and social cognition. All four individuals, that we called "expert judges," understood the task straight away from reading the guidelines.

Each clip was annotated by three expert judges.

In case of ambiguity, for example when one expert out of three considered less changes in a clip than the other experts, and made a segment last longer, we opted for leaving out the non-agreed upon segment. To reformulate, the solution was, when possible, to recut the clip to eliminate moments that led to discordance. Only moments on which judges agreed to display only one state were kept. If a state starting during a movement or a sentence was preceded by a neutral phase, a second or a second and a half might have been added to the chosen segment to enable the display of the movement development.

In two cases in which ambiguity did not allow an easy and straightforward cutting even in the above, restrictive, manner, a fourth experienced judge was asked to annotate the video clips. In both cases, two judges annotated long segments and one judge a much shorter segment. The fourth judge had a very similar segmentation to the short segmentation, for the two concerned videos. Thus, we followed this restrictive segmentation, as it enabled a definition of mental states to be extracted and presented in separate clips.

After cutting, 64 clips were obtained. Several extracts from these were excluded from the corpus, as they involved a fragment where the face was majorly obstructed or hidden behind glasses that reflected light in the view of the camera, or were presenting a situation outside of the original canvas (e.g., talking to a third

person). In the end, 39 clips were included in the study, each lasting 4–56 s, with a majority lasting between 20 and 28 s. The clips were encoded with a temporal resolution up to 1/25th of a second and showed 19 male and 20 female stimuli. The passengers presented in the clips came from a wide variety of cultural backgrounds. Preliminary analyses showed that there were no effects of gender of stimuli in terms of all of the AU and emotion/appraisal rating data (t-test, $p > 0.1$); accordingly, the factor of stimulus gender was omitted in the following analyses.

FACS Coding

As we wanted to associate short video extracts to attributions made by laypersons, it was important to code all the facial actions that could have an impact on the observers. The ANVIL software was used for the annotation, with 61 tracks for the face (FACS; Ekman et al., 2002) and 22 for the bodily action coding in time. The analysis of the latter coding is outside the scope of this article. The FACS coding was performed by a certified FACS coder and was verified by a second certified FACS coder. The second coder annotated 12 % of the videos (randomly assigned). Both coders used the FACS manual as a constant reference criterion.

In assessing the precision of scoring, we looked at the frame-by-frame agreement by computing Cohen's Kappa (k) for face action coding (Cohen, 1960). The mean agreement was observed at the $k = 0.66$ ($SD = 0.18$), which according to Cicchetti and Sparrow (1981) shows strong agreement. Each of our particular AU coding cases showed satisfactory agreement except for AU 20 (lip stretcher), where k in the 0.21–0.40 range indicated merely a weak/fair agreement.

Procedure

Participants arrived in groups of two to ten. Each participant accessed the study individually through a web browser. The guidelines provided on the first web page were sufficient for understanding the tasks. Participants were randomly attributed to rating blocks. Emotional labels were presented in two controlled orders, the same order of presentation being kept for all stimuli judged by the same participant. Participants watched and evaluated from 6 to 39 short video clip extracts, depending on their self-reported concentration level and their willingness to participate. They answered the same set of questions after each video.

On the first page after each video display, participants were asked to evaluate appraisals presented in the form of a sentence, such as "Do you have the impression that the person you saw in the video, just faced a sudden event?" (suddenness). Appraisals were presented in the chronological order defined by the Componential theory: suddenness, goal obstruction, important and incongruent event, coping potential, respect of internal standards, and violation of external standards (e.g., Scherer, 2001). Participants answered appraisal questions on a 7-point Likert scale, ranging from 0 = totally disagree to 6 = totally agree.

On the second page after the video, participants also had to judge whether the observed passenger was experiencing joy, anger, relief, sadness, contempt, fear and shame. Each emotion

was evaluated by participants on a separate 7-point Likert scale ranging from zero (no emotion) to six (strong emotion) and the emotions were not mutually exclusive. The order of presentation of emotional labels was randomized. The mean attribution of each label to each video (across participants) was the dependent variable. The independent variable was the duration of the facial action cues annotated by coders as present in videos watched by participants.

Data Analysis

For each video ($n = 39$), the annotation in terms of FACS units was quantified by computing the total duration of this AU in a video. We selected this measure as the length of videos was dependent on the duration of present AUs leading to the decoding of a mental state, and therefore the percentage of time an action is present in a video clip is not informative. Stepwise regression analyses with backward selection were performed using SPSS 16.0J (SPSS Japan, Tokyo, Japan). Stepwise regression analyses are techniques for selecting a subset of predictor variables (Ruengvirayudh and Brooks, 2016). By conducting the analyses, we tested whether and how the subset of AUs could predict the decoding of specific emotions/appraisals. Individual regression analyses were conducted for each emotion/appraisal as the dependent variable. All AUs were first entered into the model as independent variables and AUs that did not significantly predict the dependent variable were removed from the model one by one. The first model for which all AUs helped predict at least a marginally significant ($p < 0.10$) variance in the dependent variable emotion/appraisal was selected as the final model. Before the analyses, we conducted *a priori* power analyses using G*Power 3.1.9.2 (Faul et al., 2007). We used the data of Galati et al. (1997) as prior information, because only this study applied similar regression approaches and reported sufficient information for power analyses. The number of AUs associated with emotion decoding were comparable across previous studies (mean, 6.0, 5.25, and 7.1 in Galati et al., 1997; Kohler et al., 2004; and Fiorentini et al., 2012, respectively). The results showed that our regression analyses could detect the relationships between the AU sets and decoding of emotional categories reported in (Galati et al., 1997). (mean $R^2 = 0.49$) with a strong statistical power ($\alpha = 0.05$; $1 - \beta = 0.99$). Based on these data, we expected that our variable selection approach using stepwise regression analyses could detect the set of AUs similar with previous studies in terms of size. However, our analyses lacked the power to investigate full or larger sets of AUs (see discussion). For the final models, we calculated squared multiple correlation coefficients (R^2) as effect-size parameters. Also, we calculated *post hoc* statistical power ($1 - \beta$) for R^2 deviation from zero using G*Power 3.1.9.2 (Faul et al., 2007).

RESULTS

The FACS coding (total duration of AUs) and means $\pm$ SDs of attribution ratings are shown in **Tables 1**, **2**, respectively.

Stepwise regression analyses with backward selection showed that the attributions of all emotional categories and cognitive appraisals were significantly predicted by sets of AUs

TABLE 1 | Total duration of action units (AUs) observed in each video.

Videos	1 Inner brow raiser	2 Outer brow raiser	4 Brow lower	5 Upper lid raiser	6 Cheek raiser	7 Lid tighten	8 Lips toward each other	10 Upper lip raiser	11 Nasolabial deepener	12 Lip corner puller	13 Sharp lip puller	14 Dimpler	15 Lip corner depressor	16 Lower lip depressor	17 Chin raiser	18 Lip puckerer	20 Lip stretch	22 Lip funnel	23 Lip tighten	24 Lip pressor	25 Lips part	26 Jaw drop	27 Mouth stretch	28 Lip suck	29 Jaw thrust	32 Lip bite	34 Cheek puff	37 Lip wipe	43 Eyes closed	50 Speech
1	6.1	3.1	0.0	0.8	0.0	0.2	0.0	3.7	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.4	9.3	0.0	0.0	0.0	0.0	0.0	0.0	2.6	3.8
2	24.0	1.2	23.3	0.0	0.0	4.9	0.0	0.0	4.9	0.0	0.0	0.0	0.0	0.0	2.4	0.0	2.0	0.0	2.0	0.2	2.6	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	22.8
3	0.8	5.1	1.0	0.8	0.0	1.8	0.0	1.2	0.0	0.0	0.0	1.6	0.0	0.0	0.5	0.0	4.2	0.0	0.6	0.4	5.0	2.8	1.3	0.0	1.7	0.0	0.0	0.0	0.3	11.2
4	1.7	1.3	0.5	0.0	0.0	0.8	0.0	2.1	0.0	0.3	0.0	0.1	1.5	0.0	0.0	0.0	0.0	0.0	1.5	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
5	0.0	0.0	0.0	0.0	0.0	3.7	0.0	3.0	0.6	2.2	0.0	0.0	0.0	0.0	1.1	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.7
6	0.0	5.4	3.0	0.0	0.0	1.1	0.0	1.5	0.0	1.8	0.0	0.1	0.6	0.2	1.0	0.0	0.4	0.0	1.6	2.3	2.7	3.2	0.0	0.0	0.0	0.0	0.2	0.0	0.0	14.7
7	0.0	5.2	0.0	1.6	0.5	3.3	0.0	0.0	0.0	0.0	0.0	0.7	0.3	0.0	0.5	0.0	1.0	0.0	1.7	4.6	1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1
8	2.4	0.0	0.0	0.0	0.0	2.6	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.2	1.8	0.0	0.0	0.0	5.1	4.6	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9
9	0.0	0.0	0.0	0.0	0.0	12.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.8	2.5	0.0	0.0	1.8	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	1.4	0.7	1.7	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.5	2.1	0.0	6.0	0.0	1.6	0.0	0.0	4.5	2.3	1.0	1.2	0.0	0.0	0.0	0.0	0.0	1.1	2.9
11	9.2	9.2	9.1	0.0	0.0	7.6	0.0	8.5	0.0	0.0	0.0	5.8	0.6	0.0	9.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3
12	9.5	8.7	9.4	0.0	0.0	6.4	0.0	6.2	0.5	0.0	0.0	0.0	0.0	0.0	0.5	0.1	0.0	0.0	0.0	0.5	4.1	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0
13	1.4	3.1	8.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	2.8	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	7.4	3.6	6.9	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.7	0.4	3.4	0.5	0.0	1.2	0.0	0.2	0.0	12.2	12.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6
15	9.7	2.3	10.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	14.6	0.5	0.0	0.8	0.0	0.8	0.0	15.1	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	3.1	0.3	0.2	0.0	0.0	0.6	0.0	0.0	0.0	0.8	0.2	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.4
17	0.0	0.0	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.7	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	4.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	7.8	0.0	0.0	0.0	0.0	0.0	0.7	0.7	0.0	0.0	0.0	0.0	0.0	0.2	0.0	9.9
19	6.8	5.8	1.3	1.8	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	8.6	11.0	8.9	3.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
21	10.5	0.0	0.0	0.0	0.0	12.1	0.0	0.0	0.0	11.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	10.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.5
22	0.8	8.3	0.0	1.1	5.4	0.9	0.0	1.8	0.0	0.4	0.0	1.3	2.1	0.0	2.8	0.0	0.0	0.0	0.0	0.0	1.7	4.6	0.0	0.0	0.0	0.0	0.0	0.0	0.7	3.5
23	4.0	0.0	2.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	1.6	0.0	3.4	0.0	0.6	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7
24	2.7	2.6	7.4	0.3	0.0	3.1	1.2	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	6.4	0.0	0.0	1.8	0.0	0.0	0.0	0.0	8.3
25	3.0	3.2	0.2	0.0	0.0	0.1	0.0	1.6	0.0	0.0	0.0	3.4	0.0	0.0	0.2	0.0	0.1	0.0	1.0	0.0	1.6	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	3.8	2.5	1.8	3.5	0.0	3.3	0.0	3.1	0.0	0.0	0.0	2.2	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	3.3	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8
27	0.7	2.0	2.3	0.9	0.0	2.6	0.0	0.2	0.0	5.7	0.0	2.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
28	7.5	0.0	0.5	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
29	6.9	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	8.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	1.1	0.9	0.0	0.0	1.8	0.0	0.0	0.1	0.9	2.3	0.9	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31	7.0	0.6	1.9	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	17.1	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.4	1.8	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
32	7.0	0.6	1.9	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	17.1	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.4	1.8	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2
33	0.0	0.6	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	4.1	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6
34	3.2	3.5	5.2	0.9	0.7	4.5	0.0	2.0	0.0	1.7	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	2.8	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.2
35	10.4	11.1	0.0	0.0	2.1	11.4	0.0	3.1	0.0	5.7	0.0	1.8	0.0	0.0	0.3	0.0	0.0	0.0	0.2	0.0	3.3	3.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.9
36	1.7	1.8	0.8	0.0	0.0	0.9	0.0	0.2	0.0	0.0	0.0	0.6	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9
37	1.7	4.9	1.0	0.0	0.0	0.0	0.0	1.8	0.2	1.2	0.0	0.0	1.1	0.0	0.9	0.0	0.4	0.0	0.0	0.0	2.2	2.5	0.0	0.0	2.3	1.0	0.0	0.0	0.7	26.4
38	16.8	8.0	7.5	0.0	0.0	21.9	0.0	0.0	0.0	0.0	0.0	8.3	0.0	0.0	5.0	0.0	0.2	0.0	14.9	5.0	2.2	3.7	0.0	7.5	0.0	0.0	0.0	0.0	0.0	11.6
39	0.0	0.0	0.0	0.0	0.0	14.3	1.7	0.0	0.0	0.0	0.0	1.4	0.0	0.0	5.9	0.0	0.0	0.0	8.6	0.0	0.0	0.0	0.0	2.4	1.0	0.0	0.0	0.0	0.0	0.0

Only action units with >0 durations are reported.

TABLE 2 | Means ± standard deviations of participants' ratings for each video in terms of emotion and appraisal label attributions.

Videos	Joy	Anger	Relief	Sadness	Contempt	Fear	Shame	Attributions					
								Suddenness	Goal Obstructive	Important and relevant	Coping potential	External norm violation	Respect of internal standards
1	0.18 (0.5)	0.41 (0.8)	1.18 (1.44)	1.36 (1.43)	0.45 (0.86)	1.5 (1.47)	0.73 (1.08)	4.27 (1.49)	3.95 (1.33)	3.91 (1.66)	3.95 (1.79)	3 (1.41)	5.45 (1.06)
2	0.13 (0.34)	1.06 (1)	0.63 (0.89)	2 (1.63)	1.69 (1.82)	2.06 (1.98)	0.63 (0.96)	5.38 (1.59)	5.38 (1.59)	4.94 (1.69)	4.38 (1.78)	3.25 (1.57)	4.94 (1.61)
3	0.29 (0.59)	0.29 (0.47)	1.71 (1.21)	0.29 (0.47)	0.76 (1.3)	0.82 (1.29)	0.47 (0.72)	4.24 (1.64)	3.35 (1.62)	3.41 (1.5)	5.65 (1.17)	3.29 (1.49)	5.71 (1.1)
4	0.5 (0.94)	1.57 (1.91)	0.57 (1.02)	0.36 (0.74)	2.57 (1.95)	0.5 (1.09)	0.21 (0.58)	0.07 (0.27)	1.57 (1.7)	0.29 (0.47)	2 (2.11)	1.36 (2.02)	2.57 (1.99)
5	0.35 (0.86)	0.47 (0.72)	0.88 (1.17)	0.82 (0.81)	0.29 (0.77)	0.71 (0.77)	0.53 (0.72)	3.65 (1.8)	3.94 (1.95)	4.06 (1.52)	4.59 (1.37)	3.65 (1.62)	5.06 (1.14)
6	0.47 (0.9)	0.47 (0.7)	0.84 (1.34)	0.68 (1)	0.68 (0.82)	0.74 (0.87)	0.32 (0.67)	4.47 (1.71)	3.95 (1.58)	3.89 (1.52)	5.47 (1.22)	3.26 (1.52)	5.63 (1.21)
7	0.21 (0.58)	1.86 (2.18)	0.57 (0.94)	1.21 (1.48)	1.86 (1.75)	1.5 (1.79)	1.07 (1.98)	3.43 (1.65)	3.93 (1.94)	3.64 (1.82)	4.57 (1.7)	3.14 (1.46)	5 (1.52)
8	0.04 (0.19)	1.57 (1.73)	0.31 (0.67)	1.66 (1.78)	1.43 (1.79)	2.06 (1.82)	0.92 (1.16)	4.53 (1.87)	4.51 (1.78)	4.25 (1.7)	3.7 (1.77)	3.99 (1.7)	4.69 (1.32)
9	0.03 (0.16)	1.3 (1.32)	0.13 (0.4)	2.18 (1.84)	1.03 (1.42)	2.08 (2)	1 (1.28)	4.9 (1.61)	4.95 (1.54)	4.98 (1.51)	3.15 (1.55)	3.65 (1.42)	5.23 (1.17)
10	0.16 (0.5)	2.74 (2)	0.05 (0.23)	1.37 (1.71)	2.74 (2.16)	1.42 (1.89)	0.74 (0.93)	4.74 (1.73)	5.26 (1.52)	4.53 (1.68)	3.32 (1.7)	4.95 (1.72)	4.32 (1.86)
11	0 (0)	2.32 (2.38)	0 (0)	0.89 (1.49)	1.89 (2.13)	0.53 (0.96)	0.84 (1.12)	3.95 (2.27)	3.84 (2.36)	3.74 (2.38)	3.21 (1.93)	3.79 (2.27)	3.37 (1.8)
12	0 (0)	1.65 (2.18)	0 (0)	0.65 (1)	0.94 (1.6)	0.41 (0.87)	1 (1.12)	2.88 (2.18)	3 (2.21)	3.12 (2.12)	2.65 (1.9)	2.71 (1.79)	2.88 (2.03)
13	0.03 (0.16)	1.69 (1.34)	0.15 (0.43)	1.26 (1.37)	1.79 (1.79)	1.54 (1.6)	0.54 (1.1)	4.44 (1.64)	4.69 (1.56)	4.41 (1.57)	3.49 (1.47)	3.97 (1.58)	4.44 (1.43)
14	0.1 (0.31)	1 (1.41)	0.6 (0.94)	1.45 (1.47)	0.7 (1.26)	2.15 (1.5)	0.3 (0.73)	5 (1.49)	4.7 (1.84)	4.35 (1.98)	4 (1.81)	3.05 (2.01)	5.4 (1.57)
15	0 (0)	0.95 (1.47)	0.11 (0.32)	2.26 (1.94)	0.37 (0.76)	3.37 (1.61)	0.26 (0.65)	5.68 (1.29)	5.58 (1.43)	5.53 (1.31)	2.63 (1.71)	3.63 (1.64)	4.42 (1.35)
16	0.53 (0.9)	2.63 (1.54)	0.05 (0.23)	0.37 (0.5)	2.37 (1.8)	0.89 (1.05)	0.63 (0.96)	5.26 (1.41)	5.21 (1.55)	5.11 (1.41)	4.63 (1.61)	5.79 (1.4)	5.47 (1.17)
17	0.27 (0.47)	0.55 (0.82)	0.55 (0.93)	0.64 (0.5)	0.27 (0.65)	1.55 (1.21)	0.27 (0.65)	5.27 (1.49)	4.73 (1.85)	5 (1.48)	3.91 (1.81)	3.73 (2.15)	4.45 (1.97)
18	0.11 (0.32)	1.89 (1.66)	0.42 (0.84)	0.74 (0.93)	1.21 (1.55)	1.11 (1.15)	1.47 (1.54)	3.79 (1.65)	4.26 (1.37)	3.84 (1.71)	5.05 (1.39)	4.74 (1.28)	4.95 (1.81)
19	0.14 (0.36)	0.79 (1.25)	0.36 (0.63)	1.93 (1.69)	0.86 (1.41)	2.5 (1.95)	0.86 (1.03)	2.64 (1.34)	6.14 (1.03)	2.93 (1.54)	4.29 (2.09)	4.64 (1.74)	4.21 (1.53)
20	0.14 (0.67)	1.79 (1.83)	0.11 (0.46)	1.99 (1.86)	1.78 (1.95)	2.86 (2.04)	0.89 (1.33)	4.85 (1.76)	4.95 (1.73)	4.71 (1.75)	2.6 (1.53)	3.97 (1.49)	4.11 (1.4)
21	0.24 (0.56)	1.12 (1.05)	0.88 (1.45)	0.53 (0.72)	1.47 (1.91)	1.06 (1.03)	1.06 (1.34)	4.65 (1.66)	4.12 (1.58)	4.18 (1.59)	5 (1.62)	4.06 (2.01)	4.94 (1.48)
22	0 (0)	1 (1.28)	0.08 (0.29)	1.75 (2.14)	1.17 (1.95)	2.58 (2.39)	1.5 (1.57)	5 (2.04)	4.83 (2.12)	4.75 (2.01)	2 (1.21)	3.25 (1.66)	3.92 (1.93)
23	1.29 (2.11)	1 (1.28)	1.37 (2)	0.9 (1.45)	1.1 (1.41)	2.17 (1.86)	1.68 (1.75)	4.15 (1.92)	4.83 (1.26)	3.46 (2.36)	3.02 (1.42)	4.02 (1.35)	4.56 (1.53)
24	0.03 (0.16)	1.5 (1.3)	0.3 (0.76)	0.98 (1.33)	1.9 (1.93)	0.83 (1.2)	1.03 (1.58)	3.65 (1.59)	3.83 (1.66)	3.7 (1.38)	5 (1.32)	3.58 (1.75)	4.6 (1.46)
25	0.68 (1.68)	2.4 (1.73)	0.51 (1.14)	2.05 (1.63)	2.21 (1.86)	1.92 (1.68)	0.77 (1.34)	4.6 (1.61)	5.3 (1.51)	4.12 (2.05)	2.84 (1.55)	4.48 (1.8)	3.6 (1.82)
26	0.2 (0.89)	0.8 (0.89)	0.15 (0.49)	1.75 (1.77)	0.75 (1.07)	2.55 (1.82)	1.9 (2.15)	5.75 (1.12)	5.1 (1.62)	4.95 (1.39)	3.1 (1.59)	3.3 (1.75)	4.05 (1.79)
27	0.07 (0.27)	1.57 (1.7)	0.29 (0.47)	2 (2.11)	1.36 (2.02)	2.57 (1.99)	1.36 (1.95)	5.14 (1.41)	5.36 (1.28)	5.21 (1.19)	3.07 (1.44)	3.29 (2.05)	4.64 (1.22)
28	0.2 (0.73)	2.31 (1.55)	0.15 (0.59)	3.05 (1.93)	1.6 (1.77)	2.22 (1.98)	1.69 (1.95)	3.95 (2.53)	4.13 (2.56)	4.49 (2.04)	2.27 (1.73)	3.4 (2.28)	4.16 (1.71)
29	0.03 (0.16)	1.75 (1.25)	0.92 (1.91)	3.32 (1.74)	2.08 (1.92)	2.1 (1.63)	0.86 (1.32)	4.82 (2.14)	5.16 (1.61)	4.9 (1.67)	2.64 (1.56)	4.3 (1.65)	4.29 (1.5)
30	1.63 (1.99)	2 (1.89)	0.88 (1.29)	1.15 (1.47)	1.21 (1.81)	1.67 (1.69)	1.9 (1.75)	3.52 (2.36)	3.75 (2.38)	4.56 (1.69)	3.56 (2.25)	4.06 (1.81)	4.83 (1.53)
31	0 (0)	2.03 (1.83)	0.13 (0.65)	1.23 (1.44)	1.6 (1.93)	1.4 (1.53)	1.4 (1.69)	4.72 (1.57)	4.9 (1.84)	4.68 (1.75)	3.38 (1.53)	4.3 (1.83)	4.43 (1.38)
32	1.21 (1.67)	1 (1.57)	1.64 (1.95)	0.86 (0.95)	0.86 (1.83)	0.79 (0.89)	0.64 (0.84)	4.64 (1.45)	4.14 (1.56)	3.79 (1.53)	4.79 (1.67)	3.57 (1.91)	5.21 (1.76)
33	0.26 (0.73)	0.68 (0.95)	0.63 (0.96)	0.37 (0.68)	0.68 (1.06)	0.95 (1.03)	0.42 (0.96)	4.37 (1.74)	3.68 (1.45)	4.21 (1.65)	4.16 (1.5)	3.58 (1.61)	3.95 (1.65)
34	0.63 (0.76)	1.32 (1.57)	0.95 (1.27)	0.64 (0.94)	1 (1.56)	0.89 (1.2)	0.63 (1.01)	4.74 (1.59)	4.42 (1.61)	4.37 (1.67)	3.84 (1.71)	3.63 (1.64)	4.16 (1.57)
35	2.82 (1.74)	0.24 (0.66)	1.82 (1.67)	0.24 (0.44)	0.24 (0.75)	0.12 (0.33)	0.59 (1.12)	3.41 (1.58)	3.06 (1.3)	3 (1.27)	5.41 (1.33)	2.47 (1.33)	5 (1.41)
36	0.99 (1.86)	1.44 (1.22)	1.18 (1.83)	1.59 (1.66)	1.29 (1.52)	1.12 (1.4)	0.65 (1.11)	3.61 (2.04)	4.21 (1.68)	3.49 (2.07)	5.05 (1.78)	3.74 (1.59)	5.25 (1.4)
37	0.41 (0.8)	1.94 (1.82)	0.47 (0.94)	1.43 (1.47)	0.94 (1.2)	1.59 (1.46)	0.88 (1.05)	3.53 (1.62)	4.71 (1.31)	5 (1.62)	4.88 (1.62)	4.06 (1.25)	0.94 (1.82)
38	0.18 (0.45)	0.63 (0.81)	0.73 (1.22)	0.58 (1.02)	0.65 (1.08)	1.28 (1.48)	0.9 (1.41)	4.48 (1.6)	4.2 (1.77)	4.55 (1.45)	4.68 (1.69)	3.1 (1.5)	5.48 (1.3)
39	0.42 (0.84)	0.16 (0.37)	1.37 (1.5)		0.21 (0.54)	0.53 (0.77)	0.37 (0.83)	4.58 (1.57)	3.84 (1.54)	4.32 (1.25)	4.74 (1.56)	2.95 (1.08)	5.21 (1.44)

(**Table 3**). All the final regression models showed high effect-size parameters ($R^2 > 0.46$) and high statistical power ($1 - \beta > 0.99$).

When we evaluated the profiles of AUs predicting each emotion/appraisal (**Table 4**), we found that several predictions based on prior observations in the literature concerning the relation between facial actions and emotion/appraisal attributions were confirmed. Specifically, positive associations were found between joy and AU 12 (upward lip corner pulling); between anger and AU 1 (inner eyebrow raise) and AU 10 (nasolabial furrow deepening); between sadness and AU 1 (inner brow raise) between fear and AU 5 (opening of the eye/upper lid raise) and marginally AU 1 (inner brow raise); and between shame and AU 2 (outer brow raise), AU 5 (opening of the eye/upper lid raise), AU 20 (lip stretch), AU 25 (mouth opening) and marginally AU 7 (lower eyelid contraction). In terms of cognitive appraisals, goal obstruction and perception of an event as relevant but incongruent were positively associated with AU 17 (chin raise). Perception of coping potential was associated with AU 4 (brow lowering) and AU 24 (lip pressing).

At the same time, we found several unexpected positive associations between AUs and recognition of emotions/appraisals. For instance, AU 16 (lower lip depressor) was associated with fear as well as goal obstruction. It is interesting to note that there were also unexpected negative associations between facial actions and emotion/appraisal attribution (see **Table 4**). For example, the AU12 (smile) had negative associations with the attribution of some negative emotions, such as sadness and shame, but not with any appraisals. In terms of appraisal attribution, a negative association was observed for instance for AU 2 (outer brow raiser) and coping potential.

DISCUSSION

In our study we looked at the decoding of emotions and cognitive appraisals from sets of AUs seen in a naturally negative emotional setting and we addressed this question through stepwise regression analyses. Results supported our predictions and revealed the relationships between AUs and the decoding of all emotions and cognitive appraisals. These results are consistent with some previous theories postulating the relationships between decoding of emotional categories or cognitive appraisals and sets of AUs (Ekman, 1992; Scherer and Ellgring, 2007), although other theories questioned such relationships (see Barrett et al., 2018). The results are also consistent with previous empirical studies investigating these relationships (e.g., Kohler et al., 2004). However, previous studies did not test spontaneous emotional expressions in naturalistic settings, and hence, the generalizability of these relationships to real-life facial expression processing remained unclear. Extending the current theoretical and empirical knowledge, our results suggest that decoding of emotional categories and cognitive appraisals can be accomplished through the recognition of specific facial movements.

The profiles of AUs associated with the decoding of emotional categories and cognitive appraisals were at least partially consistent with those in previous theories (Ekman, 1992; Scherer and Ellgring, 2007). For instance, the duration of the AU 1 (inner brow raise) and AU 12 (upward lip corner pulling) was associated with the attribution of sadness and joy, respectively. The duration of the AU 4 (brow lowering) and AU 17 (chin raise) was associated with coping potential and goal obstruction, respectively. These findings are also consistent with previous studies with actors (e.g., Kohler et al., 2004). Our results empirically support the notion that these AUs could be the core facial movements to decode emotional categories and cognitive appraisal in natural, spontaneous facial expressions.

At the same time, our results also showed several inconsistent patterns with theoretical predictions (Ekman, 1992; Scherer and Ellgring, 2007). For example, outer brow raiser was not associated with suddenness and lower lip depressor was associated with fear as well as with goal obstruction. Further testing is required for validation purposes in dynamic naturalistic settings as it might be useful to include these AUs in the new theories regarding the relationships between emotion/appraisal decoding and AUs. Furthermore, our results revealed some negative relationships between the duration of AUs and the decoding of emotions/appraisal. This is consistent with results from one rating study of photographs of acted emotional expressions (Galati et al., 1997). In our study, for example, smiles were negatively associated with sadness and shame. These findings suggest that not only the present but also the absent facial movements can be decoded as messages of emotions or appraisals in natural, dynamic, face-to-face communication.

Our findings specifying the relationships between the decoding of emotions/appraisals and AUs in spontaneous facial expressions could have practical implications. For example, it may be possible to build artificial intelligent systems to read emotions/appraisals from emotional facial expressions in a more human-like way. Although such systems currently exist, almost all of them appear to be constructed based on theories

TABLE 3 | Results of final models in stepwise regression analyses.

Dependent variable	F	df	p	R^2	1 - β
Joy	6.54	6.34	0.000	0.54	1.00
Anger	4.83	14.26	0.000	0.72	1.00
Sadness	5.00	10.30	0.000	0.63	1.00
Contempt	4.44	14.26	0.001	0.71	1.00
Fear	11.74	8.32	0.000	0.75	1.00
Shame	3.73	22.18	0.003	0.82	1.00
Suddenness	4.43	11.29	0.001	0.63	1.00
Goal obstruction	5.88	6.34	0.000	0.51	1.00
Relevance and discrepancy	3.57	10.30	0.003	0.54	1.00
Coping potential	4.39	16.24	0.001	0.75	1.00
External standards violation	3.56	8.32	0.005	0.47	1.00
Internal standards violation	4.74	18.22	0.000	0.80	1.00

TABLE 4 | Results of significant independent variables of final models in stepwise regression analyses.

Dependent variable	Statistics	1 Inner brow raiser	2 Outer brow raiser	4 Brow lower	5 Upper lid raiser	6 Cheek raiser	7 Lid tighten	8 Lips toward each other	10 Upper lip raiser	11 Nasolabial deepener	12 Lip corner puller	13 Sharp lip puller	14 Dimpler	15 Lip corner depressor	16 Lower lip depressor	17 Chin raiser	18 Lip puckerer	20 Lip stretch	22 Lip funnel	23 Lip tighten	24 Lip pressor	25 Lips part	26 Jaw drop	27 Mouth stretch	28 Lip suck	29 Jaw thrust	32 Lip bite	34 Cheek puff	37 Lip wipe	43 Eyes closed	50 Speech
Joy	β										1.12									−1.70	0.53				2.42		−1.42				−0.80
	t										6.14									5.22	3.26				4.08		3.07				4.56
	p										0.000									0.000	0.003				0.000		0.004				0.000
Anger	β	0.58	−0.23	0.62			−0.98		0.65			0.28					0.28	−0.23		0.69		−0.34				0.41	0.30	0.38		−0.30	
	t	3.82	1.71	3.05			3.76		4.39			2.51					1.98	1.72		2.69		2.74				3.00	1.88	3.31		2.46	
	p	0.001	0.100	0.005			0.001		0.000			0.018					0.058	0.097		0.012		0.011				0.006	0.071	0.003		0.021	
Relief	β												0.28			−0.33		0.32							1.21		−1.09				
	t												1.73			1.98		2.16							2.34		2.15				
	p												0.092			0.055		0.038							0.025		0.038				
Sadness	β	1.01	−0.53			0.41		−0.21		−0.50	−0.60		−0.42	−0.25									−0.30								−0.28
	t	5.01	3.47			2.80		1.75		2.73	4.26		3.20	1.84									2.43								1.82
	p	0.000	0.002			0.009		0.090		0.011	0.000		0.003	0.076									0.021								0.079
Contempt	β	0.54	−0.35			0.27			0.50	−0.55					−0.53		0.34			0.95			−0.41					0.36			−0.43
	t	2.37	2.14			2.08			3.63	2.34					4.11		2.35			3.64			3.34					3.06			2.18
	p	0.026	0.042			0.047			0.001	0.028					0.000		0.027			0.001			0.003					0.005			0.038
Fear	β	0.23			0.29			−0.31	−0.55				−0.29		0.27	0.35														−0.40	
	t	1.85			3.03			3.20	5.66				2.79		2.71	3.42														3.36	
	p	0.074			0.005			0.003	0.000				0.009		0.011	0.002														0.002	
Shame	β		0.58	−1.11	0.42		0.69				−1.69	0.71	−0.49	−0.46	−0.36	2.20	−0.85	0.55		2.63	−2.10	1.04	−0.53	−1.00	−7.40		5.56	0.28	−1.14		−0.43
	t		2.89	4.34	3.21		1.79				3.71	2.95	3.22	2.56	1.93	5.26	3.63	3.13		4.68	4.65	4.17	2.92	5.52	4.52		4.74	2.37	4.09		1.97
	p		0.010	0.000	0.005		0.090				0.002	0.008	0.005	0.020	0.070	0.000	0.002	0.006		0.000	0.000	0.001	0.009	0.000	0.000		0.000	0.029	0.001		0.065
Suddenness	β		−0.47		0.55	0.29							0.24	−0.45		0.87				0.37	−0.41	0.42			−1.33		1.17		−0.71		
	t		3.22		4.25	2.08							1.73	3.08		4.27				1.71	2.26	3.41			2.40		2.42		4.46		
	p		0.003		0.000	0.046							0.096	0.005		0.000				0.098	0.032	0.002			0.023		0.022		0.000		
Goal obstruction	β		−0.57		0.48			−0.32				−0.21			0.26	0.44										0.28					
	t		4.19		3.71			2.56				1.70			2.13	3.47										1.84					
	p		0.000		0.001			0.015				0.098			0.040	0.001										0.075					
Relevance and discrepancy	β	−0.65	−0.63		0.46	0.41								−0.50	0.34	0.56								0.31	−0.94	0.24	1.00	−0.49			
	t	3.19	4.16		3.27	2.62								3.15	2.71	3.10								2.50	1.94	1.78	2.12	3.16			
	p	0.004	0.000		0.003	0.014								0.004	0.011	0.004								0.020	0.062	0.085	0.042	0.004			
Coping potential	β		−0.61	0.51					−0.29		1.22	−0.72				−0.40				−1.64	1.05		0.38		2.80		−1.92	0.44			0.62
	t		3.77	2.57					1.80		4.48	4.13				1.81				3.71	3.92		2.96		3.51		3.24	2.59			3.58
	p		0.001	0.017					0.084		0.000	0.000				0.083				0.001	0.001		0.007		0.002		0.004	0.016			0.002
External standards violation	β		−0.28				−0.67							−0.25		0.59											0.50				−0.48
	t		2.12				3.14							1.70		3.70											2.46				2.53
	p		0.042				0.004							0.099		0.001											0.020				0.017
Internal standards violation	β	−0.35	−1.07					−0.53	−0.43		0.97			−0.40					−0.21	−1.47	0.98		0.48	0.34	3.56		−2.60				0.31
	t	1.94	6.10					2.91	2.90		3.75			2.96					1.78	3.55	4.45		4.01	3.11	4.09		3.68				1.93
	p	0.066	0.000					0.008	0.008		0.001			0.007					0.090	0.002	0.000		0.001	0.005	0.000		0.001				0.067

or data with actors' deliberate expressions (Paleari et al., 2007; Niewiadomski et al., 2011; Ravikumar et al., 2016; Fourati and Pelachaud, 2018). Additionally, it may be possible to build humanoid virtual agents and robots (Poggi and Pelachaud, 2000; Lim and Okuno, 2015; Niewiadomski and Pelachaud, 2015) for applications in healthcare or in the long term with the elderly, with expressions, which could be recognized as showing natural human-like emotional expressions. Finally, given the importance of appropriate understanding of inner states displayed in others' faces in healthy social functioning (McGlade et al., 2008), it may be interesting to assess the relationships between decoding of emotions/appraisals and AUs using naturalistic facial expression stimuli in clinical conditions. Indeed, several clinical populations report social cognition impairments in real-life situations, while showing satisfactory performance in typical emotion recognition or theory of mind tasks, which mostly rely on the judgment of pictures of acted facial expressions or exaggerated social stories (e.g. see Bala et al., 2018). Dynamic and more naturalistic approaches might help define clinical impairments faced for example by patients with Schizophrenia (Okruszek et al., 2015; Okruszek, 2018), amygdala lesions (Bala et al., 2018) or high functioning autism (Murray et al., 2017) and eventually lead to the improvement of existing social cognition trainings.

Several limitations of the present study should be acknowledged. First, our naturalistic set was limited in the number of stimuli and included only a negatively valenced situation at a single location. In order to generalize the findings, more positive and negative situations presented in varied and controlled contexts and cultures would need to be investigated. Furthermore, although we lacked data regarding emotions experienced by the expressers and we did not monitor the internal states of the participants, it would have been interesting to investigate interactions between AUs and encoded/decoded emotions and the characteristics of the observers. Given the literature on how emotions, facial mimicry and moods of observers influence emotion perception in others (e.g., Schmid and Schmid Mast, 2010; Wood et al., 2016; Wingenbach et al., 2018) it is a valuable topic in future research on dynamic naturalistic stimuli interpretation. Second, we analyzed only male participants. Although consistent gender differences have not been reported in terms of rating-style in the decoding of emotional expressions (e.g., Duhaney and McKelvie, 1993; Biele and Grabowska, 2006; for a review, see Forni-Santos and Osório, 2015), numerous studies have reported that the gender of the decoder might influence different aspects of the processing of faces. For example, the recognition of gender of faces was enhanced (reaction time reduced) when these were presented looking away from the decoder of opposite gender, but not in the case of a same gender decoder (Vuilleumier et al., 2005). It has also been reported that exposure to angry male as opposed to angry female faces activated the visual cortex and the anterior cingulate gyrus significantly more in men than in women (Fischer et al., 2004). Similarly, although no significant differences were observed in accuracy ratings by male vs. female participants nor in the recognition of male vs. female encoder faces, higher brain activity was observed in the extrastriate body

area in reaction to threatening male faces compared to female faces, as well as in the activity of the amygdala to threatening vs. neutral female faces in male but not female participants (Kret et al., 2011). For all those reasons, the effect of gender of decoder participants needs to be carefully monitored in further studies. Third, although our final models had high statistical power, our sample size was small. In our approach we used stepwise regression analyses in order to select a subset of predictor variables. While having expected a number of predictor variables to be observed based on previous evidence, and our analyses having detected the expected number of predictor variables with high power, our analyses lacked the power to sufficiently analyse AUs not included in the final models. Future studies with a larger sample size may reveal the involvement of more AUs in the decoding of emotions/appraisals. Fourth, we coded single AUs but not the combination (i.e., simultaneous appearance) of AUs (e.g., AU 6 + 12) due to the lack of power. Because single vs. combined AUs could transmit different emotional messages (Ekman and Friesen, 1975), investigation of AU combination is an important matter for future research. Fifth, we coded AUs in a binary fashion as conducted in the previous studies testing the AUs and decoding of emotions (e.g., Kohler et al., 2004). The coding of 5-level AU intensity, which were newly added in FACS coding (Ekman et al., 2002), may provide more detailed insights regarding the relationships. Sixth, we studied only a linear additive relationship between AUs and the decoding of emotions and their components to simplify analyses. Further work could go beyond linear associations, e.g., quadratic associations. Finally, the use of naturalistic behavior in perceptive paradigms only allows for correlational studies, without the possibility of any strong claims of causality. When constructing paradigms allowing for causality testing, one aspect of interest for future investigations is the direct influence of single facial units on attributions, and future studies could carefully manipulate the presentation of AUs while keeping as much as possible of a naturalistic setting. One method to manipulate behavior one-by-one is to reproduce human behavior using a virtual humanoid or robot. Today's technology allows for dynamic and functional representations of human behavior, which can be copied from a naturalistic scene in sufficient detail in order to evoke similar reactions to the one's observed in videos of humans. Given that presenting behavior without context or one AU at a time lacks *naturalness*, AUs should be judged in sets of units they originally appear in. The manipulation of single AUs could focus on the removal of existing actions (see Hyniewska, 2013).

In conclusion, numerous studies have investigated the decoding of emotional expressions from prototypical displays and there seems to be unanimity on sets of facial AUs that provide good discriminability. However, to the authors' best knowledge, no study has looked at sets of AUs that lead to emotion and appraisal perception in naturally occurring situations. Our results show that emotional and appraisal labels can be predicted based on recorded sets of facial actions units. Interestingly, the sets of observed AUs do not coincide with what has been observed in former decoding studies.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

REFERENCES

Bala, A., Okruszek, Ł., Piejka, A., Głębicka, A., Szewczyk, E., Bosak, K., et al. (2018). Social perception in mesial temporal lobe epilepsy: interpreting social information from moving shapes and biological motion. *J. Neuropsychiatry Clin. Neurosci.* 30, 228–235. doi: 10.1176/appi.neuropsych.17080153

Barrett, L. F., Khan, Z., Dy, J., and Brooks, D. (2018). Nature of emotion categories: comment on cowen and keltner. *Trends Cogn. Sci.* 22, 97–99. doi: 10.1016/j.tics.2017.12.004

Biele, C., and Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Exp. Brain Res.* 171, 1–6. doi: 10.1007/s00221-005-0254-0

Calvo, M. G., and Nummenmaa, L. (2015). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Cicchetti, D. V., and Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am. J. Ment. Defic.* 86, 127–137.

Cohen, J. (1960). Kappa: coefficient of concordance. *Educ. Psych. Meas.* 20:37.

Duhaney, A., and McKelvie, S. J. (1993). Gender differences in accuracy of identification and rated intensity of facial expressions. *Percept. Mot. Skills* 76, 716–718. doi: 10.2466/pms.1993.76.3.716

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6:169–200. doi: 10.1080/02699939208411068

Ekman, P. (2005). "Conclusion: What we have learned by measuring facial behaviour," in What the face reveals: Basic and applied studies of spontaneous expression using the facial Action Coding System (FACS) *2nd Edn.*, eds P. Ekman, and E. L. Rosenberg (Oxford: Oxford University Press), 605–626.

Ekman, P., Friesen, W., and Hager, J. (2002). *Facial Action Coding System*. Salt Lake City: Research Nexus eBook.

Ekman, P., and Friesen, W. V. (1975). *Unmasking the Face*. Englewood Cliffs, NJ: Prentice-Hall.

Ekman, P., and Friesen, W. V. (1978). *Manual of the Facial Action Coding System (FACS)*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverb. Behav.* 6, 238–252. doi: 10.1007/BF00987191

Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Fiorentini, C., Schmidt, S., and Viviani, P. (2012). The identification of unfolding facial expressions. *Perception* 41, 532–555. doi: 10.1068/p7052

Fischer, H., Sandblom, J., Herlitz, A., Fransson, P., Wright, C. I., and Backman, L. (2004). Sex-differential brain activation during exposure to female and male faces. *Neuroreport* 15, 235–238. doi: 10.1097/00001756-200402090-00004

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., and Ellsworth,P.C. (2007). The world of emotions is not two-dimensional. *Psychol. Sci.* 18, 1050–1057. doi: 10.1111/j.1467-9280.2007.02024.x

Forni-Santos, L., and Osório, F. L. (2015). Influence of gender in the recognition of basic facial expressions: A critical literature review. *World J. Psychiatry.* 5, 342–351. doi: 10.5498/wjp.v5.i3.342

Fourati, N., and Pelachaud, C. (2018). Perception of emotions and body movement in the emilya database. *IEEE Trans. Affect Computing* 9, 90–101. doi: 10.1109/TAFFC.2016.2591039

Galati, D., Scherer, K. R., and Ricci-Bitti, P. E. (1997). Voluntary facial expression of emotion: Comparing congenitally blind with normally sighted encoders. *J. Pers. Soc. Psychol.* 73, 1363–1379. doi: 10.1037/0022-3514.73.6.1363

Hess, U., and Kleck, R. E. (1994). The cues decoders use in attempting to differentiate emotion elicited and posed facial expressions. *Eur. J. Soc. Psychol.* 24, 367–381. doi: 10.1002/ejsp.2420240306

Hyniewska, S. (2013). *Non-Verbal Expression Perception and Mental State Attribution by Third Parties*. PhD thesis, Télécom-ParisTech/University of Geneva.

Kohler, C. G., Turner, T., Stolar, N., Bilker, W., Brensinger, C. M., Gur, R. E., et al. (2004). Differences in facial expressions of four universal emotions. *Psychiatry Res.* 128, 235–244 doi: 10.1016/j.psychres.2004.07.003

Kret, M. E., Pichon, S., Grèzes, J., and de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage* 54, 1755–1762. doi: 10.1016/j.neuroimage.2010.08.012

Lim, A., and Okuno, H. G. (2015). A recipe for empathy. *Int. J. Soc. Robot.* 7, 35–49. doi: 10.1007/s12369-014-0262-y

McGlade, N., Behan, C., Hayden, J., O'Donoghue, T., Peel, R., Haq, F., et al. (2008). Mental state decoding v. mental state reasoning as a mediator between cognitive and social function in psychosis. *Br. J. Psychiatry* 193, 77–78. doi: 10.1192/bjp.bp.107.044198

Mehu, M., Mortillaro, M., Bänziger, T., and Scherer, K. R. (2012). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Emotion* 12, 701–715. doi: 10.1037/a0026717

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., et al. (2017). A new test of advanced theory of mind: the "Strange Stories Film Task" captures social processing differences in adults with autism spectrum disorders. *Autism Res.* 10, 1120–1132. doi: 10.1002/aur.1744

Niewiadomski, R., Hyniewska, S. J., and Pelachaud, C. (2011). Constraint-based model for synthesis of multimodal sequential expressions of emotions. *IEEE Trans. Affect. Comput.* 2, 134–146. doi: 10.1109/T-AFFC.2011.5

Niewiadomski, R., and Pelachaud, C. (2015). The effect of wrinkles, presentation mode, and intensity on the perception of facial actions and full-face expressions of laughter. *ACM Trans. Appl. Percept.* 12, 1–21. doi: 10.1145/2699255

Okruszek, Ł. (2018). It Is Not Just in Faces! Processing of emotion and intention from biological motion in psychiatric disorders. *Front. Hum. Neurosci.* 12:48. doi: 10.3389/fnhum.2018.00048

Okruszek, Ł., Haman, M., Kalinowski, K., Talarowska, M., Becchio, C., and Manera, V. (2015). Impaired recognition of communicative interactions from biological motion in schizophrenia. *PLoS ONE* 10:e0116793. doi: 10.1371/journal.pone.0116793

Paleari, M., Grizard, A., Lisetti, C., and Antipolis, S. (2007). "Adapting Psychologically grounded facial emotional expressions to different anthropomorphic embodiment platforms," in *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference* (Key West, FL).

Poggi, I., and Pelachaud, C. (2000). "Performative facial expressions in animated faces," in *Embodied Conversational Agents*, eds J. Cassell, J. Sullivan, S. Prevost, and E. Churchill (Cambridge: MIT Press), 155–188.

Ravikumar, S., Davidson, C., Kit, D., Campbell, N., Benedetti, L., and Cosker, D. (2016). Reading between the dots: combining 3D markers and FACS classification for high-quality blendshape facial animation. *Proc. Graphics Interface* 2016, 143–151. doi: 10.20380/GI2016.18

Reisenzein, R., and Meyer, W.U. (2009). "Surprise," in *Oxford Companion to the Affective Sciences*, eds D.Sander and K. R. Scherer (Oxford: Oxford University Press), 386–387.

Reisenzein, R., Meyer, W. U., and Niepel, M. (2012). "Surprise," in *Encyclopedia of Human Behaviour, 2nd Edn.*, ed V. S. Ramachandran (Amsterdam: Elsevier), 564–570.

Ruengvirayudh, P., and Brooks, G. P. (2016). Comparing stepwise regression models to the best-subsets models, or, the art of stepwise. *Gen Linear Model J.* 42, 1–14.

Russell, J. A., and Fernández-Dols, M. J. (1997). *The Psychology of Facial Expression*. New York, NY: Cambridge University Press.

Scherer, K., and Ceschi, G. (2000). Criteria for emotion recognition from verbal and nonverbal expression: studying baggage loss in the airport. *Pers. Soc. Psychol. Bull.* 26, 327–339. doi: 10.1177/0146167200265006

Scherer, K. R. (1984). "On the nature and function of emotion: a component process approach," in *Approaches to Emotion*, eds K. R. Scherer and P. Ekman (Hillsdale, NJ: Erlbaum), 293–318.

Scherer, K. R. (2001). "Appraisal considered as a process of multilevel sequential checking," in *Appraisal Processes in Emotion: Theory, Methods, Research*, eds. K. R. Scherer, A. Schorr, and T. Johnstone (Oxford: Oxford University Press), 92–120.

Scherer, K. R., and Ceschi, G. (1997). Lost luggage: a field study of emotion-antecedent appraisal. *Motiv. Emot.* 21, 211–235. doi: 10.1023/A:1024498629430

Scherer, K. R., and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* 7, 113–130. doi: 10.1037/1528-3542.7.1.113

Schmid, P. C., and Schmid Mast, M. (2010). Mood effects on emotion recognition. *Motiv. Emot.* 34, 288–292. doi: 10.1007/s11031-010-9170-0

Smith, C. A., and Scott, H. S. (1997). "A componential approach to the meaning of facial expressions," in *The Psychology of Facial Expression*, eds J. A. Russell and M. J. Fernández-Dols (New York, NY: Cambridge University Press), 229–254.

Topolinski, S., and Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. *Front. Psychol.* 6:134. doi: 10.3389/fpsyg.2015.00134

Vuilleumier, P., George, N., Lister, V., Armony, J., and Driver, J. (2005). Effects of perceived mutual gaze and gender on face processing and recognition memory. *Vis. Cogn.* 12, 85–101. doi: 10.1080/13506280444000120

Wingenbach, T. S., Brosnan, M., Pfaltz, M. C., Plichta, M. M., and Ashwin, C. (2018). Incongruence between observers' and observed facial muscle activation reduces recognition of emotional facial expressions from video stimuli. *Front. Psychol.* 9:864. doi: 10.3389/fpsyg.2018.00864

Wood, A., Rychlowska, M., Korb, S., and Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends Cogn. Sci.* 20, 227–240. doi: 10.1016/j.tics.2015.12.010

Yik, M., Meng, Z., and Russell, J. A. (1998). Adults' freely produced emotion labels for babies' spontaneous facial expressions. *Cogn. Emot.* 12, 723–730.

Dynamic Facial Expression of Emotion and Observer Inference

Klaus R. Scherer[1]*, Heiner Ellgring[2], Anja Dieckmann[3], Matthias Unfried[3] and Marcello Mortillaro[1]

[1] Department of Psychology and Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland,
[2] Department of Psychology, University of Würzburg, Würzburg, Germany, [3] GfK Verein, Nuremberg, Germany

Research on facial emotion expression has mostly focused on emotion recognition, assuming that a small number of discrete emotions is elicited and expressed via prototypical facial muscle configurations as captured in still photographs. These are expected to be recognized by observers, presumably via template matching. In contrast, appraisal theories of emotion propose a more dynamic approach, suggesting that specific elements of facial expressions are directly produced by the result of certain appraisals and predicting the facial patterns to be expected for certain appraisal configurations. This approach has recently been extended to emotion perception, claiming that observers first infer individual appraisals and only then make categorical emotion judgments based on the estimated appraisal patterns, using inference rules. Here, we report two related studies to empirically investigate the facial action unit configurations that are used by actors to convey specific emotions in short affect bursts and to examine to what extent observers can infer a person's emotions from the predicted facial expression configurations. The results show that (1) professional actors use many of the predicted facial action unit patterns to enact systematically specified appraisal outcomes in a realistic scenario setting, and (2) naïve observers infer the respective emotions based on highly similar facial movement configurations with a degree of accuracy comparable to earlier research findings. Based on estimates of underlying appraisal criteria for the different emotions we conclude that the patterns of facial action units identified in this research correspond largely to prior predictions and encourage further research on appraisal-driven expression and inference.

Keywords: dynamic facial emotion expression, emotion recognition, emotion enactment, affect bursts, appraisal theory of emotion expression

INTRODUCTION

A comprehensive review of past studies on facial, vocal, gestural, and multimodal emotion expression (Scherer et al., 2011) suggests three major conclusions: (1) emotion expression and emotion perception, which constitute the emotion communication process, are rarely studied in combination, (2) historically, most studies on facial expression have relied on photos of facial expressions rather than on dynamic expression sequences (with some exceptions, e.g., Krumhuber et al., 2017), and (3) the research focus was mainly on emotion recognition, particularly recognition accuracy, rather than on the production of facial expressions and the analysis of the cues used by observers to infer the underlying emotions.

There are some notable exceptions to these general trends. Hess and Kleck (1994) studied the extent to which judges rating videos of encoders' spontaneously elicited and posed emotions could identify the cues that determined their impression of spontaneity and deliberateness of the facial expressions shown. They used the Facial Action Coding System (FACS; Ekman and Friesen, 1978) to identify eye movements and the presence of action unit (AU) 6, crow's feet wrinkles, expected to differentiate spontaneous and deliberate smiles (Ekman and Friesen, 1982; Ekman et al., 1988). They found that AU6 was indeed reported as an important cue used to infer spontaneity even though it did not objectively differentiate the eliciting conditions. The authors concluded that judges overgeneralized this cue as they also used it for disgust expressions. In general, the results confirmed the importance of dynamic cues for the inference of spontaneity or deliberateness of an expression. Recent work strongly confirms the important role of dynamic cues for the judging of elicited vs. posed expressions (e.g., Namba et al., 2018; Zloteanu et al., 2018).

Scherer and Ceschi (2000) examined the inference of genuine vs. polite expressions of emotional states in a large-scale field study in a major airport. They asked 110 airline passengers who had just reported their luggage lost at the baggage claim counter, to rate their emotional state (subjective feeling criterion). The agents who had processed the claims were asked to rate the passengers' emotional state. Excerpts of the videotaped interaction for 40 passengers were rated for the underlying emotional state by judges based on (a) verbal and non-verbal cues or (b) non-verbal cues only. In addition, the video clips were objectively coded using the Facial Action Coding System (FACS; Ekman and Friesen, 1978). The results showed that "felt," but not "false" smiles [as defined by Ekman and Friesen (1982)] correlated strongly positively with a "in good humor" scale in agent ratings and both types of judges' ratings, but only weakly so with self-ratings. The video material collected by Scherer and Ceschi in this field study was used by Hyniewska et al. (2018) to study the emotion antecedent appraisals (see Scherer, 2001) and the resulting emotions of the voyagers claiming lost baggage inferred by judges on the basis of the facial expressions. The videos were annotated with the FACS system and stepwise regression was used to identify the AUs predicting specific inferences. The profiles of regression equations showed AUs both consistent and inconsistent with those found in published theoretical proposals. The authors conclude that the results suggest: (1) the decoding of emotions and appraisals in facial expressions is implemented by the perception of sets of AUs, and (2) the profiles of such AU sets could be different from previous theories.

What remains to be studied in order to better understand the underlying dynamic process and the *detailed mechanisms* involved in emotion expression *and* inference is *the nature of the morphological cues in relation to the different emotions expressed and the exact nature of the inferences of emotion categories from these cues*. In this article, we argue that the process of emotion communication and the underlying mechanisms can only be fully understood when the process of emotional expression is studied in conjunction with emotion perception and inference (decoding) based on a detailed examination of the relevant morphological cues—the facial muscle action patterns involved. Specifically, we suggest using a Brunswikian lens model approach (Brunswik, 1956) to allow a comprehensive dynamic analysis of the process of facial emotion communication. In particular, such model and its quantitative testing can provide an important impetus for future research on the dynamics of emotional expression by providing a theoretically adequate framework that allows hypothesis testing and accumulation of results (Bänziger et al., 2015).

Scherer (2013a) has formalized an extension of the lens model as a tripartite emotion expression and perception (TEEP) model (see **Figure 1**), in which the communication process is represented by four elements and three phases. The internal state of the sender (e.g., the emotion experienced) is encoded via distal cues (measured by objective, quantitative analysis); the listener perceives the vocal utterance, the facial expression and other non-verbal behavior and extracts a number of proximal cues (measured by subjective ratings obtained from naive observers), and, finally, some of these proximal cues are used by the listener to infer the internal state of the sender based on schematic recognition or explicit inference rules (measured by naive observers asked to recognize the underlying emotion). In Brunswikian terminology, the first step in this process is termed the externalization of the internal emotional state, the second step the transmission of the behavioral information and the forming of a perceptual representation of the physical non-verbal signal, and the third and last step the inferential utilization and the emergence of an emotional attribution.

Despite its recent rebirth and growing popularity, the lens model paradigm has rarely been used to study the expression and perception of emotion in voice, face, and body (with one notable exception, Laukka et al., 2016). Scherer et al. (2011) reiterated earlier proposals to use the Brunswikian lens paradigm to study the emotion communication process, as it combines both the expression and perception/inference processes in a comprehensive dynamic model of emotion communication to overcome the shortcomings of focusing on only one of the component processes. The current study was designed to demonstrate the utility of the TEEP model in the domain of facial expression research. In addition to advocating the use of a comprehensive communication process approach for the research design, we propose to directly address the issue of the mechanisms involved in the process, by using the Component Process Model (CPM) of emotion (see Scherer, 1984, 2001, 2009) as a theoretical framework.

The central assumption made by the CPM is that emotion episodes are triggered by appraisal (which can occur at multiple levels of cognitive processing, from automatic template matching to complex analytic reasoning) of events, situations, and behaviors (by oneself and others) that are of central significance for an organism's well-being, given their potential consequences and the resulting need to urgently react to the situation. The CPM assumes a sequential-cumulative mechanism, suggesting a dynamic process according to which appraisal criteria are evaluated one after another (sequence of appraisal checks)

FIGURE 1 | The tripartite emotion expression and perception (TEEP) model (based on Brunswik's lens mode). The terms "Push" and "Pull" refer to the internal and the external determinants of the emotional expression, respectively, distinguished in the lower and upper parts of the figure. ANS, autonomous nervous system; SNS, somatic nervous system. Adapted from Scherer (2013a, Figure 5.5). Pull effects refer to an expression that is shaped according to an external model (e.g., a social convention), Push effects refer to internal physiological changes that determine the nature of the expressive cues.

in that each subsequent check builds on the outcome of the preceding check and further differentiates and elaborates on the meaning and significance of the event for the organism and the potential response options. The most important appraisal criteria are novelty, intrinsic un/pleasantness, goal conducive/obstructiveness, control/power/coping potential, urgency of action and social or moral acceptability. The cumulative outcome of this sequential appraisal process is expected to determine the specific nature of the resulting emotion episode. During this process, the result of each appraisal check will cause efferent effects on the preparation of action tendencies (including physiological and motor-expressive responses), which accounts for the dynamic nature of the unfolding emotion episode (see Scherer, 2001, 2009, 2013b). Thus, the central assumption of the CPM is that the results of each individual appraisal check sequentially drive the dynamics and configuration of the facial expression of emotion (see **Figure 2**). Consequently, the sequence and pattern of movements of the facial musculature allow direct diagnosis of the underlying appraisal process and the resulting nature of the emotion episode (see Scherer, 1992; Scherer and Ellgring, 2007; Scherer et al., 2013), for further details and for similar approaches (de Melo et al., 2014; van Doorn et al., 2015).

Specific predictions for facial expression were elaborated based on several classes of determinants: (a) the effects of typical physiological changes, (b) the preparation of specific instrumental motor actions such as searching for information

or approach/avoidance behaviors, and (c) the production of signals to communicate with conspecifics (see Scherer, 1984, 1992, 2001; Lee et al., 2013). As the muscles in the face and vocal tract serve many different functions in particular situations, such predictions can serve only as approximate guidelines. An illustrative example for facial movements predicted to be triggered in the sequential order of the outcomes of individual appraisal checks in fear situations is shown in **Table 1**. The complete set of CPM predictions (following several revisions, described in Kaiser and Wehrle (2001), Scherer and Ellgring (2007), Scherer et al. (2013), and Sergi et al. (2016) as well as the pertinent empirical evidence is provided in Scherer et al. (2018), in particular Table S1 and **Appendix**. **Figure 3** shows an adaptation of the TEEP model described above to the facial expression domain, illustrating selected predictions of the CPM and empirical results. It should be noted that this is an example of the presumed mechanism and that the one-to-one mapping shown in the figure cannot be expected to hold in all cases.

It is important to note that the appraisal dimensions of pleasantness/goal conduciveness and control/power/coping potential are likely to be major determinants of the valence and power/dominance dimensions proposed by dimensional emotion theorists (see Fontaine et al., 2013, Chapter 2). While there is no direct equivalent for the arousal dimension regularly found in studies of affective feelings, it can be reasonably argued that on this dimension, emotional feeling does not vary by quality but by response activation, probably as a

FIGURE 2 | Cumulative sequential appraisal patterning as part of the Component Process Model (Scherer, 2001, 2009, 2013b). Cumulative effects were generated by additive morphing of the action unit specific photos. Adapted from Figure 19.1 in Scherer et al. (2017) (reproduced with permission from Oxford University Press).
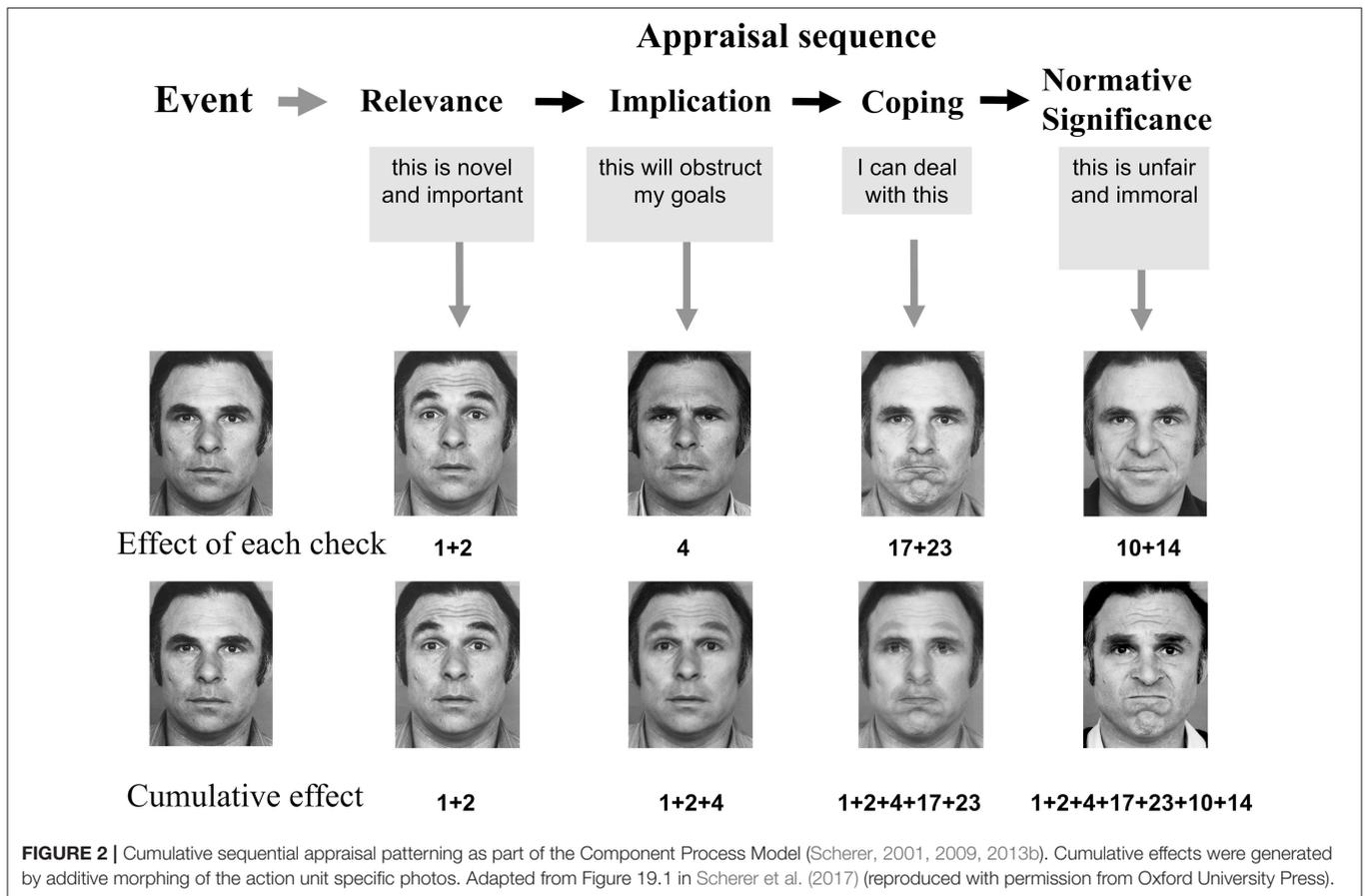
function of specific appraisal configurations, in particular the appraisals of personal relevance and urgency. A large-scale investigation of the semantic profiles of emotion words in more than 25 languages all over the world (Fontaine et al., 2013) provides strong empirical evidence for this assumption and suggests the need to add novelty/predictability as a fourth dimension (directly linked to the respective appraisals) to allow adequate differentiation of the multitude of emotion descriptions. Following this lead, we investigated the role of facial behavior in emotional communication, using both categorical and dimensional approaches (Mehu and Scherer, 2015). We used a corpus of enacted emotional expressions (GEMEP; Bänziger and Scherer, 2010; Bänziger et al., 2012) in which professional actors are instructed, with the help of scenarios, to communicate a variety of emotional experiences. The results of Study 1 in Mehu and Scherer (2015) replicated earlier findings showing that only a minority of facial action units is associated with specific emotional categories. Study 2 showed that facial behavior plays a significant role both in the detection of emotions and in the judgment of their dimensional aspects, such as valence, dominance, and unpredictability. In addition, a mediation model revealed that the association between facial behavior and recognition of the signaler's emotional intentions is mediated by perceived emotion dimensions. We concluded that, from a production perspective, facial action units convey neither specific emotions nor specific emotion dimensions, but are associated with several emotions and several dimensions. From the perceiver's perspective, facial behavior facilitated both dimensional and categorical judgments, and the former mediated the effect of facial behavior on recognition accuracy. The classification of emotional expressions into discrete categories may, therefore, rely on the perception of more general dimensions such as valence, power and arousal and, presumably, the underlying appraisals that are inferred from facial movements.

The current article extends the research approach described above in the direction of emotion enactment by professional actors, using a larger number of actors from another culture and a greater number of emotions. In Study 1, we asked professional actors to facially enact a number of major emotions and conducted a detailed, dynamic analysis of the frequency of facial actions. In Study 2 we examined to what extent emotion inferences of observers can be predicted by specific AU configurations. Finally, we estimated the appraisal criteria likely to determine the enactment of different emotions (using established semantic structure profiles of major emotion terms) and examined the relationships to the AUs coded for the actor portrayals.

Cumulative sequence of appraisal	Appraisal checks	CPM predictions for AUs generated by specific appraisal results	Appraisal results predicted for fear	AUs predicted to be produced by individual appraisal result
1	**Novelty**			
	Sudden/unpredictable	1, 2, 4, 5, 7, 26, 38	Very high	1, 2, 4, 5, 7, 26, 38
	Familiar/predictable	–	Not applicable	
2	**Intrinsic pleasantness**			
	Pleasant	5, 26, 38 or 12, 25	Open	
	Unpleasant	4, 7, 9, 10, 15, 17, 24, 39 or 16, 19, 25, 26	Open	
3	**Goal/need significance**			
	Conduciveness	12, 25	Not applicable	
	Obstructiveness	4, 7, 23, 17	Very high	4, 7, 23, 17
4	**Coping potential**			
	High power/control	4, 5 (or 7), 23, 25 (or 23, 24)	Not applicable	
	Low power/control	15, 25, 26, 41, 43 (or 1, 2, 5, 26, 20)	Very high	1, 2, 5, 15, 20, 25, 26, 41
	CPM predictions of AUs that could potentially occur for the emotion of fear as based on the accumulation of the effects of the pertinent appraisals			1, 2, 4, 5, 7, 15, 17, 20, 23, 25, 26, 38, 41, 43

Column 1 and 2: major appraisal checks postulated by the CPM (except self/norm compatibility) and the respective alternative outcomes; column 3: the Action units (AUs) predicted as potential expressions for the respective alternative results; column 4: the degree of pertinence of the specific appraisal outcome (high or very high) for the elicitation of fear ("Open—both outcome alternatives of a check can occur); column 5: the resulting AUs (from Column 3), expected to occur in the sequence shown in column 1. AU descriptions: 1, Inner brow raiser; 2, Outer brow raiser; 4, Brow lowerer; 5, Upper lid raiser; 7, Lid tightener; 15, Lip corner depressor; 17, Chin raiser; 20, Lip stretcher; 23, Lip tightener; 25, Lips part; 26, Jaw drop; 38, Nostril Dilator; 41, Lids droop; 43, Eye closure.
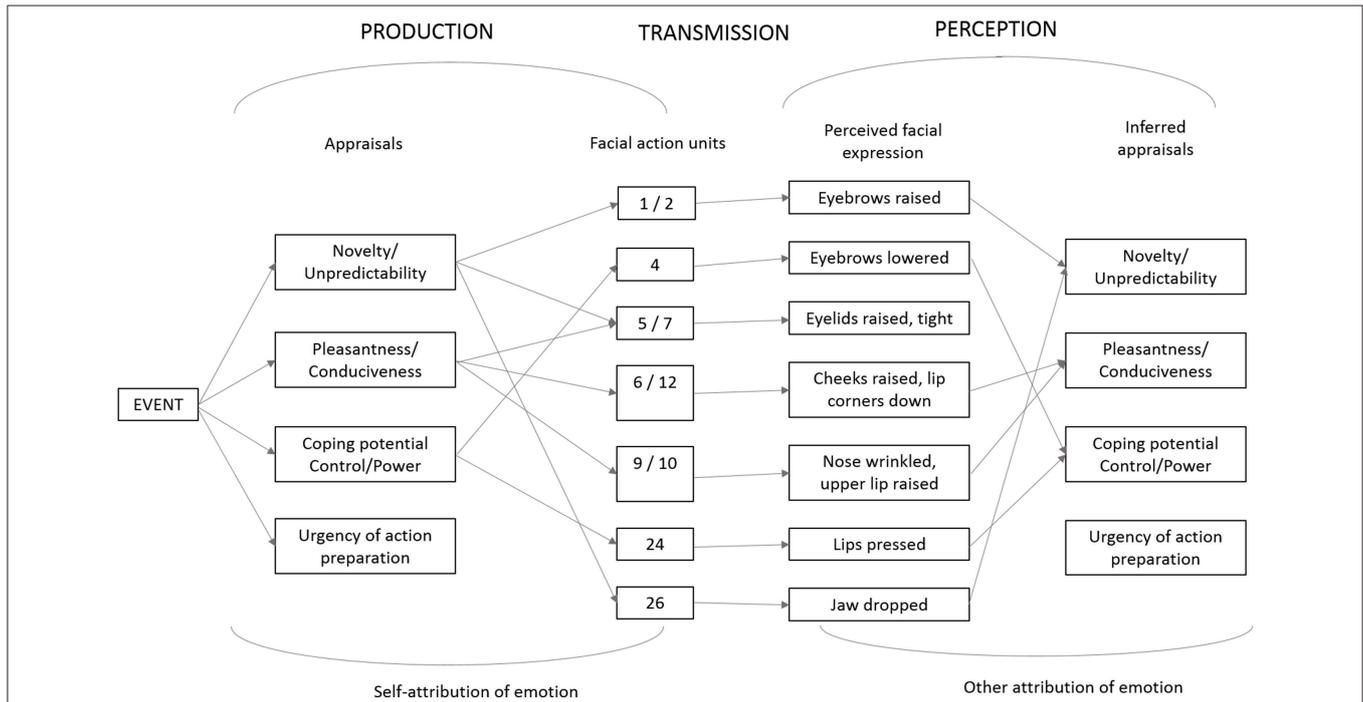
FIGURE 3 | Adaptation of the TEEP model to the domain of facial expression and inference illustrating selected predictions of the Component Process Model (CPM) and empirical results. Adapted from Scherer (2013a, Figure 5.5) and Scherer et al. (2018, Figure 2).

STUDY 1—THE ROLE OF DIFFERENT AUS IN ENACTED FACIAL EMOTION EXPRESSIONS

Aims

In the context of emotion enactment—using a Stanislavski-like method to induce an appropriate emotional state (see Scherer and Bänziger, 2010)—we wanted to investigate to what extent actors will use the AUs predicted to signal the appraisals that are constituent of the emotion being enacted.

Methods

Participants

Professional actors, 20 in total (10 males, 10 females, with an average age of 42 years, ranging from 26 to 68 years), were invited to individual recording sessions in a test studio. We recruited these actors from the Munich Artist's Employment Agency, and each received an honorarium in accordance with professional standards. The Ethics committee of the Faculty of Psychology of the University of Geneva approved the study.

Design and Stimulus Preparation

The following 13 emotions were selected to be enacted: Surprise, Fear, Anger, Disgust, Contempt, Sadness, Boredom, Relief, Interest, Enjoyment, Happiness, Pride, and Amusement. Each emotion word was illustrated by a typical eliciting situation, chosen from examples in the literature, appropriate for the daily experiences of the actors. Here is an example for pride: "A hard-to-please critic praises my outstanding performance and my interpretation of a difficult part in his review of the play for a renowned newspaper." Actors were instructed to imagine as vividly as possible that such an event happened to them and to attempt to actually feel the respective emotion and produce a realistic facial expression. To increase the ecological validity of the enactment, we asked the actors to simulate short, involuntary emotion outbreaks or affect bursts as occurring in real life (see Scherer, 1994), accompanied by a non-verbal vocalization—in this case /aah/.

Procedure

In the course of individual recording sessions, the actors were asked to perform the enacting of emotional expressions while being seated in front of a video camera. Six high power MultiLED softbox lights were set up to evenly distribute light over the actors' faces for best visibility of detailed facial activity[1].

Each recording session involved two experimenters. A certified coder and experienced expert in FACS (cf. Ekman et al., 2002) served as "face experimenter." He gave instructions to the actors and directed the "technical experimenter" who operated the camera.

The performing actor and face experimenter together read the scenario (the face experimenter aloud), before the actor gave an "ok" to signal readiness to facially express his or her emotional enactment.

Coding

To annotate the recordings with respect to the AUs shown by the actors, we recruited fifteen certified Facial Action Coding System (FACS, Ekman and Friesen, 1978) coders. To evaluate their performance, they were first given a subset of the recordings. For that purpose, the coders were divided into five groups of three coders each. All three coders in one group received eight recordings of one actor. Performance evaluation was based on coding speed and inter-coder agreement. Following the procedure proposed in the FACS manual, we first computed inter-coder agreement for each video for each coder with the other two coders who received the same set of videos. We then averaged these two values to get a single value for each coder. The agreement was calculated in terms of presence/absence of the Action Units within the coding for each target video. We did not compute agreement in terms of dynamics of the AUs (which is very hard to achieve; Sayette et al., 2001) nor in terms of intensities. Importantly, neither the dynamics nor the intensities were used in any of our analyses.

We excluded three coders because their average inter-coder reliabilities with the two other coders of their group were below 0.60. One more coder dropped out for private reasons. The reliabilities of the remaining 11 coders ranged from 0.65 to 0.87 (average = 0.75). The emotion enactment recordings were distributed among these 11 FACS coders. Each video was annotated by one coder.

Coders received a base payment of €15.00 per coding-hour, plus a bonus contingent on coding experience and their inter-coder reliability. On average, this amounted to an hourly payment of €18.00.

Coding instructions followed the FACS manual (Ekman et al., 2002; see also Cohn et al., 2007). Facial activity was coded in detail with regard to each occurrence of an AU, identifying onset, apex and offset with respect to both duration and intensity. For our current data analysis, we used occurrences and durations (between onset and offset) of single AUs. Different AUs appearing in sequence within an action unit combination were analyzed in accordance to predictions from the dynamic appraisal model. In addition to occurrence and intensity, potential asymmetry of each AU as well as a number of action descriptions (ADs, e.g., head and eye movements) were scored. To increase reliability three levels of intensity (1, 2, 3) were used instead of five, as suggested by Sayette et al. (2001), and applied successfully in several previous studies (e.g., Mortillaro et al., 2011; Mehu et al., 2012).

RESULTS

The aim of the analyses was to determine the extent to which specific AUs are used to portray specific emotions and if these correspond to the AUs that are predicted to occur (see Scherer and Ellgring, 2007; Scherer et al., 2018, and **Table 4** below) for the appraisals that are predicted as constituents for the respective emotions. While coders had scored all of the FACS categories

---

[1] The emotion enactment was the third and final part of a series of tasks which also included first, producing facial displays of specific Action Units (AUs, according to FACS), to be used as material for automatic detection, and, second, enacting a set of scenarios with different sequences of three appraisal results (to examine sequence effects). Results of these other tasks are reported elsewhere.

(a total of 57 codes), we restricted the detailed analyses (i.e., those listed in the tables) to action units (AUs) from AU1 to AU28 (see the **Appendix** for detailed illustrated descriptions) as there are only very few predictions for action descriptors (ADs). The ADs (e.g., head raising or lowering) differ from AUs in that the authors of FACS have not specified the muscular basis for the action and have not distinguished specific behaviors as precisely as they have for the AUs. In a few cases, where there are interesting findings, the statistical coefficients for ADs are included in the text. In addition, we did not analyze AUs 25 and 26 (two degrees of mouth opening) as all actors were instructed to produce an /aah/vocalization during the emotion enactment, resulting in a ubiquitous occurrence of these two AUs directly involved in vocalization.

The dynamic frame-by-frame coding allows obtaining an indication of the approximate length of the display of particular action units during a brief affect burst. **Table 2** provides a descriptive overview of the frequency of occurrence and the mean duration of different AUs for different emotions (including AUs 25 and 26, for the sake of comparison). Specifically, **Table 2A** contains the percentage of actors who use a specific AU to express different emotions, showing that actors vary with respect to the AUs they employ to express the different emotions. Only AUs 1, 2, 4, 6, 7, and 12 are regularly used by a larger percentage of actors. **Table 2B** lists the overall percentage of frames of the 78,398 frames coded in total in which the different AUs occur (column 1) and of the relative amount of time (in seconds) during which the different AUs were shown for particular emotions (the average duration across actors; columns 2–15). The table shows that average durations of AUs can vary widely, and that they are often produced for several types of emotion. AUs 1 and 2 are shown for both positive and negative emotions (possibly for greater emphasis). They are relatively brief, occurring rarely for more than 2 s. AU4 is shown for a somewhat longer period of time, mostly for negative emotions. AUs 6, 7, and 12 are primarily associated with the positive emotions, with very long durations for amusement (between 6 and 8 s) and, somewhat shorter for happiness and pride (around 3–4 s). They make briefer appearances in enactments of enjoyment and relief.

The dynamic frame-by-frame coding of the enactment videos allows to determine the temporal frames of AU combinations, i.e., frames in which two or more AUs are coded as being simultaneously present. As it would be impossible to study all possible combinations, we identified the most likely pairings in terms of claims in the literature. Thus, we computed new variables for the combinations AUs 6+12, AUs 1+2, AUs 1+4, and AUs 4+7. We also added AUs 6+7 given the discussion of the 2002 version of the FACS manual (see Cohn et al., 2007, p. 217). **Table 2C** shows the average duration per emotion for these combinations. In most cases, the simultaneous occurrence of the paired AUs is rather short—rarely exceeding 2 s.

AUs 1+2, reflecting the orientation functions of these movements, are found in surprise, as well as, even for longer duration, in interest, happiness, and fear—all of which often have an element of novelty/unexpectedness associated with them. This element can, of course, be part of many emotions, including anger, but it probably plays a less constitutive role as in interest

or fear. AUs 1+4 has the longest duration in sadness but is also found in disgust and fear. The same pattern is found for AUs 4+7, with a longer duration for disgust. AUs 6+12, but also the combination 6+7, are found for the positive emotions, in longer durations for amusement and happiness. However, 6+7 also occurs for disgust. Thus, while in some cases findings for AU combinations mirror the results for the respective individual AUs (e.g., for 6+12), in other cases (e.g., for AUs 1+4), in other cases combinations may mark rather different emotions (e.g., disgust or relief).

For the detailed statistical testing of the patterns found, we decided not to include AUs that occurred only extremely rarely, given the lack of reliability for the statistical analyses of such rare events (extremely skewed distributions). Concretely, we excluded all AUs from further analyses that occurred in <2% of the total number of frames coded (percentages ranging from 2 to 20.9%, see column 1 in **Table 2B**).

We calculated the number of frames during which each AU was shown in each of the 260 recordings (20 actors by 13 emotions). For each AU we computed a multivariate ANOVA with Emotion as independent variable (we did not include Actor as a factor because here we are interested in the group level rather than actor differences or actor-emotion interactions). The results allow determining whether an AU was present in a significantly greater number of frames for one emotion than the others. In all cases in which the Test of Between-Subject effects showed a significant ($p < 0.05$) effect for the Emotion factor, we computed *post-hoc* comparisons to identify homogeneous subgroups (no significant differences between members of a subgroup), and used the identification of non-overlapping subgroups (based on Waller-Duncan and Tukey-b criteria) to determine the emotions that had a high or a very high number of frames in which the respective AU occurred. **Table 3** shows, for both individual AUs and AU combinations, a summary of the results for which homogeneous subgroups were identified for either or both of the *post-hoc* test criteria.

To determine whether the pattern of AU differences found in this manner corresponds to expectations, we prepared **Table 4** which shows the current results in comparison with the CPM predictions, Ekman and Friesen's (1982) EMFACS predictions, and the pattern of empirical findings reported in the literature (for details and references for the latter, see Table S1 in the Supplemental Material for Scherer et al., 2018). Only the emotions covered in all of the comparison materials are shown in **Table 4**. The table shows that virtually all of the individual AUs occurring with significant frequency correspond to AUs predicted by the CPM and/or EMFACS and/or have been found in earlier studies (the CPM predictions do not include head movements). It should be noted that the current results are based on highly restrictive criteria—significant main effects for overall emotion differences and significant differences with respect to non-overlapping homogeneous subgroups. Therefore, one would expect a smaller number of AUs in comparison to the predictions, which list a large set of potentially occurring AUs or the compilation of published results from rather different studies. Many of the AUs listed for certain emotions in the three rightmost columns of **Table 4** were also shown for the

TABLE 2A | Percentage of actors displaying a particular AU when enacting a given emotion.

	Amusement	Anger	Boredom	Contempt	Disgust	Enjoy	Fear	Happiness	Interest	Pride	Relief	Sadness	Surprise
AU1	**30**	25	**40**	25	**30**	20	**60**	**50**	**50**	25	**30**	**55**	**40**
AU2	**40**	30	**35**	25	5	15	**40**	**50**	**50**	**40**	20	10	**40**
AU4	25	**45**	25	**45**	**55**	5	**70**	5	10	5	5	**75**	**40**
AU5	5	25	10	5	0	0	**50**	**40**	15	15	5	5	**40**
AU6	**100**	10	0	5	**30**	**35**	15	**80**	25	**65**	**30**	25	25
AU7	**75**	15	10	**30**	**70**	**40**	**40**	**65**	**30**	**60**	**30**	**40**	20
AU9	10	15	0	20	**50**	0	15	10	0	0	0	5	5
AU10	5	15	10	**60**	**40**	0	10	5	0	10	5	15	5
AU11	5	0	0	0	5	0	10	0	5	0	0	0	5
AU12	**100**	20	5	10	0	**70**	30	**90**	**55**	**95**	**75**	20	**40**
AU13	0	0	5	10	0	0	0	5	5	5	0	0	10
AU15	10	5	15	10	15	0	5	20	0	20	5	**30**	10
AU16	5	10	10	5	15	0	5	5	10	10	5	15	10
AU17	5	10	5	5	**30**	5	0	5	5	15	5	15	15
AU18	10	0	0	0	10	10	0	10	5	10	5	0	5
AU20	5	15	5	10	**35**	0	20	20	5	5	15	10	5
AU22	0	0	0	0	5	0	5	0	0	0	0	0	5
AU23	10	0	0	10	0	5	10	5	10	10	5	5	0
AU24	5	20	0	10	0	10	10	**30**	25	15	**30**	15	20
AU25	**100**	95	95	95	95	**100**	**100**	90	95	85	**100**	90	**100**
AU26	95	90	90	75	80	95	90	80	80	85	**100**	70	95
AU27	10	**35**	5	10	5	5	**35**	**45**	5	5	0	0	15
AU28	5	5	0	0	0	0	5	15	0	15	5	5	0

Percentages >30 are bolded to provide better visibility of the major patterns.

same emotions in the current study—but they do not reach the strong criterion we set to determine the most frequently used AUs. Another reason for the relatively small number of AUs with significant emotion effects in the current study is that we requested actors, in the interest of achieving greater spontaneity, to produce the expressions in the form of very short affect bursts (together with an/aah/vocalization), which reduced the overall time span for the expression and required AUs 25 and 26 for mouth opening. In consequence, we can assume that the AUs listed in column 1 of **Table 4** constitute essential elements of the facial expression of the respective emotions.

DISCUSSION

The results are generally in line with both the theoretical predictions and earlier empirical findings in the literature. Here we briefly review the major patterns, linking some of these to the appraisals that are considered to provide the functional basis for their production. The classic facial indicators for positive valence appraisal, AU12 (zygomaticus action, lip corners pulled up) and AU6 (cheek raiser), are present for all of the positive emotions, but we also find AUs that differentiate between them. Thus, AU7 (lid tightener) by itself and the combination AUs 6+7 are found for the expression of both pride and happiness (indicating important visual input) but not for enjoyment which is further characterized by AU43 (closing the eyes, $F = 5.97$, $p <$

0.001, $eta^2 = 0.226$), a frequently observable pose for enjoyment of auditory or sensory pleasure (Mortillaro et al., 2011). For amusement we find a pattern of exaggerated length for both AUs 6 +12 and 6+7, together with AD59 (moving the head up and down, $F = 5.19$, $p < 0.001$, $eta^2 = 0.202$), which probably is the byproduct of laughter. The major indicator for negative valence appraisal, AU4 (brow lowerer) is centrally involved in most negative emotions, but there are also many differentiating elements. Thus, AU10 (upper lip raiser) is found, as predicted as a result of unpleasantness appraisal, for disgust, often accompanied by AU9 (nose wrinkle) and sometimes by AU17 (raised chin) and AU20 (lip stretcher). A major indicator for unexpectedness appraisal, AU5 (upper eye lid raiser), is strongly involved in fear and anger, probably due to the scrutiny of threatening stimuli (Scherer et al., 2018). The pattern for sadness is the combination of AU1 (inner brow raiser) and AU4 (brow lowerer), together with AU15 (lip corner depression) and AD64 (eyes closed, $F = 2.50$, $p = 0.004$, $eta^2 = 0.109$), suggesting low power appraisals. The facial production pattern for anger is very plausible—AUs 5, 27 (mouth stretcher) and AD57 (head forward, $F = 2.87$, $p = 0.001$, $eta^2 = 0.123$): staring with the head pushed forward and mouth wide open, reminiscent of a preparation for aggression. AU4, which is generally postulated as a cue for anger as shown in the table, does not reach significance here as it is present for only short periods of time. The data for the AU combinations basically confirm the patterns found for the respective individual components, the effect sizes being rather similar. However, in

TABLE 2B | Occurrence and mean duration (s) of AU presence across actors.

Occurrence in percent of frames		AU duration for the different emotions													Total duration
Action units	In percent	Amusement	Anger	Boredom	Contempt	Disgust	Enjoy	Fear	Happiness	Interest	Pride	Relief	Sadness	Surprise	
AU1	12.7	0.7	0.6	**1.7**	**1.1**	**1.5**	0.6	**1.8**	**1.7**	**2.3**	0.7	**1.4**	**2.8**	**1.4**	18.4
AU2	9.9	1.0	**1.0**	**1.5**	**1.3**	0.3	0.3	**1.2**	**2.0**	**2.0**	1.0	0.8	0.4	**1.6**	14.3
AU4	12.9	0.9	**1.6**	**1.6**	**2.3**	**3.0**	0.2	**2.9**	0.1	0.4	0.0	0.4	**3.7**	**1.5**	18.7
AU5	3.8	0.0	0.9	0.2	0.1	0.0	0.0	**1.3**	0.7	0.3	0.3	0.1	0.3	**1.3**	5.5
AU6	15.3	**7.1**	0.1	0.0	0.4	**1.7**	**1.8**	0.6	**3.7**	0.8	**3.1**	**1.4**	**1.1**	0.5	**22.2**
AU7	18.5	**6.4**	0.4	0.5	**1.3**	**3.0**	**2.2**	**1.3**	**3.6**	1.0	**3.1**	**1.3**	**2.3**	0.3	**26.8**
AU9	2.5	0.5	0.3	0.0	0.5	**1.6**	0.0	0.2	0.3	0.0	0.0	0.0	0.1	0.0	**3.6**
AU10	5.1	0.5	0.5	0.3	**1.7**	**1.9**	0.0	0.5	0.2	0.0	0.5	0.2	1.0	0.2	**7.4**
AU11	1.0	0.3	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.3	0.0	0.0	0.0	0.3	**1.4**
AU12	20.9	**8.2**	0.1	0.4	0.3	0.0	**3.3**	0.7	**5.8**	**2.4**	**4.8**	**3.2**	0.4	0.8	**30.2**
AU13	0.6	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.4	0.1	0.1	0.0	0.0	0.0	0.9
AU15	3.2	0.3	0.1	0.6	0.3	0.9	0.0	0.1	0.5	0.0	0.3	0.1	**1.4**	0.2	**4.7**
AU16	1.8	0.0	0.1	0.3	0.0	0.6	0.0	0.2	0.2	0.2	0.1	0.2	0.4	0.3	**2.6**
AU17	2.3	0.1	0.2	0.1	0.0	**1.2**	0.1	0.0	0.1	0.1	0.2	0.2	0.4	0.6	**3.3**
AU18	0.4	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.6
AU20	2.3	0.1	0.4	0.0	0.1	**1.3**	0.0	0.4	0.4	0.0	0.0	0.2	0.3	0.0	**3.3**
AU22	0.3	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.4
AU23	0.4	0.1	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.6
AU24	1.6	0.0	0.3	0.0	0.1	0.0	0.0	0.2	0.4	0.4	0.4	0.2	0.1	0.1	**2.3**
AU25	38.9	**7.8**	**4.3**	**3.9**	**3.0**	**4.1**	**3.8**	**4.3**	**5.1**	**4.8**	**3.1**	**4.1**	**3.8**	**4.2**	56.4
AU26	29.3	**6.2**	**3.2**	**3.2**	**1.9**	**3.0**	**3.1**	**2.8**	**3.8**	**3.6**	**2.4**	**3.5**	**2.1**	**3.7**	42.5
AU27	2.9	0.6	1.0	0.1	0.2	0.0	0.1	0.8	0.8	0.0	0.1	0.0	0.0	0.5	**4.2**
AU28	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.1	0.4	0.0	0.1	0.0	0.1	0.0	0.9

Column 1—Overall percentage of video frames (of the 78,398 frames coded in total) in which the different AUs occurred; Columns 2–15—relative amount of time (in seconds) during which the different AUs were shown for particular emotions (average duration across actors). Durations exceeding 1 s are bolded to provide better visibility of the major patterns.

TABLE 2C | Mean duration (s) of the simultaneous presence of major AU combinations across actors.

Action unit combinations	Amusement	Anger	Boredom	Contempt	Disgust	Enjoy	Fear	Happiness	Interest	Pride	Relief	Sadness	Surprise	Total
AUs 1+2	0.4	0.4	**1.3**	0.7	0.3	0.2	**1.2**	**1.7**	**2.0**	0.6	0.8	0.4	**1.3**	**11.4**
AUs 1+4	0.2	0.2	0.7	0.1	**1.4**	0.2	**1.5**	0.0	0.0	0.0	0.3	**2.3**	0.3	**7.2**
AUs 4+7	0.8	0.2	0.5	0.8	**2.1**	0.0	**1.2**	0.0	0.0	0.0	0.0	**1.2**	0.2	**7.2**
AUs 6+12	**6.9**	0.0	0.0	0.0	0.0	**1.6**	0.2	**3.5**	0.5	**2.6**	**1.3**	0.1	0.4	**17.1**
AUs 6+7	**5.9**	0.1	0.0	0.4	**1.6**	**1.4**	0.6	**3.1**	0.3	**2.3**	**1.1**	1.0	0.1	**17.7**

Durations exceeding 1 s are bolded to provide better visibility of the major patterns.

some cases specific combinations attain significance although the individual components do not reach the criterion—this is notably the case for AUs 1+2 for interest and AUs 1+4 for sadness.

STUDY 2—INFERENCES FROM THE AUS SHOWN IN THE EMOTION PORTRAYALS

Aims

To investigate the emotion inferences from the actor appraisals with respect to the AU configurations used by the actors, we asked judges to recognize the emotions portrayed. However, contrary to the standard emotion recognition paradigm we are not primarily interested in the accuracy of the judgments but rather in the extent to which the emotion judgments can be explained by the theoretical predictions about appraisal inferences made from specific AUs.

Methods
Participants

Thirty four healthy, French-speaking subjects participated in the study (19 women, 15 men; age $M = 24.2$, $SD = 8.7$). They were recruited via announcements posted in a university building. The number of participants is sufficient to guarantee the stability of the mean ratings, which are the central dependent variables. A formal power analysis was not performed as no effect sizes based on a particular N were predicted.

TABLE 3 | Study 1—Compilation of the significant results in the multivariate ANOVA and associated *post-hoc* tests for homogeneous subgroups on the use of specific AU's for the portrayal of the 13 emotions.

Individual action units	F	Sig.	Partial eta^2	High	Very high
AU4	5.581	<0.001	0.214	con, fea, dis	sad
AU5	3.698	<0.001	0.153	ang	fear, sur
AU6	15.429	<0.001	0.429	pri, hap	amu
AU7	8.086	<0.001	0.283	dis, pri, hap	amu
AU9	4.951	<0.001	0.195		dis
AU10	2.616	0.003	0.113		con, dis
AU12	30.306	<0.001	0.597	int, rel, enj	pri, hap, amu,
AU15	1.879	0.037	0.084		sad
AU17	2.191	0.013	0.097		dis
AU20	2.844	0.001	0.122		dis
AU27	2.448	0.005	0.107		ang
Action unit combinations	**F**	**Sig.**	**Partial eta^2**	**High**	**Very high**
AUs 1+2	1.944	0.030	0.087		int
AUs 1+4	4.155	<0.001	0.169	dis, fea	sad
AUs 4+7	3.202	<0.001	0.135		dis
AUs 6+12	25.309	<0.001	0.552	enj, pri, hap	amu
AUs 6+7	9.684	<0.001	0.321	pri, hap	amu

sur, Surprise; fea, Fear; ang, Anger; dis, Disgust; con, Contempt; sad, Sadness; bor, Boredom; rel, Relief; int, Interest; enj, Enjoyment; hap, Happiness; pri, Pride; amu, Amusement.

Stimulus Selection and Preparation

To keep the judgment task manageable we decided to restrict the number of stimuli to be judged by using recordings for only nine of the 13 emotions portrayed, the seven listed in **Table 4** (anger, fear, sadness, disgust, pride, happiness, and enjoyment) plus two (contempt and surprise). These emotions were selected based on the frequent assumption in the literature that each of them is characterized by a prototypical expression. Again, in the interest of reducing the load for the judges, we further decided to limit the number of actors to be represented. We used two criteria for the exclusion: (1) very low degree of expressivity and (2) massive presence of potential artifacts. To examine the expressivity of each actor, we summed up the durations (in terms of number of frames) of all AUs shown by her or him and computed a univariate ANOVA with actor as a factor, followed by a *post-hoc* analysis to determine which actors had significantly shorter durations for the set of AUs coded. This measure indexes both the number of different AUs shown as well as their duration. Using this information as a guide to the degree of expressivity, together with the frequency of facial mannerisms (e.g., tics, as determined by two independent expert judges), artifacts likely to affect the ratings, we excluded actors no. 1, 2, 3, 13, 18, and 21. The remaining set of 126 stimuli (9 emotions × 14 actors) plus four example videos from the same set, not included in the analysis, were used as stimuli in the judgment task.

The 126 video clips were then trimmed (by removing some seconds unrelated to emotion enactment at the beginning and end of the videos) to have roughly the same duration

(between 4 and 6 s): with 1–1.5 s of neutral display, 2–3 s of emotional expression, and again 1–1.5 s of neutral display. All clips had a 1,624 × 1,080 resolution, with a 24 frames-per-second display rate. For the final version of the task, the 126 clips were arranged in one random sequence (the same for every subject), each followed by a screen (exposure duration 7 s) inviting subjects to answer. Video clips were presented without sound in order to avoid emotion judgements influenced by the "aah"-vocalizations during expressions.

Procedure

Three group sessions were organized on 3 different days in the same room (a computer lab) and at the same time of the day. Upon arrival, the participants were informed about the task, were reminded that, as promised on the posted announcement, the two persons with the highest scores would earn a prize, were told that they could withdraw and interrupt the study any time they wanted without penalty, and were asked to sign a written consent form. Each participant was seated in front of an individual computer and asked to read the instructions and sign the consent form. The rating instrument consisted of a digital response sheet based on Excel displayed on each participant's screen with rows corresponding to the stimulus and columns to the nine emotions. For each clip, the cell with the emotion label that in the participant's judgment best represented the facial expression seen for the respective stimulus was to be clicked. The stimuli were projected with the same resolution as their native format on a dedicated white projection surface, with an image size of 1.5 × 1.0 m. All subjects were located between 2 and 6 m from the screen, with orientation to it not exceeding 40°. Four example stimuli were used before starting to make sure everybody understood the task properly. The task then started. Halfway, a 5-min break has been made. Upon completion, after 45–50 min, participants were paid (CHF 15) and left. The two participants with the highest accuracy rate (agreement with the actor-intended emotion expression) received prizes of an additional CHF 15 each after the data analysis. The Ethics committee of the Faculty of Psychology of the University of Geneva approved the study.

RESULTS

The major aim of the analysis was to determine the pattern of inferences from the facial AUs shown by the actors in the emotion portrayal session.

We first used the classic approach of determining, with the help of a confusion matrix, how well the judges recognized the intended emotions and what types of confusions occurred. The confusion matrix is shown in **Table 5**. The raw cell entries were corrected for rater bias using the following procedure: We calculated the percentage of correct answers by dividing the number of correctly assigned labels for a given category by the overall frequency with which the respective emotion label had been used as a response by the judges. The mean percentage of accurate responses amounts to 43.7%, thus largely

TABLE 4 | Study 1—Comparison of current results on AU occurrence for the portrayal of major emotions in comparison to theoretical predictions and empirical findings reported in the literature.

Emotions	Current results	CPM predictions for specific emotions based on appraisals	Ekman and friesen EMFACS predictions for basic emotions	Empirical findings for major emotions
Anger	5, 27, 57	4, 5, 7, 17, 23, 24, 25	4+5+7+23	4, 21, 30, 53, 57 (1, 2, 16)
Fear	4, 5, (1+4)	1, 2, 4, 5, 7, 15, 17, 20, 23, 25, 26, 38, 41, 43	1+2+4+5+7+20+26	1, 4, 5, 25, 26, 53 (2, 16)
Sadness	1, 15, 64, (1+4)	1, 2, 4, 5, 7, 15, 17, 20, 23, 25, 26, 41, 43	1+4+15	45, 53 (1, 4, 15, 17)
Disgust	4, 7, 9, 10, 17, 20, (4+7), (9+10)	4, 7, 9, 10, 15, 17, 24, 39, 16, 19, 25, 26	9+15+16	10 (4, 6, 17)
Pride	6, 7, 12 (6+7), (6+12)	4, 5, 7, 12, 23, 24, 25	–	6, 7, 10, 12, 18, 25 (1, 2, 17)
Happiness	6, 7, 12 (6+7), (6+12)	1, 2, 12, 25, 26	6+12	6, 12, 25, 53 (1, 2, 26)
Enjoyment	12, 43, (6+12)	5, 12, 25, 26, 28	–	6, 7, 10, 12, 17, 18, 25, 26, 43

*The seven rows represent seven major emotions frequently studied in the literature. Colum 1 shows the current results for these seven emotions shown in detail in **Table 1**. Column 2 shows the predictions of the CPM based on postulated effects of major appraisal checks for the specific emotion on the AUs that can potentially occur (head movements, AUs 50–64, were not included). Column 3 shows the EMFACS predictions proposed by Ekman and Friesen (1982). The final column shows a summary of empirical findings obtained in a number of studies that used actors to portray the emotions. AUs in parentheses were only rarely found (see Appendix Scherer et al., 2018, for details on these studies and the methods used to summarize the results). AU descriptions: 1, Inner Brow Raiser; 2, Outer Brow Raiser; 4, Brow Lowerer; 5, Upper Lid Raiser; 6, Cheek Raiser; 7, Lid Tightener; 9, Nose Wrinkler; 10, Upper Lip Raiser; 11, Nasolabial Deepener; 12, Lip Corner Puller; 13, Cheek Puffer; 14, Dimpler; 15, Lip Corner Depressor; 16, Lower Lip Depressor; 17, Chin Raiser; 18, Lip Puckerer; 20, Lip stretcher; 22, Lip Funneler; 23, Lip Tightener; 24, Lip Pressor; 25, Lips part; 26, Jaw Drop; 27, Mouth Stretch; 28, Lip Suck; 41, Lid drop; 43, Eyes Closed; 45, Blink; 53, Head up; 57, Head forward; 64, Eyes down.*

TABLE 5 | Study 2—Confusion matrix for the judgments of the actor emotion portrayals (corrected for rater bias).

		Emotion Labels Assigned by Judges								
		Anger	Fear	Sadness	Disgust	Pride	Happiness	Enjoyment	Contempt	Surprise
	Anger	**50.1**	9.1	2.6	7.7	4.0	3.6	2.1	9.4	14.9
	Fear	8.4	**54.4**	3.8	6.0	1.4	1.6	2.4	2.7	14.9
	Sadness	3.1	6.4	**53.7**	10.4	3.1	3.6	1.5	9.8	4.9
Actor	Disgust	11.7	8.7	12.2	**39.2**	3.1	2.2	2.4	9.2	1.2
intention	Pride	2.3	1.9	3.4	3.2	**52.6**	23.9	12.9	8.2	5.1
	Happiness	1.6	2.5	3.4	2.8	12.0	**49.8**	9.1	3.3	14.0
	Enjoyment	2.1	3.2	4.1	2.7	15.7	11.5	**62.1**	9.6	5.4
	Contempt	15.1	4.0	8.1	16.1	3.1	1.6	3.2	**40.0**	4.7
	Surprise	5.5	9.8	8.7	11.8	4.9	2.0	4.4	8.0	**35.0**

Percentages for accurate judgments are bolded.

exceeding the chance hit rate of 11.1%. This is slightly lower than the average values for other studies on the recognition of the facial expression of emotions reported in the review by Scherer et al. (2011, Table 2). However, it should be noted that in this study a larger number of emotions (9) were to be judged compared to the usual five to six basic emotions generally used. Furthermore, actors had to respond to concrete scenarios rather than posing a predefined set of expressions resulting in variable and complex facial expressions. In addition, whereas in past research actors generally had to portray emotions with a longer utterance, here only a very brief affect bursts were to be produced. Given that the chance rate was largely exceeded and the frequent confusions (anger/contempt/disgust, fear/surprise, happiness/pride/enjoyment) are highly plausible, we can assume that the actor portrayals provide credible renderings of typical emotion expressions. This allows considering both the production results in Study 1 and the inference results reported

in the next section as being representative of day-to-day emotion expressions.

The central aim of this study was to examine the pattern of inferences judges draw from the occurrence of specific AU combinations. To identify these configurations, we ran a series of linear stepwise regressions of the complete set of AUs on each of the perceived emotion categories as dependent variables. The stepwise procedure (selecting variables to enter by smallest p-value of the remaining predictors at each step) determines which subset of the AUs have a significant effect on the frequency of choice of each emotion category and providing an index of the explanatory power with the help of R^2. As here we are interested in the cues that are utilized to make an inference, we computed the regressions on all occurrences of the specific category in the judgment data, independently of whether it was correct (i.e., corresponding to the intended emotion) or not. For reasons of statistical stability, we again restricted the AUs to be entered

into the regressions to those that occurred with a reasonable frequency (in this case mean occurrence >10%[2]) for the selected group of 14 actors and 9 emotions chosen for Study 2.

Table 6A summarizes the results for the individual AUs, providing for each inferred emotion category the predictors reaching significance ($p < 0.05$) in the final step of the regression, together with their beta weights (showing the direction and strength of the effects), as well as the adjusted R^2 for the final equation. In the table, the AUs that correspond to the comparable patterns of the AU production in the first column of **Table 4** (the summary of the MANOVA of emotion differences in the frequency of AUs shown in Study 1) are bolded (note that contempt and surprise had not been included in the comparison shown in **Table 4**). The results show remarkably high R^2 (>0.20) scores for five of the inferred emotions suggesting that specific AUs are indeed largely responsible for the inference of underlying emotions by observers. Although the R^2 values for anger, fear, disgust and contempt are lower, the results point in the same direction. Importantly, many of these configurations correspond to the theoretically predicted configurations (see columns 3–4 in **Table 4**). **Table 6B** shows the regression results for the selected AU combinations as predictors.

Specifically, the following AUs and AU combinations for major emotions have been theoretically predicted and empirically found to frequently occur in producing specific emotions: fear—AUs 4, 5, (1+4); sadness—1, 2, 4, 7, (1+4); disgust—4, 10, (4+7); pride—12, (6+12); happiness—6, 12, (6+12); enjoyment—12, 43, (6+12). No dominant pattern is found for anger, which is not surprising given that stable predicted patterns are very rarely found in empirical expression studies. On the other hand, anger is among the best recognized emotions as shown in **Table 5** (as well as in most recognition studies in the literature). One possible explanation for this apparent paradox is that, as there are many different types of anger (e.g., irritated, annoyed, offended, angry, enraged, and furious), there are many different ways to facially express (and recognize) this frequent emotion.

So far, we have only commented on the AUs with positive beta weights, that is, the presence of the respective AU is used as a marker for the inference of a specific emotion. As **Table 6** shows there are also many negative beta weights, indicating that the absence *of specific other AUs* rules out the inference of the respective emotion. Given space restrictions, we cannot explore the many interesting patterns contained in these data. Note that not only accurate judgments were used in the regression; rather we used all cases in which a specific emotion was inferred for the dependent variable in the regression. This strengthens our claim that the AUs that entered the regression equation are indeed utilized as cues for the emotion inference process.

The purpose of the preceding analysis was to determine which AUs are likely to have served as cues for the inference of certain emotions, independently on whether the respective emotion intended by the actor had been correctly inferred or not. One could argue that enactments that are more correctly identified

might be of particular importance to identify the AUs that are typical indicators of certain emotions. We computed the same regressions shown in **Table 6** separately for those enactments that were particularly well-recognized (using only videos that were with an accuracy percentage above the median—45%). The results of this separate analysis are shown in the two rightmost columns of **Tables 6A,B**, allowing direct comparison. Given the reduction of the N by half requires much stronger effects in order to be entered into the regression model in the stepwise procedure. For some of the emotions, none of the AU predictors made it into the equation. However, overall we find a very similar picture and—in some cases (fear and sadness)—even higher R^2s. We can assume that the AUs found to be predictors in both cases are indeed stable cues for the inference of certain emotions. As expected, the most stable predictors are AUs 1, 4, 6, and 12.

EMOTION COMMUNICATION: COMBINING EXPRESSION AND INFERENCE

We have argued in the introduction that emotion inference and recognition mirror the appraisal-driven expression process as postulated by the CPM, suggesting that judges first recognize appraisal results and then categorize specific emotions based on inference rules. To directly study the relationship between the facial expressions and the appraisals that are at the origin of the emotion experience that is expressed, ideally, one has to know the actual appraisals of the person. However, for ethical and methodological reasons it is not feasible to ask for appraisal self-report during an ongoing emotional experience without substantially altering the emotion and the appraisals themselves. An alternative approach is to use the typical appraisal profiles of the target emotions. In line with the approach used in previous publications about the relationship between appraisals and facial expressions (Mortillaro et al., 2011), here we use massive empirical evidence available on the meaning of emotion terms in many different languages to determine the typical appraisals of the target emotions.

Specifically, one large-scale study (Fontaine et al., 2013) on 24 emotion terms in 28 languages identified four dimensions that are necessary to map the semantic space of emotion words: valence, power, arousal, and novelty, in this order of importance. This cross-cultural study confirmed earlier results about affective dimensions in the literature but demonstrated that valence and arousal are not sufficient to map the major emotion terms. Furthermore, the results (based on all semantic meaning facets including appraisal) provided evidence for the strong link between affective dimensions and the major appraisal checks as postulated by the CPM—(1) valence, based on pleasantness/goal conduciveness appraisal; (2) power, based on control, power, and coping potential appraisals; (3) arousal, related to appraised personal relevance and urgency of an event; and (4) novelty, based on suddenness and predictability appraisals. In a follow-up study, Gillioz et al. (2016) confirmed this finding for 80 emotion terms in the French language. The results of this study, again a four-factorial solution with valence, power, arousal and novelty, provide us with stable appraisal coordinates for the

---

[2]Note that this threshold is higher than in Study 1, as suggested by the distribution of frequencies, due to selection of more expressive actors and prototypical emotion expressions.

TABLE 6 | Study 2—Linear Stepwise Regressions of (A) individual AUs and (B) AU combinations on emotion inference judgments.

	Regressions over all 126 videos	Adjusted R2	Regressions for videos with above median percentage of correct judgment	Adjusted R2
	Beta weights for AUs in the order of stepwise entry		Beta weights for AUs in the order of stepwise entry	
(A) INDIVIDUAL AUS AS PREDICTORS				
Anger	12 (−0.313), 1 (−0.250)	0.127	12 (−0.279)	0.060
Fear	**4 (0.248), 5 (0.226)**	0.098	**5 (0.425), 4 (0.280)**	0.256
Sadness	**1 (0.601),** 2 (−0.412), **4 (0.145),** 7 (0.201), 12 (−0.170)	0.437	**1 (0.688),** 2 (−0.346), 7 (0.280), 12 (0.260), 5 (−0.244)	0.569
Disgust	12 (−0.243), 1 (−0.266), **10 (0.195)**,4 (0.205)	0.199	12 (−0.405), 1 (−358)	0.207
Pride	**12 (0.547)**	0.324	**12 (0.316)**	0.083
Happiness	**12 (0.551),** 27 (0.177), **6 (0.157)**	0.464	**12 (0.549),** 27 (0.239)	0.428
Enjoyment*	**43 (0.561), 12 (0.252),** 1 (−0.184)	0.437	**43 (0.646), 12 (0.365),** 27 (−0.242)	0.566
Contempt	12 (−0.367), 5 (−0.188), 4 (−0.199)	0.114	No variables entered	
Surprise	5 (0.469)	0.219	4 (0.399), 12 (0.330)	0.136
(B) AU COMBINATIONS AS PREDICTORS				
Anger	6+12 (−0.253), 1+4 (−0.202)	0.070	No variables entered	
Fear	**1+4 (0.262)**	0.061	4+7 (0.332)	0.093
Sadness	**1+4 (0.626),** 1+2 (−0.207)	0.368	**1+4 (0.640)**	0.398
Disgust	**4+7 (0.377),** 6+12 (−0.219), 1+4 (−0.186)	0.176	No variables entered	
Pride	**6+12 (0.404)**	0.157	**6 +12 (0.331)**	0.079
Happiness	**6+12 (0.603),** 1+2 (0.168)	0.366	**6+12 (0.641)**	0.400
Enjoyment	**6+12 (0.176)**	0.023	No variables entered	
Contempt	6+12 (−0.240), 1+4 (−0.198)	0.063	No variables entered	
Surprise	1+2 (0.337)	0.106	6+12 (0.350)	0.106

Criteria = entry level for predictors ($p < 0.05$), R^2 change ≥ 0.10. *AU43 added as additional predictor; AUs that correspond to the comparable patterns of the AU production in the first column of **Table 4** are bolded.

target emotion terms used in Study 2—in the form of factor scores corresponding to these terms, reproduced in **Table 7**. These factor scores largely confirm the theoretical predictions of the CPM (see Table 1 in Scherer, 2001, Table 5.4): for example, surprise is characterized by average values for valence, power and arousal, but high values for novelty, and happiness is characterized by positive valence, high power and arousal and medium level of novelty. We used these dimensional coordinates in the place of the emotion words used in the enactments reported in Study 1, to test whether appraisal results could be predicted only based on the facial expressions displayed by our actors.

The group of judges in Study 2 attributed different emotion terms to the actor portrayal video clips (see **Table 5**). Based on these data, we computed a specific 4-dimensional profile for each clip by weighting the coordinates shown in **Table 7** with the respective proportion of judges that inferred a specific emotion (to give greater importance to displays that allow for stable, consensual inference). We used coordinates for French emotion terms, given that our judges were speakers of French. Thus, the coordinates of the emotion words chosen by a large number of judges would be more strongly represented in the clip-specific dimensional profile.

To address the question to what extent the coordinates of the nine emotion items can be predicted by AUs, we then used these specific dimensional profiles for each clip as dependent variables in two linear stepwise regression analyses. Specifically,

TABLE 7 | Estimated coordinates for selected French emotion words on affective dimensions [Factor scores, based on Gillioz et al. (2016)].

Emotion	Valence	Power	Arousal	Novelty
Anger	−1.24	1.93	0.52	−0.58
Fear	−0.38	−1.27	1.10	0.34
Sadness	−0.48	−1.31	−1.46	−0.16
Disgust	−0.79	0.02	−1.17	0.04
Pride	1.11	1.01	0.46	−1.49
Happiness	1.82	0.40	0.15	−0.20
Enjoyment	1.64	0.06	−0.15	−0.37
Contempt	−0.94	1.38	−0.91	−0.21
Surprise	0.24	−0.04	0.55	2.45

we regressed the AU selection used for the analyses in Study 1 to predict (a) the *expression intentions*, that is the raw coordinates for each of the four appraisal dimensions (the raw values shown in **Table 7** for each emotion) and (b) the judges' inferences (the coordinates weighted by the number of judges having inferred the respective emotions). **Table 8** shows the results, the left side of the table showing the regressions of the AUs on the raw coordinates reflecting the actors' enactment intention and the right side showing the regression of the AUs on the weighted coordinates for the inferred emotions.

On the expression intention side, Valence is best predicted with a very large adjusted R^2 of 0.559. As expected, the best

TABLE 8 | Regressions on estimated coordinates of affective dimensions for both expression (raw coordinates) and inference (weighted coordinates).

Dimensions	Expression predictors (raw coordinates)	R2	Inference predictors (weighted coordinates)	R2
Valence	AU12 (0.630), AU10 (−0.150), AU4 (−0.161)	0.559	AU12 (0.732), AU10 (−0.173)	0.600
Power	AU1 (−0.341), AU2 (0.273), AU4 (−0.214)	0.157	AU1 (−0.412), AU2 (0.197), AU4 (−0.290),	0.253
Arousal	AU1 (−0.187), AU5 (0.296), AU10 (−0.191), AU12 (0.232), AU27 (0.184)	0.242	AU1 (−0.274), AU2 (0.199), AU5 (0.310), AU12 (0.388), AU27 (0.216),	0.363
Novelty	AU12 (−0.320)	0.098	AU12 (−0.170)	0.093

Criteria = entry level for predictors (p < 0.05), R^2 change ≥ 0.10.

predictor for positive valence expression is AU12. AUs 4 and 10 predict negative valence (as one would expect from their predominance in disgust expressions). Power is not very well-predicted with an R^2 of only 0.157. Only AU2 seems to imply high power, and AUs 1 and 4 low power. Arousal also shows a relatively low $R^2 = 0.242$, with AUs 12, 5, and 27 implying high arousal, AUs 10 and 1 low arousal. The novelty dimension is the least well-predicted ($R^2 = 0.098$) with AU12 for low novelty.

On the inference side, valence is again best predicted with a very large adjusted R^2 of 0.600. As expected, the best predictor for positive valence inference is AU12. AU10 predicts negative valence inference. Power inference has a slightly higher prediction success on the inference side with an R^2 of 0.253. Again, only AU2 signals high power, and AUs 1 and 4 low power. For inference, arousal also shows a somewhat higher $R^2 = 0.363$, with AUs 12, 5, 27, and 2 leading to the inference of high arousal, AU1 to low arousal. As for expression intention, novelty is least well-predicted ($R^2 = 0.093$) with AU12 for low novelty.

The main outcome of this analysis is the very high degree of equivalence in the respective AU patterns on both the expression and inference sides, which explains the accuracy results shown in **Table 5**. The low prediction success for power suggests that the face may not be a primary channel to communicate control, power, or coping potential, contrary to the voice (see Goudbeek and Scherer, 2010). For novelty, the low proportion of variance explained is most likely due to the low variability in novelty for the emotions studied here with the exception of surprise, and some degree, pride (as shown in **Table 7**). The respective predictor, AU12 for pride corresponds very well with the production side.

DISCUSSION AND CONCLUSION

It should be noted that the TEEP model that served as the theoretical framework for our empirical studies, represents a structural account of the emotion communication architecture and processes. It does not specify the detailed mechanisms, on neither the expression nor perception/inference side. It remains for further theoretical and empirical work to address exactly what mechanisms are operative on the neuromotor and neurosensory levels. Thus, with respect to inference, the model does not predict whether this happens in the form of classical perception mechanisms involving templates or discrete cue combinations and (more or less conscious) inference rules, or whether the process works in an embodied fashion with the

observer covertly mimicking the observed movement to derive an understanding (see Hess and Fischer, 2016). In both cases, correct communication relies on the nature of the AUs produced in expression that are objectively measurable and that serve as the input for perceived and embodied mimicry. The research reported here addresses only the issue of the nature of the AUs involved.

Based on the theoretical assumptions about the nature of the appraisal combinations that produce specific emotions, the CPM also predicts expression patterns for specific emotions (see column 2 in **Table 4**). Study 1 was designed to test these predictions in an enactment study using professional actors with very brief, affect-burst like non-verbal vocal utterances (see Scherer, 1994). This differs from earlier portrayal studies where generally longer verbal utterances are used, which may affect the facial expression due to the articulation movements around the mouth as well as involuntary prosodic signals in the eye and forehead regions. As shown in **Table 4**, the AUs consistently shown by the actors for certain emotions are in line with the theoretical predictions of the model.

Study 2 used the video stimuli with the enactments of major emotions by actors in a recognition design to obtain independent judgments as to the perceived or inferred emotions expressed. This approach served two purposes: (1) Obtaining evidence as to the representativeness of the enactments of specific emotions. The results show that this is indeed the case, hit rates exceeding chance level by a factor of 4–5 times and confusions being in line with similar patterns found in other studies; (2) Allowing us to investigate which cues are consistently utilized as markers for the inference of certain emotions.

This demonstration also supports the hypothesis described in the introduction (see also Mortillaro et al., 2012; Scherer et al., 2017, 2018), namely that the emotion inference and recognition process mirrors the production process. Specifically, our results suggest that observers use the facial expression to identify the nature of the underlying appraisals or dimensions and use inference rules to categorize and label the perceived emotion (in line with the semantic profiles of the emotion words; see Fontaine et al., 2013). We estimated the coordinates of the emotion terms used for the enactments in Study 1 on the four major affective dimensions valence, power, arousal, novelty (directly linked to the appraisal criteria of pleasantness/goal conduciveness, control/power, urgency of action, and suddenness/predictability) and then regressed the observed AU frequencies on these

estimates. The results shown in **Table 8** are consistent with the expectations generated by the production/perception mirroring hypothesis.

The approach we have chosen to obtain information on which appraisal dimensions are most likely to be inferred from certain AU configurations is somewhat unorthodox, using weighted estimates of the dimension coordinates for the expressive stimuli generated in Study 1 of this study as dependent variables, rather than direct ratings of appraisal dimensions. However, the latter approach would have the disadvantage of strong demand characteristics encouraging judges to consciously construct relationships between the facial expression and particular dimensions. Another major disadvantage with such a design is that the ratings of the valence dimension strongly affect all other dimensions with a powerful halo effects (see the strong evidence for these halos in Sergi et al. (2016) and Scherer et al. (2018). The advantage of our indirect method of examining the issue is that judges were focused on the emotions expressed and did not consider the appraisal dimensions explicitly, thus avoiding the occurrence of valence halos.

Overall, the results of the two studies presented here strongly confirm the utility and promise of further research on the mechanisms underlying the dynamic process of emotion expression and emotion inference using a unified theoretical framework. We suggest that further research be extended by including additional cues that may be relevant in the process of inferring emotions from facial cues. Recently, Calvo and Nummenmaa (2016) published a comprehensive integrative review on the perceptual and affective mechanisms in facial expression recognition. They conclude that (1) behavioral, neurophysiological, and computational measures indicate that basic expressions are reliably recognized and discriminated from one another, (2) affective content along the dimensions of valence and arousal is extracted early from facial expressions (but play a minimal role for categorical recognition), and (3) morphological structure of facial configurations and the visual saliency of distinctive facial cues contribute significantly to expression recognition. It seems promising to examine the interaction of such cues with the classic facial action units typically used in this research.

DATA AVAILABILITY

The datasets generated for this study (coding and rating data) are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

KS and MM conceived the research and study 1. KS conceived study 2 and directed the research. All authors contributed to the data collection (led by AD and MU). HE and MM were responsible for facial expression production and coding. KS analyzed the data and wrote the first draft. All authors contributed in several rounds of revision.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Bänziger, T., Hosoya, G., and Scherer, K. R. (2015). Path models of vocal emotion communication. *PLoS ONE* 10:e0136675. doi: 10.1371/journal.pone.0136675

Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the Geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* 12, 1161–1179. doi: 10.1037/a0025827

Bänziger, T., and Scherer, K. R. (2010). "Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) corpus," in *Blueprint for Affective Computing: A Sourcebook*, eds K. R. Scherer, T. Bänziger, and E. B. Roesch (Oxford: Oxford University Press), 271–294.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: University of California Press.

Calvo, M. G., and Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: an integrative review. *Cogn. Emot.* 30, 1081–1106. doi: 10.1080/02699931.2015.1049124

Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). "Observer-based measurement of facial expression with the Facial Action Coding System," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. Allen (Oxford: Oxford University Press), 203–221.

de Melo, C. M., Carnevale, P. J., Read, S. J., and Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *J. Personality Soc. Psychol.* 106, 73–88. doi: 10.1037/a0034251

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., and Friesen, W. V. (1982). *EMFACS*. San Francisco, CA.

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *The Facial Action Coding System, 2nd Edn.* Salt Lake City, UT: Research Nexus eBook.

Ekman, P., Friesen, W. V., and O'Sullivan, M. (1988). Smiles when lying. *J. Personality Soc. Psychol.* 54, 414–420.

Fontaine, J. R. J., Scherer, K. R., and Soriano, C. (eds.). (2013). *Components of Emotional Meaning: A Sourcebook*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199592746.001.0001

Gillioz, C., Fontaine, J. R., Soriano, C., and Scherer, K. R. (2016). Mapping emotion terms into affective space. *Swiss J. Psychol.* 75, 141–148. doi: 10.1024/1421-0185/a000180

Goudbeek, M., and Scherer, K. R. (2010). Beyond arousal: valence and potency/control in the vocal expression of emotion. *J. Acoust. Soc. Am.* 128, 1322–1336. doi: 10.1121/1.3466853

Hess, U., and Fischer, A. (2016). *Emotional Mimicry in Social Context*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107587595

Hess, U., and Kleck, R. E. (1994). The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions. *Eur. J. Soc. Psychol.* 24, 367–381.

Hyniewska, S., Sato, W., Kaiser, S., and Pelachaud, C. (2018). Naturalistic emotion decoding from facial action sets. *Front. Psychol.* 9:2678. doi: 10.3389/fpsyg.2018.02678

Kaiser, S., and Wehrle, T. (2001). "Facial expressions as indicators of appraisal processes," in *Appraisal Processes in Emotions: Theory, Methods, Research*, eds K. R. Scherer, A. Schorr, and T. Johnstone (New York, NY: Oxford University Press), 285–300.

Krumhuber, E. G., Skora, L., Küster, D., and Fou, L. (2017). A review of dynamic datasets for facial expression research. *Emot. Rev.* 9, 280–292. doi: 10.1177/1754073916670022

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., et al. (2016). The expression and recognition of emotions in the voice across five nations: a lens model analysis based on acoustic features. *J. Personality Soc. Psychol.* 111, 686–705. doi: 10.1037/pspi0000066

Lee, D. H., Susskind, J. M., and Anderson, A. K. (2013). Social transmission of the sensory benefits of eye widening in fear expressions. *Psychol. Sci.* 24, 957–965. doi: 10.1177/0956797612464500

Mehu, M., Mortillaro, M., Bänziger, T., and Scherer, K. R. (2012). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Emotion* 12, 701–715. doi: 10.1037/a0026717

Mehu, M., and Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: an integrated approach. *Emotion* 15, 798–811. doi: 10.1037/a0039416

Mortillaro, M., Mehu, M., and Scherer, K. R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. *Soc. Psychol. Personality Sci.* 2, 262–271. doi: 10.1177/1948550610389080

Mortillaro, M., Meuleman, B., and Scherer, K. R. (2012). Advocating a componential appraisal model to guide emotion recognition. *Int. J. Synth. Emot.* 3, 18–32. doi: 10.4018/jse.2012010102

Namba, S., Kabir, R. S., Miyatani, M., and Nakao, T. (2018). Dynamic displays enhance the ability to discriminate genuine and posed facial expressions of emotion. *Front. Psychol.* 9:672. doi: 10.3389/fpsyg.2018.00672

Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., and Parrott, D. J. (2001). A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *J. Nonverbal Behav.* 25, 167–185. doi: 10.1023/A:1010671109788

Scherer, K. R. (1984). "On the nature and function of emotion: a component process approach," in *Approaches to Emotion*, eds K. R. Scherer and P. Ekman (Hillsdale, NJ: Erlbaum), 293–317.

Scherer, K. R. (1992). "What does facial expression express?," in *International Review of Studies on Emotion*, Vol. 2, ed K. Strongman (Chichester: Wiley), 139–165.

Scherer, K. R. (1994). "Affect bursts," in *Emotions: Essays on Emotion Theory*, eds S. van Goozen, N. E. van de Poll, and J. A. Sergeant (Hillsdale, NJ: Erlbaum), 161–196.

Scherer, K. R. (2001). "Appraisal considered as a process of multi-level sequential checking," in *Appraisal Processes in Emotion: Theory, Methods, Research*, eds K. R. Scherer, A. Schorr, and T. Johnstone (New York, NY: Oxford University Press), 92–120.

Scherer, K. R. (2009). The dynamic architecture of emotion: evidence for the component process model. *Cogn. Emot.* 23, 1307–1351. doi: 10.1080/02699930902928969

Scherer, K. R. (2013a). "Emotion in action, interaction, music, and speech," in *Language, Music, and the Brain: A Mysterious Relationship*, ed M. A. Arbib (Cambridge, MA: MIT Press), 107–139.

Scherer, K. R. (2013b). The nature and dynamics of relevance and valence appraisals: theoretical advances and recent evidence. *Emot. Rev.* 5, 150–162. doi: 10.1177/1754073912468166

Scherer, K. R., and Bänziger, T. (2010). "On the use of actor portrayals in research on emotional expression," in *Blueprint for Affective Computing: A Sourcebook*, eds K. R. Scherer, T. Bänziger, and E. B. Roesch (Oxford: Oxford University Press), 166–178.

Scherer, K. R., and Ceschi, G. (2000). Studying affective communication in the airport: the case of lost baggage claims. *Personality Soc. Psychol. Bull.* 26, 327–339. doi: 10.1177/0146167200265006

Scherer, K. R., Clark-Polner, E., and Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *Int. J. Psychol.* 46, 401–435. doi: 10.1080/00207594.2011.626049

Scherer, K. R., and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* 7, 113–130. doi: 10.1037/1528-3542.7.1.113

Scherer, K. R., Mortillaro, M., and Mehu, M. (2013). Understanding the mechanisms underlying the production of facial expression of emotion: a componential perspective. *Emot. Rev.* 5, 47–53. doi: 10.1177/1754073912451504

Scherer, K. R., Mortillaro, M., and Mehu, M. (2017). "Facial expression is driven by appraisal and generates appraisal inference," in *The Science of Facial Expression*, eds J.-M. Fernández-Dols and J. A. Russell (New York, NY: Oxford University Press), 353–373.

Scherer, K. R., Mortillaro, M., Rotondi, I., Sergi, I., and Trznadel, S. (2018). Appraisal-driven facial actions as building blocks for emotion recognition. *J. Personality Soc. Psychol.* 114, 358–379. doi: 10.1037/pspa0000107

Sergi, I., Fiorentini, C., Trznadel, S., and Scherer, K. R. (2016). Appraisal inference from synthetic facial expressions. *Int. J. Synth. Emot.* 7, 46–63. doi: 10.4018/IJSE.2016070103

van Doorn, E. A., van Kleef, G. A., and van der Pligt, J. (2015). Deriving meaning from others' emotions: attribution, appraisal, and the use of emotions as social information. *Front. Psychol.* 6:1077. doi: 10.3389/fpsyg.2015.01077

Zloteanu, M., Krumhuber, E. G., and Richardson, D. C. (2018). Detecting genuine and deliberate displays of surprise in static and dynamic faces. *Front. Psychol.* 9:1184. doi: 10.3389/fpsyg.2018.01184

APPENDIX

Labels and visual illustrations of the major AUs investigated (modified from Mortillaro et al., 2011).

AU	Action Name	Illustration
AU1	Inner brow raise	
AU2	Outer brow raise	
AU4	Brow lower	
AU5	Upper lid raiser	
AU6	Cheek raiser	
AU7	Lid tightener	
AU9	Nose wrinkler	
AU10	Upper lip raiser	
AU12	Lip corner puller	
AU15	Lower corner depressor	
AU16	Lower lip depressor	
AU17	Chin raiser	
AU18	Lip puckerer	
AU20	Lip stretcher	
AU22	Lip funneler	
AU23	Lip tightener	
AU24	Lip pressor	
AU25	Lips part	
AU26	Jaw drops	
AU28	Lips suck	
AU43	Eye closure	
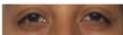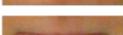
Empathy in Facial Mimicry of Fear and Disgust: Simultaneous EMG-fMRI Recordings During Observation of Static and Dynamic Facial Expressions

Krystyna Rymarczyk[1]*, Łukasz Żurawski[2]*, Kamila Jankowiak-Siuda[1] and Iwona Szatkowska[2]

[1] Department of Experimental Psychology, Institute of Cognitive and Behavioural Neuroscience, SWPS University of Social Sciences and Humanities, Warsaw, Poland, [2] Laboratory of Psychophysiology, Department of Neurophysiology, Nencki Institute of Experimental Biology, Polish Academy of Sciences (PAS), Warsaw, Poland

Real-life faces are dynamic by nature, particularly when expressing emotion. Increasing evidence suggests that the perception of dynamic displays enhances facial mimicry and induces activation in widespread brain structures considered to be part of the mirror neuron system, a neuronal network linked to empathy. The present study is the first to investigate the relations among facial muscle responses, brain activity, and empathy traits while participants observed static and dynamic (videos) facial expressions of fear and disgust. During display presentation, blood-oxygen level-dependent (BOLD) signal as well as muscle reactions of the corrugator supercilii and levator labii were recorded simultaneously from 46 healthy individuals (21 females). It was shown that both fear and disgust faces caused activity in the corrugator supercilii muscle, while perception of disgust produced facial activity additionally in the levator labii muscle, supporting a specific pattern of facial mimicry for these emotions. Moreover, individuals with higher, compared to individuals with lower, empathy traits showed greater activity in the corrugator supercilii and levator labii muscles; however, these responses were not differentiable between static and dynamic mode. Conversely, neuroimaging data revealed motion and emotional-related brain structures in response to dynamic rather than static stimuli among high empathy individuals. In line with this, there was a correlation between electromyography (EMG) responses and brain activity suggesting that the Mirror Neuron System, the anterior insula and the amygdala might constitute the neural correlates of automatic facial mimicry for fear and disgust. These results revealed that the dynamic property of (emotional) stimuli facilitates the emotional-related processing of facial expressions, especially among whose with high trait empathy.

Keywords: facial mimicry, EMG, fMRI, mirror neuron system, emotional expressions, dynamic, disgust, fear

INTRODUCTION

Empathy and Facial Mimicry

In the last decade, researchers have focused on empathy as an essential component of human social interaction. The term 'empathy' – derived from Greek *empatheia* – 'passion' – is a multifaceted construct that is thought to involve both cognitive (i.e., understanding of another's beliefs and feelings) and affective (i.e., ability to share another's feelings) components (Jankowiak-Siuda et al., 2011; Betti and Aglioti, 2016). It is believed that people empathize with others by simulating their mental states or feelings. According to the Perception-Action Model (PAM) of empathy, simulative processes discovered and defined in the domain of actions "result from the fact that the subject's representations of the emotional state are automatically activated when the subject pays attention to the emotional state of the object" (Preston and de Waal, 2002, p. 1; de Waal and Preston, 2017). Paying attention to the other's emotional state, in turn, leads to the related autonomic and somatic responses (Preston, 2007). Consistent with this model, a positive association between emotional empathy and somatic response was observed for both skin conductance (Levenson and Ruef, 1992; Blair, 1999; Hooker et al., 2008) and cardiac activation (Krebs, 1975; Hastings et al., 2000). This might indicate that more empathic persons react with stronger affective sharing. Recent studies suggest also that empathic traits relate to variation in facial mimicry (FM) (Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2003; Dimberg et al., 2011; Balconi and Canavesio, 2013a, 2014; Rymarczyk et al., 2016b).

Facial mimicry is spontaneous unconscious mirroring of others' emotional facial expressions, which leads to congruent facial muscles activity (Dimberg, 1982). This phenomenon usually is measured by electromyography (EMG; e.g., Dimberg, 1982; Larsen et al., 2003). Evidence for FM has been most consistently reported when viewing happy (Dimberg and Petterson, 2000; Weyers et al., 2006; Rymarczyk et al., 2011) and angry (Dimberg et al., 2002; Sato et al., 2008) facial expressions. Interestingly, angry facial expressions induce greater activity than happy faces in the corrugator supercilii (CS, muscle involved in frowning), whereas happy facial expressions induce greater activity in the zygomaticus major (ZM, the muscle involved in smiling) and decreased CS activity. In addition, few EMG studies support also the phenomenon of FM for other emotions, i.e., fear with increased activity of CS (e.g., van der Schalk et al., 2011a) or frontalis muscle (e.g., Rymarczyk et al., 2016b) and for disgust with increased activity of CS (e.g., Lundquist and Dimberg, 1995) or levator labii (LL) (e.g., Vrana, 1993). Furthermore, the magnitude of FM has been shown to depend on many factors (for a review see Seibt et al., 2015), including empathic traits (Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2003; Dimberg et al., 2011; Balconi and Canavesio, 2013a; Balconi et al., 2014; Rymarczyk et al., 2016b). For example, Dimberg et al. (2011) have found that more empathic individuals showed greater CS contraction to angry faces and greater ZM contraction to happy faces, as compared to less empathic individuals. Similar patterns were observed in response to fearful facial expressions, where in more empathic individuals exhibited larger CS reactions

(Balconi and Canavesio, 2016). Recently, Rymarczyk et al. (2016b) found that emotional empathy moderates activity in other muscles, for instance levator labii in response to disgust and lateral frontalis in response to fearful facial expressions. Results of these studies suggest that more empathic individuals are more sensitive to the emotions expressed by others at the level of facial mimicry. It has been suggested that FM has important consequences for social behavior (Kret et al., 2015) because it facilitates understanding of emotion by inducing an appropriate empathic response (Adolphs, 2002; Preston and de Waal, 2002; Decety and Jackson, 2004).

Emotional Facial Expression, Mirror Neuron System and Limbic Structures

On the neuronal level, the PAM assumes that observing the actions of another individuals stimulates the same action in the observers by activating the brain structures that are involved in executing the same behavior (Preston, 2007). It has been suggested that the Mirror Neuron System (MNS) represents the neural basis of the PAM (Gallese et al., 1996; Rizzolatti and Sinigaglia, 2010). Indeed, the first evidence of mirror neurons (localized in monkeys in the ventral sector of the F5 area) came from experiments where monkeys performed a goal-directed action (e.g., holding, grasping or manipulating objects) or when they observed another individual (monkeys or human) execute the same action (Gallese et al., 1996; Rizzolatti and Craighero, 2004; Gallese et al., 2009). Similarly, studies in humans have shown that the MNS is activated during imagination or imitation of simple or complex hand movements (Ruby and Decety, 2001; Iacoboni et al., 2005; Iacoboni and Dapretto, 2006). Furthermore, neuroimaging studies have shown that pure observation and imitation of emotional facial expressions engaged the MNS, particularly regions of the inferior frontal gyrus (IFG) and the inferior parietal lobule (IPL) (Rizzolatti et al., 2001; Carr et al., 2003; Rizzolatti and Craighero, 2004; Iacoboni and Dapretto, 2006), which are considered core regions of the MNS in human.

Apart from core regions of the MNS, the insula and the amygdala, limbic system's structures, are proposed to be involved in processing of emotional facial expressions (Iacoboni et al., 2005). For example the amygdala activation was shown for fear expressions (Carr et al., 2003; Ohrmann et al., 2007; van der Zwaag et al., 2012), while the anterior insula (AI) for disgust expressions (Jabbi and Keysers, 2008; Seubert et al., 2010). Recently, the insula and dorsal part of anterior cingulate cortex together with a set of limbic and subcortical structures (including the amygdala), constitute the brain's salience network (Seeley et al., 2007). The salience network is thought to mediate the detection and integration of behaviorally relevant stimuli (Menon and Uddin, 2010) including stimuli that elicit fear (Liberzon et al., 2003; Zheng et al., 2017).

Taking into account involvement of the MNS in social mirroring and phenomenon of facial mimicry, the interactions between the MNS and limbic system is postulated (Iacoboni et al., 2005). It is proposed that during observation and imitation of emotional expressions, the core regions of the MNS (i.e., IFG and IPL) activate the insula, which further activate other structure

of limbic system, i.e., amygdala (Jabbi and Keysers, 2008). However, it should be emphasize that the specific function of the amygdala in affective resonance is still under debate (Adolphs, 2010). For example, van der Gaag et al. (2007) found bilateral anterior insula activation during perception of happy, disgusted and fearful facial expressions compared to non-emotional facial expressions, however, they did not find any amygdala activation. The amount of studies revealed that amygdala is activated rather during conscious imitation than pure observation of emotional facial expressions (Lee et al., 2006; van der Gaag et al., 2007; Montgomery and Haxby, 2008). Moreover, it was shown that extent of amygdala activation could be predicted by extent of movement during imitation of facial expressions (Lee et al., 2006). Some authors proposed that amygdala activation during imitation, but not observation, of emotional facial expressions might reflect increased autonomic activity or feedback from facial muscles to the amygdala (Pohl et al., 2013).

To sum up, there is general agreement that exists among researchers that the insula is involved in affective resonance. Furthermore, the insula and the amygdala were proposed to be a part of an emotional perception-action matching system (Iacoboni and Dapretto, 2006; Keysers and Gazzola, 2006) and therefore to "extend" the classical MNS during emotion processing (van der Gaag et al., 2007; Likowski et al., 2012; Pohl et al., 2013). It is believed that the mirror mechanism might be responsible for motor simulation of facial expressions (core MNS, i.e., IFG and IPL) (Carr et al., 2003; Wicker et al., 2003; Grosbras and Paus, 2006; Iacoboni, 2009), and for affective imitation (extended MNS, i.e., insula) (van der Gaag et al., 2007; Jabbi and Keysers, 2008). However, the exactly role of the amygdala in these processes is not clear.

MNS, FM and Empathy

According to the Perception-Action Model, the facial mimicry is an automatic matched motor response, based on a perception-behavior link (Chartrand and Bargh, 1999; Preston and de Waal, 2002). However, other authors proposed that that FM is not only a simple motor reaction, but also a result of a more generic processes of interpreting the expressed emotion (Hess and Fischer, 2013, 2014). Some evidence for this proposition comes from two studies that used simultaneous measures of blood-oxygen level-dependent (BOLD) and facial electromyography (EMG) signals in an MRI scanner (Likowski et al., 2012; Rymarczyk et al., 2018). Likowski et al. (2012) have found that, for emotional facial expressions of happiness, sadness, and anger, facial EMG correlated with BOLD activity localized to parts of the core MNS (i.e., IFG), as well as areas responsible for processing of emotion (i.e., AI). Similar results were obtained in a separate study that additionally utilized videos of happiness and anger facial expressions were also used (Rymarczyk et al., 2018). In that study, Rymarczyk et al. (2018) showed that activation in core MNS and MNS-related structures were more frequently observed when dynamic emotional expressions were presented as compared to static emotional expressions presentations. The authors concluded that dynamic emotional facial expressions might be a clearer signal to induce motor simulation processes in the core the MNS as well as the affective resonance processes

in limbic structure, i.e., insula. It is worth noting that dynamic stimuli, as compared to static, selectively activated structures related to motion and biological motion perception (Arsalidou et al., 2011; Foley et al., 2012; Furl et al., 2015), as well as MNS brain structures (Sato et al., 2004; Kessler et al., 2011; Sato et al., 2015). Results of aforementioned EMG-fMRI studies suggest that the core MNS and MNS-related limbic structures (e.g., insula) may constitute neuronal correlates of FM. Furthermore, it appears that FM phenomenon contains a motor and an emotional component, each represented by a specific neural network of active brain structures that correlated with facial muscle responses during perception of emotions. Responsible for the motor component are structures thought to be the one constituting the core MNS (e.g., inferior frontal gyrus), involved in observation and execution of motor actions. The insula, MNS-related limbic structure, is involved in emotional-related processes. It should be noted that this assumption is restricted to FM for happiness, sadness, and anger emotion, based on the results of EMG-fMRI studies.

Furthermore, several studies have linked empathic traits to neural activity in the MNS indicating that individuals who have higher activity in the MNS also score higher on emotional aspects of empathy (Kaplan and Iacoboni, 2006; Jabbi et al., 2007; Pfeifer et al., 2008). For example, Jabbi et al. (2007) found positive correlation between the bilateral anterior insula and the frontal operculum activation when subjects observed video clips displaying pleased or disgusted facial expressions. To sum up, there is some evidence that the MNS is underpinning of empathy and that subsystems of MNS is supporting motor and affective simulation. However, till now there is no empirical evidence for link between the MNS, empathy and simulation processes.

Aims of the Study

In our study simultaneous recording of EMG and BOLD signal during perception of facial stimuli were used. We selected natural, static and dynamic facial expressions (neutral, fear, and disgust) from the Amsterdam Dynamic Facial Expression Set (ADFES) (van der Schalk et al., 2011b), based on studies showing that dynamic stimuli are a truer reflection of real-life situations (Krumhuber et al., 2013; Sato et al., 2015; Rymarczyk et al., 2016a). Empathy levels were assessed with the Questionnaire Measure of Emotional Empathy (QMEE), wherein empathy is defined as a "vicarious emotional response to the perceived emotional experiences of others" (Mehrabian and Epstein, 1972, p. 1). According to the reasoning outlined above, our EMG-fMRI investigation had two main goals.

Firstly, we wanted to explore whether the neuronal bases for FM, established for socially related stimuli, i.e., anger and happiness, would be the same for more biologically relevant ones, i.e., fear and disgust. We predicted, that similarly to anger and happiness, the core MNS (i.e., IFG and IPL) and MNS-related limbic structures (i.e., insula, amygdala) would be involved in perception of emotional facial expression. Since, that there is evidence that perception of dynamic emotional stimuli elicits greater brain activity as compared to static stimuli (Arsalidou et al., 2011; Kessler et al., 2011; Foley et al., 2012; Furl et al., 2015), we expected the stronger activation in all

structures of MNS subsystems for dynamic compared to static emotional facial expression.

Secondly, based on the evidence that empathy traits modulate facial mimicry for fear (Balconi and Canavesio, 2016) and disgust (Balconi and Canavesio, 2016; Rymarczyk et al., 2016b), as well as based on the assumption that MNS is the underpinning of empathy processes, we wanted to test whether there are a relations between facial mimicry, empathy and the mirror neuron system. We predicted that highly empathic people would be characterized with greater activation of extended MNS sites, i.e., insula and amygdala, and that these activations would be correlated with stronger facial reactions. Next, according to neuroimaging evidence that the dynamic compared to static emotional stimuli are stronger signal for social communication (Bernstein and Yovel, 2015; Wegrzyn et al., 2015), we explored whether the relations between facial mimicry, empathy and subsystems of MNS could be also be dependant on the modality of the stimuli.

MATERIALS AND METHODS

Subjects

Forty-six healthy individuals (25 males, 21 females, mean ± standard deviation age = 23.8 ± 2.5 years) participated in this study. The subjects had normal or corrected to normal eyesight and none reported neurological diseases. This study was carried out in accordance with the recommendations of Ethics Committee of Faculty of Psychology at the University of Social Sciences and Humanities with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee at the SWPS University of Social Sciences and Humanities. An informed consent form was signed by each participant after the experimental procedures had been clearly explained. After the scanning session, subjects were informed of the aims of the study.

Empathy

Empathy scores were measured with Questionnaire Measure of Emotional Empathy (QMEE), wherein empathy is defined as *a* "vicarious emotional response to the perceived emotional experiences of others" (Mehrabian and Epstein, 1972, p. 1). The QMEE contains 33-items to be completed using a 9-point ratings from −4 (=very strong disagreement) to +4 (=very strong agreement) and was selected given that the questionnaire has a Polish adaptation (Rembowski, 1989) and has been shown to be a useful measure in FM research (Sonnby-Borgström, 2002; Dimberg et al., 2011). For analysis purposes subjects were split into High Empathy (HE) and Low Empathy (LE) groups based on the median score on the QMEE questionnaire.

Facial Stimuli and Apparatus

Facial expressions of disgust and fear were taken from The Amsterdam Dynamic Facial Expression Set (van der Schalk et al., 2011b). Additionally, neutral conditions of the same human actors were used, showing no visible action units specific

to emotional facial expression. Stimuli (F02, F04, F05, M02, M08, and M12) consisted of forward-facing facial expressions presented as static and dynamic displays. Stimuli in the static condition consisted of a single frame from the dynamic video clip, corresponding to its condition. For static fear and disgust, the selected frame represented the peak moment of facial expression. In the case of neutral dynamic expressions, motion was still apparent because actors were either closing their eyes or slightly changing the position of their head. Stimuli were 576 pixels in height and 720 pixels in width. All expressions were presented on a gray background. For an overview of procedure and stimuli see **Figure 1**.

EMG Acquisition

Electromyography data were acquired using an MRI-compatible Brain Products' BrainCap consisting of 2 bipolar and one reference electrode. The electrodes with a diameter of 2 mm were filled with electrode paste and positioned in pairs over the CS and LL on the left side of the face (Cacioppo et al., 1986; Fridlund and Cacioppo, 1986). A reference electrode, 6 mm in diameter, was filled with electrode paste and attached to the forehead. Before the electrodes were attached, the skin was cleaned with alcohol. This procedure was repeated until electrode impedance was reduced to 5 kΩ or less. The digitized EMG signals were recorded using a BrainAmp MR plus ExG amplifier and BrainVision Recorder. The signal was low-pass filtered at 250 Hz during acquisition. Finally, data were digitized using a sampling rate of 5 kHz, and stored on a computer running MS Windows 7 for offline analysis.

Image Acquisition

The MRI data were acquired on a Siemens Trio 3 T MR-scanner equipped with a 12-channel phased array head coil. Functional MRI images were collected using a T2*-weighted EPI gradient-echo pulse sequence with the following parameters: TR = 2,000 ms, TE = 25 ms; 90° flip angle, FOV = 250 mm, matrix = 64 × 64, voxel size = 3.5 mm × 3.5 mm × 3.5 mm, interleaved even acquisition, slice thickness = 3.5 mm, 39 slices.

Procedure

Each volunteer was introduced to the experimental procedure and signed a consent form. To conceal the true purpose, facial electromyography recordings, participants were told that sweat gland activity was being recorded while watching the faces of actors selected for commercials by an external marketing company. Following the attachment of the electrodes of the FaceEMGCap-MR, participants were reminded to carefully observe the actors presented on the screen and were positioned in the scanner. The subjects were verbally encouraged to feel comfortable and behave naturally.

The scanning session started with a reminder of the subject's task. In the session subjects were presented with 72 trials that lasted approximately 15 min. Each trial started with a white fixation cross, 80 pixels in diameter, which was visible for 2 s in the center of the screen. Next, one of the stimuli with a facial expression (disgusts, fear or neutral, each presented as static image or dynamic video clip) was presented for 6 s. The expression was followed by a blank gray screen presented

FIGURE 1 | Scheme of procedure used in the study. Images used in the figure were obtained and published with permission of the copyright holder of the Amsterdam Dynamic Facial Expression Set (van der Schalk et al., 2011b).
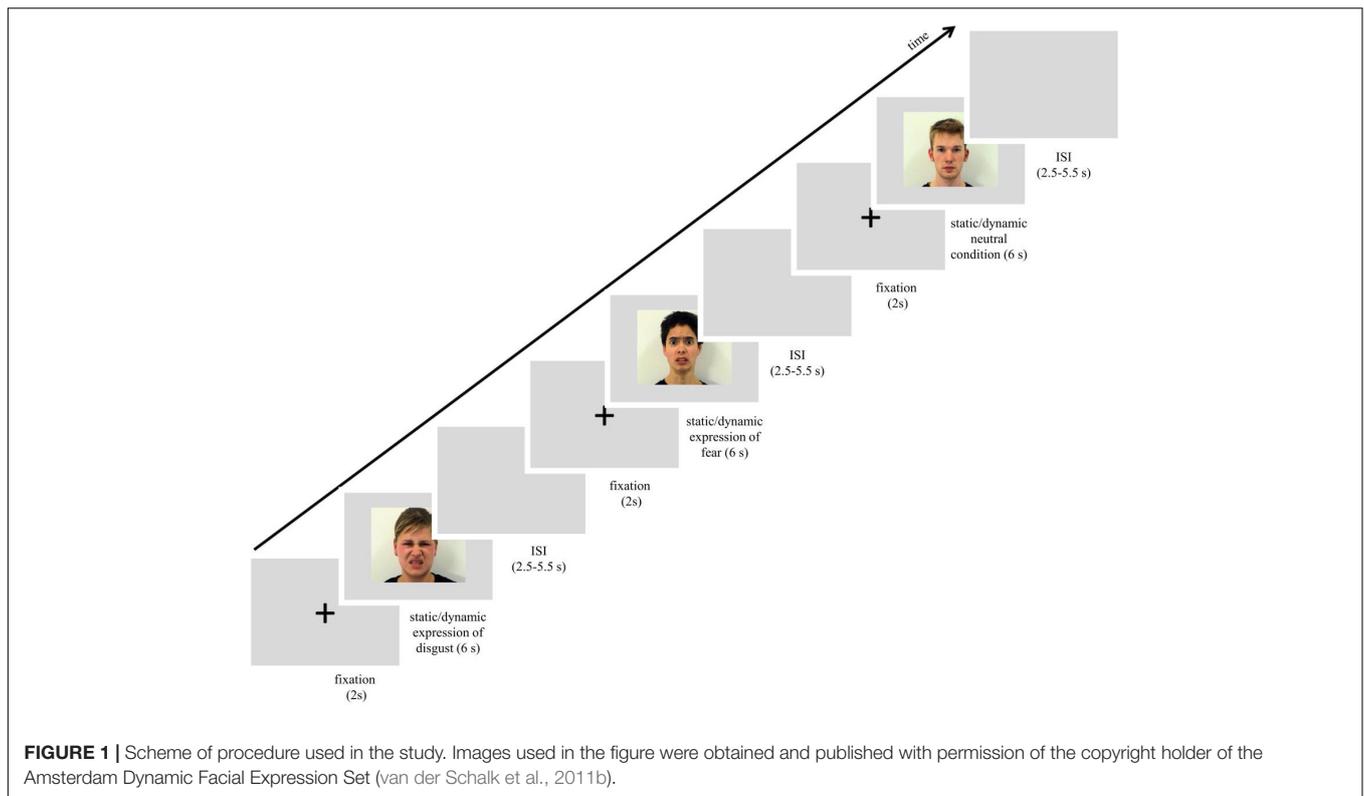
for 2.75–5.25 s (see **Figure 1**). All stimuli were presented in the center of the screen. In summary, each stimulus was repeated once, for a total of 6 presentations within a type of expression (e.g., 6 dynamic presentations of happiness). The stimulus appeared in an event-related manner, pseudo-randomized trial by trail with constraints in rand no facial expression from the same actor, and no more than 2 actors of the same sex or the same emotion were presented consecutively. In total, 6 randomized event-related sessions with introduced constraints were balanced between subjects. The procedure was controlled using Presentation® software running on a computer with Microsoft Windows operating system and was displayed on a 32-inch NNL LCD MRI-compatible monitor with a mirroring system (1920 pixels × 1080 pixels resolution; 32 bit color rate; 60 Hz refresh rate) from a viewing distance of approximately 140 cm.

Data Analysis

EMG Analysis

Pre-processing was carried out using BrainVision Analyzer 2 (version 2.1.0.327). First, EPI gradient-echo pulse artifacts were removed using the average artifact subtraction AAS method (Allen et al., 2000) implemented in the BrainVision Analyzer. This method is based on the sliding average calculation, and consists of 11 consecutive functional volumes marked in the data logs. Synchronization hardware and MR trigger markers allowed for the use of the AAS method for successfully removing MR-related artifacts from the data. Next, standard EMG processing was carried out, which included a signal transformation with

30 Hz high-pass filter. The EMG data were subsequently rectified and integrated over 125 ms and resampled to 10 Hz. Artifacts related to EMG were detected using two methods. First, when single muscle activity was above 8 μV at baseline (i.e., visibility of the fixation cross) (Weyers et al., 2006; Likowski et al., 2008, 2011), the trial was classified as an artifact and excluded from further analysis. All remaining trials were blind-coded and visually checked for artifacts. In the next step, trials were baseline corrected such that the EMG response was measured as the difference of averaged signal activity between the stimuli duration (6 s) and baseline period (2 s). Finally, the signal was averaged for each condition, for each participant. These averaged values were subsequently imported into SPSS 21 for statistical analysis.

Differences in EMG responses were examined using a three-way mixed-model ANOVA with expression (disgust, fear, and neutral) and stimulus mode (dynamic and static) as within-subjects factors and empathy group [low empathy (LE), high empathy (HE)] as the between-subjects factor[1]. Separate ANOVAs were calculated for responses from each muscle, and reported with a Bonferroni correction and with Greenhouse-Geisser correction, when the sphericity assumption was violated. In order to confirm that EMG activity changed from baseline and that FM occurred, EMG data for each significant effect were tested for a difference from zero (baseline) using one-sample, two-tailed t-tests.

---

[1] Additionally robust 3 way ANOVA based on trimmed means (trim = 0.1) was calculated using WRS2 package (*t3way* function) in R software (version 3.4.0).

fMRI Processing and Analysis

Image processing and analysis was carried out using SPM12 software (6470) run in MATLAB 2013b (The Mathworks Inc, 2013). Standard pre-processing steps were applied to functional images, i.e., motion-correction and co-registration to the mean functional image. The independent SPM segmentation module was used to divide structural images into different tissue classes [gray matter, white matter, and non-brain (cerebrospinal fluid, skull)]. Next, based on previously segmented structural images, a study-specific template was created and affine registered to MNI space using the DARTEL algorithm. In particular, the functional images were warped to MNI space based on DARTEL priors, resliced to 2 mm × 2 mm × 2 mm isotropic voxels and later smoothed with an 8 mm × 8 mm × 8 mm full-width at half maximum Gaussian kernel. Single subject design matrices were constructed with six experimental conditions, corresponding to dynamic and static trials for each of the three expression conditions (disgust, fear, and neutral). These conditions were modeled with a standard hemodynamic response function, as well as, other covariates including head movements and parameters that excluded other fMRI artifacts produced by Artifact Detection Toolbox (ART). Later, the same sets of contrasts of interest (listed under "Results" section, i.e., fMRI data) were calculated for each subject and used in group level analysis (i.e., one-sample t-test) for statistical Regions of Interest (ROIs) analysis. The analysis was performed using the MarsBar toolbox (Brett et al., 2002) for the individual ROIs. ROIs consisted of anatomical masks derived from the WFU Pickatlas (Wake Forest University, 2014), and SPM Anatomy Toolbox (Eickhoff, 2016). The STS was defined as an overlapping set of peaks with a radius of 8 mm based on activation peaks reported in literature (Van Overwalle, 2009). Each ROI was extracted as the mean value from the mask. Statistics of brain activity in each contrast were reported with Bonferroni correction.

Correlation Analysis

To understand mutual relationship between brain activity and the facial muscle activity and reveal which ROIs are directly related to FM, bootstrapped (BCa, samples = 1000) Pearson correlation coefficients were calculated between contrasts of brain activity (disgust dynamic, disgust static, fear dynamic, and fear static) and corresponding mimicry.

Each ROI was represented by a single value, which was the mean of all the voxels in that anatomical mask in each hemisphere. Muscle activity was defined as baseline corrected EMG trials of the same muscle and type. The correlations were performed in pairs of variables of muscle and EMG activity, e.g., CS response to static disgust faces with fMRI response in the left insula to static disgust faces.

RESULTS

Empathy Scores

The QMEE scores of the two groups were significantly different [$t_{(44)} = 9.583$, $p < 0.001$; $M_{HE} = 69.4$, $SE_{HE} = 3.7$; $M_{LE} = 14,64$, $SE_{LE} = 4.3$]. The HE group included 13 males ($M = 61.38$,

$SE = 4.86$) and 11 females ($M = 78.91$, $SE = 4.42$) and the LE group consisted of 12 males ($M = 12.83$, $SE = 6.35$) and 10 females ($M = 16.8$, $SE = 6.18$).

EMG Measures

M. Corrugator Supercilii

ANOVA[2] showed a significant main effect of expression [$F_{(2,72)} = 26.527$, $p < 0.001$, $\eta^2 = 0.424$], indicating that activity of the CS for disgust ($M = 0.217$, $SE = 0.025$) was similar to fear [$M = 0.216$, $SE = 0.020$; $t_{(36)} = 0.036$, $p > 0.999$] and higher for both fear and disgust as compared to neutral expressions [$M = 0.028$, $SE = 0.018$; disgust vs. neutral: $t_{(36)} = 5.559$, $p < 0.001$; fear vs. neutral: $t_{(36)} = 6.714$, $p < 0.001$]. Between-subject effect of empathy were also significant [$F_{(1,36)} = 24.813$, $p < 0.001$, $\eta^2 = 0.408$] with the activity of CS generally higher for HE ($M = 0.215$, $SE = 0.016$) than LE ($M = 0.092$, $SE = 0.019$) group.

The significant interaction of expression × empathy group [$F_{(2,72)} = 4.583$, $p = 0.013$, $\eta^2 = 0.113$] revealed that activity of the CS in the HE group for disgust ($M = 0.307$, $SE = 0.032$) was similar to fear [$M = 0.300$, $SE = 0.026$; $t_{(36)} = 0.194$, $p > 0.999$] and higher for both emotions compared to neutral expressions [$M = 0.037$, $SE = 0.024$; disgust vs. neutral: $t_{(36)} = 6.136$, $p < 0.001$; fear vs. neutral: $t_{(35)} = 7.306$, $p < 0.001$]. In the LE, in contrast, higher CS activity was observed for fear ($M = 0.131$, $SE = 0.030$) compared to neutral expressions [$M = 0.019$, $SE = 0.028$; $t_{(36)} = 2.690$, $p = 0.034$] and no other pair differences were observable [$M_{LE \text{ disgust}} = 0.126$, $SE_{LE \text{ disgust}} = 0.038$; LE: disgust vs. neutral: $t_{(36)} = 2.118$, $p = 0.128$; LE: disgust vs. fear: $t_{(36)} = 0.119$, $p > 0.999$]. Higher CS activity was observed in the HE group as compared to the LE group for disgust [$t_{(36)} = 3.620$, $p = 0.001$] and fearful faces [$t_{(36)} = 4.225$, $p < 0.001$]. No group differences observed for neutral expressions [$t_{(36)} = 0.486$, $p = 0.621$] (see **Figure 2**).

There was no significant main effect of modality [$F_{(1,36)} = 0.169$, $p = 0.683$, $\eta^2 = 0.005$] and the following interactions did not reach significance: modality × empathy [$F_{(1,36)} = 0.044$, $p = 0.834$, $\eta^2 = 0.001$], expression × modality [$F_{(2,72)} = 0.013$, $p = 0.987$, $\eta^2 = 0.000$] and expression × modality × empathy [$F_{(2,72)} = 0.039$, $p = 0.962$, $\eta^2 = 0.001$].

One-sample t-tests in HE and LE groups revealed a significant increase in CS activity for all disgust and fear conditions (see **Table 1**). There was no significant difference in CS activity from baseline in response to neutral expressions.

---

[2]Robust ANOVA based on trimmed means confirmed the parametric ANOVA results of CS. There were significant main effects of expression ($Q = 67.741$, $p < 0.001$), empathy ($Q = 31.549$, $p < 0.01$) and significant expression × empathy interaction ($Q = 12.417$, $p < 0.01$). No other effects or interactions were significant ($Q_{modality} = 0.140$, $p > 0.05$; $Q_{emotion \times modality} = 0.144$, $p > 0.05$; $Q_{modality \times empathy} = 0.036$, $p > 0.05$; $Q_{emotion \times modality \times empathy} = 0.043$, $p > 0.05$). Shapiro–Wilk test was used to test normality distribution assumption of levator activity for parametric ANOVA ($W_{HE:disgustdynamic} = 0.983$, $p > 0.05$; $W_{HE:disguststatic} = 0.940$, $p > 0.05$; $W_{HE:feardynamic} = 0.958$, $p > 0.05$; $W_{HE:fearstatic} = 0.976$, $p > 0.05$; $W_{HE:neutraldynamic} = 0.956$, $p > 0.05$; $W_{HE:neutralstatic} = 0.880$, $p < 0.05$; $W_{LE:disgustdynamic} = 0.955$, $p > 0.05$; $W_{LE:disguststatic} = 0.951$, $p > 0.05$; $W_{LE:feardynamic} = 0.945$, $p > 0.05$; $W_{LE:fearstatic} = 0.992$, $p > 0.05$; $W_{LE:neutraldynamic} = 0.958$, $p > 0.05$; $W_{LE:neutralstatic} = 0.966$, $p > 0.05$).

FIGURE 2 | Mean (±SE) EMG activity changes and corresponding statistics for corrugator supercilii during presentation conditions. Separate asterisks indicate significant differences between conditions (simple effects) in EMG responses: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SE, standard error.

M. Levator Labii

ANOVA[3] showed a significant main effect of expression [$F_{(2,76)} = 33.989$, $p < 0.001$, $\eta^2 = 0.486$], indicating that activity of the LL was higher for disgust ($M = 0.170$, $SE = 0.022$) as compared to both fear [$M = -0.073$, $SE = 0.031$; $t_{(36)} = 6.914$, $p < 0.001$] and neutral expressions [$M = -0.073$, $SE = 0.025$; $t_{(36)} = 8.483$, $p < 0.001$]. There was no difference in LL activity between fear and neutral conditions [$t_{(36)} = 0.105$, $p > 0.999$]. The between-subject effect of empathy was significant [$F_{(1,36)} = 6.579$, $p = 0.015$, $\eta^2 = 0.155$], such that activity of LL was higher for HE ($M = 0.052$, $SE = 0.023$) compared to LE ($M = -0.038$, $SE = 0.026$) groups.

The significant interaction of expression × empathy group [$F_{(2,72)} = 3.980$, $p = 0.023$, $\eta^2 = 0.100$] revealed that, for HE group, activity of the LL was higher for disgust ($M = 0.270$, $SE = 0.028$) compared to both fear [$M = -0.053$, $SE = 0.040$; $t_{(36)} = 7.022$, $p < 0.001$] and neutral expressions [$M = -0.062$, $SE = 0.033$; $t_{(36)} = 8.973$, $p < 0.001$]. Similarly, in the LE group, higher LL activity was observed for disgust ($M = 0.070$, $SE = 0.033$) compared to fear [$M = -0.092$, $SE = 0.030$; $t_{(36)} = 2.981$, $p = 0.014$] and neutral expressions [$M = -0.090$, $SE = 0.039$; $t_{(36)} = 3.636$, $p = 0.003$] (see **Figure 3**). Within groups, there was no difference in LL between fear and neutral expressions [HE: $t_{(36)} = 0.184$, $p > 0.999$; LE: $t_{(36)} = 0.034$, $p > 0.999$].

---

[3]Robust ANOVA based on trimmed means confirmed the parametric ANOVA results of levator. There were significant main effects of expression ($Q = 102.411$, $p < 0.001$), empathy ($Q = 11.668$, $p < 0.01$) and significant expression × empathy interaction ($Q = 12.923$, $p < 0.01$). No other effects or interactions were significant ($Q_{modality} = 2.102$, $p > 0.05$; $Q_{emotion \times modality} = 0.902$, $p > 0.05$; $Q_{modality \times empathy} = 0.225$, $p > 0.05$; $Q_{emotion \times modality \times empathy} = 0.070$, $p > 0.05$). Shapiro–Wilk test was used to test normality distribution assumption of levator activity for parametric ANOVA ($W_{HE:disgustdynamic} = 0.984$, $p > 0.05$; $W_{HE:disguststatic} = 0.941$, $p > 0.05$; $W_{HE:feardynamic} = 0.936$, $p > 0.05$; $W_{HE:fearstatic} = 0.925$, $p > 0.05$; $W_{HE:neutraldynamic} = 0.885$, $p < 0.05$; $W_{HE:neutralstatic} = 0.949$, $p > 0.05$; $W_{LE:disgustdynamic} = 0.937$, $p > 0.05$; $W_{LE:disguststatic} = 0.836$, $p < 0.01$; $W_{LE:feardynamic} = 0.975$, $p > 0.05$; $W_{LE:fearstatic} = 973$, $p > 0.05$; $W_{LE:neutraldynamic} = 0.874$, $p < 0.05$; $W_{LE:neutralstatic} = 0.836$, $p < 0.01$).

TABLE 1 | Descriptive statistics for corrugator supercilii activity.

		M	*SE*	*t*	*p*
Disgust	HE	0.31	0.04	8.77	0.000
	LE	0.14	0.03	5.16	0.000
Fear	HE	0.29	0.03	9.06	0.000
	LE	0.13	0.02	7.78	0.000
Neutral	HE	0.04	0.02	1.48	0.154
	LE	0.02	0.02	0.86	0.405

HE, high empathy group; LE, low empathy group; M, mean; SE, standard error; t, value of one sample t-test if value in column M differs from baseline; p, p-value, indicating if one sample t-test is significant, i.e., if value in column M significantly differs from baseline.

FIGURE 3 | Mean (±SE) EMG activity changes and corresponding statistics for levator labii during presentation conditions. Separate asterisks indicate significant differences between conditions (simple effects) in EMG responses: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. SE, standard error.

TABLE 2 | Descriptive statistics for levator labii activity.

		M	*SE*	*t*	*p*
Disgust	HE	0.27	0.03	8.53	0.000
	LE	0.09	0.02	3.52	0.003
Fear	HE	−0.06	0.04	−1.44	0.164
	LE	−0.08	0.04	−2.13	0.048
Neutral	HE	−0.06	0.03	−2.05	0.053
	LE	−0.08	0.04	−1.76	0.097

HE, high empathy group; LE, low empathy group; M, mean; SE, standard error; t, value of one sample t-test if value in column M differs from baseline; p, p-value, indicating if one sample t-test is significant, i.e., if value in column M significantly differs from baseline.

The main effect of modality was not significant [$F_{(1,36)} = 1.315$, $p = 0.259$, $\eta^2 = 0.035$] and the following interactions did not reach significance: modality × empathy [$F_{(1,36)} = 0.000$, $p = 0.995$, $\eta^2 = 0.000$], expression × modality [$F_{(2,72)} = 0.458$, $p = 0.634$, $\eta^2 = 0.013$] and expression × modality × empathy [$F_{(2,72)} = 0.238$, $p = 0.789$, $\eta^2 = 0.007$].

One-sample t-tests in HE and LE groups revealed higher LL activity for all disgust conditions as compared to baseline (see **Table 2**). There was no differences in LL activity from baseline in response to fear and neutral expressions.

fMRI Data

Regions of interest analyses were carried out for the contrasts that compare brain activation while viewing dynamic versus static facial expressions, resulting in eleven contrasts of interest: disgust dynamic > disgust static, fear dynamic > fear static, neutral dynamic > neutral static, emotion dynamic > emotion static (emotion dynamic – pooled dynamic disgust, and fear conditions; emotion static – similar pooling), all dynamic > all static (all dynamic – pooled dynamic disgust, fear and neutral conditions; all static – similar pooling), disgust dynamic > neutral dynamic, disgust static > neutral static, fear dynamic > neutral dynamic, fear static > neutral static, emotion dynamic > neutral dynamic, emotion static > neutral static. The aforementioned contrasts were calculated in order to investigate two types of questions. The contrast emotion/disgust/fear/all dynamic/static > neutral dynamic/static addresses neural correlates of FM of emotional/disgust/fear/all expressions. The other contrasts (i.e., emotion/disgust/fear/all dynamic > emotion/disgust/fear/all static) relate to the difference in processing between dynamic and static stimuli. Due to no group differences between HE and LE subjects, we report only fMRI ROI results for all subjects (for corresponding whole brain analysis see **Supplementary Tables**).

Regions of interest analyses identified activation in the right hemisphere for the disgust dynamic > disgust static contrast (see **Table 3**).

Bilateral activation was observed in the V5/MT+ and STS for the fear dynamic > fear static contrast. Moreover in the right hemisphere BA45, amygdala and AI were activated (see **Table 4**).

For the neutral dynamic > neutral static contrast, only V5/MT+ and STS were activated bilaterally (see **Table 5**).

Regions of interest analysis for the emotion dynamic > emotion static contrast, revealed bilateral activations

TABLE 4 | Summary statistics for activation in each ROI across all participants for fear dynamic > fear static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.401	8.114	0.000***	0.748	12.919	0.000***
Primary Motor Cortex	−0.079	−2.084	0.979	−0.051	−1.405	0.917
Premotor Cortex	−0.049	−1.610	0.943	−0.012	−0.406	0.657
Inferior Parietal Lobule	−0.033	−0.869	0.805	0.025	0.624	0.268
Superior Temporal Sulcus	0.245	7.639	0.000***	0.362	11.236	0.000***
BA44	0.055	1.404	0.084	0.001	0.032	0.487
BA45	0.113	2.438	0.009	0.121	3.196	0.001*
Amygdala	0.088	2.458	0.009	0.102	2.882	0.003+
Anterior Cingulate Cortex	−0.003	−0.106	0.542	0.006	0.210	0.417
Anterior Insula	0.073	2.536	0.007	0.075	2.958	0.002+
Caudate Head	0.009	0.266	0.396	0.035	0.986	0.165
Putamen	0.020	0.719	0.238	0.010	0.411	0.341
Globus Pallidus	0.008	0.351	0.363	0.027	1.407	0.083

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

in V5/MT+, STS, AI and amygdala. Other structures activated by this contrast were right BA45 and left AI (see **Table 6**).

The all dynamic > all static contrast, indicated bilateral activations in V5/MT+, STS, amygdala and AI. The right BA45 was also activated (see **Table 7**).

Regions of interest analysis for the disgust dynamic > neutral dynamic contrast, revealed bilateral activations in V5/MT+, STS and BA45. Other structures revealed by this contrast were left BA44 and left AI (see **Table 8**).

Regions of interest analysis for the disgust static > neutral static contrast, showed activations in left IPL and right BA45 (see **Table 9**).

For the fear dynamic > neutral dynamic contrast, activations were visible bilaterally in V5/MT+, STS, BA45, amygdala and AI. Activation was also noted in the left BA44 and right putamen (see **Table 10**).

For the fear static > neutral static contrast, activations were observed in the left IPL and left AI (see **Table 11**).

The emotion dynamic > neutral dynamic contrast indicated bilateral activations in V5/MT+, STS, BA45, amygdala and AI. Activation was also observed in the left BA44 and right putamen for this contrast (see **Table 12**).

The emotion static > neutral static contrast was associated with activation in left premotor cortex, left IPL, and right BA45 (see **Table 13**).

Correlation Analysis

Muscle-Brain Correlations of Dynamic and Static Disgust Conditions in All Subjects

Correlation analyses in all subjects revealed linear relationships in the disgust dynamic condition between left AI and LL. In the disgust static condition, a positive relationship was present between the LL and activation of the right premotor cortex, and right caudate head. In the left hemisphere, positive relationships

TABLE 3 | Summary statistics for activation in each ROI across all participants for disgust dynamic > disgust static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.445	11.098	0.000***	0.732	9.419	0.000***
Primary Motor Cortex	−0.043	−0.976	0.833	−0.038	−0.927	0.821
Premotor Cortex	−0.003	−0.079	0.531	0.020	0.615	0.271
Inferior Parietal Lobule	0.000	0.006	0.498	0.025	0.708	0.241
Superior Temporal Sulcus	0.255	7.797	0.000***	0.301	8.313	0.000***
BA44	0.034	0.996	0.162	0.022	0.786	0.218
BA45	0.051	1.371	0.089	0.084	2.969	0.002+
Amygdala	0.083	2.435	0.009	0.085	2.774	0.004
Anterior Cingulate Cortex	−0.044	−1.590	0.941	−0.048	−1.745	0.956
Anterior Insula	0.065	2.733	0.004	0.044	2.133	0.019
Caudate Head	0.014	0.399	0.346	0.020	0.565	0.288
Putamen	0.020	0.765	0.224	0.006	0.244	0.404
Globus Pallidus	0.031	1.517	0.068	0.029	1.354	0.091

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: $^{+}p < 0.1$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 5 | Summary statistics for activation in each ROI across all participants for neutral dynamic > neutral static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.153	3.212	0.001*	0.341	5.245	0.000***
Primary Motor Cortex	−0.011	−0.255	0.600	−0.012	−0.267	0.605
Premotor Cortex	0.022	0.700	0.244	0.021	0.628	0.267
Inferior Parietal Lobule	0.062	1.679	0.050	0.089	2.156	0.018
Superior Temporal Sulcus	0.103	3.126	0.002*	0.168	4.055	0.000**
BA44	−0.026	−0.575	0.716	0.016	0.416	0.340
BA45	0.036	0.671	0.253	0.079	2.015	0.025
Amygdala	−0.021	−0.620	0.731	0.016	0.614	0.271
Anterior Cingulate Cortex	−0.007	−0.228	0.589	0.001	0.041	0.484
Anterior Insula	0.001	0.043	0.483	0.018	0.585	0.281
Caudate Head	0.061	1.488	0.072	0.094	2.566	0.007
Putamen	0.010	0.339	0.368	0.014	0.544	0.295
Globus Pallidus	0.030	1.474	0.074	0.012	0.568	0.286

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: $^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 6 | Summary statistics for activation in each ROI across all participants for emotion dynamic > emotion static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.846	11.680	0.000***	1.481	12.925	0.000***
Primary Motor Cortex	−0.121	−1.954	0.972	−0.089	−1.505	0.930
Premotor Cortex	−0.052	−1.029	0.846	0.008	0.176	0.431
Inferior Parietal Lobule	−0.032	−0.570	0.714	0.051	0.896	0.187
Superior Temporal Sulcus	0.500	9.325	0.000***	0.663	12.771	0.000***
BA44	0.090	1.581	0.060	0.023	0.427	0.336
BA45	0.164	2.531	0.007	0.205	3.623	0.000*
Amygdala	0.170	3.074	0.002+	0.187	3.773	0.000**
Anterior Cingulate Cortex	−0.047	−1.090	0.859	−0.042	−0.912	0.817
Anterior Insula	0.138	3.654	0.000*	0.119	3.381	0.001*
Caudate Head	0.023	0.415	0.340	0.056	0.918	0.182
Putamen	0.040	0.897	0.187	0.016	0.405	0.344
Globus Pallidus	0.039	1.275	0.104	0.056	1.808	0.039

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: $^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 7 | Summary statistics for activation in each ROI across all participants for all dynamic > all static expressions contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.999	11.740	0.000***	1.822	11.546	0.000***
Primary Motor Cortex	−0.132	−1.994	0.974	−0.102	−1.415	0.918
Premotor Cortex	−0.029	−0.552	0.708	0.029	0.553	0.291
Inferior Parietal Lobule	0.029	0.459	0.324	0.140	2.247	0.015
Superior Temporal Sulcus	0.603	10.616	0.000***	0.830	12.137	0.000***
BA44	0.064	0.929	0.179	0.039	0.587	0.280
BA45	0.200	2.631	0.006	0.283	4.016	0.000**
Amygdala	0.149	2.909	0.003+	0.203	4.526	0.000***
Anterior Cingulate Cortex	−0.054	−1.126	0.867	−0.040	−0.864	0.804
Anterior Insula	0.139	3.111	0.002+	0.136	2.900	0.003+
Caudate Head	0.084	1.268	0.106	0.150	2.273	0.014
Putamen	0.050	0.986	0.165	0.030	0.639	0.263
Globus Pallidus	0.068	2.135	0.019	0.068	2.289	0.013

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: $^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 8 | Summary statistics for activation in each ROI across all participants for disgust dynamic > neutral dynamic contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.379	7.154	0.000***	0.440	6.654	0.000***
Primary Motor Cortex	0.060	1.021	0.156	0.072	1.268	0.106
Premotor Cortex	0.057	1.283	0.103	0.074	1.652	0.053
Inferior Parietal Lobule	0.070	1.560	0.063	0.036	0.745	0.230
Superior Temporal Sulcus	0.226	5.144	0.000***	0.180	4.452	0.000***
BA44	0.145	3.149	0.001*	0.084	2.594	0.006
BA45	0.153	3.281	0.001*	0.099	3.157	0.001*
Amygdala	0.114	2.761	0.004	0.096	2.652	0.006
Anterior Cingulate Cortex	−0.018	−0.533	0.702	−0.026	−0.775	0.779
Anterior Insula	0.131	3.905	0.000**	0.055	2.083	0.021
Caudate Head	0.008	0.175	0.431	−0.003	−0.052	0.521
Putamen	0.061	1.975	0.027	0.057	2.210	0.016
Globus Pallidus	0.007	0.327	0.373	0.035	1.481	0.073

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: $^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

were found between the LL and activation in BA44, BA45, and AI (see **Table 14**).

Positive relationships between CS and brain activity was found in the right hemisphere in the caudate head and globus pallidus as well as in various regions of the left hemisphere (IPL, STS, ACC, AI, caudate head, globus pallidus) (see **Table 14**).

Muscle-Brain Correlations of Dynamic and Static Fear Conditions in All Subjects

Correlation analyses in all subjects revealed a positive relationship between CS in activation in the left BA44, right BA45, and AI for the static fear condition.

In the dynamic fear condition, there was a positive relationship between CS and activation in the left globus pallidus (see **Table 15**).

Muscle-Brain Correlations of Dynamic and Static Disgust Conditions in High Empathic Subjects

Correlation analyses of dynamic disgust in HE subjects revealed a positive relationship between LL and brain activity in several region of the right (STS, amygdala, AI, caudate head, putamen, globus pallidus) and left hemispheres (amygdala, AI, caudate head, putamen). For static disgust in HE subjects, the relationship between LL and brain activity

TABLE 9 | Summary statistics for activation in each ROI across all participants for disgust static > neutral static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.087	2.050	0.023	0.048	0.831	0.205
Primary Motor Cortex	0.092	2.133	0.019	0.098	2.065	0.022
Premotor Cortex	0.082	2.317	0.013	0.075	2.044	0.023
Inferior Parietal Lobule	0.132	3.582	0.000*	0.099	2.110	0.020
Superior Temporal Sulcus	0.073	2.300	0.013	0.047	1.385	0.086
BA44	0.084	1.961	0.028	0.078	2.323	0.012
BA45	0.138	2.861	0.003	0.094	3.058	0.002+
Amygdala	0.010	0.266	0.396	0.027	0.811	0.211
Anterior Cingulate Cortex	0.020	0.625	0.268	0.023	0.680	0.250
Anterior Insula	0.067	2.353	0.012	0.028	1.324	0.096
Caudate Head	0.056	1.700	0.048	0.071	1.943	0.029
Putamen	0.050	1.729	0.045	0.064	2.358	0.011
Globus Pallidus	0.005	0.211	0.417	0.018	0.718	0.238

Asterisks indicate significant, Bonferroni corrected, activations of each ROI:
$^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 10 | Summary statistics for activation in each ROI across all participants for fear dynamic > neutral dynamic contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.310	5.995	0.000***	0.470	8.129	0.000***
Primary Motor Cortex	0.010	0.250	0.402	0.020	0.486	0.315
Premotor Cortex	0.015	0.467	0.321	0.046	1.501	0.070
Inferior Parietal Lobule	0.033	0.842	0.202	0.016	0.409	0.342
Superior Temporal Sulcus	0.213	5.166	0.000***	0.269	7.614	0.000***
BA44	0.153	3.569	0.000*	0.073	2.084	0.021
BA45	0.184	3.181	0.001*	0.134	3.510	0.001*
Amygdala	0.187	4.699	0.000***	0.165	5.454	0.000***
Anterior Cingulate Cortex	0.014	0.464	0.323	0.010	0.360	0.360
Anterior Insula	0.171	4.755	0.000***	0.098	3.806	0.000**
Caudate Head	0.008	0.203	0.420	0.016	0.436	0.332
Putamen	0.075	2.691	0.005	0.072	3.260	0.001*
Globus Pallidus	0.034	1.457	0.076	0.055	2.538	0.007

Asterisks indicate significant, Bonferroni corrected, activations of each ROI:
$^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 11 | Summary statistics for activation in each ROI across all participants for fear static > neutral static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.062	1.509	0.069	0.062	1.438	0.079
Primary Motor Cortex	0.078	1.988	0.026	0.059	1.495	0.071
Premotor Cortex	0.087	2.680	0.005	0.078	2.688	0.005
Inferior Parietal Lobule	0.128	3.402	0.001*	0.079	2.063	0.022
Superior Temporal Sulcus	0.071	2.287	0.013	0.075	2.218	0.016
BA44	0.072	1.544	0.065	0.087	2.663	0.005
BA45	0.107	1.855	0.035	0.092	2.420	0.010
Amygdala	0.078	2.199	0.017	0.079	2.447	0.009
Anterior Cingulate Cortex	0.010	0.347	0.365	0.005	0.144	0.443
Anterior Insula	0.099	2.929	0.003+	0.041	1.573	0.061
Caudate Head	0.060	1.313	0.098	0.075	1.665	0.051
Putamen	0.065	2.030	0.024	0.075	2.481	0.008
Globus Pallidus	0.055	2.102	0.021	0.040	1.577	0.061

Asterisks indicate significant, Bonferroni corrected, activations of each ROI:
$^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 12 | Summary statistics for activation in each ROI across all participants for emotion dynamic > neutral dynamic contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.689	7.126	0.000***	0.909	7.935	0.000***
Primary Motor Cortex	0.070	0.748	0.229	0.092	0.996	0.162
Premotor Cortex	0.073	1.030	0.154	0.120	1.731	0.045
Inferior Parietal Lobule	0.104	1.342	0.093	0.051	0.660	0.256
Superior Temporal Sulcus	0.438	5.496	0.000***	0.449	6.412	0.000***
BA44	0.298	3.748	0.000**	0.156	2.612	0.006
BA45	0.337	3.682	0.000**	0.232	3.805	0.000**
Amygdala	0.301	4.085	0.000**	0.261	4.335	0.000**
Anterior Cingulate Cortex	−0.004	−0.068	0.527	−0.016	−0.302	0.618
Anterior Insula	0.302	4.765	0.000***	0.152	3.248	0.001*
Caudate Head	0.016	0.208	0.418	0.013	0.170	0.433
Putamen	0.136	2.523	0.008	0.128	2.978	0.002+
Globus Pallidus	0.040	1.055	0.149	0.090	2.235	0.015

Asterisks indicate significant, Bonferroni corrected, activations of each ROI:
$^+p < 0.1$, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

was significant for the left AI, right caudate head, and bilateral amygdalae (see **Table 16**).

Correlation analyses of dynamic disgust in HE subjects revealed no relationship between CS and brain activations. For the static disgust in HE subjects, the relationship between CS and brain activity was significant in the regions of the right (caudate head, putamen, globus pallidus) and left hemispheres (amygdala, caudate head, putamen, globus pallidus) (see **Table 16**).

Muscle-Brain Correlations of Dynamic and Static Fear Conditions in High Empathic Subjects
Correlation analyses of dynamic fear in HE subjects revealed a positive relationship between CS and brain activity in amygdalae

bilaterally and left globus pallidus. For static fear in HE subjects a significant relationship between CS and brain activity was significant for the bilateral amygdalae and putamen and right AI (see **Table 17**).

Muscle-Brain Correlations of Dynamic and Static Disgust Conditions in Low Empathic Subjects
In LE subjects, the relationship between LL and brain activity was found only in the disgust static condition, for left BA44, putamen and globus pallidus bilaterally (see **Table 18**).

Correlation analyses of dynamic disgust in LE subjects revealed a positive relationship between CS and activity in right amygdala, and negative relationship between this muscle and

TABLE 13 | Summary statistics for activation in each ROI across all participants for emotion static > neutral static contrast.

Region of Interest	Left Hemisphere			Right Hemisphere		
	M	t	p	M	t	p
V5/MT+	0.149	2.041	0.024	0.111	1.296	0.101
Primary Motor Cortex	0.170	2.474	0.009	0.156	2.096	0.021
Premotor Cortex	0.169	2.915	0.003[+]	0.153	2.786	0.004
Inferior Parietal Lobule	0.259	4.121	0.000**	0.178	2.496	0.008
Superior Temporal Sulcus	0.144	2.624	0.006	0.122	2.240	0.015
BA44	0.156	1.898	0.032	0.166	2.733	0.004
BA45	0.245	2.443	0.009	0.185	2.898	0.003[+]
Amygdala	0.088	1.371	0.089	0.106	1.889	0.033
Anterior Cingulate Cortex	0.030	0.558	0.290	0.028	0.479	0.317
Anterior Insula	0.167	2.935	0.003[+]	0.069	1.685	0.049
Caudate Head	0.116	1.723	0.046	0.146	2.071	0.022
Putamen	0.115	2.107	0.020	0.139	2.754	0.004
Globus Pallidus	0.061	1.300	0.100	0.058	1.278	0.104

Asterisks indicate significant, Bonferroni corrected, activations of each ROI: [+]p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001.

left globus pallidus. For the static disgust condition, there was a positive relationship between CS and brain activity in the right (IPL, STS, ACC, and caudate head) and in the left hemisphere (V5/MT+, premotor cortex, IPL, STS, ACC, AI, caudate head, globus pallidus) among LE subjects (see **Table 18**).

Muscle-Brain Correlations of Dynamic and Static Fear Conditions in Low Empathic Subjects

In LE subjects, there was a relationship between CS and brain activity only in static fear condition, for left BA44 (see **Table 19**).

TABLE 15 | Muscles-brain correlations of dynamic and static fear conditions in all subjects.

Region of Interest	Fear Dynamic		Fear Static	
	CS		CS	
	LH	RH	LH	RH
V5/MT+	−0.099	0.042	−0.049	−0.112
Primary Motor Cortex	0.063	0.034	−0.038	−0.003
Premotor Cortex	−0.019	−0.024	0.128	0.126
Primary Somatosensory Cortex	0.050	0.050	−0.038	−0.007
Inferior Parietal Lobule	−0.034	−0.016	0.059	0.120
Superior Temporal Sulcus	0.061	−0.019	0.052	0.029
BA44	−0.049	−0.001	0.281[+]	0.102
BA45	−0.119	−0.087	0.200	0.200
Amygdala	0.171	0.183	0.213	0.295[+]
Anterior Cingulate Cortex	−0.028	−0.027	−0.141	−0.087
Anterior Insula	0.007	−0.006	0.201	0.301[+]
Caudate Head	−0.048	−0.087	0.130	0.152
Putamen	0.060	0.044	0.202	0.246
Globus Pallidus	0.268[+]	0.005	0.201	0.213

Post-number asterisks indicate significant Pearson correlations of muscle-ROI pairs: [+]p < 0.1. CS, corrugator supercilii. LH, left hemisphere; RH, right hemisphere.

DISCUSSION

In the present study, static and dynamic stimuli were used to investigate facial reactions and brain activation in response to emotional facial expressions. To assess neuronal structures involved in automatic, spontaneous mimicry during perception of fear and disgust facial expressions, we collected simultaneous

TABLE 14 | Muscles-brain correlations of dynamic and static disgust conditions in all subjects.

Region of Interest	Disgust Dynamic				Disgust Static			
	CS		LL		CS		LL	
	LH	RH	LH	RH	LH	RH	LH	RH
V5/MT+	−0.094	−0.219	0.033	0.045	0.162	0.003	−0.057	0.141
Primary Motor Cortex	0.169	0.204	0.115	−0.024	0.161	0.153	0.181	0.177
Premotor Cortex	0.156	0.139	0.193	0.160	0.236	0.223	0.317*	0.276[+]
Primary Somatosensory Cortex	0.165	0.069	0.153	−0.035	0.213	0.156	0.182	0.161
Inferior Parietal Lobule	0.141	−0.031	0.181	0.067	0.358*	0.216	0.193	0.009
Superior Temporal Sulcus	0.086	−0.067	0.194	0.248	0.337*	0.176	0.107	−0.040
BA44	−0.014	−0.070	0.052	−0.176	0.025	0.113	0.299[+]	0.187
BA45	0.062	−0.046	0.031	−0.028	0.147	0.086	0.308[+]	0.219
Amygdala	0.163	0.214	0.156	0.154	0.246	0.114	0.235	0.142
Anterior Cingulate Cortex	−0.140	−0.141	0.086	0.105	0.279[+]	0.241	−0.009	−0.011
Anterior Insula	0.259	0.072	0.285[+]	0.228	0.317*	0.214	0.306[+]	0.235
Caudate Head	0.020	0.108	0.189	0.210	0.437**	0.476**	0.253	0.276[+]
Putamen	0.063	0.090	0.191	0.132	0.150	0.215	0.156	0.173
Globus Pallidus	−0.076	0.138	0.059	0.220	0.339*	0.274[+]	0.240	0.259

Post-number asterisks indicate significant Pearson correlations of muscle-ROI pairs: [+]p < 0.1, *p < 0.05, **p < 0.01. CS, corrugator supercilii; LL, levator labii; LH, left hemisphere; RH, right hemisphere.

TABLE 16 | Muscles-brain correlations of dynamic and static disgust conditions in high empathic subjects.

Region of Interest	Disgust Dynamic				Disgust Static			
	CS		LL		CS		LL	
	LH	RH	LH	RH	LH	RH	LH	RH
V5/MT+	−0.169	−0.257	−0.011	0.054	−0.071	−0.241	−0.194	0.154
Primary Motor Cortex	0.333	0.379[+]	0.236	0.093	0.013	0.097	0.286	0.173
Premotor Cortex	0.284	0.232	0.297	0.236	0.181	0.195	0.351	0.319
Primary Somatosensory Cortex	0.407[+]	0.364[+]	0.371[+]	0.202	0.187	0.108	0.269	0.213
Inferior Parietal Lobule	0.131	0.020	0.289	0.296	0.245	0.199	0.209	0.143
Superior Temporal Sulcus	−0.034	0.058	0.150	0.462*	0.215	0.181	0.046	0.145
BA44	0.062	0.013	0.149	−0.147	0.016	0.143	0.314	0.240
BA45	0.033	−0.097	0.077	−0.011	0.036	0.018	0.344	0.302
Amygdala	0.403[+]	0.391[+]	0.401[+]	0.375[+]	0.395[+]	0.328	0.426*	0.368[+]
Anterior Cingulate Cortex	−0.017	−0.099	0.339	0.323	0.327	0.281	0.156	0.186
Anterior Insula	0.318	0.053	0.377[+]	0.376[+]	0.332	0.258	0.379[+]	0.278
Caudate Head	0.114	0.126	0.422[+]	0.441*	0.560**	0.555**	0.322	0.390[+]
Putamen	0.255	0.313	0.426*	0.423[+]	0.368[+]	0.497*	0.287	0.294
Globus Pallidus	0.115	0.247	0.315	0.428*	0.363[+]	0.371[+]	0.193	0.301

*Post-number asterisks indicate significant Pearson correlations of muscle-ROI pairs: [+]$p < 0.1$, *$p < 0.05$, **$p < 0.01$. CS, corrugator supercilii; LL, levator labii. LH, left hemisphere; RH, right hemisphere.*

recordings of the EMG signal and BOLD response during the perception of stimuli. Additionally, to explore whether empathic traits are linked with facial muscle and brain activity, we divided participants into low and high empathy groups (i.e., LE and HE) based on the median score on a validated questionnaire.

The EMG analysis revealed activity in the CS muscle while viewing both fear and disgust facial displays, while perception of disgust induced facial activity specifically in the LL muscle. Moreover, the HE group showed a larger responses in the CS and LL muscles as compared to the LE group, however, these responses were not differentiable between static and dynamic mode of stimuli.

For BOLD data, we used ROI analyses. We found that dynamic emotional expressions elicited higher activation in the bilateral STS, V5/MT+, bilateral amygdalae, and right BA45 as compared to emotional static expression. For the opposite contrast (static > dynamic), as expected, no significant activations emerged.

Using combined EMG-fMRI analysis, we found significant correlations between brain activity and facial muscle reactions for perception of dynamic as well as static emotional stimuli. The correlated brain structures, e.g., amygdala and AI, were more frequent in the HE compared to LE group.

EMG Response to Fear and Disgust

The main result from EMG recording is that both fear and disgust emotions increased corrugator muscle reactions, whereas levator labii muscle activity was more pronounced in response to disgust than to fearful expressions. Before discussing this result, it should be emphasized that fear and disgust expressions have an opposite biological function, fear is thought to enhance perception to danger and disgust dampens

TABLE 17 | Muscles-brain correlations of dynamic and static fear conditions in high empathic subjects.

Region of Interest	Fear Dynamic		Fear Static	
	CS		CS	
	LH	RH	LH	RH
V5/MT+	−0.227	0.015	−0.122	0.026
Primary Motor Cortex	0.150	0.077	0.010	0.022
Premotor Cortex	−0.040	−0.008	0.177	0.191
Primary Somatosensory Cortex	0.153	0.261	0.018	0.071
Inferior Parietal Lobule	−0.010	0.184	0.082	0.199
Superior Temporal Sulcus	−0.173	0.016	0.018	0.195
BA44	−0.004	0.130	0.276	0.122
BA45	−0.150	−0.039	0.233	0.190
Amygdala	0.368[+]	0.406[+]	0.393[+]	0.411[+]
Anterior Cingulate Cortex	0.159	0.245	−0.141	−0.113
Anterior Insula	0.045	0.144	0.233	0.369[+]
Caudate Head	−0.041	−0.095	0.131	0.199
Putamen	0.165	0.217	0.383[+]	0.426*
Globus Pallidus	0.405[+]	0.090	0.284	0.332

Post-number asterisks indicate significant Pearson correlations of muscle-ROI pairs: [+]$p < 0.1$. CS, corrugator supercilii. LH, left hemisphere; RH, right hemisphere.

it (Susskind et al., 2008). Accordingly, both emotions are characterized by opposite visible surface features, e.g., faster eye movements or velocity inspiration during perception of fear in comparison to perception of disgust (Susskind et al., 2008). It is suggested that fear and disgust involve opposite psychological mechanisms at the physiological level (Krusemark and Li, 2011). Based on the above-mentioned findings, we

TABLE 18 | Muscles-brain correlations of dynamic and static disgust conditions in low empathic subjects.

Region of Interest	Disgust Dynamic				Disgust Static			
	CS		LL		CS		LL	
	LH	RH	LH	RH	LH	RH	LH	RH
V5/MT+	−0.111	−0.275	−0.023	0.036	0.523*	0.403	0.139	0.385
Primary Motor Cortex	0.247	0.118	0.206	−0.075	0.389	0.207	0.098	0.166
Premotor Cortex	0.258	0.202	0.344	0.284	0.468[+]	0.401	0.356	0.316
Primary Somatosensory Cortex	0.226	−0.013	0.231	−0.020	0.381	0.396	0.176	0.340
Inferior Parietal Lobule	0.373	0.207	0.149	−0.120	0.653**	0.669**	0.135	0.052
Superior Temporal Sulcus	0.195	−0.126	0.152	0.175	0.538*	0.508*	0.142	−0.158
BA44	0.069	0.220	0.109	0.123	−0.018	0.221	0.443[+]	0.321
BA45	0.226	0.098	−0.014	−0.035	0.377	0.277	0.177	0.116
Amygdala	0.378	0.412[+]	0.330	0.211	0.372	0.135	0.352	0.181
Anterior Cingulate Cortex	0.234	0.354	0.299	0.278	0.610**	0.646**	−0.022	−0.098
Anterior Insula	0.289	0.230	0.247	0.041	0.459[+]	0.258	0.221	0.324
Caudate Head	0.166	0.383	0.017	0.017	0.454[+]	0.548*	0.257	0.225
Putamen	0.249	0.338	0.344	0.219	0.139	0.056	0.423[+]	0.394
Globus Pallidus	−0.478*	0.026	−0.333	−0.156	0.424[+]	0.379	0.486*	0.610**

Post-number asterisks indicate significant Pearson correlations of muscle-ROI pairs: [+]$p < 0.1$, *$p < 0.05$, **$p < 0.01$. CS, corrugator supercilii; LL, levator labii. LH, left hemisphere; RH, right hemisphere.

anticipated different patterns of facial muscle reaction for the evaluated emotions. Our results concerning the CS contraction for both negative emotions are congruent with earlier studies reporting CS activity during perception of anger (Sato et al., 2008; Dimberg et al., 2011), fear, and disgust emotions (Murata et al., 2016; Rymarczyk et al., 2016b). Moreover, Topolinski and Strack (2015) demonstrated that perception of highly surprising events, compared to lower-level ones, elicited CS activity specifically.

In addition, Neta et al. (2009) suggested that CS activity could reflect the participants' bias, i.e., tendency to rate surprise as either positive or negative. Thus, it is proposed that CS reactions could be an indicator of a global negative affect (Bradley et al., 2001; Larsen et al., 2003) as well as a tool to measure individual differences in emotion regulatory ability (Lee et al., 2012).

Furthermore, we found increased LL activity for disgust facial expressions, but no evidence of activity for fear presentation. There is some evidence that perception of disgust faces (Vrana, 1993; Lundquist and Dimberg, 1995; Cacioppo et al., 2007; Rymarczyk et al., 2016b), a disgusting picture related to contamination (Yartz and Hawk, 2002) or tasting an unpleasant substance (Chapman et al., 2009) leads to the specific contraction of the LL muscle. Moreover, it was shown that reaction of the LL muscle occurred not only for biological but also moral disgust, i.e., during violation of moral norms (Whitton et al., 2014). Taken together, these results demonstrate the reliability of LL as an indicator of disgust experience (Armony and Vuilleumier, 2013, p. 62).

As far as modality of the stimulus is concerned, we did not observe any differences in the magnitude of facial reactions between static and dynamic stimuli. Similar results were found in our earlier study (Rymarczyk et al., 2016b), wherein reaction of the CS, LL and also lateral frontalis muscles were measured.

TABLE 19 | Muscles-brain correlations of dynamic and static fear conditions in low empathic subjects.

Region of Interest	Fear Dynamic		Fear Static	
	CS		CS	
	LH	RH	LH	RH
V5/MT+	0.049	0.072	0.176	−0.382
Primary Motor Cortex	0.313	0.305	0.191	0.279
Premotor Cortex	0.119	0.178	0.263	0.230
Primary Somatosensory Cortex	0.281	0.207	0.148	0.145
Inferior Parietal Lobule	0.002	0.039	−0.062	0.089
Superior Temporal Sulcus	0.275	0.137	0.117	−0.187
BA44	0.028	0.063	0.461[+]	0.319
BA45	−0.171	−0.055	0.139	0.196
Amygdala	0.386	0.322	0.398	0.337
Anterior Cingulate Cortex	−0.104	−0.129	0.271	0.389
Anterior Insula	−0.220	−0.235	0.330	0.371
Caudate Head	−0.027	0.126	0.276	0.288
Putamen	0.188	0.186	0.224	0.110
Globus Pallidus	0.025	0.100	0.030	0.086

Post-number asterisks indicate significant Pearson correlations of muscle-ROI pairs: *$p < 0.05$. CS, corrugator supercilii. LH, left hemisphere; RH, right hemisphere.

We showed only a weak impact of dynamic stimuli on the strength of facial reactions for fear expressions. These reactions were apparent only in the lateral frontalis muscle, which was not measured in the present study. It should be noted that most studies have reported higher EMG response during perception of dynamic than static emotional facial expressions (Weyers et al., 2006; Sato et al., 2008; Rymarczyk et al., 2011); however, most

of these studies tested the role of dynamic mode on the FM phenomenon for happiness and anger. Together, the role of dynamic stimuli in the FM phenomenon for more biologically embedded emotions needs further research.

Facial Mimicry and Empathy

Our data also provide some evidence for the relationship between the intensity of FM and trait emotional empathy. We found that HE compared to LE subjects showed stronger activity in CS and LL muscles for fear and disgust. However, the pattern of FM was the same in HE in LE groups. Our results are in agreement with previous EMG studies, wherein researchers have shown that HE subjects show greater mimicry of emotional expressions for happiness and anger (Sonnby-Borgström, 2002; Sonnby-Borgström et al., 2003; Dimberg et al., 2011), as well as, for fear (Balconi and Canavesio, 2016; Rymarczyk et al., 2016b) and disgust (Balconi and Canavesio, 2013b; Rymarczyk et al., 2016b) expressions as compared to LE subjects. Together, these results suggest that FM and emotional empathy are interrelated phenomena (Hatfield et al., 1992; McIntosh, 2006). Moreover, the magnitude of FM may be a strong predictor of empathy. According to PAM (de Waal, 2008), HE people exhibit stronger FM for emotional stimuli because on a neuronal level they engage brain areas related to the representation of their own feelings, for, e.g., the AI (Preston, 2007).

Neural Network for Fear and Disgust

Neuroimaging data revealed that, observation of dynamic emotional, compared to dynamic neutral stimuli, triggered a distributed brain network that consisted of bilateral STS, V5/MT+, amygdala, AI, and BA45. The left BA44 and right putamen were also activated. In contrast, the perception of static emotional faces as compared to static neutral faces elicited activity in the left IPL, right BA45, and left AI, and left premotor cortex.

Apart from STS and V5/MT+, greater activity for contrast dynamic vs. static fear was found in the right BA45, right amygdala, and right AI. Dynamic versus static disgust faces induced greater activity in the right BA45. Our findings concerning the bilateral visual area V5/MT+ and STS corroborate previous results confirming the importance of these structures in motion and biological motion perception, respectively (Robins et al., 2009; Arsalidou et al., 2011; Foley et al., 2012; Furl et al., 2015). It has been suggested that, due to their complex features dynamic facial characteristics require enhanced visual analysis in V5/MT+, which might result in wide-spread activation patterns (Vaina et al., 2001).

Previous studies have reported activations in the STS for facial motion due to speech production (Hall et al., 2005), or facial emotional expressions for happiness and anger (Kilts et al., 2003; Rymarczyk et al., 2018), fear (LaBar et al., 2003) and disgust (Trautmann et al., 2009). Moreover, STS activation was reported during detection of movements of natural faces (Schultz and Pilz, 2009), but not computer-generated faces (Sarkheil et al., 2013). According to the neurocognitive model for face processing (Haxby et al., 2000), STS activity could be related to

enhanced perceptual and/or cognitive processing for dynamic characteristics of faces (Sato et al., 2004). To summarize our results, together with those of others, support the use of dynamic stimuli to study the neuronal correlates of emotional facial expressions (Fox et al., 2009; Zinchenko et al., 2018).

In our study, we found activity in brain areas typically implicated in simulative process, namely the IFG and IPL (Carr et al., 2003; Jabbi and Keysers, 2008). It has been proposed that understating the behavior of others is based on direct mirroring of somatosensory or motor representations of the observed action in the observer's brain (Gazzola et al., 2006; van der Gaag et al., 2007; Jabbi and Keysers, 2008). For example, activation of these MNS structures was found during observation and imitation of others actions, i.e., during hand movement (Gallese et al., 1996; Rizzolatti and Craighero, 2004; Molnar-Szakacs et al., 2005; Vogt et al., 2007). Moreover, activity in the IFG was greater during the observation of action-related context as opposed to context-free actions, suggesting this structure plays a role not only in recognition but also in coding the intentions of others (Iacoboni et al., 2005) and contemplating others' mental states (for meta-analysis see Mar, 2011). Neuroimaging studies have shown involvement of the IFG and IPL during observation of both dynamic and static (Carr et al., 2003) facial stimuli, for example, when comparing dynamic faces to dynamic objects (Fox et al., 2009), dynamic faces to dynamic scrambled faces (Sato et al., 2004; Schultz and Pilz, 2009) and dynamic faces to static faces (Arsalidou et al., 2011; Foley et al., 2012; Rymarczyk et al., 2018). It is interesting that in our study we also found that static compared to neutral images activated IPL and IFG. It is possible that the brain areas involved in the process of motor imagery could be activated also in absence of biological movement, which is typical for emotional but not neutral facial expressions. Accordingly, Kilts et al. (2003) reported that judgment of emotion intensity during perception of both angry and happy static expressions compared to neutral expressions activate motor and premotor cortices. Those authors proposed that during perception of static emotional images "*decoding for emotion content is accomplished by the covert motor simulation of the expression prior to attempts to match the static percept to its dynamic mental representation*" (Kilts et al., 2003, p. 165). To summarize, growing neuroimaging evidence confirms the role of frontal and parietal dorsal streams in the processing of both static (Carr et al., 2003) as well as dynamic emotional stimuli (Sarkheil et al., 2013), also for fear (Schaich Borg et al., 2008) and disgust emotions (Schaich Borg et al., 2008). Since facial emotional expressions are a strong cue in social interactions, it is proposed that natural stimuli (Schultz and Pilz, 2009), especially dynamic ones, may be powerful signals for activating simulation processes within the MNS.

Relationships Between Facial Muscle Reactions and Neural Activity

In our study, we found that activity in several regions correlated with facial reactions. For fear expressions, CS reactions correlated with activation in the right amygdala, right AI and left BA44 for static displays, and in the left pallidus for dynamic ones. A similar

pattern of correlated structures was observed for disgust displays, such that CS reactions correlated with activation in the left AI, left IPL, pallidus, and caudate head bilaterally, for static displays. Moreover, for disgust static displays LL reaction correlated with activation in the left BA44, and left BA45, left AI and bilateral premotor cortex, LL correlations with dynamic displays were primarily observed in the left AI (see **Table 14**).

In almost all conditions (i.e., during perception of fear and disgust as well as static and dynamic stimuli) facial reactions correlated with activity of brain regions related to motor simulation of facial expressions (i.e., IFG and IPL), as discussed above, as well as in the AI. Similar results were obtained in other studies, wherein simultaneous recording of the EMG signal and BOLD response during perception of stimuli was applied (Likowski et al., 2012; Rymarczyk et al., 2018). For example, Likowski et al. (2012) found that ZM reactions to static happy expressions and CS reactions to static angry faces correlated with activations in the right IFG. Moreover, Rymarczyk et al. (2018) observed such correlations mainly for dynamic stimuli. All together, these studies emphasize the role of the IFG and IPL in intentional imitation of emotional expressions and suggest that these regions, that are sensitive to goal-directed actions, may constitute the neuronal correlates of FM [for a review see, Bastiaansen et al., 2009].

The activation of the AI observed in our study during perception of disgust and fear is in line with the results of other studies (Phan et al., 2002). For example, the AI has been shown to respond during experiences of unpleasant odors (Wicker et al., 2003), tastes (Jabbi et al., 2007), and perception of disgust-inducing pictures (Shapira et al., 2003) as well as disgusted faces (Chen et al., 2009). However, the AI seems to be engaged in processing not only negative but also positive emotions, for, e.g., during smile execution (Hennenlotter et al., 2005). Furthermore, most researchers agree that the AI, which is considered to be structure extending MNS, may underlie a simulation of emotional feeling states (van der Gaag et al., 2007; Jabbi and Keysers, 2008). These assumptions correspond with other findings of simultaneous EMG-fMRI studies that show correlations between insula activity with facial reactions during perception of emotional expressions. For example, Likowski et al. (2012) showed that CS muscle reactions to angry faces were associated with the right insula, while Rymarczyk et al. (2018) found such relationships for happiness expressions with ZM and orbicularis oculi responses. It should be noted that, more recently, the AI is considered to be a key brain region involved in the experience of emotions (Menon and Uddin, 2010), among other processes like judgments of trustworthiness or sexual arousal [for a review see (Bud) Craig, 2009].

Next, in our study we found correlations between activity of the amygdala and facial reactions in the CS muscle during perception of fear stimuli. These results are parallel to other findings of neuroimaging studies that revealed activity of the amygdala during observation (Carr et al., 2003) as well as execution of fear and other negative facial expressions (van der Gaag et al., 2007). A number of studies emphasize the role of the amygdala in social-emotional recognition (Adolphs, 2002; Adolphs and Spezio, 2006), and in particular, in the processing

of salient face stimuli during unpredictable circumstances (Adolphs, 2010). Moreover, it has been suggested that the amygdala contributes to relevant stimuli detection (Sander et al., 2003). Therefore, it is possible that, due an increased vigilance in observing the dynamically changing salient features of faces, the processing of dynamic aspects of faces requires amygdala activation.

Furthermore, our EMG-fMRI analysis revealed correlations between activity of the basal ganglia (i.e., globus pallidus and caudate head) and facial reactions for fear and disgust expressions. One interpretation of this result might be that the caudate nucleus and the globus pallidus, which are involved in motor control (Salih et al., 2009), also play a role in motor control during automatic FM. On the other hand, clinical studies (Sprengelmeyer et al., 1996; Calder et al., 2016) and neuroimaging data (Sprengelmeyer et al., 1998) suggest that both the globus pallidus and the caudate nuclei play an important role in processing of disgust expressions. Moreover, the globus pallidus seems to be involved in aversive responses to fear and anxiety (Talalaenko et al., 2008), as well as in affect regulation (Murphy et al., 2003).

Relationships Between Facial Mimicry, Neural Activity and Empathy

A further innovative feature of our study was to test whether empathy traits modulate the neuronal correlates of FM. As discussed above, the high empathy group as compared to the low empathic one presented a distinct pattern of EMG response that is consistent with a typical FM, i.e., greater CS reactions for fear and disgust and greater LL reactions for disgust. What is important to note here is the FM activity in emotion-related brain structures (e.g., AI, amygdala) was more evident in the HE group. Our finding of the anterior insula activity is partially consistent with few neuroimaging studies where disgust stimuli were used (for a review see Baird et al., 2011). For example it was shown that an observation of film clips of people drinking liquids and displaying disgusted faces evoked activity in a neural circuit consisting of the AI, IFG and cingulate cortex, but only in high empathic persons. It seems that activations related to disgust were more frequently observed for high-arousing stimuli, like pictures of painful situations (Jackson et al., 2006) or facial pain expressions (Botvinick et al., 2005; Saarela et al., 2007). However, in our study, we found no differences in brain activity during perception of fear and disgust facial expressions when comparing low and high empathic subjects. This may be the result from different kind of stimuli used in our and other studies. While most studies used the high-arousing stimuli like the pain-inducing situations, our study applied low arousing stimuli. In other words, the perception of emotional facial expressions, compared to perception of pain-inducing situations may be not sufficient to detect brain differences related to low and high empathic characteristics of subjects.

In relation to correlation between facial reaction for fear and disgust stimuli and activity of the amygdala, our result stay in agreement of the assumption, that the amygdala, next to AI, IFG,

and IPL constitute the neuronal structures required for complex empathic processes (Bzdok et al., 2012; Decety et al., 2012; Marsh, 2018). Taken together, it is proposed that activity of the amygdala, together with activity of the insula may constitute the neuronal bases of affective simulation, however, the specificity of role of the amygdala in affective resonance requires further clarification. As noted by Preston and de Waal (2002): "So, if the mirror neurons represent emotional behavior, then the insula may relay information from the premotor mirror neurons to the amygdala" (see Augustine, 1996).

Summary and Conclusion

Our results from study using simultaneously recorded EMG and BOLD signals during perception of fear and disgust have confirmed that, similarly to anger and happiness (Likowski et al., 2012; Rymarczyk et al., 2018), the MNS may constitute the neuronal bases of FM. In particular, the core MNS structures (i.e., IFG and IPL) are thought to be responsible for motor simulation, while MNS-related limbic regions (e.g., AI) seem to be related to affective resonance. In line with this, it is suggested that FM includes both motor and emotional component; however, their mutual relations required further studies. For example, it is possible that motor imitation leads to emotional contagion or vice versa, among other factors, which play an important role in social interactions.

Our study is the first attempt when the relation between facial mimicry, activity of subsystems of the MNS, and level of emotional empathy was explored. We have found that high empathic people demonstrated the stronger facial reactions and what is worth noting, these reactions were correlated with stronger activation of structures of core MNS and MNS-related limbic structures. In other words, it appears that high empathic people imitate emotions of others more than low empathic ones. Additionally, we have shown that the processes of motor imitation and affective contagion were more evident for dynamic, more natural, than static emotional facial expressions.

As far as modality of the stimuli is concerned, our study confirmed the general agreement that exists among researchers that dynamic facial expressions are a valuable source of information in social communication. The evidence was visible in greater neural network activations during dynamic compared to static facial expressions of fear and disgust. Moreover, it appeared that presentation of stimulus dynamics is an important factor for elicitation of emotion, especially for fear.

Limitations

As it was noted in the introduction, the increased activity of CS or LL in response to emotional facial expressions are not distinct to single emotions, i.e., neither for fear nor for disgust. Some studies confirmed increased CS activity during perception of various negative emotions (Murata et al., 2016). Accordingly LL increased activity was found not only for in disgust mimicry but also in pain expression, together with increased activity of CS (Prkachin and Solomon, 2008). Therefore, our inference about brain-muscle relationships are limited due to non-specificity of the CS and LL which are indicators of FM for fear and disgust.

Next, there is some evidence that increased activity of other facial muscle, i.e., the lateral frontalis, could be related to fear expression (Van Boxtel, 2010). In our previous work we showed that fear presentations induced activity in this muscle (Rymarczyk et al., 2016b). However, in the current work we did not measure activity of this muscle because the cap intended for EMG measurements in MRI environment was not designed for that purpose.

AUTHOR CONTRIBUTIONS

KR, KJ-S, and ŁŻ conceived and designed the experiments. KR and ŁŻ performed the experiments, analyzed the data, and contributed materials. KR, ŁŻ, KJ-S, and IS wrote the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00701/full#supplementary-material

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* 12, 169–177. doi: 10.1016/S0959-4388(02)00301-X

Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Ann. N. Y. Acad. Sci.* 1191, 42–61. doi: 10.1111/j.1749-6632.2010.05445.x

Adolphs, R., and Spezio, M. (2006). Role of the amygdala in processing visual social stimuli. *Prog. Brain Res.* 156, 363–378. doi: 10.1016/S0079-6123(06)56020-0

Allen, P. J., Josephs, O., and Turner, R. (2000). A method for removing imaging artifact from continuous EEG recorded during functional MRI. *Neuroimage* 12, 230–239. doi: 10.1006/nimg.2000.0599

Armony, J., and Vuilleumier, P. (eds). (2013). *The Cambridge Handbook of Human Affective Neuroscience.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511843716

Arsalidou, M., Morris, D., and Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain Topogr.* 24, 149–163. doi: 10.1007/s10548-011-0171-4

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res. Rev.* 22, 229–244. doi: 10.1016/S0165-0173(96)00011-2

Baird, A. D., Scheffer, I. E., and Wilson, S. J. (2011). Mirror neuron system involvement in empathy: a critical look at the evidence. *Soc. Neurosci.* 6, 327–335. doi: 10.1080/17470919.2010.547085

Balconi, M., and Canavesio, Y. (2013a). Emotional contagion and trait empathy in prosocial behavior in young people: the contribution of autonomic (facial feedback) and balanced emotional empathy scale (BEES) measures. *J. Clin. Exp. Neuropsychol.* 35, 41–48. doi: 10.1080/13803395.2012.742492

Balconi, M., and Canavesio, Y. (2013b). High-frequency rTMS improves facial mimicry and detection responses in an empathic emotional task. *Neuroscience* 236, 12–20. doi: 10.1016/j.neuroscience.2012.12.059

Balconi, M., and Canavesio, Y. (2014). Is empathy necessary to comprehend the emotional faces? The empathic effect on attentional mechanisms (eye movements), cortical correlates (N200 event-related potentials) and facial behaviour (electromyography) in face processing. *Cogn. Emot.* 30, 210–224. doi: 10.1080/02699931.2014.993306

Balconi, M., and Canavesio, Y. (2016). Empathy, approach attitude, and rTMs on left DLPFC affect emotional face recognition and facial feedback (EMG). *J. Psychophysiol.* 30, 17–28. doi: 10.1027/0269-8803/a000150

Balconi, M., Vanutelli, M. E., and Finocchiaro, R. (2014). Multilevel analysis of facial expressions of emotion and script: self-report (arousal and valence) and psychophysiological correlates. *Behav. Brain Funct.* 10:32. doi: 10.1186/1744-9081-10-32

Bastiaansen, J. A. C. J., Thioux, M., and Keysers, C. (2009). Evidence for mirror systems in emotions. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 364, 2391–2404. doi: 10.1098/rstb.2009.0058

Bernstein, M., and Yovel, G. (2015). Two neural pathways of face processing: a critical evaluation of current models. *Neurosci. Biobehav. Rev.* 55, 536–546. doi: 10.1016/j.neubiorev.2015.06.010

Betti, V., and Aglioti, S. M. (2016). Dynamic construction of the neural networks underpinning empathy for pain. *Neurosci. Biobehav. Rev.* 63, 191–206. doi: 10.1016/j.neubiorev.2016.02.009

Blair, R. J. (1999). Psychophysiological responsiveness to the distress ofothers in children with autism. *Pers. Individ. Dif.* 26, 477–485. doi: 10.1016/S0191-8869(98)00154-8

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., and Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *Neuroimage* 25, 312–319. doi: 10.1016/j.neuroimage.2004.11.043

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation I: defensive and appetitive reactions in picture processing. *Emotion* 1, 276–298. doi: 10.1037//1528-3542.1.3.276

Brett, M., Anton, J. L., Valabregue, R., and Poline, J. B. (2002). "Region of interest analysis using an SPM toolbox," in *Presented at the 8th International Conference on Functional Mapping of the Human Brain*, Sendai.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., et al. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct. Funct.* 217, 783–796. doi: 10.1007/s00429-012-0380-y

Cacioppo, J. T., Petty, R. E., Losch, M. E., and Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *J. Pers. Soc. Psychol.* 50, 260–268. doi: 10.1037//0022-3514.50.2.260

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (2007). "Handbook of psychophysiology," in *Dreaming*, Vol. 44, eds J. T. Cacioppo, L. G. Tassinary, and G. Berntson (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511546396

Calder, A. J., Keane, J., Manes, F., Antoun, N., and Young, A. W. (2016). *Facial Expression Recognition. Facial Expression Recognition: Selected works of Andy Young*. Abingdon: Routledge. doi: 10.4324/9781315715933

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502. doi: 10.1073/pnas.0935845100

Chapman, H. A., Kim, D. A., Susskind, J. M., and Anderson, A. K. (2009). In bad taste: evidence for the oral origins of moral disgust. *Science* 323, 1222–1226. doi: 10.1126/science.1165565

Chartrand, T. L., and Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* 76, 893–910. doi: 10.1037/0022-3514.76.6.893

Chen, Y. H., Dammers, J., Boers, F., Leiberg, S., Edgar, J. C., Roberts, T. P. L., et al. (2009). The temporal dynamics of insula activity to disgust and happy facial expressions: a magnetoencephalography study. *Neuroimage* 47, 1921–1928. doi: 10.1016/j.neuroimage.2009.04.093

Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555

de Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annu. Rev. Psychol.* 59, 279–300. doi: 10.1146/annurev.psych.59.103006.093625

de Waal, F. B. M., and Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nat. Rev. Neurosci.* 18, 498–509. doi: 10.1038/nrn.2017.72

Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71–100. doi: 10.1177/1534582304267187

Decety, J., Norman, G. J., Berntson, G. G., and Cacioppo, J. T. (2012). A neurobehavioral evolutionary perspective on the mechanisms underlying empathy. *Prog. Neurobiol.* 98, 38–48. doi: 10.1016/j.pneurobio.2012.05.001

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology* 19, 643–647. doi: 10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U., Andréasson, P., and Thunberg, M. (2011). Emotional empathy and facial reactions to facial expressions. *J. Psychophysiol.* 25, 26–31. doi: 10.1027/0269-8803/a000029

Dimberg, U., and Petterson, M. (2000). Facial reactions to happy and angry facial expressions: evidence for right hemisphere dominance. *Psychophysiology* 37, 693–696. doi: 10.1111/1469-8986.3750693

Dimberg, U., Thunberg, M., and Grunedal, S. (2002). Facial reactions to emotional stimuli: automatically controlled emotional responses. *Cogn. Emot.* 16, 449–471. doi: 10.1080/02699930143000356

Eickhoff, S. (2016). *SPM Anatomy Toolbox*. Berlin: Springer.

Foley, E., Rippon, G., Thai, N. J., Longe, O., and Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. *J. Cogn. Neurosci.* 24, 507–520. doi: 10.1162/jocn_a_00120

Fox, C. J., Iaria, G., and Barton, J. J. S. (2009). Defining the face processing network: optimization of the functional localizer in fMRI. *Hum. Brain Mapp.* 30, 1637–1651. doi: 10.1002/hbm.20630

Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x

Furl, N., Henson, R. N., Friston, K. J., and Calder, A. J. (2015). Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. *Cereb. Cortex* 25, 2876–2882. doi: 10.1093/cercor/bhu083

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119(Pt 2), 593–609. doi: 10.1093/brain/119.2.593

Gallese, V., Rochat, M., Cossu, G., and Sinigaglia, C. (2009). Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Dev. Psychol.* 45, 103–113. doi: 10.1037/a0014436

Gazzola, V., Aziz-Zadeh, L., and Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Curr. Biol.* 16, 1824–1829. doi: 10.1016/j.cub.2006.07.072

Grosbras, M.-H., and Paus, T. (2006). Brain networks involved in viewing angry hands or faces. *Cereb. Cortex* 16, 1087–1096. doi: 10.1093/cercor/bhj050

Hall, D. A., Fussell, C., and Summerfield, A. Q. (2005). Reading fluent speech from talking faces: typical brain networks and individual differences. *J. Cogn. Neurosci.* 17, 939–953. doi: 10.1162/0898929054021175

Hastings, P. D., Zahn-Waxler, C., Robinson, J., Usher, B., and Bridges, D. (2000). The development of concern for others in children with behavior problems. *Dev. Psychol.* 36, 531–546. doi: 10.1037/0012-1649.36.5.531

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1992). "Primitive emotional contagion," in *Emotion and Social Behavior*, ed. M. S. Clark (Thousand Oaks, CA: Sage Publications), 151–177.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0

Hennenlotter, A., Schroeder, U., Erhard, P., Castrop, F., Haslinger, B., Stoecker, D., et al. (2005). A common neural basis for receptive and expressive communication of pleasant facial affect. *Neuroimage* 26, 581–591. doi: 10.1016/j.neuroimage.2005.01.057

Hess, U., and Fischer, A. (2013). Emotional mimicry as social regulation. *Pers. Soc. Psychol. Rev.* 17, 142–157. doi: 10.1177/1088868312472607

Hess, U., and Fischer, A. (2014). Emotional mimicry: why and when we mimic emotions. *Soc. Pers. Psychol. Compass* 8, 45–57. doi: 10.1111/spc3.12083

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., and D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Soc. Cogn. Affect. Neurosci.* 3, 204–217. doi: 10.1093/scan/nsn019

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.* 60, 653–670. doi: 10.1146/annurev.psych.60.110707.163604

Iacoboni, M., and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* 7, 942–951. doi: 10.1038/nrn2024

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol.* 3:e79. doi: 10.1371/journal.pbio.0030079

Jabbi, M., and Keysers, C. (2008). Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion* 8, 775–780. doi: 10.1037/a0014194

Jabbi, M., Swart, M., and Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage* 34, 1744–1753. doi: 10.1016/j.neuroimage.2006.10.032

Jackson, P. L., Brunet, E., Meltzoff, A. N., and Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia* 44, 752–761. doi: 10.1016/j.neuropsychologia.2005.07.015

Jankowiak-Siuda, K., Rymarczyk, K., and Grabowska, A. (2011). How we empathize with others: a neurobiological perspective. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* 17, RA18–RA24. doi: 10.12659/MSM.881324

Kaplan, J. T., and Iacoboni, M. (2006). Getting a grip on other minds: mirror neurons, intention understanding, and cognitive empathy. *Soc. Neurosci.* 1, 175–183. doi: 10.1080/17470910600985605

Kessler, H., Doyen-Waldecker, C., Hofer, C., Hoffmann, H., Traue, H. C., and Abler, B. (2011). Neural correlates of the perception of dynamic versus static facial expressions of emotion. *Psychosoc. Med.* 8:Doc03. doi: 10.3205/psm000072

Keysers, C., and Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Prog. Brain Res.* 156, 379–401. doi: 10.1016/S0079-6123(06)56021-2

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., and Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *Neuroimage* 18, 156–168. doi: 10.1006/nimg.2002.1323

Krebs, D. (1975). Empathy and altruism. *J. Pers. Soc. Psychol.* 32, 1134–1146. doi: 10.1037//0022-3514.32.6.1134

Kret, M. E., Fischer, A. H., and De Dreu, C. K. W. (2015). Pupil mimicry correlates with trust in in-group partners with dilating pupils. *Psychol. Sci.* 26, 1401–1410. doi: 10.1177/0956797615588306

Krumhuber, E. G., Kappas, A., and Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: a review. *Emot. Rev.* 5, 41–46. doi: 10.1177/1754073912451349

Krusemark, E. A., and Li, W. (2011). Do all threats work the same way? Divergent effects of fear and disgust on sensory perception and attention. *J. Neurosci.* 31, 3429–3434. doi: 10.1523/JNEUROSCI.4394-10.2011

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., and McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cereb. Cortex* 13, 1023–1033. doi: 10.1093/cercor/13.10.1023

Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* 40, 776–785. doi: 10.1111/1469-8986.00078

Lee, H., Heller, A. S., van Reekum, C. M., Nelson, B., and Davidson, R. J. (2012). Amygdala–prefrontal coupling underlies individual differences in emotion regulation. *Neuroimage* 62, 1575–1581. doi: 10.1016/j.neuroimage.2012.05.044

Lee, T. W., Josephs, O., Dolan, R. J., and Critchley, H. D. (2006). Imitating expressions: emotion-specific neural substrates in facial mimicry. *Soc. Cogn. Affect. Neurosci.* 1, 122–135. doi: 10.1093/scan/nsl012

Levenson, R. W., and Ruef, A. M. (1992). Empathy: a physiological substrate. *J. Pers. Soc. Psychol.* 63, 234–246. doi: 10.1037/0022-3514.63.2.234

Liberzon, I., Phan, K. L., Decker, L. R., and Taylor, S. F. (2003). Extended amygdala and emotional salience: a PET activation study of positive and negative affect. *Neuropsychopharmacology* 28, 726–733. doi: 10.1038/sj.npp.1300113

Likowski, K. U., Mühlberger, A., Gerdes, A. B. M., Wieser, M. J., Pauli, P., and Weyers, P. (2012). Facial mimicry and the mirror neuron system:

simultaneous acquisition of facial electromyography and functional magnetic resonance imaging. *Front. Hum. Neurosci.* 6:214. doi: 10.3389/fnhum.2012.00214

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., and Weyers, P. (2008). Modulation of facial mimicry by attitudes. *J. Exp. Soc. Psychol.* 44, 1065–1072. doi: 10.1016/j.jesp.2007.10.007

Likowski, K. U., Mühlberger, A., Seibt, B., Pauli, P., and Weyers, P. (2011). Processes underlying congruent and incongruent facial reactions to emotional facial expressions. *Emotion* 11, 457–467. doi: 10.1037/a0023162

Lundquist, L.-O., and Dimberg, U. (1995). Facial expressions are contagious. *J. Psychophysiol.* 9, 203–211.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annu. Rev. Psychol.* 62, 103–134. doi: 10.1146/annurev-psych-120709-145406

Marsh, A. A. (2018). The neuroscience of empathy. *Curr. Opin. Behav. Sci.* 19, 110–115. doi: 10.1016/j.cobeha.2017.12.016

McIntosh, D. N. (2006). Spontaneous facial mimicry, liking and emotional contagion. *Pol. Psychol. Bull.* 37, 31–42. doi: 10.3389/fpsyg.2016.00458

Mehrabian, A., and Epstein, N. (1972). A measure of emotional empathy. *J. Pers.* 40, 525–543. doi: 10.1111/j.1467-6494.1972.tb00078.x

Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0

Molnar-Szakacs, I., Iacoboni, M., Koski, L., and Mazziotta, J. C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cereb. Cortex* 15, 986–994. doi: 10.1093/cercor/bhh199

Montgomery, K. J., and Haxby, J. V. (2008). Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 20, 1866–1877. doi: 10.1162/jocn.2008.20127

Murata, A., Saito, H., Schug, J., Ogawa, K., and Kameda, T. (2016). Spontaneous facial mimicry is enhanced by the goal of inferring emotional states: evidence for moderation of "automatic" mimicry by higher cognitive processes. *PLoS One* 11:e0153128. doi: 10.1371/journal.pone.0153128

Murphy, F. C., Nimmo-Smith, I., and Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 207–233. doi: 10.3758/CABN.3.3.207

Neta, M., Norris, C. J., and Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion* 9, 640–648. doi: 10.1037/a0016819

Ohrmann, P., Rauch, A. V., Bauer, J., Kugel, H., Arolt, V., Heindel, W., et al. (2007). Threat sensitivity as assessed by automatic amygdala response to fearful faces predicts speed of visual search for facial expression. *Exp. Brain Res.* 183, 51–59. doi: 10.1007/s00221-007-1022-0

Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., and Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. *Neuroimage* 39, 2076–2085. doi: 10.1016/j.neuroimage.2007.10.032

Phan, K. L., Wager, T., Taylor, S. F., and Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16, 331–348. doi: 10.1006/nimg.2002.1087

Pohl, A., Anders, S., Schulte-Rüther, M., Mathiak, K., and Kircher, T. (2013). Positive facial affect – an fMRI study on the involvement of insula and amygdala. *PLoS One* 8:e69886. doi: 10.1371/journal.pone.0069886

Preston, S. D. (2007). "A perception-action model for empathy," in *Empathy in Mental Illness*, eds T. F. D. Farrow and P. W. R. Woodruff (Cambridge: Cambridge University Press), 428–447. doi: 10.1017/CBO9780511543753.024

Preston, S. D., and de Waal, F. B. M. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20; discussion 20–71.

Prkachin, K. M., and Solomon, P. E. (2008). The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. *Pain* 139, 267–274. doi: 10.1016/j.pain.2008.04.010

Rembowski, J. (1989). *Empatia - Studium Psychologiczne*. Warsaw: Państwowe Wydawnictwo Naukowe.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Rizzolatti, G., and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* 11, 264–274. doi: 10.1038/nrn2805

Robins, D. L., Hunyadi, E., and Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain Cogn.* 69, 269–278. doi: 10.1016/j.bandc.2008.08.007

Ruby, P., and Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550. doi: 10.1038/87510

Rymarczyk, K., Biele, C., Grabowska, A., and Majczynski, H. (2011). EMG activity in response to static and dynamic facial expressions. *Int. J. Psychophysiol.* 79, 330–333. doi: 10.1016/j.ijpsycho.2010.11.001

Rymarczyk, K., Żurawski, Ł., Jankowiak-Siuda, K., and Szatkowska, I. (2016a). Do dynamic compared to static facial expressions of happiness and anger reveal enhanced facial mimicry? *PLoS One* 11:e0158534. doi: 10.1371/journal.pone.0158534

Rymarczyk, K., Żurawski, Ł., Jankowiak-Siuda, K., and Szatkowska, I. (2016b). Emotional empathy and facial mimicry for static and dynamic facial expressions of fear and disgust. *Front. Psychol.* 7:1853. doi: 10.3389/fpsyg.2016.01853

Rymarczyk, K., Żurawski, Ł., Jankowiak-Siuda, K., and Szatkowska, I. (2018). Neural correlates of facial mimicry: simultaneous measurements of EMG and BOLD responses during perception of dynamic compared to static facial expressions. *Front. Psychol.* 9:52. doi: 10.3389/fpsyg.2018.00052

Saarela, M. V., Hlushchuk, Y., Williams, A. C. D. C., Schürmann, M., Kalso, E., and Hari, R. (2007). The compassionate brain: humans detect intensity of pain from another's face. *Cereb. Cortex* 17, 230–237. doi: 10.1093/cercor/bhj141

Salih, F., Sharott, A., Khatami, R., Trottenberg, T., Schneider, G., Kupsch, A., et al. (2009). Functional connectivity between motor cortex and globus pallidus in human non-REM sleep. *J. Physiol.* 587, 1071–1086. doi: 10.1113/jphysiol.2008.164327

Sander, D., Grafman, J., and Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. *Rev. Neurosci.* 14, 303–316. doi: 10.1515/REVNEURO.2003.14.4.303

Sarkheil, P., Goebe, R., Schneider, F., and Mathiak, K. (2013). Emotion unfolded by motion: a role for parietal lobe in decoding dynamic facial expressions. *Soc. Cogn. Affect. Neurosci.* 8, 950–957. doi: 10.1093/scan/nss092

Sato, W., Fujimura, T., and Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *Int. J. Psychophysiol.* 70, 70–74. doi: 10.1016/j.ijpsycho.2008.06.001

Sato, W., Kochiyama, T., and Uono, S. (2015). Spatiotemporal neural network dynamics for the processing of dynamic facial expressions. *Sci. Rep.* 5:12432. doi: 10.1038/srep12432

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., and Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cogn. Brain Res.* 20, 81–91. doi: 10.1016/j.cogbrainres.2004.01.008

Schaich Borg, J., Lieberman, D., and Kiehl, K. A. (2008). Infection, incest, and iniquity: investigating the neural correlates of disgust and morality. *J. Cogn. Neurosci.* 20, 1529–1546. doi: 10.1162/jocn.2008.20109

Schultz, J., and Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Exp. Brain Res.* 194, 465–475. doi: 10.1007/s00221-009-1721-9

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007

Seibt, B., Mühlberger, A., Likowski, K. U., and Weyers, P. (2015). Facial mimicry in its social setting. *Front. Psychol.* 6:1122. doi: 10.3389/fpsyg.2015.01122

Seubert, J., Kellermann, T., Loughead, J., Boers, F., Brensinger, C., Schneider, F., et al. (2010). Processing of disgusted faces is facilitated by odor primes: a functional MRI study. *Neuroimage* 53, 746–756. doi: 10.1016/j.neuroimage.2010.07.012

Shapira, N. A., Liu, Y., He, A. G., Bradley, M. M., Lessig, M. C., James, G. A., et al. (2003). Brain activation by disgust-inducing pictures in obsessive-compulsive disorder. *Biol. Psychiatry* 54, 751–756. doi: 10.1016/S0006-3223(03)00003-9

Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scand. J. Psychol.* 43, 433–443. doi: 10.1111/1467-9450.00312

Sonnby-Borgström, M., Jönsson, P., and Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *J. Nonverbal Behav.* 27, 3–23. doi: 10.1023/A:1023608506243

Sprengelmeyer, R., Rausch, M., Eysel, U. T., and Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* 265, 1927–1931. doi: 10.1098/rspb.1998.0522

Sprengelmeyer, R., Young, A., Calder, A., Karnat, A., Lange, H., Homberg, V., et al. (1996). Loss of Disgust. Perception of faces and emotions in Huntington's disease. *Brain* 119(Pt 5), 1647–1665. doi: 10.1093/brain/119.5.1647

Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., and Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nat. Neurosci.* 11, 843–850. doi: 10.1038/nn.2138

Talalaenko, A. N., Krivobok, G. K., Bulgakova, N. P., and Pankrat'ev, D. V. (2008). Functional roles of the monoaminergic and aminoacidergic mechanisms of the dorsal pallidum in anxiety of different aversive origins. *Neurosci. Behav. Physiol.* 38, 115–118. doi: 10.1007/s11055-008-0016-0

The Mathworks Inc (2013). *Matlab 2013b*. Natick, MA: The Mathworks Inc.

Topolinski, S., and Strack, F. (2015). Corrugator activity confirms immediate negative affect in surprise. *Front. Psychol.* 6:134. doi: 10.3389/fpsyg.2015.00134

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., and Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11656–11661. doi: 10.1073/pnas.191374198

Van Boxtel, A. (2010). "Facial EMG as a tool for inferring affective states," in *Proceedings of the Measuring Behavior*, eds A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, and P. H. Zimmerman (Eindhoven: Eindhoven University of Technology).

van der Gaag, C., Minderaa, R. B., and Keysers, C. (2007). Facial expressions: what the mirror neuron system can and cannot tell us. *Soc. Neurosci.* 2, 179–222. doi: 10.1080/17470910701376878

van der Schalk, J., Fischer, A., Doosje, B., Wigboldus, D., Hawk, S., Rotteveel, M., et al. (2011a). Convergent and divergent responses to emotional displays of ingroup and outgroup. *Emotion* 11, 286–298. doi: 10.1037/a0022582

van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011b). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion* 11, 907–920. doi: 10.1037/a0023853

van der Zwaag, W., Da Costa, S. E., Zürcher, N. R., Adams, R. B., and Hadjikhani, N. (2012). A 7 Tesla fMRI study of amygdala responses to fearful faces. *Brain Topogr.* 25, 125–128. doi: 10.1007/s10548-012-0219-0

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* 30, 829–858. doi: 10.1002/hbm.20547

Vogt, S., Buccino, G., Wohlschläger, A. M., Canessa, N., Shah, N. J., Zilles, K., et al. (2007). Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. *Neuroimage* 37, 1371–1383. doi: 10.1016/j.neuroimage.2007.07.005

Vrana, S. R. (1993). The psychophysiology of disgust: differentiating negative emotional contexts with facial EMG. *Psychophysiology* 30, 279–286. doi: 10.1111/j.1469-8986.1993.tb03354.x

Wake Forest University (2014). *WFU PickAtlas 3.0.3*. New York, NY: School of Medicine.

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., et al. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* 69, 131–140. doi: 10.1016/j.cortex.2015.05.003

Weyers, P., Muhlberger, A., Hefele, C., and Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology* 43, 450–453. doi: 10.1111/j.1469-8986.2006.00451.x

Whitton, A. E., Henry, J. D., Rendell, P. G., and Grisham, J. R. (2014). Disgust, but not anger provocation, enhances levator labii superioris activity during exposure to moral transgressions. *Biol. Psychol.* 96, 48–56. doi: 10.1016/j.biopsycho.2013.11.012

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of

seeing and feeling disgust. *Neuron* 40, 655–664. doi: 10.1016/S0896-6273(03)00679-2

Yartz, A. R., and Hawk, L. W. (2002). Addressing the specificity of affective startle modulation: fear versus disgust. *Biol. Psychol.* 59, 55–68. doi: 10.1016/S0301-0511(01)00121-1

Zheng, J., Anderson, K. L., Leal, S. L., Shestyuk, A., Gulsen, G., Mnatsakanyan, L., et al. (2017). Amygdala-hippocampal dynamics during salient information processing. *Nat. Commun.* 8:14413. doi: 10.1038/ncomms14413

Zinchenko, O., Yaple, Z. A., and Arsalidou, M. (2018). Brain responses to dynamic facial expressions: a normative meta-analysis. *Front. Hum. Neurosci.* 12:227. doi: 10.3389/fnhum.2018.00227

Check for updates

Strategy Shift Toward Lower Spatial Frequencies in the Recognition of Dynamic Facial Expressions of Basic Emotions: When It Moves It Is Different

Marie-Pier Plouffe-Demers[1,2], Daniel Fiset[1], Camille Saumure[1], Justin Duncan[1,2] and Caroline Blais[1]*

[1]Département de Psychologie, Universtité du Québec en Outaouais, Gatineau, QC, Canada, [2]Département de Psychologie, Université du Québec à Montréal, Montreal, QC, Canada

Facial expressions of emotion play a key role in social interactions. While in everyday life, their dynamic and transient nature calls for a fast processing of the visual information they contain, a majority of studies investigating the visual processes underlying their recognition have focused on their static display. The present study aimed to gain a better understanding of these processes while using more ecological dynamic facial expressions. In two experiments, we directly compared the spatial frequency (SF) tuning during the recognition of static and dynamic facial expressions. Experiment 1 revealed a shift toward lower SFs for dynamic expressions in comparison to static ones. Experiment 2 was designed to verify if changes in SF tuning curves were specific to the presence of emotional information in motion by comparing the SF tuning profiles for static, dynamic, and shuffled dynamic expressions. Results showed a similar shift toward lower SFs for shuffled expressions, suggesting that the difference found between dynamic and static expressions might not be linked to informative motion *per se* but to the presence of motion regardless its nature.

Keywords: facial expressions, basic emotion, perceptual strategy, spatial frequency tuning, dynamic advantage

INTRODUCTION

In social settings, the human face represents one of the richest nonverbal sources of information. It is thus an essential skill for humans to continually monitor the facial expressions of others in order to appropriately tailor their behavior throughout social interactions. The ability to accurately extract emotional information plays a major role in prosociality (Marsh et al., 2007), and this capacity is often found to be altered in numerous psychiatric conditions characterized by impaired social functioning, such as schizophrenia (Mandal et al., 1998; Edwards et al., 2002; Lee et al., 2010; Clark et al., 2013; Kring and Elis, 2013) and autism spectrum disorder (Baron-Cohen and Wheelwright, 2004; Harms et al., 2010).

Until recently, a majority of studies investigating the visual processes underlying facial emotion recognition have relied on static pictures displaying facial emotions at their apex (i.e., highest intensity). However, facial emotions are dynamic and transient by nature; thus, the visual

information necessary to recognize a facial expression in everyday life must be extracted quickly. The present study was aimed at gaining a better understanding of this process by investigating the mechanisms subtending this important endeavor, using more ecological dynamic facial expressions. More specifically, we were interested in utilization of spatial frequencies (SF), considered the "atom" upon which primary visual cortex neurons base their world representation (DeValois and DeValois, 1990), during recognition of static and dynamic facial expressions. Simply put, lower SFs code coarser visual information, such as global face shape or facial feature location, while higher SFs code finer visual information, such as facial feature shape or details like wrinkles.

Behavioral, neuroimaging, and lesion data suggest that static and dynamic facial expressions rely on partially nonoverlapping perceptual mechanisms. For instance, dynamic expressions are associated with enhanced onlooker facial muscular reactions (Weyers et al., 2006; Rymarczyk et al., 2011), and they are also better recognized than static expressions (Wehrle et al., 2000; Kamachi et al., 2001; Ambadar et al., 2005; Bould and Morris, 2008; Hammal et al., 2009; Cunningham and Wallraven, 2009a; Chiller-Glaus et al., 2011; Recio et al., 2011; see however Kätsyri and Sams, 2008; Fiorentini and Viviani, 2011; Gold et al., 2013; Widen and Russell, 2015). In addition, neuroimaging studies have shown that dynamic expressions, compared to static ones, lead to a greater activation of many structures involved in facial emotion processing (Kilts et al., 2003; LaBar et al., 2003; Sato and Yoshikawa, 2004; Schultz and Pilz, 2009; Trautmann et al., 2009; Recio et al., 2011). Crucially, dynamic expressions engage areas of the magnocellular-dorsal pathway to a greater extent than static ones (e.g., area MT; Schultz and Pilz, 2009). This parallels the findings from studies performed on patients with ventral visual stream lesions, whom exhibit dramatically impaired recognition of static emotions (Adolphs et al., 1994; Humphreys et al., 2007; Fiset et al., 2017), but a relatively preserved ability to recognize dynamic emotions (Humphreys et al., 1993; Adolphs et al., 2003; Richoz et al., 2015).

Interestingly, the magnocellular-dorsal pathway is associated with processing of motion and shows a higher sensitivity to lower SFs (Livingstone and Hubel, 1988), which might explain these various findings pertaining to dynamic emotion recognition. In contrast, the parvocellular-ventral pathway, which encompasses most of the areas involved in static face processing, is associated with processing of typically higher SF information (Livingstone and Hubel, 1988). Seeing as static and dynamic emotion recognition may rely on partially nonoverlapping cortical structures, one might expect this to be reflected in different visual information extraction strategies, namely a reliance on lower SFs during the processing of dynamic expressions compared to static ones.

Previous work by our team and others also feeds this hypothesis according to which dynamic facial emotion recognition might rely on comparatively lower SFs – though this prediction has not been explicitly tested. Indeed, although diagnostic (i.e., relevant) facial features are mostly the same for static and dynamic expressions (namely, the eyes and mouth), eye fixation patterns underlying the extraction of these features differ. Specifically, participants spend more time directly fixating

diagnostic features for static expressions, whereas they spend more time fixating the center of the face (i.e., nose) for dynamic expressions (Buchan et al., 2007; Blais et al., 2012, 2017; see however, for videos of longer duration, Calvo et al., 2018). Seeing as diagnostic features will be processed in parafoveal vision for dynamic expressions viewed at a conversational distance (i.e., face span of approx. 6–14°; Yang et al., 2014) and that sensitivity to high SFs monotonically decreases with foveal eccentricity (Hilz and Cavonius, 1974), viewing dynamic (vs. static) expressions is likely to induce a shift away from higher SFs and toward lower SFs.

The finding of different patterns of eye fixations for static and dynamic expressions also begs the question of what the underlying cause might be for such an outcome. One possibility is that dynamic expressions convey additional information through motion, thereby reducing the need to extract precise feature representations coded in higher SFs – which requires foveal processing, and thus, direct fixation. A role for motion has been supported by computational studies showing that information it conveys drastically increases performance of artificial vision systems (e.g., Jiang et al., 2011, 2014). The fact that human performance during dynamic facial emotion recognition is resistant to spatial information degradation (e.g., texture and shape) as long as motion contained within expressions is preserved (e.g., exhibited by point-light displays; Cunningham and Wallraven, 2009b), and that performance is reduced when the emotion unfolding sequence (i.e., video frame order) is shuffled or reversed (Cunningham and Wallraven, 2009a), is also a strong argument in favor of motion conveying crucial information for emotion recognition.

Although many studies have supported the importance of motion for expression processing, it is possible that the different patterns of eye fixations observed for static and dynamic expressions are not necessarily for the purpose of using emotion information that is conveyed by motion. Another possibility is instead that the mere presence of motion could activate mechanisms aimed at processing it, regardless of the emotion information it may or may not convey. Such mechanisms may involve changes in eye fixation patterns, since retinal periphery is more efficient at processing temporal variations and motion (Takeuchi et al., 2004; Thompson et al., 2007; Gurnsey et al., 2008).

In other words, fixating dynamic emotional faces in their center may serve the purpose of optimizing the processing of emotion information conveyed through motion by projecting this content in parafoveal regions of the retina. Or, the change in eye fixation pattern may instead be reflexive and caused by the mere presence of motion – irrespective of the information it might convey. In turn, the SF shift hypothesized above could very well be a consequence of fixation optimization for motion processing.

The objective of the present study was twofold. First, we wished to verify the hypothesis according to which the recognition of dynamic and static facial expressions relies on partially nonoverlapping SFs by comparing tuning profiles for both types of expressions (Experiments 1, 2). Second, we wanted to verify if changes in SF tuning curves are specific to the presence of informative motion by comparing the SF tuning

profiles for static, dynamic, and shuffled dynamic expressions (Experiment 2).

EXPERIMENT 1

The SF Bubbles method (Willenbockel et al., 2010a, 2012, 2013; Thurman and Grossman, 2011; Tadros et al., 2013; Royer et al., 2017) was used in order to compare SF utilization in two different facial emotion recognition conditions: static and dynamic expressions. Although filtering faces may create stimuli that differ from what observers consciously perceive in everyday life, it directly manipulates the visual information considered as the atom of visual perception according to the dominant theory in the field of vision (DeValois and DeValois, 1990).

The SF Bubbles method consists in creating, trial-by-trial, random SF filters that are applied to an image – here, one depicting a facial expression. Participant accuracy with each filtered image is then used to infer which SF increases the likelihood of a correct answer (see *Stimuli* section for more details). This method presents important advantages in comparison with the fixed low-pass and high-pass filters that are frequently used to tackle the SF processing during facial emotion recognition (e.g., Vuilleumier et al., 2003). First, instead of simply comparing performance with low vs. high SFs, it allows to measure the complete SF tuning curve of participants. This is particularly important for tasks involving face processing, since it has been shown that sensitivity peaks at SFs between 8 and 16 cycles per face (Näsänen, 1999; Gaspar et al., 2008). Removing those frequencies from the stimuli, as is often done with low-pass and high-pass filter, may thus tap into visual mechanisms that are not specialized for face processing. Relatedly to this last point, a second important advantage of the SF Bubbles method is that, contrary to fixed filters, it does not require an (often arbitrary) decision on where the cutoffs should be applied for the low-pass and high-pass filters; in other words, what SFs should be included in the low-pass (or high-pass filters). Such decision may have a huge impact on the results. SF Bubbles make no *a priori* decision regarding such cutoffs; it simply randomly samples all of the SFs contained in a stimulus and measure performance with all of these random filters.

Materials and Methods

Participants

Twenty participants (4 males; 22.8 years old on average; SD = 3.24) took part in Experiment 1. The number of participants was chosen based on previous experiments using similar methods (Willenbockel et al., 2010a; Royer et al., 2017; Tardif et al., 2017). Because the method relies on random sampling of visual information, a high number of trials are required to obtain a reasonable signal-to-noise ratio. Studies using SF Bubbles have typically relied on a high total number of trials (i.e., across participants) ranging between 10,800 (Tadros et al., 2013) and 34,500 trials (Estéphan et al., 2018) per condition (see also Tardif et al., 2017, 33,000 trials and Royer et al., 2017, 19,200 trials). The present experiment contained a total of 39,200 trials per condition thus having enough trials to

obtain very stable SF tuning for each condition. All participants had normal or corrected-to-normal visual acuity and were naïve to the purpose of the experiment.

Stimuli

The stimuli consisted of videos and photos of 10 actors (5 males) expressing the six basic emotions (i.e., anger, disgust, fear, joy, sadness, surprise; Ekman and Friesen, 1975) as well as neutrality. Stimuli were taken from the STOÏC database (Roy et al., 2007). Videos had a duration of 450 ms and were composed of 15 frames with a duration of 30 ms each. They started with a neutral facial expression and ended at the apex of the expression. Photo stimuli were generated by extracting the last frame from the videos (i.e., the apex). Static and dynamic stimuli were spatially aligned on the main internal features (eyes, nose, mouth) across facial expressions and across actors using linear manipulations such as translation, rotation, and scaling. Additionally, dynamic stimuli were temporally aligned. Faces were cropped to exclude non-facial cues, and they were equated on mean luminance using the SHINE toolbox (Willenbockel et al., 2010b).

On each trial, a stimulus was generated by randomly sampling the SFs of the photo or the frames of the video using the SF Bubbles technique (Willenbockel et al., 2010a). This technique involves the following steps, also depicted in **Figure 1**. First and foremost, in order to reduce edge artifacts, the stimulus is padded with a uniform gray background (**Figure 1A**). A fast Fourier transform is then applied to the padded stimulus (**Figure 1B**), resulting in the base image amplitude spectrum to which a random SF filter is later applied. This filter is created by first generating a random binary vector of X ones among 10,240 zeros, where X is the number of bubbles (**Figure 1C**). This vector is then convolved with a Gaussian kernel with a standard deviation of 1.5 cycles per image (**Figure 1D**). The smoothed sampling vector (**Figure 1E**) is then log-transformed in order to fit the human contrast sensitivity function (**Figure 1F**; see DeValois and DeValois, 1990). The resulting vector is used to generate a two-dimensional isotropic SF filter (**Figure 1G**) by rotating it 360° on its origin. A pointwise multiplication is performed between the base image amplitude spectrum and the SF filter (**Figure 1H**). The result is then back-transformed into the image domain by submitting it to an inverse fast Fourier transform (**Figure 1I**) and cropped to its original size (**Figure 1J**). The resulting "SF bubblized" image contains a random subset of the base image's SF content. Note that with videos, the same filter was applied to all the frames within a trial. Examples of stimuli are presented in **Figure 2**.

Apparatus

The faces in all pictures and videos were presented within a square subtending 256 × 256 pixels and were displayed on a calibrated LCD monitor (51 × 28.5 cm; resolution of 1,920 × 1,080) with a refresh rate of 100 Hz. All participants were asked to place their head on a chin rest at a viewing distance of 38 cm; face width (about 176 pixels) subtended ≈7° of visual angle. The experimental program was written in Matlab (MathWorks, 2012), using functions from the

FIGURE 1 | Example of the creation of one stimulus with the SF Bubbles method. **(A)** Padded stimulus. **(B)** Fast Fourrier transformed base image amplitude spectrum. **(C)** Random binary vector. **(D)** Spatial frequency Bubble. **(E)** Smoothed sampling vector. **(F)** Log-transformed sampling vector. **(G)** Two-dimensional isotropic spatial frequency filter. **(H)** Pointwise multiplication of the Fast Fourrier transformed base image amplitude spectrum and the spatial frequency filter. **(I)** Filtered stimulus. **(J)** Final cropped stimulus. Written informed consent was obtained from the individual for the publication of this image.

Psychophysics Toolbox (Brainard and Vision, 1997; Pelli, 1997; Kleiner et al., 2007).

Procedure

Each participant completed 14 blocks of 140 trials per condition (i.e., Static and Dynamic), for a total of 3,920 trials. The experiment took on average 4 h per participant that was divided into two sessions taking place on separate days. During each session, the


FIGURE 2 | Example of the sequence of events in the Static **(A)**, Dynamic **(B)**, and Shuffled **(C)** conditions. Note that only three frames out of 15 are represented in the Dynamic and Shuffled conditions.
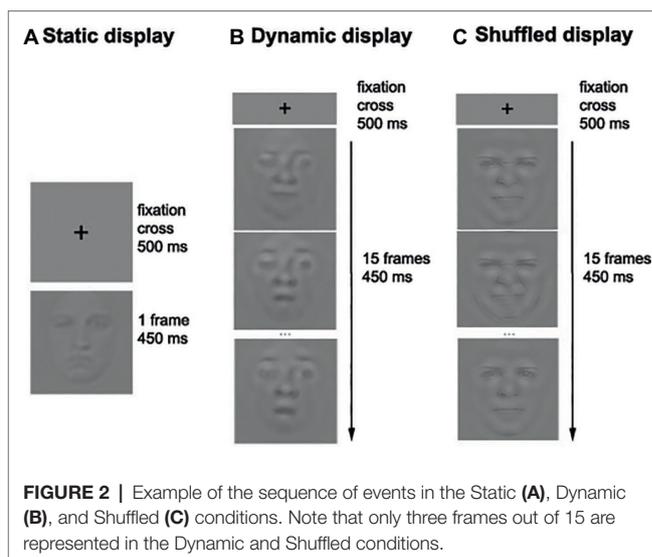
participants were encouraged to take breaks whenever they felt some fatigue. On each trial, a fixation cross was first displayed in the middle of the screen for 500 ms, followed by the stimulus (picture or video) for a duration of 450 ms. A uniform gray background was then displayed until the participant's response. Participants were asked to categorize the emotion displayed by static or dynamic facial expressions by pressing the button associated with each of the six basic emotions as well as neutrality (e.g., "A" for anger, "D" for disgust, "F" for fear, etc.). **Figure 2** shows the sequence of events within one trial.

All participants started with a block containing dynamic expressions and alternated between conditions thereafter. This order was kept for all participants for a specific reason. When using SF Bubbles method, the number of bubbles is manipulated with the objective of maintaining the performance between ceiling and floor. In fact, the analysis procedure allows to infer the SF utilization by comparing the SFs that were available in the stimuli on correct and incorrect trials – hence, it is imperative that a significant number of mistakes is made. In the present experiment, we decided to use the same number of bubbles with dynamic and static expressions in order to ensure that any difference found in SF tuning could not be attributable to a between-condition difference in the number of sampled SFs on each trial. We also decided to adjust the number of bubbles based on the average accuracy with dynamic expressions to minimize the likelihood of a ceiling effect, as previous studies have revealed better performance with these vs. static ones. Thus, for each participant, the number of bubbles was adjusted on a trial basis with QUEST (Watson and Pelli, 1983), but only during the blocks that contained dynamic expressions. The target average accuracy was set to 70%. The number of bubbles used on a given Static block was set to the last output of QUEST in the immediately preceding Dynamic block.

The protocol of this experiment was approved by the Research Ethics Committee of Université du Québec en Outaouais and was conducted in accordance with the Code of Ethics of the

World Medical Association (Declaration of Helsinki). All participants provided informed written consent.

Results

Accuracy

An average of 14.4 (SD = 13.5) bubbles was necessary to maintain an approximate accuracy of 70% during the recognition of dynamic expressions. The number of bubbles reflects the quantity of SF information (and, as a result, the total amount of energy contained in the stimulus) needed by the participants.

An average accuracy of 62.6% (SD = 4.8%) and 68.1% (SD = 5.5%) was found in the Static and Dynamic conditions, respectively. The average accuracy with each emotion in each condition is displayed in **Figure 3**. A 7 (Emotions) × 2 (Conditions) repeated-measure ANOVA was conducted on accuracy. The results indicated significant main effects of the factors of Emotion [$F(1, 19) = 72.6$, $p < 0.001$; $\eta^2 = 0.79$] and Condition [$F(6, 114) = 30.8$, $p < 0.001$; $\eta^2 = 0.62$]. There was also an interaction effect between both factors [$F(6, 114) = 5.63$, $p < 0.001$; $\eta^2 = 0.23$]. A dynamic advantage was found for most facial expressions: anger [$t(19) -8.7049$; $p < 0.001$; 95% CI (−10.30 to −6.31%)], fear [$t(19) -3.3401$; $p = 0.0034$; 95% CI (−6.77 to −1.55%)], sadness [t(19) −5.2577; $p < 0.001$; 95% CI (−11.94 to −5.14%)], and surprise [$t(19) -7.4219$; $p < 0.001$; 95% CI (−10.94 to −6.13%)]. The effect for disgust did not resist the Bonferroni adjustment (p must be <0.007) [$t(19)$ −2.6413; $p = 0.0161$; 95% CI (−9.09 to −1.05%)]. No significant effect was found for happiness [$t(19) -2.0472$; $p = 0.0547$; 95% CI (−3.84 to 0.04%)]. There was also no significant difference with neutrality [$t(19) -1.8383$; $p = 0.0817$; 95% CI (−4.69 to 0.30%)], which is normal considering the absence of motion even in the dynamic stimuli.

Spatial Frequency Tuning

SF tunings for static and dynamic expressions were obtained separately for each participant by calculating a weighted sum of all the unsmoothed SF vectors that were used during testing (see **Figure 1C**), using accuracies transformed into z-scores as weights (see Willenbockel et al., 2010a; Royer et al., 2017; Tardif et al., 2017; for a similar procedure). Thus, positive weights were granted to SF vectors that led to correct responses and negative weights were given to SF vectors that led to incorrect responses. The resulting classification vectors were smoothed using a Gaussian kernel with a standard deviation of 2.5 cycles per image and then log-transformed. Finally, they were transformed into z scores using a permutation procedure whereby weights were randomly redistributed across trials and random classification vectors were created using these weights. This procedure was repeated 20 times, and the average and standard deviation for each SF across these random classification vectors were used to standardize the coefficients obtained for each SF in the participant's classification vector.

Group classification vectors were then produced for each condition by summing individual vectors across participants and dividing the outcome by the square root of the number of observers. The statistical threshold was determined with the Pixel test from the Stat4Ci toolbox (Zcrit = 3.1, $p < 0.025$; Chauvin et al., 2005). This threshold corrects for the multiple comparisons across SFs, while also taking into account the non-independence between contiguous SFs.

Group classification vectors are displayed in **Figure 4**. A SF tuning peaking at 18.0 cycles per face (cpf) with a full width at half maximum (FWHM) of 30.3 cpf was found in the Static condition, and a SF tuning peaking at 17.3 cpf with a FWHM of 29.3 cpf was found in


FIGURE 3 | Average accuracy for the static and dynamic facial expressions in Experiment 1. Error bars represent the standard deviations. *Significant at a p <0.007.

FIGURE 4 | Association between the availability of a given SF and participants accuracy for recognizing static (in black) and dynamic (in gray) expressions. This association is averaged across all participants and emotions. The dotted red line represents the difference between the Dynamic and Static conditions. The SFs that are significantly more used in the Dynamic than Static condition are indicated by the shaded orange area between the curves.

Dynamic condition. Most importantly, a significant difference in tuning was found between 3 and 7 cpf, indicating that this information was used more efficiently in the Dynamic vs. Static condition.

Discussion

The results of Experiment 1 show a shift toward lower SFs for dynamic compared to static expressions. This shift was expected based on the differences previously observed in the eye fixation pattern used with dynamic and static expressions. Experiment 2 aimed at verifying if the difference observed in the SF tuning is related to the presence of informative motion in dynamic expressions.

EXPERIMENT 2

Materials and Methods
Participants

Twenty-eight participants (9 males; 23 years old on average; SD = 5.77), none of whom participated in Experiment 1, were tested in Gatineau (Quebec, Canada). The number of participants was selected in order to match the total number of trials per condition in Experiment 1. However, to avoid an excessive increase in the duration of the experiment due to the addition of a third condition, we decreased the number of trials that a participant needed to complete in each condition and increased the number of participants. All participants had normal or corrected-to-normal visual acuity.

Stimuli

The same stimuli as in Experiment 1 were used in the Static and Dynamic conditions. In the Shuffled condition, the stimuli were created by randomizing the order of the 15 frames contained in the original dynamic stimuli.

Apparatus

Same as in Experiment 1.

Procedure

Each participant completed 10 blocks of 140 trials in each condition, for a total of 4,200 trials. The unfolding of events in a trial was the same as in Experiment 1 (see **Figure 2**). The participant's task was also the same as in Experiment 1.

All participants started with a block from the Dynamic condition, followed by a block from the Static condition and by a block from the Shuffled condition. The three conditions were then interleaved, and the same order was kept for the rest of the experiment. As was done in Experiment 1, the number of bubbles was adjusted on a trial basis, using QUEST during the Dynamic condition; the same number of bubbles was then applied for the following Static and Shuffled blocks.

The protocol of this experiment was approved by the Research Ethics Committee of Université du Québec en Outaouais and was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants provided informed written consent.

Results
Accuracy

An average of 13.6 (SD = 4.15) bubbles was necessary to maintain an approximate accuracy rate of 70% in the Dynamic condition. An average accuracy of 66.5% (SD = 2.4%), 71.7% (SD = 2.2%), and 64.3% (SD = 2.7%) was found in the Static, Dynamic, and Shuffled conditions, respectively. The average accuracy with each emotion in each condition is presented in **Figure 5**. A 7 (Emotions) × 3 (Conditions) repeated-measure ANOVA was conducted on accuracy. The results indicated significant main effects of the factors of Emotion [$F_{(6, 162)} = 37.4$, $p < 0.001$; $\eta^2 = 0.58$] and Condition [$F_{(2, 64)} = 201.2$, $p < 0.001$; $\eta^2 = 0.88$]. These were characterized by the presence of an interaction effect between both factors [$F_{(12, 324)} = 37.1$, $p < 0.001$; $\eta^2 = 0.58$]. One-way ANOVAs were then performed for each emotion. A significant effect of condition was found for disgust [$F_{(2)} = 53.8$, $p < 0.001$; $\eta^2 = 0.57$], happiness [$F_{(2)} = 13.3$, $p < 0.001$; $\eta^2 = 0.25$], sadness [$F_{(2)} = 5.3$, $p = 0.007$; $\eta^2 = 0.12$], and surprise [$F_{(2)} = 14.6$, $p < 0.001$; $\eta^2 = 0.27$]. With anger, the effect of Condition did not resist the Bonferroni adjustment (p must be <0.007) [$F_{(2)} = 4.2$, $p = 0.019$; $\eta^2 = 0.09$]. No significant effect of condition was found for fear [$F_{(2)} = 0.20$, $p = 0.82$] or neutrality [$F_{(2)} = 1.2$, $p = 0.31$]. For the four emotions showing a significant effect of Condition, as well as for anger (for which there was an effect prior to the Bonferroni adjustment), paired sample t-tests were carried to contrast accuracy for
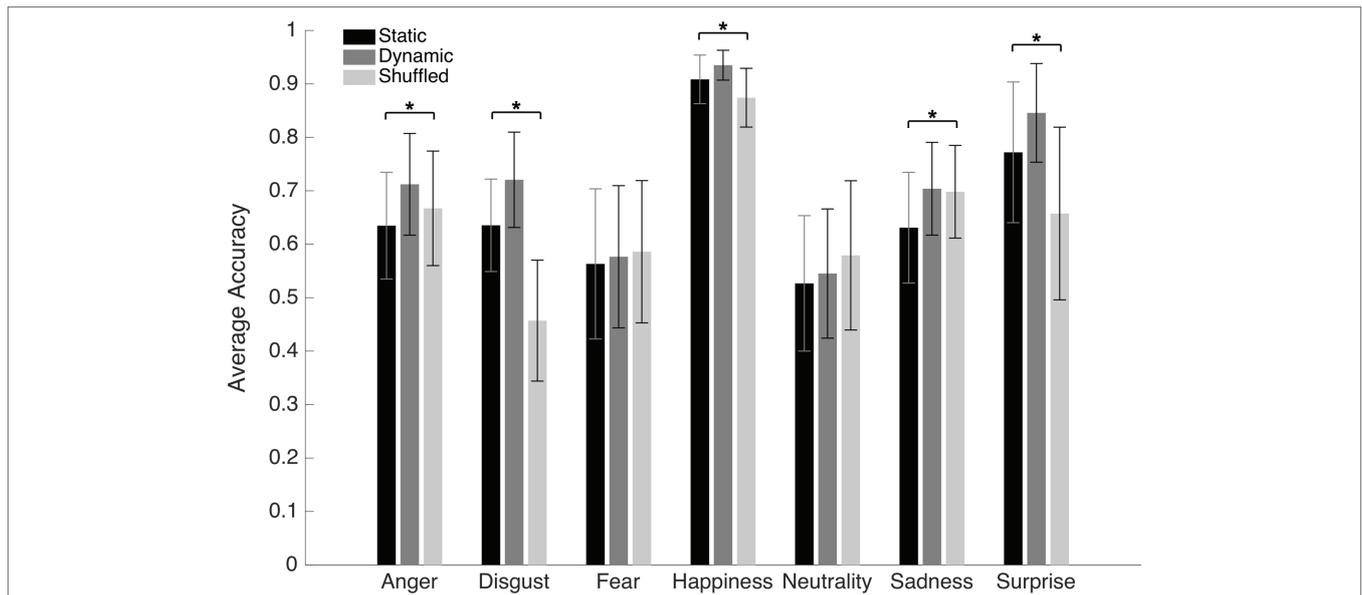
FIGURE 5 | Average accuracy for the static, dynamic, and shuffled facial expressions in Experiment 2. The error bars represent the standard deviations. *Significant at a $p < 0.007$.

TABLE 1 | Paired t-test results comparing the accuracy in the Dynamic and the Static conditions, the Dynamic and the Shuffled conditions, and the Static and Shuffled conditions.

		Dynamic/static	Dynamic/shuffled	Static/shuffled
Anger	t	6.5	3.4	−2.66
	p	<0.001*	0.0024*	0.0129
	d	1.24	0.63	−0.5
Disgust	t	8.2	13.8	9
	p	<0.001*	<0.001*	<0.001*
	d	1.55	2.6	1.69
Happiness	t	4.3	7.3	4.9
	p	<0.001*	<0.001*	<0.001*
	d	0.8	1.4	0.91
Sadness	t	5	0.4	−5.71
	p	<0.001*	0.67	<0.001*
	d	0.95	0.08	−1.08
Surprise	t	5.6	8.7	6.8
	p	<0.001*	<0.001*	< 0.001*
	d	1.06	1.64	1.29

*Paired t-tests were not performed for fear and neutrality since no significant effect of Condition was found in the one-way ANOVAs. *Significant at a p < 0.003.
d = Cohen's d.*

Dynamic vs. Static, Dynamic vs. Shuffled, and Static vs. Shuffled. The detailed results are provided in **Table 1**. Overall, participants were significantly more accurate in the Dynamic (vs. Static) condition for the emotions of anger, disgust, happiness, surprise, and sadness. They were also significantly more accurate in the Dynamic (vs. Shuffled) condition for the emotions of anger, disgust, happiness, and surprise, but not sadness. Finally, participants were significantly more accurate in the Static (vs. Shuffled) condition for the emotions of disgust, happiness, and surprise and less accurate for the emotions of anger and sadness.

Spatial Frequency Tuning

The group classification vectors obtained in the Static, Dynamic, and Shuffled conditions were produced using the same procedure as described in Experiment 1. The results are displayed in **Figure 6**. SF tunings peaking at 17.0, 14.3, and 16.0 cpf with FWHMs of 32.0, 26.7, and 21.0 cpf were found in the Static, Dynamic, and Shuffled conditions, respectively ($Z_{Crit} = 3.1$, $p < 0.025$).

A significant tuning difference was found between the tunings of the Static and Dynamic conditions: mid-to-high SFs ranging between 18.9 and 37.7 cpf were significantly more useful for static expressions. Significant differences were also found between the Static and Shuffled conditions, whereby low SFs ranging between 3.2 and 4.2 cpf were significantly more useful in the Shuffled condition and SFs higher than 18.9 cpf were significantly more useful in the Static condition. Moreover, no significative differences were found between the SF tuning of Dynamic and Shuffled conditions.

Discussion

Although the higher reliance on lower SFs with dynamic than with static expressions observed in Experiment 1 was not replicated, we did find a decreased reliance on higher SFs. This is consistent with the idea of a shift in SF tuning between static and dynamic expressions which will be further discussed in the next section.

A shift toward lower SFs was also observed for shuffled expressions. This suggests that the differences observed in the SF tunings for static and dynamic expressions are not caused by the presence of informative motion. In fact, contrary to what was expected, eliminating or reducing the amount of information contained in the motion by altering the natural sequence of facial changes led to a SF tuning significantly lower than the one observed in the Static condition and similar to the one observed in the Dynamic condition.
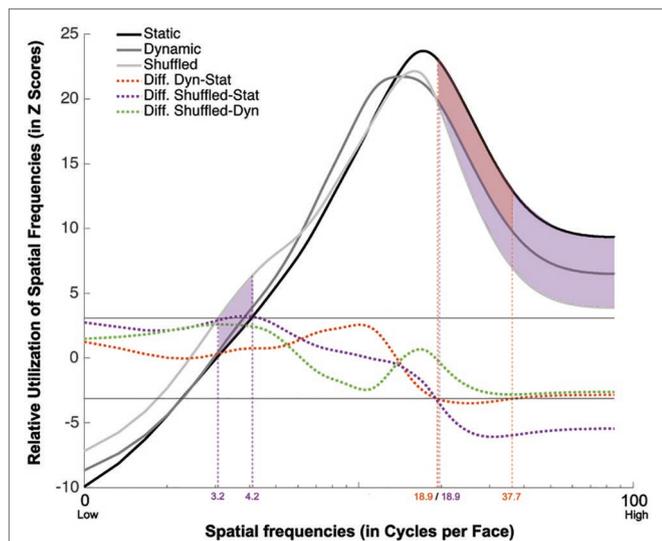
FIGURE 6 | Association between the availability of a given SF and participants accuracy for recognizing static (in black), dynamic (in dark gray), and shuffled (in pale gray) expressions. This association is averaged across all participants and emotions. The dotted red line represents the difference between the Dynamic and Static conditions. The dotted purple line represents the difference between the Shuffled and Static conditions. The dotted green line represents the difference between the Shuffled and Dynamic conditions. The red shaded area indicates the SFs that were significantly less useful in the Dynamic than the Static condition. The purple shaded area indicated the SFs that are significantly more used in the Shuffled than in the Static condition.
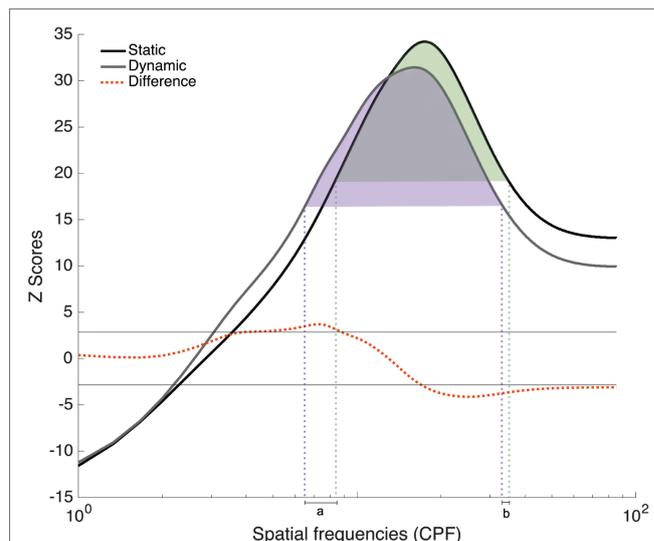

FIGURE 7 | Classification vectors representing the SFs used by the 48 participants tested in Experiments 1 and 2 for static and dynamic facial expressions, averaged across all emotions. The horizontal gray lines represent the statistical thresholds. The dotted red line represents the difference between the Dynamic and Static conditions. The vertical dotted lines corresponded to the beginning and end of each tuning curve at its half maximum (Dynamic in purple and static in green).

ANALYSIS OF EXPERIMENTS 1 AND 2 COMBINED

Since participants in Experiments 1 and 2 all completed trials with static and dynamic expressions, additional analyses combining all 48 participants were conducted in order to verify the robustness of the SF tuning shift between these conditions. Group classification vectors based on the 48 participants tested in Experiments 1 and 2 were produced for the Static and Dynamic conditions using the same procedure as described in Experiment 1. The results are presented in **Figure 7**. A SF tuning peak at 17.3 cpf with a FWHM of 31.3 cpf and a SF tuning peak at 16.0 cpf with a FWHM of 28.3 cpf were found in the Static and Dynamic conditions, respectively. Low SFs ranging between 5.6 and 8.3 cpf were significantly more useful in the Dynamic condition and mid-to-high SFs ranging between 17.6 and 85.3 cpf were significantly more useful in the Static condition. Note that the presence of extremely high SFs (i.e., >25 cpf) in the significant clusters is most likely due to the logarithmic SF sampling mentioned in the Materials and Methods; this impacts the resolution of the high SFs, as we have previously demonstrated in a previous study (see supplementary material in Estéphan et al., 2018).

In order to better quantify the tuning shift, we conducted a permutation analysis in which we randomly reassigned the Static and Dynamic conditions during the creation of the group classification vectors. More specifically, on each

iteration of the permutation analysis, the Static and Dynamic classification vectors of each participant were randomly assigned to either group classification vector. This procedure was repeated 10,000 times, which allowed us to estimate differences that may have occurred by chance. Two measures were taken: the distance between the tuning peaks for static and dynamic expressions and the translation between the two curves. This last measure was calculated in three steps. First, we indexed the SFs that corresponded to the beginning and end of each tuning curve at its half maximum [**Figure 7**; purple (Dynamic) and green (Static) dotted lines]. Second, SF values delineating the beginning of the static tuning curve were subtracted from those delineating the beginning of the dynamic curve (see value *a* in **Figure 7**); and SF values delineating the end of the static tuning curve were subtracted from those delineating the end of the dynamic curve (see value *b* in **Figure 7**). Finally, these two values, *a* and *b*, were added together. This measure therefore captures differences in the global shape of the tuning curves, as well as their relative position on the SF spectrum, whereas peak displacement reveals differences in SF values to which participants are most sensitive between static and dynamic expressions. For both of these measures, the value corresponding to the 5th percentile across these 10,000 pairs of random classification vectors was used as threshold. In terms of peak displacement, the difference observed between static and dynamic expressions (1.33 cpf) was marginally significant [95% CI (−1.66, 1.66), $p = 0.0759$]. In terms of tuning curve displacement on the SF spectrum, SF tuning for dynamic expressions was significantly translated toward lower SFs (6.33 cpf), relative

to static expressions [95% CI (−5.33, 5.66), p = 0.02]. Note that this permutation analysis only revealed a significant effect on peaks in Exp. 2 [average of 3 cpf; 95% CI (−3, 3), p = 0.05]. There was no significant difference in tuning peaks in Exp. 1 [average of 0.67 cpf; 95% CI (−2, 2), p = 0.33]. The tuning translation was neither significant in Exp. 1 [average translation of 4.33 cpf; 95% CI (−6.33, 6.33), p = 0.11] nor in Exp. 2 [average translation of 8.33 cpf; 95% CI (−48, 48), p = 0.38].

We also conducted an analysis to verify if the shift in SF tuning between dynamic and static expressions is related to the increased accuracy observed with dynamic expressions. We calculated the dynamic advantage in terms of accuracy (i.e., accuracy$_{Dynamic}$ − accuracy$_{Static}$) for each participant separately. We then measured the correlation between the individual dynamic advantage and the shifts in SF tunings (Peak$_{Dynamic}$ − Peak$_{Static}$) and the correlation between the individual dynamic advantage and the magnitude of translation between their tunings. The results indicate that the dynamic advantage was not correlated with any of these two measures: $r(46)$ = −0.026, p = 0.86 and $r(46)$ = −0.015, p = 0.92 were obtained for the shift in peaks and the translation of tunings, respectively. Finally, we conducted a preliminary analysis to verify if the SF tuning curves differed between men and women. The results indicated no significant effect of sex on the distance between the tuning peaks for static and dynamic expressions and the translation between the two curves. However, the sample was unbalanced with regards to sex and more research will be necessary to confirm this result.

GENERAL DISCUSSION

The present study investigated the SFs used during static and dynamic facial emotion recognition. In Experiment 1, we found higher reliance on lower SFs for dynamic expressions, whereas we found a decrease in higher SF utilization in Experiment 2. Taken together, these results are consistent with the hypothesized SF tuning shift, i.e., away from higher SFs and toward lower SFs for dynamic emotions.

The SF tuning shift was further assessed in a subsequent analysis that combined data from Experiments 1 and 2, using a permutation procedure. This revealed a marginally significant shift in the peak of the tuning curve for dynamic expressions, as well as a significant translation of the tuning curve itself. However, the fact that this result was nonsignificant when datasets of Experiments 1 and 2 were considered separately suggests that the difference is in fact quite small; hence, this last result should be interpreted with caution until replicated again. In the context of the replication crisis that is often discussed nowadays, new practices have been proposed with regard to how statistical results should be reported and interpreted (Amrhein et al., 2017). When interpreting the result of a replication study, as was done here with Exp. 2, it is recommended to base the comparison on the qualitative profile of the results rather than on the p-values or the traditional significance status. That said, the present study described two distinct experiments

that generated a similar pattern of results, and this pattern was expected based on the higher sensitivity of the magnocellular pathway to both low SF and motion (Livingstone and Hubel, 1988) and also based on previous eye-tracking results (Buchan et al., 2007; Blais et al., 2017). This, we argue, increases the likelihood that dynamic emotions induce a real shift in SF tuning, however small this shift may be.

Experiment 2 explored if the presence of informative motion in dynamic expressions may be the source of the shift toward lower SFs. In contrast with this hypothesis, the results revealed that altering the information provided by the naturally unfolding motion (i.e., shuffled dynamic emotions) did not eliminate this shift toward lower SFs. In fact, while there was no significant difference in SF tuning for dynamic and shuffled dynamic emotions, there was a significant difference in SF tuning for static and shuffled dynamic stimuli. Specifically, lower SFs were significantly more useful for shuffled dynamic expressions than they were for static expressions, and higher SFs were significantly more useful for static expressions than they were for shuffled dynamic expressions. This suggests that motion increases reliance on low SFs, irrespective of whether the natural unfolding of the expression is preserved or not. This is not however to say that motion was not used to gain an advantage during the recognition of dynamic expressions; in fact, higher accuracy for dynamic expressions may be related to utilization of such information.

As for why a shift toward lower SFs might be induced by motion, one possible – though speculative – explanation pertains to the undoubtedly high importance of motion perception from an evolutionary perspective. As such, the brain has likely developed mechanisms that protect and prioritize processing of motion signals, irrespective of whether this motion conveys information pertinent to a given context or not. Several findings from the literature support this idea. For example, studies have revealed the existence of subcortical pathways, in addition to cortical routes of motion processing, that allow motion perception. Such pathways would explain how visual motion perception can sometimes occur in the cortically blind (Tamietto and Morrone, 2016). Among these subcortical structures is the superior colliculus, a structure known for its role in guiding eye movements (Spering and Carrasco, 2015).

There are also studies indicating that motion processing is suppressed during ocular saccades (Ross et al., 1996), that saccades are suppressed prior to motion processing (Burr et al., 1999), and that rapid motion is better processed in peripheral vision (Tynan and Sekuler, 1982). These mechanisms can inform us as to how prioritizing motion processing should affect eye movements. Indeed, they predict that prioritization of motion processing should lead to saccade suppression (i.e., longer fixations), and a fixation location that allows for parafoveal processing of this information, when motion is detected. As such, fixating a face in its center when viewing dynamic expressions is consistent with prioritizing motion processing. This would also predict central face fixations when viewing shuffled dynamic expressions. In turn, parafoveal processing of diagnostic features may lower the spatial resolution of the visual information extracted.

Finally, it was also shown that processing of low SFs is suppressed during saccades (Burr et al., 1994). Thus, in addition to the fact that features are directly fixated (i.e., processed with highest spatial resolution in the fovea) during the processing of static expressions, the larger number of saccades that is also observed in such conditions may also play a role in lowering visual processing of low SFs and increasing reliance on higher SFs.

A second possible and straightforward explanation for the shift toward lower SFs might be the visual percept itself. Indeed, rapid local changes in time might blur higher SFs as a result of temporal averaging in visual short-term memory (Dubé and Sekuler, 2015). Thus, it may be that high SF information is simply not available to later processing stages in the visual system, leading to a decrease in their use and a commensurate increase in lower SF utilization – i.e., the observed SF tuning shift.

As previously stated, our analysis of accuracies supports the idea that informative motion is beneficial to the recognition of facial expressions. Consistent with this is our observation of a dynamic advantage over a majority of static expressions in both Experiments 1 and 2. Taken together the behavioral results of both experiments add to a growing body of evidence showing that dynamic expressions are often better recognized (Wehrle et al., 2000; Kamachi et al., 2001; Ambadar et al., 2005; Bould and Morris, 2008; Hammal et al., 2009; Cunningham and Wallraven, 2009a; Chiller-Glaus et al., 2011; Recio et al., 2011).

Several studies have found that the dynamic advantage was particularly evident when the physical information contained in the stimuli was either limited in terms of intensity (i.e., expressions not at apex) (Ambadar et al., 2005; Bould and Morris, 2008) or deteriorated in terms of shape, texture, or realism (e.g., photo vs. sketch) (Ehrlich et al., 2000; Wallraven et al., 2008; Cunningham and Wallraven, 2009b). In the present experiment, in addition to physical deterioration associated with the filtering procedure, the presentation time was also constrained (450 ms) in order to respect the natural unfolding of dynamic expressions. This may have favored the emergence of a dynamic advantage. One could even argue that the time restriction is involved in the observation of a dynamic advantage, as most studies that failed to find such an advantage presented their stimuli for more than a 1,000 ms (Gold et al., 2013, 1,059 ms; Fiorentini and Viviani, 2011, ~3,000 ms; Bould and Morris, 2008 ~1,500 ms; Widen and Russell, 2015, ~5,000 ms; Kätsyri and Sams, 2008, until answer). Indeed, such an extended presentation duration might allow a deeper exploration of static stimuli and therefore reducing the relative advantage found for dynamic stimuli.

The results of the second experiment also suggest better recognition of dynamic expressions over shuffled dynamic ones for almost all expressions, with the exception of fear and sadness, for which no significant difference was found. This absence of effect for shuffled expressions of fear and sadness corroborates previous results (Cunningham and Wallraven, 2009a; Richoz et al., 2018). One explanation to this increased accuracy found in shuffled fear and sadness might be attributable to the properties of the stimuli themselves. As reported by various participants, the shuffling of frames might have given the impression that actors performing were either having tremors (in the case of fear) or had their lower lip quivering (in the case of sadness). Again, this general advantage of dynamic expressions over shuffled ones supports the idea that motion containing information facilitates the recognition of dynamic facial expressions. However, our results suggest that the mere presence of motion is nonetheless associated with a shift toward lower SFs and that such shift is not associated with the size of the dynamic advantage.

Despite the obvious limits on ecological validity imposed by an artificial laboratory setting, dynamic expressions such as those used in the present study nonetheless represent a more ecological form of facial expressions compared to the static expressions used in previous research. However, the facial expressions depicted in our stimuli were posed by actors, and posed expressions have been shown to differ from spontaneous expressions with respect to clarity (Matsumoto et al., 2009), achieved intensity (Kayyal and Russell, 2013), and, most importantly, temporal unfolding (Ross et al., 2007; Ross and Pulusu, 2013). As it turns out, these differences between posed and spontaneous static expressions translate as differences in visual strategies in facial feature utilization (Saumure et al., 2018). Future studies should therefore examine the impact of motion on visual strategy variations across posed and spontaneous dynamic expressions.

It should also be mentioned that the samples for both studies were unbalanced with regard to gender. Although there is no clear evidence to suggest that sensitivity to motion differs between females and males (Vanston and Strother, 2017), some anatomical and functional differences have been found in regions of the visual cortex known for motion processing (Amunts et al., 2007; Anderson et al., 2013). Moreover, visual acuity has systematically been shown to be better in males (Burg, 1966; McGuinness, 1976; Ishigaki and Miyao, 1994; Abramov et al., 2012), and males also exhibit higher contrast sensitivity across the entire spatiotemporal domain, especially at higher SFs (Abramov et al., 2012). On the other hand, impact of sex on emotional recognition ability has also been studied, and the evidence favors females over males (e.g., Jenness, 1932; Hall, 1978; Collignon et al., 2010; Derntl et al., 2010; Kret and De Gelder, 2012). It would thus be important for future research to test the impact of sex on SF tuning and on the shift found for dynamic vs. static facial expressions – though our preliminary analysis did not corroborate the presence of sex differences in the SF tuning.

Finally, future studies should be conducted with larger stimuli in order to evaluate the impact of changing the visual eccentricity at which diagnostic information falls on the SF tuning. More specifically, it would be interesting to see if such a change in size would magnify the rather small SF peak shift that was obtained in the present study. It is however important to note that stimulus size alone cannot explain this outcome. In fact, one of our prior work on cross-cultural differences in face identification did reveal a considerably larger SF peak shift (as much as 6.68 cpf) as a function of culture, using face stimuli of similar size (i.e., 256 × 256 pixels) (Tardif et al., 2017).

CONCLUSION

Although much neuroanatomical and behavioral evidence suggest that dynamic and static facial expressions of emotion could rely on different perceptual mechanisms, little research has directly compared the visual strategies underlying the recognition of both kinds of expressions. The present research sought to address this shortfall by investigating SF tuning underlying the recognition of both types of expressions. Consistent with our hypothesis, our results suggested a shift toward lower SFs for dynamic expressions in comparison to static ones. This shift is not linked to the presence of natural and informative motion *per se*, but instead appears to be caused by the very presence of motion, notwithstanding the information it conveys. Nevertheless, natural motion does seem to be beneficial to the recognition of facial expressions, since both experiments revealed a dynamic recognition advantage over static or shuffled dynamic expressions. More research will be necessary to better understand the observed shift in SF tuning. One promising avenue is the idea that the mere presence of motion activates mechanisms aimed at prioritizing motion processing and that this in turn affects eye movements and SF processing.

REFERENCES

Abramov, I., Gordon, J., Feldman, O., and Chavarga, A. (2012). Sex & vision I: spatio-temporal resolution. *Biol. Sex Differ.* 3:20. doi: 10.1186/2042-6410-3-20

Adolphs, R., Tranel, D., and Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain Cogn.* 52, 61–69. doi: 10.1016/s0278-2626(03)00009-5

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* 372, 669.

Ambadar, Z., Schooler, J. W., and Cohn, J. F. (2005). Deciphering the enigmatic face: the importance of facial dynamics in interpreting subtle facial expressions. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

Amrhein, V., Korner-Nievergelt, F., and Roth, T. (2017). The earth is flat ($p > 0.05$): significance thresholds and the crisis of unreplicable research. *PeerJ* 5:e3544. doi: 10.7717/peerj.3544

Amunts, K., Armstrong, E., Malikovic, A., Hömke, L., Mohlberg, H., Schleicher, A., et al. (2007). Gender-specific left–right asymmetries in human visual cortex. *J. Neurosci.* 27, 1356–1364. doi: 10.1523/JNEUROSCI.4753-06.2007

Anderson, L., Bolling, D. Z., Schelinski, S., Coffman, M., Pelphrey, K. A., and Kaiser, M. D. (2013). Sex differences in the development of brain mechanisms for processing biological motion. *NeuroImage* 83, 751–760. doi: 10.1016/j.neuroimage.2013.07.040

Baron-Cohen, S., and Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *J. Autism Dev. Disord.* 34, 163–175. doi: 10.1023/B:JADD.0000022607.19833.00

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., and Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion* 17:1107. doi: 10.1037/emo0000283

Blais, C., Roy, C., Fiset, D., Arguin, M., and Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia* 50, 2830–2838. doi: 10.1016/j.neuropsychologia.2012.08.010

Bould, E., and Morris, N. (2008). Role of motion signals in recognizing subtle facial expressions of emotion. *Br. J. Psychol.* 99, 167–189. doi: 10.1348/000712607X206702

Brainard, D. H., and Vision, S. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

ETHICS STATEMENT

AUTHOR CONTRIBUTIONS

CB, DF, MPPD, and CS conceived and designed the experiments. MPPD and CS performed the experiments. MPPD, CB, DF, and JD analyzed the data. MPPD, CB, and DF drafted the manuscript. CS and JD reviewed the manuscript.

FUNDING

Buchan, J. N., Paré, M., and Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Soc. Neurosci.* 2, 1–13. doi: 10.1080/17470910601043644

Burg, A. (1966). Visual acuity as measured by dynamic and static tests: a comparative evaluation. *J. Appl. Psychol.* 50, 460–466. doi: 10.1037/h0023982

Burr, D. C., Morgan, M. J., and Morrone, M. C. (1999). Saccadic suppression precedes visual motion analysis. *Curr. Biol.* 9, 1207–1209. doi: 10.1016/S0960-9822(00)80028-7

Burr, D. C., Morrone, M. C., and Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature* 371, 511–513. doi: 10.1038/371511a0

Calvo, M. G., Fernández-Martín, A., Gutiérrez-García, A., and Lundqvist, D. (2018). Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Sci. Rep.* 8:17039. doi: 10.1038/s41598-018-35259-w

Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., and Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J. Vis.* 5, 659–667. doi: 10.1167/5.9.1

Chiller-Glaus, S. D., Schwaninger, A., Hofer, F., Kleiner, M., and Knappmeyer, B. (2011). Recognition of emotion in moving and static composite faces. *Swiss J. Psychol. Schweizerische Zeitschrift für Psychologie / Revue Suisse de Psychologie* 70, 233–240. doi: 10.1024/1421-0185/a000061

Clark, C. M., Gosselin, F., and Goghari, V. M. (2013). Aberrant patterns of visual facial information usage in schizophrenia. *J. Abnorm. Psychol.* 122, 513–519. doi: 10.1037/a0031944

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., and Lassonde, M. (2010). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia* 48, 220–225. doi: 10.1016/j.neuropsychologia.2009.09.007

Cunningham, D. W., and Wallraven, C. (2009a). Dynamic information for the recognition of conversational expressions. *J. Vis.* 9, 1–17. doi: 10.1167/9.13.7

Cunningham, D. W., and Wallraven, C. (2009b). *The interaction between motion and form in expression recognition*. Paper presented at the Proceedings of the 6th symposium on applied perception in graphics and visualization.

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., et al. (2010). Multidimensional assessment of empathic abilities: neural correlates and gender differences. *Psychoneuroendocrinology* 35, 67–82. doi: 10.1016/j.psyneuen.2009.10.006

DeValois, R. L., and DeValois, K. K. (1990). *Spatial vision*. New York: Oxford University Press.

Dubé, C., and Sekuler, R. (2015). Obligatory and adaptive averaging in visual short-term memory. *J. Vis.* 15:13. doi: 10.1167/15.4.13

Edwards, J., Jackson, H. J., and Pattison, P. E. (2002). Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. *Clin. Psychol. Rev.* 22, 789–832. doi: 10.1016/S0272-7358(02)00130-7

Ehrlich, S. M., Schiano, D. J., and Sheridan, K. (2000). *Communicating facial affect: it's not the realism, it's the motion*. Paper presented at the CHI'00 Extended Abstracts on Human Factors in Computing Systems.

Ekman, P., and Friesen, W. V. (1975). Unmasking the face: a guide to recognizing emotions from facial cues. Englewood Cliffs, NJ: Prentice Hall.

Estéphan, A., Fiset, D., Saumure, C., Plouffe-Demers, M.-P., Zhang, Y., Sun, D., et al. (2018). Time course of cultural differences in spatial frequency use for face identification. *Sci. Rep.* 8, 1–14. doi: 10.1038/s41598-018-19971-1

Fiorentini, C., and Viviani, P. (2011). Is there a dynamic advantage for facial expressions? *J. Vis.* 11:17. doi: 10.1167/11.3.17

Fiset, D., Blais, C., Royer, J., Richoz, A. R., Dugas, G., and Caldara, R. (2017). Mapping the impairment in decoding static facial expressions of emotion in prosopagnosia. *Soc. Cogn. Affect. Neurosci.* 12, 1334–1341.

Gaspar, C., Sekuler, A. B., and Bennett, P. J. (2008). Spatial frequency tuning of upright and inverted face identification. *Vis. Res.* 48, 2817–2826. doi: 10.1016/j.visres.2008.09.015

Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bromfield, W. D., Chu, N., et al. (2013). The efficiency of dynamic and static facial expression recognition. *J. Vis.* 13:23. doi: 10.1167/13.5.23

Gurnsey, R., Roddy, G., Ouhnana, M., and Troje, N. F. (2008). Stimulus magnification equates identification and discrimination of biological motion across the visual field. *Vis. Res.* 48, 2827–2834. doi: 10.1016/j.visres.2008.09.016

Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychol. Bull.* 85, 845–857. doi: 10.1037/0033-2909.85.4.845

Hammal, Z., Gosselin, F., and Fortin, I. (2009). How efficient are the recognition of dynamic and static facial expressions? *J. Vis.* 9:499. doi: 10.1167/9.8.499

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychol. Rev.* 20, 290–322. doi: 10.1007/s11065-010-9138-6

Hilz, R., and Cavonius, C. (1974). Functional organization of the peripheral retina: sensitivity to periodic stimuli. *Vis. Res.* 14, 1333–1337. doi: 10.1016/0042-6989(74)90006-6

Humphreys, G. W., Donnelly, N., and Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: neuropsychological evidence. *Neuropsychologia* 31, 173–181. doi: 10.1016/0028-3932(93)90045-2

Humphreys, K., Avidan, G., and Behrmann, M. (2007). A detailed investigation of facial expression processing in congenital prosopagnosia as compared to acquired prosopagnosia. *Exp. Brain Res.* 176, 356–373.

Ishigaki, H., and Miyao, M. (1994). Implications for dynamic visual acuity with changes in age and sex. *Percept. Mot. Skills* 78, 363–369.

Jenness, A. (1932). The recognition of facial expressions of emotion. *Psychol. Bull.* 29, 324–350. doi: 10.1037/h0074342

Jiang, B., Valstar, M. F., Martinez, B., and Pantic, M. (2014). A dynamic appearance descriptor approach to facial actions temporal modeling. *IEEE Trans. Cybern.* 44, 161–174. doi: 10.1109/TCYB.2013.2249063

Jiang, B., Valstar, M. F., and Pantic, M. (2011). Action unit detection using sparse appearance descriptors in space-time video volumes. Paper presented at the 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011).

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., and Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception* 42, 1266–1278. doi: 10.1068/p3131n

Kätsyri, J., and Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *Int. J. Human-Comput. Stud.* 66, 233–242. doi: 10.1016/j.ijhcs.2007.10.001

Kayyal, M. H., and Russell, J. A. (2013). Americans and Palestinians judge spontaneous facial expressions of emotion. *Emotion* 13, 891–904. doi: 10.1037/a0033244

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., and Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage* 18, 156–168. doi: 10.1006/nimg.2002.1323

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What's new in psychtoolbox-3. *Perception* 36, 1–16.

Kret, M. E., and De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia* 50, 1211–1221. doi: 10.1016/j.neuropsychologia.2011.12.022

Kring, A. M., and Elis, O. (2013). Emotion deficits in people with schizophrenia. *Annu. Rev. Clin. Psychol.* 9, 409–433. doi: 10.1146/annurev-clinpsy-050212-185538

LaBar, K. S., Crupain, M. J., Voyvodic, J. T., and McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cereb. Cortex* 13, 1023–1033. doi: 10.1093/cercor/13.10.1023

Lee, J., Gosselin, F., Wynn, J. K., and Green, M. F. (2010). How do schizophrenia patients use visual information to decode facial emotion? *Schizophr. Bull.* 37, 1001–1008. doi: 10.1093/schbul/sbq006

Livingstone, M., and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240, 740–749. doi: 10.1126/science.3283936

Mandal, M. K., Pandey, R., and Prasad, A. B. (1998). Facial expressions of emotions and schizophrenia: a review. *Schizophr. Bull.* 24, 399–412. doi: 10.1093/oxfordjournals.schbul.a033335

Marsh, A. A., Kozak, M. N., and Ambady, N. (2007). Accurate identification of fear facial expressions predicts prosocial behavior. *Emotion* 7, 239–251. doi: 10.1037/1528-3542.7.2.239

MathWorks, I. (2012). *MATLAB and statistics toolbox release 2012b*. Natick, Massachusetts, United States: The MathWorks, Inc.

Matsumoto, D., Olide, A., Schug, J., Willingham, B., and Callan, M. (2009). Cross-cultural judgments of spontaneous facial expressions of emotion. *J. Nonverbal Behav.* 33, 213–238. doi: 10.1007/s10919-009-0071-4

McGuinness, D. (1976). Away from a unisex psychology: individual differences in visual sensory and perceptual processes. *Perception* 5, 279–294.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vis. Res.* 39, 3824–3833. doi: 10.1016/S0042-6989(99)00096-6

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Recio, G., Sommer, W., and Schacht, A. (2011). Electrophysiological correlates of perceiving and evaluating static and dynamic facial emotional expressions. *Brain Res.* 1376, 66–75. doi: 10.1016/j.brainres.2010.12.041

Richoz, A. R., Jack, R. E., Garrod, O. G., Schyns, P. G., and Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex* 65, 50–64. doi: 10.1016/j.cortex.2014.11.015

Richoz, A.-R., Lao, J., Pascalis, O., and Caldara, R. (2018). Tracking the recognition of static and dynamic facial expressions of emotion across the life span. *J. Vis.* 18:5. doi: 10.1167/18.9.5

Ross, J., Burr, D., and Morrone, C. (1996). Suppression of the magnocellular pathway during saccades. *Behav. Brain Res.* 80, 1–8. doi: 10.1016/0166-4328(96)00012-5

Ross, E. D., Prodan, C. I., and Monnot, M. (2007). Human facial expressions are organized functionally across the upper-lower facial axis. *Neuroscientist* 13, 433–446. doi: 10.1177/1073858407305618

Ross, E. D., and Pulusu, V. K. (2013). Posed versus spontaneous facial expressions are modulated by opposite cerebral hemispheres. *Cortex* 49, 1280–1291. doi: 10.1016/j.cortex.2012.05.002

Roy, S., Roy, C., Fortin, I., Ethier-Majcher, C., Belin, P., and Gosselin, F. (2007). A dynamic facial expression database. *J. Vis.* 7:944. doi: 10.1167/7.9.944

Royer, J., Willenbockel, V., Blais, C., Gosselin, F., Lafortune, S., Leclerc, J., et al. (2017). The influence of natural contour and face size on the spatial frequency tuning for identifying upright and inverted faces. *Psychol. Res.* 81, 13–23. doi: 10.1007/s00426-015-0740-3

Rymarczyk, K., Biele, C., Grabowska, A., and Majczynski, H. (2011). EMG activity in response to static and dynamic facial expressions. *Int. J. Psychophysiol.* 79, 330–333. doi: 10.1016/j.ijpsycho.2010.11.001

Sato, W., and Yoshikawa, S. (2004). Brief report the dynamic aspects of emotional facial expressions. *Cognit. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Saumure, C., Plouffe-Demers, M.-P., Estéphan, A., Fiset, D., and Blais, C. (2018). The use of visual information in the recognition of posed and spontaneous facial expressions. *J. Vis.* 18:21. doi: 10.1167/18.9.21

Schultz, J., and Pilz, K. S. (2009). Natural facial motion enhances cortical responses to faces. *Exp. Brain Res.* 194, 465–475. doi: 10.1007/s00221-009-1721-9

Spering, M., and Carrasco, M. (2015). Acting without seeing: eye movements reveal visual processing without awareness. *Trends Neurosci.* 38, 247–258. doi: 10.1016/j.tins.2015.02.002

Tadros, K., Dupuis-Roy, N., Fiset, D., Arguin, M., and Gosselin, F. (2013). Reading laterally: the cerebral hemispheric use of spatial frequencies in visual word recognition. *J. Vis.* 13:4. doi: 10.1167/13.1.4

Takeuchi, T., Yokosawa, K., and De Valois, K. K. (2004). Texture segregation by motion under low luminance levels. *Vis. Res.* 44, 157–166. doi: 10.1016/j.visres.2003.09.005

Tamietto, M., and Morrone, M. C. (2016). Visual plasticity: blindsight bridges anatomy and function in the visual system. *Curr. Biol.* 26, R70–R73. doi: 10.1016/j.cub.2015.11.026

Tardif, J., Fiset, D., Zhang, Y., Estéphan, A., Cai, Q., Luo, C., et al. (2017). Culture shapes spatial frequency tuning for face identification. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 294–306. doi: 10.1037/xhp0000288

Thompson, B., Hansen, B. C., Hess, R. F., and Troje, N. F. (2007). Peripheral vision: good for biological motion, bad for signal noise segregation? *J. Vis.* 7:12. doi: 10.1167/7.10.12

Thurman, S. M., and Grossman, E. D. (2011). Diagnostic spatial frequencies and human efficiency for discriminating actions. *Atten. Percept. Psychophys.* 73, 572–580. doi: 10.3758/s13414-010-0028-z

Trautmann, S. A., Fehr, T., and Herrmann, M. (2009). Emotions in motion: dynamic compared to static facial expressions of disgust and happiness reveal more widespread emotion-specific activations. *Brain Res.* 1284, 100–115. doi: 10.1016/j.brainres.2009.05.075

Tynan, P. D., and Sekuler, R. (1982). Motion processing in peripheral vision: reaction time and perceived velocity. *Vis. Res.* 22, 61–68. doi: 10.1016/0042-6989(82)90167-5

Vanston, J. E., and Strother, L. (2017). Sex differences in the human visual system. *J. Neurosci. Res.* 95, 617–625. doi: 10.1002/jnr.23895

Vuilleumier, P., Armony, J. L., Driver, J., and Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat. Neurosci.* 6, 624–631. doi: 10.1038/nn1057

Wallraven, C., Breidt, M., Cunningham, D. W., and Bülthoff, H. H. (2008). Evaluating the perceptual realism of animated facial expressions. *ACM Trans. Appl. Percept.* 4, 1–22. doi: 10.1145/1278760.1278764

Watson, A. B., and Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Percept. Psychophys.* 33, 113–120. doi: 10.3758/BF03202828

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *J. Pers. Soc. Psychol.* 78, 105–119. doi: 10.1037/0022-3514.78.1.105

Weyers, P., Mühlberger, A., Hefele, C., and Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology* 43, 450–453. doi: 10.1111/j.1469-8986.2006.00451.x

Widen, S. C., and Russell, J. A. (2015). Do dynamic facial expressions convey emotions to children better than do static ones? *J. Cogn. Dev.* 16, 802–811. doi: 10.1080/15248372.2014.916295

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J. W., et al. (2010a). Does face inversion change spatial frequency tuning? *J. Exp. Psychol. Hum. Percept. Perform.* 36, 122–135. doi: 10.1037/a0016465

Willenbockel, V., Lepore, F., Bacon, B. A., and Gosselin, F. (2013). The informational correlates of conscious and nonconscious face-gender perception. *J. Vis.* 13:10. doi: 10.1167/13.2.10

Willenbockel, V., Lepore, F., Nguyen, D. K., Bouthillier, A., and Gosselin, F. (2012). Spatial frequency tuning during the conscious and non-conscious perception of emotional facial expressions–an intracranial ERP study. *Front. Psychol.* 3:237. doi: 10.3389/fpsyg.2012.00237

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010b). Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* 42, 671–684. doi: 10.3758/brm.42.3.671

Yang, N., Shafai, F., and Oruc, I. (2014). Size determines whether specialized expert processes are engaged for recognition of faces. *J. Vis.* 14:17. doi: 10.1167/14.8.17

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Atypical Amygdala–Neocortex Interaction During Dynamic Facial Expression Processing in Autism Spectrum Disorder

Wataru Sato[1]*[†], Takanori Kochiyama[2][†], Shota Uono[3], Sayaka Yoshimura[3], Yasutaka Kubota[4], Reiko Sawada[5,6], Morimitsu Sakihama[7] and Motomi Toichi[5,6]

[1]Kokoro Research Center, Kyoto University, Kyoto, Japan, [2]Brain Activity Imaging Center, ATR-Promotions, Inc., Kyoto, Japan, [3]Department of Neurodevelopmental Psychiatry, Habilitation and Rehabilitation, Graduate School of Medicine, Kyoto University, Kyoto, Japan, [4]Health and Medical Services Center, Shiga University, Hikone, Japan, [5]Faculty of Human Health Science, Graduate School of Medicine, Kyoto University, Kyoto, Japan, [6]The Organization for Promoting Developmental Disorder Research, Kyoto, Japan, [7]Rakuwa-kai Otowa Hospital, Kyoto, Japan

Atypical reciprocal social interactions involving emotional facial expressions are a core clinical feature of autism spectrum disorder (ASD). Previous functional magnetic resonance imaging (fMRI) studies have demonstrated that some social brain regions, including subcortical (e.g., amygdala) and neocortical regions (e.g., fusiform gyrus, FG) are less activated during the processing of facial expression stimuli in individuals with ASD. However, the functional networking patterns between the subcortical and cortical regions in processing emotional facial expressions remain unclear. We investigated this issue in ASD ($n = 31$) and typically developing (TD; $n = 31$) individuals using fMRI. Participants viewed dynamic facial expressions of anger and happiness and their corresponding mosaic images. Regional brain activity analysis revealed reduced activation of several social brain regions, including the amygdala, in the ASD group compared with the TD group in response to dynamic facial expressions vs. dynamic mosaics ($p < 0.05$, $\eta_p^2 = 0.19$). Dynamic causal modeling (DCM) analyses were then used to compare models with forward, backward, and bi-directional effective connectivity between the amygdala and neocortical networks. The results revealed that: (1) the model with effective connectivity from the amygdala to the neocortex best fit the data of both groups; and (2) the same model best accounted for group differences. Coupling parameter (i.e., effective connectivity) analyses showed that the modulatory effects of dynamic facial processing were substantially weaker in the ASD group than in the TD group. These findings suggest that atypical modulation from the amygdala to the neocortex underlies impairment in social interaction involving dynamic facial expressions in individuals with ASD.

Keywords: amygdala, autism spectrum disorder (ASD), dynamic causal modeling (DCM), dynamic facial expressions of emotion, functional magnetic resonance imaging (fMRI)

INTRODUCTION

Individuals with autism spectrum disorder (ASD) exhibit atypical social interactions (American Psychiatric Association, 2013). One of the most evident features of their social atypicality is deficient communication *via* emotional facial expressions (Hobson, 1993). Previous observational studies have reported that individuals with ASD exhibited attenuated emotional behaviors (e.g., Corona et al., 1998) and reduced and/or inappropriate facial reactions (e.g., Yirmiya et al., 1989) in response to others' facial expressions in social interactions compared with typically developing (TD) individuals. Experimental studies suggested that individuals with ASD are specifically impaired in the processing of dynamic, compared with static, facial expressions. For example, previous studies reported that ASD groups showed atypical perceptual (e.g., Palumbo et al., 2015; Uono et al., 2014), cognitive (e.g., Kessels et al., 2010; Sato et al., 2013), and motor (e.g., Rozga et al., 2013; Yoshimura et al., 2015) reactions during observations of dynamic facial expressions.

Several functional magnetic resonance imaging (fMRI) studies have investigated the neural mechanisms underlying atypical processing of dynamic facial expressions in individuals with ASD (Pelphrey et al., 2007; Rahko et al., 2012; Sato et al., 2012b). Although the results are not consistent across studies, some studies consistently reported that the observation of dynamic facial expressions evoked less activation in ASD groups than in TD groups of some subcortical brain regions, such as the amygdala, and some neocortical regions, such as the fusiform gyrus (FG) and superior temporal sulcus (STS) region (including the adjacent middle and superior temporal gyri; see Allison et al., 2000), and the inferior frontal gyrus (IFG; Pelphrey et al., 2007; Sato et al., 2012b). Abundant neuroimaging and neuropsychological evidence from TD individuals suggests that these brain regions are involved in the specific processing of social stimuli, such as emotional processing in the amygdala (for a review, see Calder et al., 2001), visual analysis of faces in the FG and STS region (for a review, see Haxby et al., 2000), and motor resonance in the IFG (for a review, see Rizzolatti et al., 2001). These regions have been called the "social brain" regions (Brothers et al., 1990; Adolphs, 2003; Blakemore, 2008) and were proposed to be impaired in individuals with ASD (Baron-Cohen et al., 2000; Emery and Perrett, 2000; Johnson et al., 2005; Bachevalier and Loveland, 2006; Frith, 2007; Pelphrey and Carter, 2008). One previous study further investigated functional coupling patterns in the neocortical network during the processing of dynamic facial expressions (Sato et al., 2012b). That study tested the bi-directional network connecting the primary visual cortex (V1), STS region, and IFG using dynamic causal modeling (DCM; Friston et al., 2003). The results showed that the modulatory effects of dynamic expressions on all connections were weaker in the ASD group than in the TD group. Together, these data suggest that a reduction in the activity of subcortical and neocortical social brain regions and their neocortical network may underlie atypical processing of dynamic facial expressions in individuals with ASD.

However, functional networking patterns between the subcortical and neocortical regions during the processing of dynamic facial expressions in individuals with ASD remain unclear, as these previous studies tested the neocortical network only in individuals with ASD. A recent neuroimaging study systematically investigated this issue in TD individuals (Sato et al., 2017b). That study analyzed fMRI data during the observation of dynamic facial expressions using DCM and compared models of the modulatory effects of dynamic facial expressions from the amygdala to the neocortex, from the neocortex to the amygdala, and bi-directionally. The results supported the model of the modulatory effect from the amygdala to the neocortex. This finding is consistent with anatomical evidence in animals that the amygdala receives visual input *via* subcortical pathways bypassing neocortical visual areas (Day-Brown et al., 2010), and sends widespread projections to neocortical regions, including the visual and motor areas (for a review, see Amaral et al., 1992). Several neuroscientific studies in TD individuals have also suggested that the amygdala conducts rapid emotional processing of facial expressions and modulates activities in the neocortical regions (for a review, see Vuilleumier and Pourtois, 2007). Based on these data, together with the aforementioned behavioral findings reporting impaired rapid processing of dynamic facial expressions in individuals with ASD (e.g., perception: Uono et al., 2014), we hypothesized that the modulatory effect from the amygdala to the neocortex may be weaker during the processing of dynamic facial expressions in individuals with ASD than in TD individuals.

In this fMRI study, we tested this hypothesis in a group of individuals with ASD and TD controls while they viewed dynamic facial expressions and their corresponding mosaic images. We analyzed group differences in regional brain activity in response to dynamic facial expressions vs. dynamic mosaics to determine differences in activity in the social brain regions between the ASD and TD groups. We prepared facial expressions of both negative (anger) and positive (happy) valences, though we did not expect different effects across emotions based on previous findings (Sato et al., 2012b). We then conducted DCM and compared models with the modulatory effects of dynamic facial expressions from the amygdala to the neocortex, from the neocortex to the amygdala, and bi-directionally, to determine which model optimally accounted for group commonalities and differences. We predicted that the model with the modulatory effect from the amygdala to the neocortex would be optimal for both purposes.

MATERIALS AND METHODS

Participants

The study included 31 Japanese adults in the ASD group (nine female, 22 male; mean $\pm$ *SD* age, 27.2 $\pm$ 8.5 years). This group consisted of 23 individuals with Asperger's disorder (six female, 17 male) and eight with pervasive developmental disorder not otherwise specified (PDD-NOS; three female, five male). Both diagnoses are included within the ASD category in the Diagnostic and Statistical Manual (DSM)-5 (American Psychiatric Association, 2013). PDD-NOS can include the

heterogeneous subtypes of ASD, as defined in the DSM-IV-Text Revision (TR; American Psychiatric Association, 2000); only high-functioning PDD-NOS participants with milder symptoms than those associated with Asperger's disorder were included in this study. The diagnosis was made by at least two psychiatrists with expertise in developmental disorders using the DSM-IV-TR *via* a strict procedure in which every item of the ASD diagnostic criteria was investigated in interviews with participants and their parents (and professionals who helped them, if any). Only participants who met at least one of the four social impairment items without satisfying any items of the criteria of autistic disorder were included. Each participant's developmental history was assessed through comprehensive interviews. Neurological and psychiatric problems other than those associated with ASD were ruled out. The participants were not taking medication. The intelligence quotients (IQs) of all participants in the ASD group had been assessed at other facilities and were reported to be within the normal range. Participants who agreed to newly undergo IQ tests ($n = 28$) were assessed using the revised Wechsler Adult Intelligence Scale, third edition (Nihon Bunka Kagakusha, Tokyo, Japan) and were confirmed to be in the normal range (full-scale IQ, mean $\pm$ *SD*, 110.0 $\pm$ 13.4). The symptom severity of the participants who were willing to undergo a further detailed interview ($n = 25$) was assessed quantitatively using the Childhood Autism Rating Scale (Schopler et al., 1986); the scores (mean $\pm$ *SD*, 24.4 $\pm$ 3.7) were comparable to those from previous studies that included high-functioning individuals with ASD (Koyama et al., 2007; Sato et al., 2012b; Uono et al., 2014; Yoshimura et al., 2015; *t*-test, $p > 0.1$).

The TD control group was comprised of 31 Japanese adults (nine female, 22 male; mean $\pm$ *SD* age, 24.2 $\pm$ 1.0 years). TD participants had no neurological or psychiatric problems and were matched with the ASD group for age (*t*-test, $p > 0.1$) and sex (χ^2-test, $p > 0.1$). Some of the TD participants agreed to participate in IQ tests ($n = 27$) using the revised Wechsler Adult Intelligence Scale, third edition (Nihon Bunka Kagakusha, Tokyo, Japan) and were confirmed to be in the normal range (full-scale IQ, mean $\pm$ *SD*, 121.8 $\pm$ 9.7), which was significantly higher than that of the ASD group ($t = 3.73$, $p < 0.001$).

All participants had normal or corrected-to-normal visual acuity and were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). After the procedures were fully explained, all participants provided written informed consent for participation. This study was approved by the Ethics Committee of the Primate Research Institute, Kyoto University (H2011–05), and was conducted in accordance with the ethical guidelines of the institution.

Stimuli

Angry and happy facial expressions of eight Japanese models (four female, four male) were presented as video clips. These stimuli were selected from our video database of facial expressions of emotion, which includes 65 Japanese models. The stimulus model looked straight ahead. All faces in the clips were unfamiliar to the participants.

The dynamic expression stimuli consisted of 38 frames ranging from neutral to emotional expressions. Each frame was

presented for 40 ms, and each clip was presented for 1,520 ms. The stimuli subtended a visual angle of approximately 15° vertically and 12° horizontally. The validity of these stimuli was supported by previous behavioral findings. Specifically, the speed of these stimuli was demonstrated to sufficiently represent natural changes in dynamic facial expressions (Sato and Yoshikawa, 2004). The stimuli were appropriately recognized as angry and happy expressions (Sato et al., 2010) and elicited appropriate subjective emotional reactions (Sato and Yoshikawa, 2007b) and spontaneous facial mimicry (Sato and Yoshikawa, 2007a; Sato et al., 2008) in TD individuals, but reduced spontaneous facial mimicry in individuals with ASD (Yoshimura et al., 2015).

The dynamic mosaic image stimuli were made from the same materials. All face images were divided into 50 vertical $\times$ 40 horizontal squares, which were randomly reordered using a fixed algorithm. This rearrangement made each image unrecognizable as a face. A set of 38 images, corresponding to the original dynamic facial expression stimuli, were presented as a clip at a speed identical to that of the dynamic expression stimuli.

Apparatus

Experiments were controlled using the Presentation 16.0 software (Neurobehavioral Systems, Albany, CA, USA). Stimuli were projected using a liquid crystal projector (DLA-HD10K; Japan Victor Company, Yokohama, Japan) onto a mirror that was positioned in a scanner in front of the participants. Responses were made using a response box (Response Pad; Current Designs, Philadelphia, PA, USA).

Procedure

Each participant completed the experimental scanning session, consisting of 20 epochs of 20 s each separated by 20 rest periods (a blank screen) of 10 s each. Each of the four stimulus conditions was presented in different epochs in a pseudorandomized order and the stimuli within each epoch were presented in a randomized order. Each epoch consisted of eight trials; a total of 160 trials were completed by each participant. Stimulus trials were replaced by target trials in eight trials.

During each stimulus trial, a fixation point (a small gray cross on a white background the same size as the stimulus) was presented in the center of the screen for 980 ms. The stimulus was then presented for 1,520 ms. During each target trial, a red cross (1.2° $\times$ 1.2°) was presented instead of the stimulus. Participants were instructed to detect the red cross and indicate that they had seen it by pressing a button with the right forefinger as quickly as possible. These dummy tasks ensured that the participants were attending to the stimuli but did not involve any controlled processing of the stimuli. Performance on the dummy target-detection task was perfect (correct identification rate = 100.0%).

Image Acquisition

Images were acquired using a 3-T scanning system (MAGNETOM Trio, A Tim System; Siemens, Malvern, PA, USA) with a 12-channel head coil. Lateral foam pads were used to stabilize the head position. The functional images consisted of 40 consecutive slices parallel to the anterior–posterior

commissure plane, and covered the whole brain. A T2*-weighted gradient-echo echo-planar imaging sequence was used with the following parameters: repetition time (TR) = 2,500 ms; echo time (TE) = 30 ms; flip angle = 90°; matrix size = 64 × 64; voxel size = 3 × 3 × 4 mm. After the acquisition of the functional images, a T1-weighted high-resolution anatomical image was acquired using a magnetization-prepared rapid-acquisition gradient-echo sequence (TR = 2,250 ms; TE = 3.06 ms; TI = 1,000 ms; flip angle = 9°; field of view = 256 × 256 mm; voxel size = 1 × 1 × 1 mm).

Image Analysis

Image analyses were accomplished using the statistical parametric mapping package SPM12[1], implemented in the MATLAB R2017b (MathWorks, Natick, MA, USA).

Preprocessing

For preprocessing, functional images were realigned using the first scan as a reference to correct for head motion. The realignment parameters revealed only a small (<3 mm) motion correction and no significant difference between the ASD and TD groups ($p > 0.1$ for x, y, z-translation and x, y, z-rotation). Next, all functional images were corrected for slice timing. The functional images were then coregistered to the anatomical image and all anatomical and functional images were normalized to Montreal Neurological Institute space using the anatomical image-based unified segmentation-spatial normalization approach (Ashburner and Friston, 2005). Finally, the normalized functional images were resampled to a voxel size of 2 × 2 × 2 mm and smoothed with an isotopic Gaussian kernel of 8 mm full width at half maximum.

Regional Brain Activity Analysis

To ensure that our paradigm engaged the functional anatomy of dynamic facial expression processing—for subsequent dynamic causal modeling, we performed two sets of activation analyses (**Supplementary Figure S1**). These included a region of interest (ROI) analysis within predefined ROIs and a mass-univariate, whole-brain analysis using statistical parametric mapping.

For these analyses, we performed a two-stage random effects analysis to identify significantly activated voxels at the population level (Holmes and Friston, 1998). First, a subject-level analysis was performed using a general linear model (GLM) framework (Friston et al., 1995). Boxcar functions encoded the main conditions, and Delta or stick functions modeled the target condition. These functions were convolved with a canonical hemodynamic response function. The realignment parameters were used as covariates to account for motion-related noise. We used a high-pass filter with a cut-off period of 128 s to eliminate the artifactual low-frequency trend. Serial autocorrelation was accounted for using a first-order autoregressive model.

Next, second group-level analyses were counducted. Based on our primary interest in analyzing group differences in functional networking patterns, we selected regions previously reported to be activated as ROIs for use in constructing the functional network during the processing of dynamic facial

expressions in TD individuals (Sato et al., 2017b). The ROIs specifically included the amygdala, fifth visual area (V5)/middle temporal area (MT), FG, STS, and IFG in the right hemisphere. Although a previous study reported that V5/MT activity during the observation of dynamic facial expressions did not differ between ASD and TD groups (Pelphrey et al., 2007), we included this region because: (1) data from another study testing the observation of dynamic facial expressions suggested reduced activity in this region in the ASD group (Sato et al., 2012b); (2) several fMRI studies testing different types of dynamic social stimuli reported reduced activity in this region in ASD individuals (Herrington et al., 2007; Brieber et al., 2010; Borowiak et al., 2018); and (3) a previous DCM study indicated that the functional network for processing dynamic facial expressions in TD individuals includes this region (Sato et al., 2017b). The coordinates in the Montreal Neurological Institute space of each ROI were derived from the results of this previous study (Sato et al., 2017b) and were identical to those used in the subsequent DCM analysis (**Supplementary Figure S2**).

The beta value for the effect of interest for each participant was extracted as the first eigenvariate of all voxels within a sphere of 4-mm radius around the participant-specific activation foci. The beta values for all ROIs were then subjected to a multivariate analysis of covariance (MANCOVA) with group (ASD vs. TD) as a between-subject factor, stimulus type (expression vs. mosaic) and emotion (anger vs. happiness) as within-subject factors, and sex and age as effect-of-no-interest covariates. Wilks' λ criterion was used. Significant effects were further tested using t-tests for single ROIs. Statistical significance was determined at a level of $p < 0.05$. To investigate possible confounding factors, including full-scale IQ and ASD subgroups, we preliminarily conducted the same multivariate analyses: (1) using full-scale IQ as a covariate among participants for whom we collected IQ data; or (2) substituting one ASD subgroup (Asperger or PDD-NOS) for the full ASD group. Because these analyses obtained similarly significant results, we omitted these factors in the reported results.

We then conducted exploratory analyses for the whole brain. Based on the results of the above ROI analysis, the effects of stimulus type (expression vs. mosaic) were analyzed using a two-sample t-test with group (ASD, TD) as an effect of interest and sex (male, female) and age as effects of no interest. Significantly activated voxels were identified if they reached an extent threshold of $p < 0.05$, corrected for multiple comparisons, with a cluster-forming threshold of $p < 0.001$ (uncorrected).

Brain structures were labeled anatomically and identified according to Brodmann's areas using the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) and Brodmann maps (Brodmann.nii), respectively, with the MRIcron tool[2].

DCM

For DCM analysis, we conducted group-level inference using a parametric empirical Bayesian (PEB) approach with the SPM12/DCM12 software (Friston et al., 2016; Zeidman et al.,

---

[1]https://www.fil.ion.ucl.ac.uk/spm

[2]http://www.mccauslandcenter.sc.edu/mricro/mricron/

2019a,b; see **Supplementary Figure S1**). PEB-DCM involved specifying a hierarchical model with two levels: individual subject and group. At the individual subject level, DCM parameters including neuronal interaction and a hemodynamic model of neurovascular coupling in each region was estimated from the fMRI time series data using variational Bayes under the Laplace approximation (Friston et al., 2003). At the group level, first-level (connectivity) parameters were entered into the second-level GLM to evaluate group effects and between-subjects parameter variability. We adopted the PEB-DCM approach because it offers several advantages over previously applied methods. Theoretically, PEB-DCM allows us to conduct more accurate and robust group inference by taking into account the posterior expectations (i.e., means) of the parameters and their posterior covariance; thus, parameter estimates at the individual subject level are adaptively weighted according to precision. Practically, this approach provides a direct and efficient method of performing group-level Bayesian model comparisons (BMCs) and Bayesian parameter inference to determine which model and connections best explain group differences. PEB-DCM was performed in the following four steps: (1) re-specification of the GLM to construct factor-specific regressors or DCM inputs and extraction of an fMRI time series from each participant; (2) specification of the neural network model space; (3) model estimation [steps (1–3) were performed at the individual subject level]; and (4) model comparison and parameter inference at the group level.

DCM allows for the modeling of three different types of effects in a neural network: (1) driving input, which represents the influence of exogenous input on neural states; (2) fixed connections, which represent baseline (i.e., applicable to all experimental conditions) connectivity among neural states; and (3) modulation of extrinsic (between-region) connections by experimental manipulation. Based on our research questions, we investigated the modulatory effect of dynamic facial expression. To construct driving and modulatory inputs for our DCM analysis, we remodeled the single-subject analyses. The design matrix contained the following two experimental factor-specific regressors: visual input (i.e., dynamic facial expressions and dynamic mosaic images) was the driving input in the DCM, and the dynamic facial expression condition was the modulatory input. Based on the results of the above regional brain activity analysis, emotion (anger vs. happiness) and target detection were included as effects of no interest. Other nuisance regressors (realignment parameters and constant terms), high-pass filters, and serial autocorrelations were applied using the settings described above for whole-brain statistical parametric mapping.

To investigate the direction of amygdala–neocortex functional interaction, seven brain regions in the right hemisphere were selected: the pulvinar (x14, y-30, z0), amygdala (x24, y-8, z-12), primary visual cortex (V1; x18, y-86, z-6), V5/MT (x48, y-60, z0), FG (x44, y-66, z-10), STS (x58, y-38, z14), and IFG (x50, y18, z26). The center coordinates of each ROI were derived from the results of the previous study (Sato et al., 2017b). ROIs were restricted to the right hemisphere because some ROIs showed significant activity only in the right hemisphere

(Sato et al., 2017b). The time series for each participant was extracted as the first eigenvariate of all voxels within a sphere of 4-mm radius around participant-specific activation foci, within the above ROIs. Participant-specific maxima for each region were selected using the following anatomical and functional criteria. The coordinates for the pulvinar were derived from within a sphere of 4-mm radius around the center coordinates used in the previous study. The coordinates for the amygdala were derived from within the intersection of a sphere of 8-mm radius around the center coordinates used in the previous study and the anatomically defined amygdala mask (Amygdala R in AAL atlas). The coordinates for the V1 were derived from the intersection of a sphere of 16-mm radius around the center coordinates used in the previous study and the anatomically defined calcarine sulcus (Calcarine R in the AAL atlas). The coordinates for the V5/MT, FG, STS and IFG were all derived within a sphere of 8-mm radius around the center coordinates used in the previous study. If no participant-specific maxima were identified, the center coordinates used in the previous study were used as the individual coordinates for that participant. Time-series data were adjusted for effects of no interest and nuisance regressors, high-pass filtered, and corrected for serial correlation.

Next, hypothesized models (**Figure 1**) were constructed for each participant. As a first assumption, the neocortical network, which had a driving input into the V1, and the bi-directional (i.e., forward and backward) extrinsic (between-region) connections of V1–V5/MT, V1–FG, V5/MT–STS, FG–STS, and STS–IFG were all estimated, and the modulatory effect of dynamic facial expression on all extrinsic connections was modeled. This neocortical network was constructed based on the theoretical proposals in the two-pathway model (Oram and Perrett, 1996) and the mirror neuron system model (Hamilton, 2008) for processing dynamic social signals. This neocortical network was validated in the previous study in TD individuals (Sato et al., 2017b), and a similar (partially simplified) model was also validated in ASD individuals (Sato et al., 2012b). As a second assumption, the subcortical network, which had a driving input into the pulvinar, and the forward extrinsic connection of the pulvinar–amygdala were estimated, and the modulatory effect of dynamic presentation on this extrinsic connection was estimated. This subcortical network was constructed based on theoretical (e.g., Vuilleumier, 2005) and empirical (e.g., Morris et al., 1999) evidence for processing emotional facial expressions. Although these studies posited that the superior colliculus sends input to the pulvinar, we did not include the superior colliculus in our model because this region was located adjacent to the pulvinar, making these regions difficult to dissociate using the defined ROI selection method. As a third assumption, we tested the connectivity between the amygdala and the V5/MT, FG, STS, and IFG neocortical regions. We made this assumption because several previous fMRI studies reported a functional interaction between the amygdala and these regions, which was consistent with the results of a previous study in TD individuals (Foley et al., 2012; Sato et al., 2017b). Based on the direction of modulatory effects, we constructed the three models

FIGURE 1 | Dynamic causal models. The analyzed brain regions are rendered on spatially normalized brains. Arrows indicate extrinsic connections between brain regions. Red points indicate modulatory effects of dynamic expression. Blue and orange regions indicate subcortical and neocortical subnetworks, respectively, both of which have the same structure across models. Amy, amygdala; FG, fusiform gyrus; IFG, inferior frontal gyrus; STS, superior temporal sulcus; Pul, pulvinar; V1, primary visual cortex; V5, fifth visual area/middle temporal area.

(**Figure 1**): Model 1 had modulatory connectivity from the amygdala to the neocortex; Model 2 had modulatory connectivity from the neocortex to the amygdala; and Model 3 had bi-directional modulatory connectivity between the amygdala and neocortex.

DCM models were estimated using the FULL + BMR option, which is the default estimation type for DCM12. We estimated only the full-model (in this case, Model 3) parameters for each subject; those of the reduced models (Models 1 and 2) were rapidly computed from the estimated parameters of the full model using a Bayesian model reduction (BMR; Friston et al., 2016).

To examine the direction of amygdala–neocortex connectivity, which best accounted for commonalities and differences across groups, we performed BMC among the three hypothesized models using the second-level PEB-DCM framework (Friston et al., 2016). Based on our research questions, we entered the eight modulatory parameters of amygdala–neocortex interaction from the B matrix of each DCM into the second-level GLM. The second level design matrix consisted of four regressors: the first regressor was a constant term representing commonalities across subjects and second regressor encoded group differences. Two covariates, sex and age, were added to the design matrix as effects of no interest. All regressors except for the first were mean-centered, allowing interpretation of the first regressor as a group mean across subjects. Posterior probability, a BMC evaluation measure, was computed for the three different models with a combination of the two group effects (commonalities and differences) using Bayesian model reduction.

To evaluate the group mean and differences in effective connectivity, we additionally calculated parameter estimates of the averaged model resulting from Bayesian model averaging (BMA). We used the entire model space for averaging, computing weighted averages of each model parameter for which the weighting was provided by the posterior probability for each model (Penny et al., 2010). We thus obtained eight parameter estimates for the modulatory connection of amygdala–neocortex interaction, which were evaluated using the posterior probability of models with and without each parameter.


FIGURE 2 | Mean (± SE) beta value for the main effect of stimulus type (dynamic expression vs. dynamic mosaic) in autism spectrum disorder (ASD) and typically developing (TD) groups in the social brain regions, including the amygdala (AMY), fifth visual area/middle temporal area (V5), fusiform gyrus (FG), superior temporal sulcus region (STS), and inferior frontal gyrus (IFG).
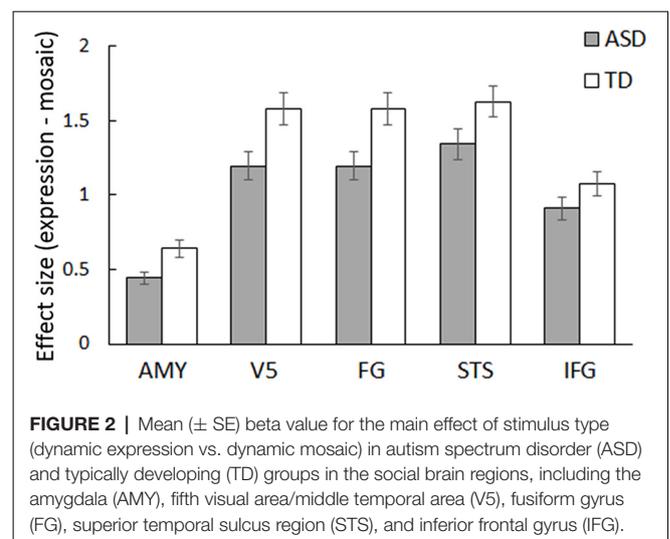
RESULTS

Regional Brain Activity

ROI analyses were conducted for predefined social brain regions, including the amygdala, V5, FG, STS region, and IFG, using a MANCOVA with group, stimulus type, and emotion as factors and sex and age as covariates. The results revealed a significant interaction between group and stimulus type ($F_{(5,54)} = 2.58$, $p < 0.05$, $\eta_p^2 = 0.19$). Besides, only the main effect of stimulus type was significant ($F_{(5,54)} = 4.28$, $p < 0.005$, $\eta_p^2 = 0.28$); other main effects and interactions were not significant ($p > 0.1$, $\eta_p^2 < 0.13$). The interaction between group and stimulus type indicates that activity in response to dynamic facial expressions vs. dynamic mosaics in these regions differed between the ASD and TD groups; the activity profile showed reduced activity in the ASD group (**Figure 2**). Follow-up univariate t-tests for the difference between dynamic facial expressions vs. dynamic mosaics confirmed significantly reduced activity in the ASD group compared with that in the TD group in the amygdala, V5/MT, and FG ($t_{(60)} > 2.08$; $p < 0.05$), although group

FIGURE 3 | Posterior probabilities of the models, accounting for group commonalities (left) and differences (right) between the ASD and TD groups determined by dynamic causal modeling (DCM) analysis. Model 1, which incorporates the modulatory effect of dynamic expression from the amygdala to the neocortex, best accounted for both commonalities and differences.
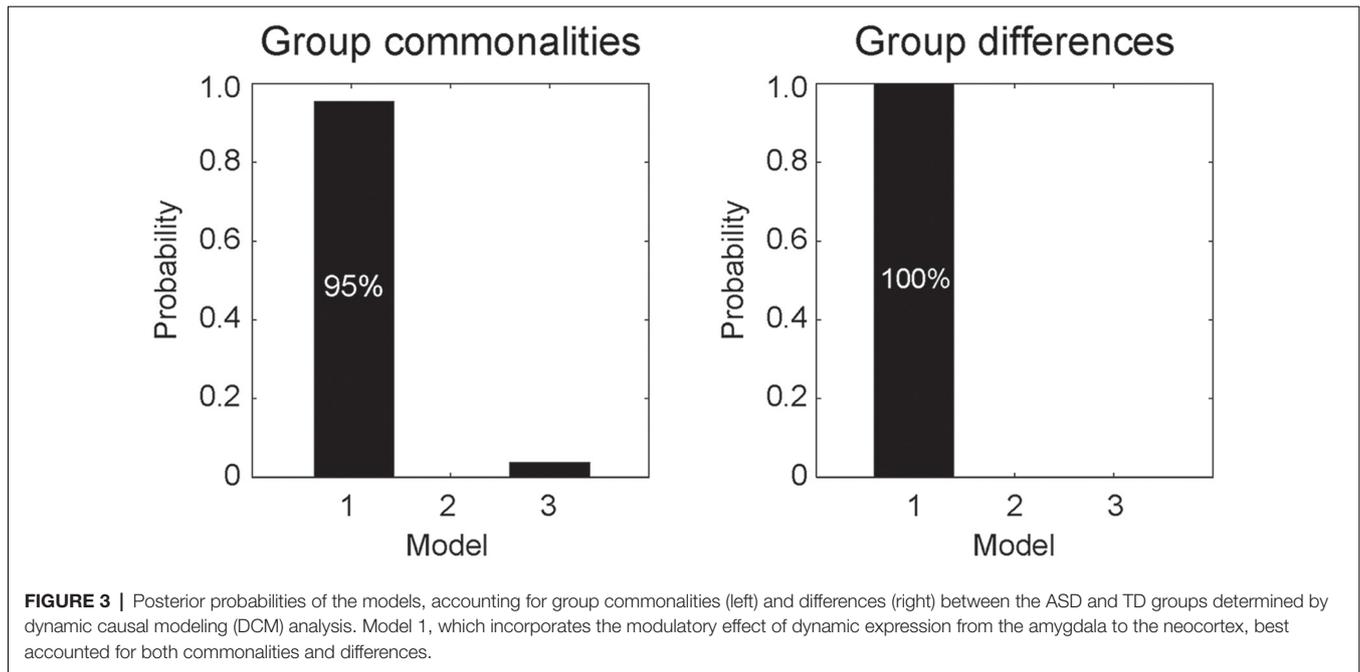
differences were marginally significant in the STS region ($t_{(60)}$ = 1.54; p < 0.1) and not significant in the IFG ($t_{(60)}$ = 0.61; p > 0.1; **Supplementary Figure S2**). Whole-brain analyses detected no other significant activation associated with group differences.

DCM

DCM analyses were conducted to compare the three network models having different modulatory effects of dynamic expression between the amygdala and neocortical regions (**Figure 1**). The posterior probability of PEB-DCM analysis indicated that Model 1 with the modulatory effect from the amygdala to the neocortex best accounted for both commonalities and differences among the ASD and TD groups (**Figure 3**).

BMA analysis was conducted to inspect profiles of the modulatory effect of dynamic expression. The resultant posterior means of modulatory effect parameters (**Figure 4**) showed that, with respect to commonalities across groups, the modulatory effects of dynamic facial expression were evident from the amygdala to the neocortex compared with connectivity from the neocortex to the amygdala. Modulatory effects from the amygdala were negative for connectivity to the V5, FG, and STS region and positive for connectivity to the IFG. For all connections from the amygdala to the neocortex, the modulatory effects of dynamic facial expression were weaker (i.e., near zero) in the ASD group than in the TD group.

DISCUSSION

Our regional brain activity analyses revealed that activity in the social brain regions was collectively lower in the ASD group

FIGURE 4 | Mean coupling parameters of the modulatory effect of dynamic expression for group commonalities (upper) and differences (lower) between the ASD and TD groups determined by DCM analysis. Pink bars indicate 90% Bayesian credible intervals. Amy, amygdala; FG, fusiform gyrus; IFG, inferior frontal gyrus; STS, superior temporal sulcus; V5, fifth visual area/middle temporal area.

than in the TD group in response to dynamic facial expressions vs. dynamic mosaic images. The reduced activity of social brain

regions in response to dynamic facial expressions in individuals with ASD was largely consistent with the findings of previous studies (Pelphrey et al., 2007; Sato et al., 2012b). However, our results did not show clear group differences in activity in the STS region or IFG, which was not consistent with previous findings, perhaps due to methodological differences. Specifically, the participants in this study were all high-functioning, not taking medication, and without severe symptoms; hence, their ASD traits may have been weaker than those in typical ASD individuals. Together with previous findings, our results suggest that social brain region activity during the processing of dynamic facial expressions is reduced in individuals with ASD.

More important, our DCM analysis provided interesting information regarding the functional networking patterns between the amygdala and neocortical regions during the processing of dynamic facial expressions in individuals with ASD. First, the model with the modulatory effect of dynamic expression from the amygdala to the neocortex best accounted for group commonalities. The results are consistent with previous findings in TD individuals (Sato et al., 2017b). Second, the same model best accounted for differences between the ASD and TD groups. Coupling parameter profiles revealed that the ASD group had weaker modulatory effects than the TD group. Differences in the functional networking patterns observed in ASD individuals were consistent with the previous finding that the modulatory effect of dynamic expression in the neocortical network was weaker in the ASD group than in the TD group (Sato et al., 2012b). However, the previous study did not investigate the functional network between the amygdala and neocortical regions. To the best of our knowledge, these results represent the first evidence that modulatory effects from the amygdala to the neocortex are reduced in individuals with ASD during the processing of dynamic facial expressions.

The coupling parameter profiles showed that the modulatory effects of dynamic facial expressions relative to dynamic mosaics were negative from the amygdala to the V5, FG, and STS region and positive from the amygdala to the IFG in both the ASD and TD groups. These patterns are not necessarily consistent with those reported in the previous study of TD individuals, which showed positive modulatory effects of dynamic facial expressions from the amygdala to all neocortical regions (Sato et al., 2017b). We speculate that this discrepancy may be due to methodological differences between studies, such as the use of stimulus facial expressions of same-race models rather than other-race models, or the use of dynamic mosaic images rather than static facial expressions as the control condition. Similarly, numerous previous studies have investigated functional coupling between the amygdala and posterior neocortical regions during the facial and/or emotional tasks and reported mixed findings, including positive (e.g., Foley et al., 2012; Diano et al., 2017; Jansma et al., 2014) and negative (e.g., Das et al., 2005; Williams et al., 2006; Pantazatos et al., 2012) modulation. These data suggest that the modulatory influence from the amygdala to the posterior neocortical regions may change depending on experimental conditions.

The findings of the present study, together with other neuroscientific evidence, may provide a mechanistic understanding of behavioral problems involving facial expression processing in individuals with ASD. Previous neuroimaging and electrophysiological findings in TD individuals have suggested that the amygdala rapidly conducts emotional processing of facial expressions, because the amygdala is activated by visual input *via* subcortical pathways prior to conscious awareness of the expressions (Morris et al., 1999; Pasley et al., 2004; Williams et al., 2006) specifically at about 100 ms (Bayle et al., 2009; Hung et al., 2010; Sato et al., 2011). A previous DCM analysis of electrophysiological data in TD individuals indicated that the modulation of dynamic facial expressions from the amygdala to the neocortex occurs rapidly at around 200 ms (Sato et al., 2017b). Together with these data, our observation of the reduced modulatory effect from the amygdala to the neocortex in ASD individuals may indicate impaired rapid emotional modulation in widespread neocortical processing for facial expressions, which may partly account for previous behavioral findings that individuals with ASD showed atypical perceptual, cognitive, and motor processing for emotional facial expressions (e.g., Yoshimura et al., 2015).

Our findings may have theoretical implications for the neural mechanisms of social atypicalities in ASD. Several researchers have proposed the theory that individuals with ASD have atypical activity and connectivity in the social brain regions (e.g., Baron-Cohen et al., 2000). However, empirical support for this remains controversial (for reviews, see Müler and Fishman, 2018; Sato and Uono, 2019). Several neuroimaging studies have provided positive evidence for reduced activity in the social brain regions during social stimulus processing. For example, Ciaramidaro et al. (2018) investigated brain activity during implicit and explicit processing of photographs of emotional facial expressions in ASD and TD groups. Implicit, but not explicit, processing of emotional facial expressions were associated with weaker activity in several social brain regions, including the FG, STS region, and amygdala in the ASD group than in the TD group. Sato et al. (2017a) reported reduced activation of the amygdala in response to subliminally presented averted eye gaze in the ASD group. However, other studies reported null or contradictory patterns of social brain region activity. For example, Tottenham et al. (2014) found stronger amygdala activity during the observation of facial expression photographs in the ASD group than in the TD group. Therefore, it may be difficult to draw conclusions about activity in the social brain regions in individuals with ASD. In contrast, a relatively small number of studies have accumulated a positive evidence for reduced functional coupling of the social brain regions in ASD. For example, Ciaramidaro et al. (2015) measured brain activity in response to social films in ASD and TD groups and found reduced functional connectivity between the FG and STS region in the ASD group. Borowiak et al. (2018) reported several reduced functional connections, including between the V5/MT and STS region, during the observation of visual speech in the ASD group. Together with these data, our data suggest that further investigation of the atypical social brain network theory of ASD may be worthwhile, specifically regarding atypical networking patterns in individuals with ASD.

Our finding of atypical amygdala modulation of the widespread neocortical network also has a practical implication. These data suggest the possibility that improvement in amygdala activity may have positive effects on various types of perceptual, cognitive, or motor processing for facial expressions. One previous study has reported that electrical stimulation of the amygdala in individuals with ASD modified their autistic symptoms and face-to-face interactions (Sturm et al., 2013). The effect of oxytocin on ASD symptoms may also be relevant. Previous behavioral studies in individuals with ASD have shown that intranasal administration of oxytocin improved their facial expression processing, including rapid perceptual processing (Xu et al., 2015; Domes et al., 2016). Because neuroimaging studies in TD individuals showed that administration of oxytocin modulates amygdala activity during the processing of emotional facial expressions (Domes et al., 2007; Kanat et al., 2015), we speculate that the modulatory effect from the amygdala to the neocortex may account for the behavioral effect of oxytocin in individuals with ASD. Future research might further examine the effect of electric or pharmacological intervention on amygdala activity to influence various types of social processing *via* modulation of neocortical activity in individuals with ASD.

Several limitations of the present study should be acknowledged. First, IQ was not assessed in all participants. Although we acquired IQ data from most members of the ASD and TD groups and our preliminary analyses suggested that IQ was not related to the patterns of regional brain activity, this finding is not conclusive. This issue could be critical, as previous behavioral studies have suggested that IQ differences between ASD and TD groups may affect differences in the recognition of emotional facial expressions (Harms et al., 2010). Second, the ASD group included heterogeneous subgroups (i.e., Asperger's disorder and PDD-NOD). Although our preliminary analyses suggested similar patterns of regional brain activity across these subgroups, our sample was too small to investigate this issue. Third, dynamic mosaic images were presented as control stimuli; it remains unclear which types of information processing might reveal group differences in activity and connectivity of the social brain regions. Although dynamic mosaic stimuli could act as control stimuli for dynamic facial expressions in terms of low-level visual properties, such as brightness and motion, and have been used in several previous neuroimaging studies (e.g., De Winter et al., 2015), different types of control stimuli are required to identify specific cognitive or emotional factors associated with group differences in social brain network functioning. Finally, only angry and happy facial expressions were tested. To demonstrate the generalizability of the present findings, investigations of facial expressions of various types of emotions (e.g., Tottenham et al., 2014) are needed. Furthermore, because the amygdala is active during the processing of emotionally neutral faces (Ishai et al., 2005; Sato et al., 2012a), we speculate that the atypical amygdala–neocortex modulation in individuals with ASD may be related to their atypical processing of non-emotional facial actions (e.g., Williams et al., 2004). Additional studies investigating these unsettled issues are required to deepen our understanding of the functioning of the social brain network in individuals with ASD.

In conclusion, our regional brain activity analysis revealed a reduced activity of several social brain regions in response to dynamic facial expressions vs. dynamic mosaic images in the ASD group relative to the TD group. Our DCM analyses revealed that the model with effective connectivity from the amygdala to the neocortex best accounted for commonalities and differences between groups. Modulatory effects were weaker in the ASD group than in the TD group. These results suggest that atypical modulation from the amygdala to the neocortex underlie impairment in social interaction involving dynamic facial expressions in individuals with ASD.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study was approved by the Ethics Committee of the Primate Research Institute, Kyoto University, and was conducted in accordance with the ethical guidelines of the institution. After the experimental procedures had been fully explained, written informed consent was obtained from all participants.

AUTHOR CONTRIBUTIONS

WS, TK, SU and MT designed the research. WS, TK, SY and MT analyzed the data. All authors obtained the data, wrote the manuscript, read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00351/full#supplementary-material.

FIGURE S1 | Flowchart for data analysis. fMRI data were acquired from the subjects **(A)**. A general linear model (GLM) with four conditions of interest [i.e., dynamic facial expressions of anger (DyAn) and happiness (DyHa) and their corresponding mosaic images (MoAn, MoHa)] was estimated for each individual subject **(B)**. A multivariate analysis of covariance (MANCOVA) was conducted on

the beta estimates of five regions of interest and four conditions **(C)**. In the effective connectivity analysis, time series data **(F)** were extracted from seven volumes of interest (VOIs; **E**) using the rearranged GLM **(D)**. Dynamic causal modeling (DCM) was conducted on three network models encoding the sub–neocortex interaction for individual subjects **(G)**. The estimated DCMs for all subjects were entered into the second-level parametric empirical Bayesian (PEB)-DCM engine **(H)**. Bayesian model comparisons **(J)** and parameter inferences **(K)** were accomplished to evaluate group effects (commonalities and differences across groups) based on the second-level design matrix **(I)**.

FIGURE S2 | Regions of interest (ROIs) for regional brain activity and dynamic causal modeling analyses for the amygdala (AMY), fifth visual area/middle temporal (V5), fusiform gyrus (FG), superior temporal sulcus region (STS), and inferior frontal gyrus (IFG) rendered on the spatially normalized brain of a representative participant. The coordinates of each ROI are in the Montreal Neurological Institute space and were derived from the results of Sato et al. (2017b). Activation indicating an interaction between group (TD vs. ASD) and stimulus type (expression vs. mosaic), based on follow-up univariate *t*-tests of a multivariate analysis of covariance, is overlaid in the red–yellow color scale (see the "Results" section).

REFERENCES

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat. Rev. Neurosci.* 4, 165–178. doi: 10.1038/nrn1056

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278. doi: 10.1016/s1364-6613(00)01501-1

Amaral, D. G., Price, J. L., Pitkanen, A., and Carmichael, S. T. (1992). "Anatomical organization of the primate amygdaloid complex," in *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, ed. J. P. Aggleton (New York, NY: Wiley-Liss), 1–66.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.)*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Bachevalier, J., and Loveland, K. A. (2006). The orbitofrontal-amygdala circuit and self-regulation of social-emotional behavior in autism. *Neurosci. Biobehav. Rev.* 30, 97–117. doi: 10.1016/j.neubiorev.2005.07.002

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., and Williams, S. C. (2000). The amygdala theory of autism. *Neurosci. Biobehav. Rev.* 24, 355–364. doi: 10.1016/S0149-7634(00)00011-7

Bayle, D. J., Henaff, M. A., and Krolak-Salmon, P. (2009). Unconsciously perceived fear in peripheral vision alerts the limbic system: a MEG study. *PLoS One* 4:e8207. doi: 10.1371/journal.pone.0008207

Blakemore, S. J. (2008). The social brain in adolescence. *Nat. Rev. Neurosci.* 9, 267–277. doi: 10.1038/nrn2353

Borowiak, K., Schelinski, S., and von Kriegstein, K. (2018). Recognizing visual speech: reduced responses in visual-movement regions, but not other speech regions in autism. *Neuroimage Clin.* 20, 1078–1091. doi: 10.1016/j.nicl.2018.09.019

Brieber, S., Herpertz-Dahlmann, B., Fink, G. R., Kamp-Becker, I., Remschmidt, H., and Konrad, K. (2010). Coherent motion processing in autism spectrum disorder (ASD): an fMRI study. *Neuropsychologia* 48, 1644–1651. doi: 10.1016/j.neuropsychologia.2010.02.007

Brothers, L., Ring, B., and Kling, A. (1990). Response of neurons in the macaque amygdala to complex social stimuli. *Behav. Brain Res.* 41, 199–213. doi: 10.1016/0166-4328(90)90108-q

Calder, A. J., Lawrence, A. D., and Young, A. W. (2001). Neuropsychology of fear and loathing. *Nat. Rev. Neurosci.* 2, 352–363. doi: 10.1038/35072584

Ciaramidaro, A., Bölte, S., Schlitt, S., Hainz, D., Poustka, F., Weber, B., et al. (2015). Schizophrenia and autism as contrasting minds: neural evidence for the hypo-hyper-intentionality hypothesis. *Schizophr. Bull.* 41, 171–179. doi: 10.1093/schbul/sbu124

Ciaramidaro, A., Bölte, S., Schlitt, S., Hainz, D., Poustka, F., Weber, B., et al. (2018). Transdiagnostic deviant facial recognition for implicit negative emotion in autism and schizophrenia. *Eur. Neuropsychopharmacol.* 28, 264–275. doi: 10.1016/j.euroneuro.2017.12.005

Corona, R., Dissanayake, C., Arbelle, S., Wellington, P., and Sigman, M. (1998). Is affect aversive to young children with autism? Behavioral and cardiac responses to experimenter distress. *Child Dev.* 69, 1494–1502. doi: 10.2307/1132127

Das, P., Kemp, A. H., Liddell, B. J., Brown, K. J., Olivieri, G., Peduto, A., et al. (2005). Pathways for fear perception: modulation of amygdala activity by thalamo-cortical systems. *Neuroimage* 26, 141–148. doi: 10.1016/j.neuroimage.2005.01.049

Day-Brown, J. D., Wei, H., Chomsung, R. D., Petry, H. M., and Bickford, M. E. (2010). Pulvinar projections to the striatum and amygdala in the tree shrew. *Front. Neuroanat.* 4:143. doi: 10.3389/fnana.2010.00143

De Winter, F. L., Zhu, Q., Van den Stock, J., Nelissen, K., Peeters, R., de Gelder, B., et al. (2015). Lateralization for dynamic facial expressions in human superior temporal sulcus. *Neuroimage* 106, 340–352. doi: 10.1016/j.neuroimage.2014.11.020

Diano, M., Tamietto, M., Celeghin, A., Weiskrantz, L., Tatu, M. K., Bagnis, A., et al. (2017). Dynamic changes in amygdala psychophysiological connectivity reveal distinct neural networks for facial expressions of basic emotions. *Sci. Rep.* 7:45260. doi: 10.1038/srep45260

Domes, G., Heinrichs, M., Gläscher, J. B., Büchel, C., Braus, D. F., and Herpertz, S. C. (2007). Oxytocin attenuates amygdala responses to emotional faces regardless of valence. *Biol. Psychiatry* 62, 1187–1190. doi: 10.1016/j.biopsych.2007.03.025

Domes, G., Normann, C., and Heinrichs, M. (2016). The effect of oxytocin on attention to angry and happy faces in chronic depression. *BMC Psychiatry* 16:92. doi: 10.1186/s12888-016-0794-9

Emery, N. J., and Perrett, D. I. (2000). "How can studies of the monkey brain help us understand "theory of mind" and autism in humans," in *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, eds 2nd Edn. S. Baron-Cohen, H. Tager-Flusberg and D. J. Cohen (Oxford: Oxford University Press), 274–305.

Foley, E., Rippon, G., Thai, N. J., Longe, O., and Senior, C. (2012). Dynamic facial expressions evoke distinct activation in the face perception network: a connectivity analysis study. *J. Cogn. Neurosci.* 24, 507–520. doi: 10.1162/jocn_a_00120

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/s1053-8119(03)00202-7

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S. J., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage* 2, 45–53. doi: 10.1006/nimg.1995.1007

Friston, K. J., Litvak, V., Oswal, A., Razi, A., Stephan, K. E., van Wijk, B. C. M., et al. (2016). Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage* 128, 413–431. doi: 10.1016/j.neuroimage.2015.11.015

Frith, C. D. (2007). The social brain? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 671–678. doi: 10.1098/rstb.2006.2003

Hamilton, A. F. D. C. (2008). Emulation and mimicry for social interaction: a theoretical approach to imitation in autism. *Q. J. Exp. Psychol.* 61, 101–115. doi: 10.1080/17470210701508798

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychol. Rev.* 20, 290–322. doi: 10.1007/s11065-010-9138-6

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0

Herrington, S., Baron-Cohen, S. J., Wheelwright, K. D., Singh, E. D., Bullmore, E. T., Brammer, M., et al. (2007). The role of MT+/V5 during biological motion perception in Asperger Syndrome: an fMRI study. *Res. Autism Spectr. Disord.* 14, 14–27. doi: 10.1016/j.rasd.2006.07.002

Hobson, R. P. (1993). *Autism and the Development of Mind*. Hove, UK: Lawrence Erlbaum Associates.

Holmes, A. P., and Friston, K. J. (1998). Generalisability, random effects and population inference. *Neuroimage* 7:S754. doi: 10.1016/s1053-8119(18)31587-8

Hung, Y., Smith, M. L., Bayle, D. J., Mills, T., Cheyne, D., and Taylor, M. J. (2010). Unattended emotional faces elicit early lateralized amygdala-frontal and fusiform activations. *Neuroimage* 50, 727–733. doi: 10.1016/j.neuroimage. 2009.12.093

Ishai, A., Schmidt, C. F., and Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain Res. Bull.* 67, 87–93. doi: 10.1016/j. brainresbull.2005.05.027

Jansma, H., Roebroeck, A., and Münte, T. F. (2014). A network analysis of audiovisual affective speech perception. *Neuroscience* 256, 230–241. doi: 10.1016/j.neuroscience.2013.10.047

Johnson, M. H., Griffin, R., Csibra, G., Halit, H., Farroni, T., de Haan, M., et al. (2005). The emergence of the social brain network: evidence from typical and atypical development. *Dev. Psychopathol.* 17, 599–619. doi: 10.1017/s0954579405050297

Kanat, M., Heinrichs, M., Schwarzwald, R., and Domes, G. (2015). Oxytocin attenuates neural reactivity to masked threat cues from the eyes. *Neuropsychopharmacology* 40, 287–295. doi: 10.1038/npp.2014.183

Kessels, R. P. C., Spee, P., and Hendriks, A. W. (2010). Perception of dynamic facial emotional expressions in adolescents with autism spectrum disorders (ASD). *Trans. Neurosci.* 1, 228–232. doi: 10.2478/v10134-010-0033-8

Koyama, T., Tachimori, H., Osada, H., Takeda, T., and Kurita, H. (2007). Cognitive and symptom profiles in Asperger's syndrome and high-functioning autism. *Psychiatry Clin. Neurosci.* 61, 99–104. doi: 10.1111/j.1440-1819.2007.01617.x

Morris, J. S., Ohman, A., and Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc. Natl. Acad. Sci. U S A* 96, 1680–1685. doi: 10.1073/pnas.96.4.1680

Müler, R. A., and Fishman, I. (2018). Brain connectivity and neuroimaging of social networks in autism. *Trends Cogn. Sci.* 22, 1103–1116. doi: 10.1016/j.tics. 2018.09.008

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Oram, M. W., and Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J. Neurophysiol.* 76, 109–126. doi: 10.1152/jn.1996.76.1.109

Palumbo, L., Burnett, H. G., and Jellema, T. (2015). Atypical emotional anticipation in high-functioning autism. *Mol. Autism* 6:47. doi: 10.1186/s13229-015-0039-7

Pantazatos, S. P., Talati, A., Pavlidis, P., and Hirsch, J. (2012). Cortical functional connectivity decodes subconscious, task-irrelevant threat-related emotion processing. *Neuroimage* 61, 1355–1363. doi: 10.1016/j.neuroimage.2012.03.051

Pasley, B. N., Mayes, L. C., and Schultz, R. T. (2004). Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42, 163–172. doi: 10.1016/s0896-6273(04)00155-2

Pelphrey, K. A., and Carter, E. J. (2008). Charting the typical and atypical development of the social brain. *Dev. Psychopathol.* 20, 1081–1102. doi: 10.1017/s0954579408000515

Pelphrey, K. A., Morris, J. P., McCarthy, G., and Labar, K. S. (2007). Perception of dynamic changes in facial affect and identity in autism. *Soc. Cogn. Affect. Neurosci.* 2, 140–149. doi: 10.1093/scan/nsm010

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., et al. (2010). Comparing families of dynamic causal models. *PLoS Comput. Biol.* 6:e1000709. doi: 10.1371/journal.pcbi.1000709

Rahko, J., Paakki, J. J., Starck, T., Nikkinen, J., Pauls, D., Kätsyri, J. V., et al. (2012). Valence scaling of dynamic facial expressions is altered in high-functioning subjects with Autism Spectrum Disorders: an fMRI Study. *J. Autism Dev. Disord.* 42, 1011–1024. doi: 10.1007/s10803-011-1332-8

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670. doi: 10.1038/35090060

Rozga, A., King, T. Z., Vuduc, R. W., and Robins, D. L. (2013). Undifferentiated facial electromyography responses to dynamic, audio-visual emotion displays in individuals with autism spectrum disorders. *Dev. Sci.* 16, 499–514. doi: 10.1111/desc.12062

Sato, W., Fujimura, T., and Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *Int. J. Psychophysiol.* 70, 70–74. doi: 10.1016/j.ijpsycho.2008.06.001

Sato, W., Kochiyama, T., Uono, S., Matsuda, K., Usui, K., Inoue, Y., et al. (2011). Rapid amygdala γ oscillations in response to fearful facial expressions. *Neuropsychologia* 49, 612–617. doi: 10.1016/j.neuropsychologia.2010. 12.025

Sato, W., Kochiyama, T., Uono, S., Matsuda, K., Usui, K., Inoue, Y., et al. (2012a). Temporal profile of amygdala γ oscillations in response to faces. *J. Cogn. Neurosci.* 24, 1420–1433. doi: 10.1162/jocn_a_00142

Sato, W., Toichi, M., Uono, S., and Kochiyama, T. (2012b). Impaired social brain network for processing dynamic facial expressions in autism spectrum disorders. *BMC Neurosci.* 13:99. doi: 10.1186/1471-2202-13-99

Sato, W., Kochiyama, T., Uono, S., and Yoshikawa, S. (2010). Amygdala integrates emotional expression and gaze direction in response to dynamic facial expressions. *Neuroimage* 50, 1658–1665. doi: 10.1016/j.neuroimage.2010.01. 049

Sato, W., Kochiyama, T., Uono, S., Yoshimura, S., and Toichi, M. (2017a). Neural mechanisms underlying conscious and unconscious gaze-triggered attentional orienting in autism spectrum disorder. *Front. Hum. Neurosci.* 11:339. doi: 10.3389/fnhum.2017.00339

Sato, W., Kochiyama, T., Uono, S., Yoshikawa, S., and Toichi, M. (2017b). Direction of amygdala-neocortex interaction during dynamic facial expression processing. *Cereb. Cortex* 27, 1878–1890. doi: 10.1093/cercor/bhw036

Sato, W., and Uono, S. (2019). The atypical social brain network in autism: advances in structural and functional MRI studies. *Curr. Opin. Neurol.* 32, 617–621. doi: 10.1097/wco.0000000000000713

Sato, W., Uono, S., and Toichi, M. (2013). Atypical recognition of dynamic changes in facial expressions in autism spectrum disorders. *Res. Autism Spectr. Disord.* 7, 906–912. doi: 10.1016/j.rasd.2013.04.008

Sato, W., and Yoshikawa, S. (2004). The dynamic aspects of emotional facial expressions. *Cogn. Emot.* 18, 701–710. doi: 10.1080/02699930341000176

Sato, W., and Yoshikawa, S. (2007a). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition* 104, 1–18. doi: 10.1016/j.cognition.2006. 05.001

Sato, W., and Yoshikawa, S. (2007b). Enhanced experience of emotional arousal in response to dynamic facial expressions. *J. Nonverb. Behav.* 31, 119–135. doi: 10.1007/s10919-007-0025-7

Schopler, E., Reichler, R. J., and Renner, B. R. (1986). *The Childhood Autism Rating Scale (CARS): For Diagnostic Screening and Classification of Autism.* New York, NY: Irvington.

Sturm, V., Fricke, O., Bührle, C. P., Lenartz, D., Maarouf, M., Treuer, H., et al. (2013). DBS in the basolateral amygdala improves symptoms of autism and related self-injurious behavior: a case report and hypothesis on the pathogenesis of the disorder. *Front. Hum. Neurosci.* 6:341. doi: 10.3389/fnhum. 2012.00341

Tottenham, N., Hertzig, M. E., Gillespie-Lynch, K., Gilhooly, T., Millner, A. J., and Casey, B. J. (2014). Elevated amygdala response to faces and gaze aversion in autism spectrum disorder. *Soc. Cogn. Affect. Neurosci.* 9, 106–117. doi: 10.1093/scan/nst050

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Uono, S., Sato, W., and Toichi, M. (2014). Reduced representational momentum for subtle dynamic facial expressions in individuals with autism spectrum disorders. *Res. Autism Spectr. Disord.* 8, 1090–1099. doi: 10.1016/j.rasd.2014. 05.018

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci.* 9, 585–594. doi: 10.1016/j.tics.2005.10.011

Vuilleumier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45, 174–194. doi: 10.1016/j.neuropsychologia. 2006.06.003

Williams, L. M., Das, P., Liddell, B. J., Kemp, A. H., Rennie, C. J., and Gordon, E. (2006). Mode of functional connectivity in amygdala pathways dissociates level of awareness for signals of fear. *J. Neurosci.* 26, 9264–9271. doi: 10.1523/JNEUROSCI.1016-06.2006

Williams, J. H., Massaro, D. W., Peel, N. J., Bosseler, A., and Suddendorf, T. (2004). Visual-auditory integration during speech imitation in autism. *Res. Dev. Disabil.* 25, 559–575. doi: 10.1016/j.ridd.2004.01.008

Xu, L., Ma, X., Zhao, W., Luo, L., Yao, S., and Kendrick, K. M. (2015). Oxytocin enhances attentional bias for neutral and positive expression faces in individuals with higher autistic traits. *Psychoneuroendocrinology* 62, 352–358. doi: 10.1016/j.psyneuen.2015.09.002

Yirmiya, N., Kasari, C., Sigman, M., and Mundy, P. (1989). Facial expressions of affect in autistic, mentally retarded and normal children. *J. Child Psychol. Psychiatry* 30, 725–735. doi: 10.1111/j.1469-7610.1989.tb00785.x

Yoshimura, S., Sato, W., Uono, S., and Toichi, M. (2015). Impaired overt facial mimicry in response to dynamic facial expressions in high-functioning autism spectrum disorders. *J. Autism Dev. Disord.* 45, 1318–1328. doi: 10.1007/s10803-014-2291-7

Zeidman, P., Jafarian, A., Seghier, M. L., Litvak, V., Cagnan, H., Price, C. J., et al. (2019a). A guide to group effective connectivity analysis, part 2: second level analysis with PEB. *Neuroimage* 200, 12–25. doi: 10.1016/j.neuroimage.2019.06.032

Zeidman, P., Jafarian, A., Corbin, N., Seghier, M. L., Razi, A., Price, C. J., et al. (2019b). A guide to group effective connectivity analysis, part 1: first level analysis with DCM for fMRI. *Neuroimage*. 200, 174–190. doi: 10.1016/j.neuroimage.2019.06.031

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership