ADVANCES AND OBSTACLES IN CONTEMPORARY NONVERBAL COMMUNICATION RESEARCH

EDITED BY: Miles L. Patterson, Norah E. Dunbar, Marianne Schmid Mast and José-Miguel Fernández-Dols

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-359-2
DOI 10.3389/978-2-88971-359-2

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: frontiersin.org/about/contact

ADVANCES AND OBSTACLES IN CONTEMPORARY NONVERBAL COMMUNICATION RESEARCH

Topic Editors:
Miles L. Patterson, University of Missouri–St. Louis, United States
Norah E. Dunbar, University of California Santa Barbara, United States
Marianne Schmid Mast, University of Lausanne, Switzerland
José-Miguel Fernández-Dols, Autonomous University of Madrid, Spain

Citation: Patterson, M. L., Dunbar, N. E., Mast, M. S., Fernández-Dols, J.-M., eds. (2022). Advances and Obstacles in Contemporary Nonverbal Communication Research. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-359-2

Table of Contents

05 ***Editorial: Advances and Obstacles in Contemporary Nonverbal Communication Research***
Miles L. Patterson, Norah E. Dunbar, Marianne Schmid Mast and J. M. Fernandez-Dols

08 ***Unraveling the Misconception About Deception and Nervous Behavior***
Aldert Vrij and Ronald P. Fisher

16 ***How Our Gaze Reacts to Another Person's Tears? Experimental Insights Into Eye Tracking Technology***
Alfonso Picó, Raul Espert and Marien Gadea

29 ***More Data, Please: Machine Learning to Advance the Multidisciplinary Science of Human Sociochemistry***
Jasper H. B. de Groot, Ilja Croijmans and Monique A. M. Smeets

38 ***Misconceptions About Nonverbal Cues to Deception: A Covert Threat to the Justice System?***
Vincent Denault

42 ***Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport***
Gary Bente, Eric Novotny, Daniel Roth and Ahmad Al-Issa

58 ***Corrigendum: Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport***
Gary Bente, Eric Novotny, Daniel Roth and Ahmad Al-Issa

60 ***A Novel Test of the Duchenne Marker: Smiles After Botulinum Toxin Treatment for Crow's Feet Wrinkles***
Nancy Etcoff, Shannon Stock, Eva G. Krumhuber and Lawrence Ian Reed

67 ***Anger and Sadness Expressions Situated in Both Positive and Negative Contexts: An Investigation in South Korea and the United States***
Sunny Youngok Song, Alexandria M. Curtis and Oriana R. Aragón

90 ***Expression Authenticity: The Role of Genuine and Deliberate Displays in Emotion Perception***
Mircea Zloteanu and Eva G. Krumhuber

96 ***Is Technology Enhancing or Hindering Interpersonal Communication? A Framework and Preliminary Results to Examine the Relationship Between Technology Use and Nonverbal Decoding Skill***
Mollie A. Ruben, Morgan D. Stosic, Jessica Correale and Danielle Blanch-Hartigan

106 ***Nonverbal Behaviors "Speak" Relational Messages of Dominance, Trust, and Composure***
Judee K. Burgoon, Xinran Wang, Xunyu Chen, Steven J. Pentland and Norah E. Dunbar

123 ***Who's Leading This Dance?: Theorizing Automatic and Strategic Synchrony in Human-Exoskeleton Interactions***
Gavin Lawrence Kirkwood, Christopher D. Otmar and Mohemmad Hansia

137 ***Implications for Emotion: Using Anatomically Based Facial Coding to Compare Emoji Faces Across Platforms***
Jennifer M. B. Fugate and Courtny L. Franco

158 ***The Expressive Triad: Structure, Color, and Texture Similarity of Emotion Expressions Predict Impressions of Neutral Faces***
Daniel N. Albohn and Reginald B. Adams Jr.

176 ***Non-verbal Adaptation to the Interlocutors' Inner Characteristics: Relevance, Challenges, and Future Directions***
Valerie Carrard

182 ***Capturing Behavior in Small Doses: A Review of Comparative Research in Evaluating Thin Slices for Behavioral Measurement***
Nora A. Murphy and Judith A. Hall

195 ***Nonverbal Social Sensing: What Social Sensing Can and Cannot Do for the Study of Nonverbal Behavior From Video***
Laetitia Aurelie Renier, Marianne Schmid Mast, Nele Dael and Emmanuelle Patricia Kleinlogel

frontiers in Psychology

EDITORIAL
published: 12 August 2021
doi: 10.3389/fpsyg.2021.731334

Editorial: Advances and Obstacles in Contemporary Nonverbal Communication Research

Miles L. Patterson[1]*, *Norah E. Dunbar*[2], *Marianne Schmid Mast*[3] and *J. M. Fernandez-Dols*[4]

[1] Department of Psychological Sciences, University of Missouri-St. Louis, St. Louis, MO, United States, [2] Department of Communication, University of California, Santa Barbara, Santa Barbara, CA, United States, [3] Department of Organizational Behavior, Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland, [4] Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain

Keywords: nonverbal communication, facial expression, culture, technology, deception

Editorial on the Research Topic

Advances and Obstacles in Contemporary Nonverbal Communication Research

OPEN ACCESS

Edited and reviewed by:
M. Teresa Anguera,
University of Barcelona, Spain

Correspondence:
Miles L. Patterson
miles_patterson@umsl.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *26 June 2021*
Accepted: *13 July 2021*
Published: *12 August 2021*

Citation:
Patterson ML, Dunbar NE, Mast MS and Fernandez-Dols JM (2021) Editorial: Advances and Obstacles in Contemporary Nonverbal Communication Research. Front. Psychol. 12:731334. doi: 10.3389/fpsyg.2021.731334

For centuries, speculation about the meaning and impact of nonverbal behavior has been common in literature, philosophy, and science (see Knapp, 2006 for a historical review). In the latter half of the nineteenth century, Darwin's 1872 *The expression of the emotions in man and animals* work was particularly instrumental in focusing attention on expressive behavior. Nevertheless, sustained and systematic empirical research on nonverbal communication was not widespread until the middle of the twentieth century. Examples of its diverse roots can be found in anthropology (Birdwhistell, 1955, 1970; Hall, 1959, 1966), sociology (Goffman, 1959, 1963), and psychology (Sommer, 1959, 1962; Exline, 1963; Ekman, 1964, 1965). Since that time, literally tens of thousands of articles and hundreds of scholarly books have expanded our knowledge of the nonverbal communication and prompted new and interesting questions about its scope and functions. This acceleration of publications, especially in recent years, provides an appropriate opportunity to examine the current landscape of nonverbal communication research and to provide an outlook into future areas and topics.

In laying the foundation for our "Advances and Obstacles" issue, it is worth noting some of the important topics addressed in current research. For example, we are learning more about the accuracy of pervasive automatic judgments of others' appearance and behavior (Todorov, 2017; Murphy et al., 2019). But automatic judgments can also facilitate prejudice and discrimination, as studies of implicit bias show (Richeson and Shelton, 2005). The long-held view that facial expressions necessarily reflect underlying emotions (Ekman, 1982) is now being challenged. One alternative view proposes that facial behaviors are adaptive and adaptable tools for social influence, rather than universal uniform expressions of basic emotions (Crivelli and Fridlund, 2018). The relative merits of these opposing views also have relevance for understanding nonverbal communication in a variety of settings, including the justice system (e.g., detecting deception), policy decisions, national security, and clinical settings (Denault et al., 2020). Research on cultural differences in nonverbal communication provides insight into cultural dynamics and is relevant for reducing inter-group conflict and facilitating cooperation (Matsumoto and Hwang, 2016). Exciting recent work in behavioral neuroscience examines the neural correlates of nonverbal communication (e.g., Jacob et al., 2014, Lindenberg et al., 2012; Arioli and Canessa, 2019).

In the present digital age, rapidly-evolving communication technologies might seem to displace the more mundane role of face-to-face nonverbal communication in everyday life. The continuing expansion of social media, artificial intelligence systems, virtual reality, and social robots, however, is not replacing, but rather extending nonverbal communication to new platforms (see also von der Pütten et al., 2010; Hasler and Friedman, 2012; Küster et al., 2015; Patterson, 2019; Blunden and Brodsky, 2021). As a result, this is a time of expanding research and theory into new domains. Nevertheless, the opportunities provided by the new technologies must be weighed against the ease of spreading misleading and deceptive images that affect our trust in their content (e.g., Tolosana et al., 2020).

Consequently, this is an appropriate time to (1) examine more fully the questions driving current research and theory, (2) weigh the obstacles to a broader understanding of nonverbal communication, and (3) consider the potential opportunities for advancing future research on nonverbal communication. The collection of articles here is testimony to the diversity of nonverbal communication research in addressing these goals.

Many of the 17 articles in this issue focus in some fashion on methodological advances and their potential limitations in new directions for research. Murphy and Hall review the thin-slice method with a particular focus on its reliability and validity in representing sustained behavioral sequences. The article proposes that deciding if and when to employ thin-slice measurement should focus on its broader representativeness for behavior, predictive validity for variables or constructs beyond the sampled behavior, and assessing how the length of the sampled thin-slices affects the accuracy of interpersonal judgments.

Three articles deal with new technologies that include machine learning and the application of algorithms to the scoring and evaluation of nonverbal stimuli. Albohn and Adams applied computer vision algorithms to the structure, color, and texture of faces to predict gender-stereotypic impressions. In addition, the computer impressions were similar to those made by human participants. The broader issue of the opportunities and limitations of machine learning were addressed in two other articles. Burgoon et al. used machine learning and automated analysis to examine the role of dominance-submission, composure-nervousness, and trust-mistrust in relational communication. They also discussed the potential benefits of the new techniques in simplifying the study of nonverbal communication. Renier et al. also recognize the utility of applying algorithms in machine learning techniques in analyzing nonverbal behavior. Nevertheless, they caution that automated nonverbal coding can be as biased as human coding and can be limited to the particular context for the behavior.

Several empirical articles focus on a variety of issues related to the encoding and decoding of expressive displays. Bente et al. developed a motion capture and character animation method eliminating cultural and gender appearance cues that can precipitate stereotypic biased judgments. In the absence of visual culture and gender cues, they found that female dyads were rated significantly higher on rapport and that this difference was greater in Arab dyads than in German dyads. Song et al. examined anger and sadness expressions in South Korean and American samples. They found that in both cultures, anger and sadness displays signaled both negative and positive underlying states. Fugate and Franco studied the correspondence between human facial expressions and analogous emoji faces. They found that the majority of emoji faces did not conform to human emotional expressions, even though the anatomical codes for the two types of faces were generally shared. Etcoff et al. investigated the effects of botulinum toxin treatments on the perceptions of pre- and post-treatment smiles. Pre-treatment smiles were rated as more felt, more spontaneous, and happier than post-treatment smiles. Although post-treatment patients were rated as looking younger, they were not judged as more attractive than pre-treatment patients. The effects of tears on visual attention to faces and on subsequent judgments of emotional intensity were the focus of an experimental study by Pico et al. An eye tracking method provided evidence for tears being a magnet for visual attention that, in turn, facilitated perceptions of greater emotional intensity. Ruben et al. addressed the issue of whether technology use enhanced or hindered nonverbal decoding skill. Overall screen time was unrelated to objective measures of decoding skill, but how participants used their screen time was related to decoding skill. Active users (e.g., posting content) performed worse on decoding skill measures, but passive users performed better.

Various issues dealing with authenticity/deception in expressive behavior are the focus of three other articles. Zloteanu and Krumhuber discuss different perspectives on facial displays in the context of increasing evidence contradicting the traditional view that reliable facial muscle movements signal distinct emotional experiences. They discuss spontaneous vs. posed expressions and advocate a functional approach to expressions as neurophysiological states and communicative signals. Vrij and Fisher's article addresses the common assumption that liars display more nervous behaviors than truth tellers. They provide evidence that liars do not show more nervous behaviors. Consequently, observers who focus on such nervous behaviors are likely to do poorly in detecting deception. On a similar theme, Denault discusses the negative consequences of depending on unreliable nonverbal cues for detecting deception. Specifically, in the justice system, judges, and juries are vulnerable to the common, but scientifically discredited, assumption of valid nonverbal indicators of deception. As a result, assessments of witness credibility can be distorted, with detrimental effects on trial outcomes.

The last four articles provide a range of commentaries on approaches to future research. Matsumoto and Hwang advocated for a multimodal approach to research and theory. That is, increased attention to clusters of nonverbal behavior, rather than a single channel at a time, can facilitate our understanding of underlying mental states. Carrard addresses a similar theme of linking interactants' inner preferences and expectations to patterns of nonverbal behavior. That is, nonverbal communication should be viewed as an adaptive process driven by actors' inner characteristics. DeGroot et al. focus on the emerging and important research on the diverse effects of olfaction on a wide variety of interpersonal processes, including identity, emotion, and mate selection. The authors argue that pursuing effectively the wide range of important issues in olfaction requires an integration of the psychology and chemistry disciplines into a new field of "sociochemistry." Finally, Kirkwood et al. extend the process of interpersonal synchrony from the nonverbal mimicry between partners to individuals' synchrony with wearable exoskeletons. Recent technological advances in wearable robots are designed to augment a user's strength and mobility. The authors discuss the utility of the Interpersonal Adaptation Theory in facilitating research maximizing human-exoskeleton synchrony.

In conclusion, we hope that this interesting set of articles provides an informative window into some of the diverse issues driving current research on nonverbal communication. The advances in research discussed in many of these articles are often responses to existing obstacles or discrepancies in research. Other articles are focused more on identifying the new obstacles yet to receive attention that, in turn, will stimulate new research. Thus, the present issue provides a vehicle for facilitating our understanding of nonverbal communication and appreciating where future research may be headed.

AUTHOR CONTRIBUTIONS

All of the authors have contributed to this work and approved its publication.

ACKNOWLEDGMENTS

The authors are grateful for the efforts of all of the contributors to this special issue and for the very helpful comments and insights of the manuscript reviewers.

REFERENCES

Arioli, M., and Canessa, N. (2019). Neural processing of social interaction: coordinate-based meta-analytic evidence from human neuroimaging studies. *Hum. Brain Mapp.* 40, 5354–5369. doi: 10.1002/hbm.24627

Birdwhistell, R. L. (1955). Background to kinesics. *ETC Rev. General Semant.* 13, 10–15.

Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia, PA: University of Pennsylvania Press. doi: 10.9783/9780812201284

Blunden, H., and Brodsky, A. (2021). Beyond the emoticon: are there unintentional cues of emotion in email? *Pers. Soc. Psychol. Bull.* 47, 565–579. doi: 10.1177/0146167220936054

Crivelli, C., and Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends iCognitive Sci.* 22, 388–399. doi: 10.1016/j.tics.2018.02.006

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

Denault, V., Plusquellec, P., Jupe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., et al. (2020). The analysis of nonverbal communication: the dangers of pseudoscience in security and justice contexts. *Anuario de Psicol. Jurídica* 30, 1–12. doi: 10.5093/apj2019a9

Ekman, P. (1964). Body positions, facial expression and verbal behavior during interviews. *J. Abnorm. Soc. Psychol.* 68, 295–301. doi: 10.1037/h0040225

Ekman, P. (1965). Differential communication of affect by head and body cues. *J. Pers. Soc. Psychol.* 2, 726–735. doi: 10.1037/h0022736

Ekman, P. (1982). *Emotion in the Human Face (2nd ed.)*. New York, NY: Cambridge University Press.

Exline, R. V. (1963). Explorations in the process of person perception: visual interaction in relation to competition, sex, and need for affiliation. *J. Pers.* 31, 1–20. doi: 10.1111/j.1467-6494.1963.tb01836.x

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Anchor.

Goffman, E. (1963). *Behavior in Public Places*. New York, NY: Free Press.

Hall, E. T. (1959). *The Silent Language*. New York, NY: Doubleday.

Hall, E. T. (1966). *The Hidden Dimension*. New York, NY: Doubleday.

Hasler, B. S., and Friedman, D. A. (2012). Sociocultural conventions in avatar-mediated nonverbal communication: a cross-cultural analysis of virtual proxemics. *J. Intercultural Commun. Res.* 41, 238–259. doi: 10.1080/17475759.2012.728764

Jacob, H., Brück, C., Domin, M., Lotze, M., and Wildgruber, D. (2014). I can't keep your face and voice out of my head: neural correlates of an attentional bias toward nonverbal emotional cues. *Cerebral Cortex* 24, 1460–1473. doi: 10.1093/cercor/bhs417

Knapp, M. L. (2006). "An historical overview of nonverbal research," in *The Sage Handbook of Nonverbal Communication*, eds V. Manusov and M. L. Patterson (Thousand Oaks, CA: Sage Publications), 3–19.

Küster, D., Krumhuber, E., and Kappas, A. (2015). "Nonverbal behavior online: a focus on interactions with and via artificial agents and avatars," in *The Social Psychology of Nonverbal Communication*, eds A. Kostíc and D. Chadee (London: Palgrave Macmillan), 272–302. doi: 10.1057/9781137345868_13

Lindenberg, R., Uhlig, M., Scherfeld, D., Schlaug, G., and Seitz, R. J. (2012). Communication with emblematic gestures: shared and distinct neural correlates of expression and reception. *Human Brain Mapp.* 33, 812–823. doi: 10.1002/hbm.21258

Matsumoto, D., and Hwang, H. C. (2016). "The cultural bases of nonverbal communication," in *APA Handbook of Nonverbal Communication*, eds D. Matsumoto, H. C. Hwang, and M. G. Frank (Washington, DC: American Psychological Association), 77–101. doi: 10.1037/14669-004

Murphy, N. A., Hall, J. A., Ruben, M. A., Frauendorfer, D., Schmid Mast, M., Johnson, K. E., et al. (2019). Predictive validity of thin-slice behavior nonverbal behavior from social interactions. *Pers. Soc. Psychol. Bull.* 45, 983–993. doi: 10.1177/0146167218802834

Patterson, M. L. (2019). A systems model of dyadic nonverbal communication. *J. Nonverbal Behav.* 43, 111–132. doi: 10.1007/s10919-018-00292-w

Richeson, J. A., and Shelton, J. N. (2005). Brief report: thin slices of racial bias. *J. Nonverbal Behav.* 29, 75–86. doi: 10.1007/s10919-004-0890-2

Sommer, R. (1959). Studies in personal space. *Sociometry* 22, 247–260. doi: 10.2307/2785668

Sommer, R. (1962). The distance for comfortable conversation: a further study. *Sociometry* 25, 111–116. doi: 10.2307/2786041

Todorov, A. (2017). *Face Value: The Irresistible Influence of First Impressions*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400885725

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., and Ortega-Garcia, J. (2020). Deepfakes and beyond: a survey of face manipulation and fake detection. *Informat. Fusion* 64, 131–148. doi: 10.1016/j.inffus.2020.06.014

von der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S.-H. (2010). "It doesn't matter what you are!" Explaining social effects of agents and avatars. *Comput. Human Behav.* 26, 1641–1650. doi: 10.1016/j.chb.2010.06.012

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

PERSPECTIVE
published: 18 June 2020
doi: 10.3389/fpsyg.2020.01377

Unraveling the Misconception About Deception and Nervous Behavior

Aldert Vrij[1] and Ronald P. Fisher[2]*

[1] Department of Psychology, University of Portsmouth, Portsmouth, United Kingdom, [2] Department of Psychology, Florida International University, Miami, FL, United States

In this article, we attempt to unravel the misconception about deception and nervous behavior. First we will cite research demonstrating that observers believe lie tellers display more nervous behaviors than truth tellers; that observers pay attention to nervous behaviors when they attempt to detect deception; and that lie tellers actually feel more nervous than truth tellers. This is all in alignment with a lie detection approach based on spotting nervous behaviors. We then will argue that the next, vital, step is missing: Research has found that lie tellers generally do *not* display more than truth tellers the nervous behaviors laypersons and professionals appear to focus on. If observers pay attention to nervous behaviors but lie tellers do not come across as being nervous, lie detection performance is expected to be poor. Research has supported this claim. We finally discuss ideas for research into lie detection based on non-verbal behaviors.

Keywords: deception, non-verbal behavior, lie detection, cues of nervousness, illusion of transparency, perceived correlates of deception, actual correlates of deception

OPEN ACCESS

Edited by:
Miles L. Patterson,
University of Missouri–St. Louis,
United States

Reviewed by:
Elise Fenn,
California State University, Northridge,
United States
Dino Krupiæ,
Josip Juraj Strossmayer University
of Osijek, Croatia

****Correspondence:***
Aldert Vrij
aldert.vrij@port.ac.uk

Specialty section:
This article was submitted to
Personality and Social Psychology,
a section of the journal
Frontiers in Psychology

Received: *26 March 2020*
Accepted: *22 May 2020*
Published: *18 June 2020*

Citation:
Vrij A and Fisher RP (2020)
Unraveling the Misconception About
Deception and Nervous Behavior.
Front. Psychol. 11:1377.
doi: 10.3389/fpsyg.2020.01377

UNRAVELING THE MISCONCEPTION ABOUT DECEPTION AND NERVOUS BEHAVIOR

Distinguishing between truth tellers and lie tellers is an important task for a wide range of practitioners, including police officers, intelligence officers, and security personnel. It can be achieved through measuring (1) physiological responses, (2) brain activity, (3) non-verbal behaviors or (4) speech content. Of these methods, non-verbal lie detection is particularly popular, amongst other reasons, because it can be carried out all the time. It does not require equipment (needed for measuring physiological responses or brain activity) and does not require the target person to speak (needed for measuring speech content). The overwhelming view amongst practitioners (and laypersons) is that lie tellers display more nervous behaviors than truth tellers. In this article we provide evidence that this is a misconception, which could explain why people typically obtain poor accuracy rates when they make veracity assessments based on nervous behaviors.

We often refer to Mann et al. (2020) in this article, because the two experiments reported in that article demonstrate several of the points we want to make. See **Appendix 1** for a synopsis of Mann et al. (2020) procedure and results relevant for the present article.

BELIEF: LIE TELLERS DISPLAY MORE NERVOUS BEHAVIORS THAN TRUTH TELLERS

The belief that lie tellers display more nervous behaviors than truth tellers is well established. The most thorough investigation into beliefs about deception was carried out by Charles Bond (The Global Deception Team, 2006). Researchers from 58 countries collected data from 20 males and 20 females of their country. The participants were asked to answer the question: "How can you tell when people are lying?" They mentioned 103 different beliefs, four of which were given by more than 25% of the participants. Most people (64% of the participants) believed that lie tellers display gaze aversion and this belief was the most frequently reported in 51 out of 58 countries. The second strongest belief was "nervousness," which was mentioned by 28% of the participants, followed by incoherent speech (25%) and body movements (25%). All four beliefs relate to nervousness and two beliefs (gaze aversion and body movements) relate exclusively to non-verbal behavior.

Apart from laypersons, also practitioners often associate nervous behaviors with deception (Strömwall et al., 2004; Vrij and Granhag, 2007; Vrij et al., 2018). In one study, 99 British police officers were asked to answer the question: "What verbal or non-verbal cues do you use to decide whether another person is lying or telling the truth?" (Mann et al., 2004). A total of 30 different beliefs emerged, of which two were mentioned by at least 25% of the police officers: Gaze aversion (mentioned by 73% of the police officers) and making body movements (mentioned by 25% of participants).

After the 9/11 terrorist attacks, the United States Transportation Security Administration (TSA) introduced SPOT (Screening of Passengers by Observation Techniques). In SPOT, trained individuals called Behavior Detection Officers (BDOs) observe passengers at airports with the aim to identify security threats. Cues that BDOs were taught to pay attention to included cues to nervousness such as avoiding eye contact, looking down, emitting a strong body odor, and covering the mouth with the hand when speaking (Denault et al., 2020).

The belief that lie tellers display more nervous behaviors than truth tellers also appears in police manuals (Vrij and Granhag, 2007). In these manuals deceptive behavior has been described as: Problem with eye contact, touching the nose, and restless foot and leg movements (Gordon and Fleisher, 2011); avoiding eye contact, frequent posture changes, grooming gestures, and placing hand over mouth/eyes (Inbau et al., 2013); rubbing the eyes, avoiding eye contact, and covering/rubbing the ears (Macdonald and Michaud, 1992); and moving the chair, abrupt and jerky behavior, problem with fine motor coordination, cold and clammy hands, using hands to cover mouth, and failure to maintain eye contact (Zulawski and Wicklander, 1993). See Vrij and Granhag (2007) for a more detailed discussion of the views expressed in police manuals.

Sometimes the beliefs of professionals and laypersons were investigated within the same study so that their answers could be compared directly (e.g., Akehurst et al., 1996; Vrij and Semin, 1996; Vrij et al., 2006). None of these studies found consistent differences among different groups of professionals; neither did the beliefs of professionals differ from those of laypersons. However, a different picture emerged for prisoners whose beliefs differed somewhat from the beliefs of both professionals and laypersons. Generally speaking, prisoners endorse the "lie tellers display nervous behaviors" beliefs less than non-prisoners do (Vrij and Semin, 1996; Granhag et al., 2004). For example, non-prisoners thought that deception is associated with more hand/finger movements, more trunk movements and more position shifts, whereas prisoners thought that such behaviors are not associated with deception (Vrij and Semin, 1996).

Finally, in two surveys, laypersons (Masip et al., 2012b) and law enforcement personnel (Masip et al., 2012a) completed a "beliefs about cues to deception" questionnaire based on the Behavior Analysis Interview, a lie detection method that relies, in part, on non-verbal cues of nervousness. It is popular amongst practitioners in many parts of the world (Inbau et al., 2013). The laypersons and law enforcement personnel expressed similar views (Masip et al., 2012a).

DO PEOPLE PAY ATTENTION TO NERVOUS BEHAVIORS WHEN THEY TRY TO DETECT DECEIT?

The belief that lie tellers display more nervous behaviors than truth tellers is relevant for lie detection, only if people actually pay attention to nervous behaviors when they try to detect deceit. In fact, they do. Vrij (2008) summarized the results of more than 30 studies analyzing the relationship between displaying nervous behaviors and being judged as deceptive. In those studies, truth tellers and lie tellers are videotaped and their non-verbal and verbal behavior is coded. Observers are shown those videotapes with the request to indicate after each fragment whether they think the person was telling the truth or lying. The observers' judgments are then correlated with the target persons' actual non-verbal and verbal cues displayed in the video fragments. The results showed that judged deception was associated with more gaze aversion and more movements (more fidgeting, hand/finger, leg/foot and trunk movements, shifting position) again suggesting that people pay attention to nervous behaviors when they try to detect deception.

A meta-analysis addressing the relationship between displaying nervous behaviors and being judged as deceptive showed similar results as Vrij's (2008) review (Hartwig and Bond, 2011). Again, deception was associated with more gaze aversion, more fidgeting and more postural shifts. It was also associated with the general concept "nervousness," a concept not examined by Vrij (2008)[1].

In Mann et al. (2020) observers (laypersons) saw videotapes of laboratory experiments in which participants did or did

[1] Hartwig and Bond (2011, Table 1) provide for 66 cues the actual and perceived relationship with deception, but they do not identify which of them are cues to nervousness.

not smuggle an object during a ferry-crossing. The observers were asked, amongst other questions, to indicate for each videotape whether the person was smuggling (yes/no) and to what extent the person came across as feeling nervous. Strong positive correlations between judging the person as a smuggler and perceiving the person as feeling nervous were found in Experiment 1 ($r = 0.67$) and Experiment 2 ($r = 0.62$)[2].

DO LIE TELLERS FEEL MORE NERVOUS THAN TRUTH TELLERS?

The implication of observers paying attention to nervous behaviors is that they think that lie tellers will feel more nervous than truth tellers. This assumption is supported. In Mann et al. (2020) the smugglers and non-smugglers were asked after completing their mission how nervous they felt. Smugglers felt considerably more nervous than non-smugglers in Experiment 1 ($d = 0.73$) and Experiment 2 ($d = 0.62$).

In a typical laboratory experiment, truth tellers and lie tellers are interviewed about an alleged experience. Sometimes they are asked afterward how nervous they felt during the interview. In his review of ten studies, Vrij (2008) concluded that lie tellers typically reported that they felt more nervous than truth tellers. Ekman (1985/2001; Ekman and Friesen, 1969) provided three explanatory mechanisms to account for lie tellers' feelings of nervousness: Lie tellers, more than truth tellers, experience fear (of getting caught), guilt (of committing a morally disputable act) or duping delight (excitement of the opportunity to fool someone).

The finding that lie tellers generally feel more nervous than truth tellers does not mean that truth tellers do not feel nervous. They may feel nervous in interview settings, because they may experience fear (Bond and Fahey, 1987; Ofshe and Leo, 1997). Rather than experiencing fear of being detected (lie tellers), truth tellers may experience fear of not being believed because that could have serious consequences for them (further interrogations by police, taken aside by investigators at ferries or airports etc.).

DO LIE TELLERS DISPLAY MORE NERVOUS BEHAVIORS THAN TRUTH TELLERS?

Both arguments presented so far – (1) observers pay attention to signs of nervousness when they try to detect deception and (2) lie tellers feel more nervous than truth tellers- work in favor of lie detection based on spotting nervous behaviors. However, one more step is required for such a method to work: Do lie tellers actually display more nervous behaviors than truth tellers? Only when lie tellers do so, can a lie detection method based on spotting nervous behaviors be successful. Research has found that lie tellers generally do *not* display more than truth tellers the nervous behaviors laypersons and professionals appear to focus on. A meta-analysis of deceptive behavior has shown that truth tellers and lie tellers display similar gaze behavior patterns and that lie tellers make *fewer* rather than *more* movements than truth tellers (DePaulo et al., 2003). This finding is sometimes challenged by practitioners or scientists who claim that these findings are based on laboratory-based studies where the stakes are low (Buckley, 2012; Frank and Svetieva, 2012). That is, there are for lie tellers no strong negative consequences associated with being detected or strong positive consequences associated with remaining undetected. They claim that in high-stakes situations the findings will be different. However, research does not support this claim. In their meta-analysis, Hartwig and Bond (2014) compared the behaviors displayed by truth tellers and lie tellers in (1) low-stakes situations and in (2) high-stakes situations. The same pattern of results emerged in both situations. It seems reasonable that lie tellers will display more nervous behaviors in high-stakes settings than in low-stakes settings; however, so will truth tellers, making the difference between the two groups unchanged.

In a truly high-stakes field-study, videotaped interviews with interviewees suspected of murder, rape, and arson were analyzed (Mann et al., 2002). Similar to the DePaulo et al. (2003) findings, the suspects showed no different gaze patterns when truth telling or lying, but they moved less when they lied. A selection of these videotapes was shown to police officers who were asked to indicate, amongst other factors, to what extent the suspects appeared nervous. Suspects who told the truth appeared more nervous than those who lied (Mann and Vrij, 2006). A similar finding was obtained in Mann et al. (2020). In Experiment 1, smugglers and non-smugglers made an equally nervous impression on observers, but in Experiment 2 smugglers made a *less* nervous impression than non-smugglers, representing a substantial effect size ($d = 0.60$).

Why are lie tellers more nervous than truth tellers, but are perceived by others as being equally or less nervous? In brief, we suggest that lie tellers actively attempt to alter their overt behavior to appear truthful (e.g., by minimizing signs of nervousness) whereas truth tellers are less concerned about others' perceptions of them, and so they do not alter their behavior (Hocking and Leathers, 1980). In addition, lie tellers experience higher cognitive load in interviews than truth tellers (Vrij, 2014) and increased cognitive demand automatically reduces the amount of movements people make (Shallice and Burgess, 1994).

Both truth tellers and lie tellers believe their inner states shine through (Kassin, 2005; Granhag et al., 2007) and are knowable by others, the illusion of transparency (Gilovich et al., 1998). As a result, lie tellers cannot take their credibility for granted. They develop strategies to control their non-verbal behavior (Buller and Burgoon, 1996; Colwell et al., 2006; Hartwig et al., 2010) by attempting to avoid displaying behaviors they perceive as suspicious (Hocking and Leathers, 1980). In Mann et al. (2020, Experiment 2) two confederates approached the (non)smugglers

[2]The terms truth tellers and lie tellers refer to people who speak. Since the non-smugglers and smugglers did not speak we do not refer to them as truth tellers and lie tellers but they are the equivalent to non-smugglers and smugglers, respectively. That is, the notion that lie tellers experience more guilt, fear or duping delight than truth tellers (see the next section) equally applies to smugglers versus non-smugglers. Similarly, in the deception literature the terms *guilty* and *innocent* suspects are sometimes used. They are also the equivalent of lie tellers and truth tellers.

on the ferry pretending to be looking for someone. This may have made both smugglers and non-smugglers nervous, but the smugglers –as a result for not taking their credibility for granted- may have tried to suppress displaying signs of nervousness more than the non-smugglers. Consequently, they made a less nervous impression on observers than non-smugglers. Attempting to control behavior is a mentally taxing strategy and could, as such, also automatically result in lie tellers making fewer movements than truth tellers and in a decrease in displaying nervous behaviors. Increased cognitive load leads to fewer hand and arm movements and inhibits fidgety movements (Ekman and Friesen, 1972; Shallice and Burgess, 1994; Ekman, 1997), because cognitive demand results in a neglect of non-verbal behavior, which subsequently reduces overall animation.

The previous and present section can thus be summarized as follows. Although lie tellers feel more nervous than truth tellers, lie tellers' nervousness become less apparent in their behavior than truth tellers' nervousness because (i) lie tellers actively try to avoid displaying signs of nervousness and (ii) cognitive demand automatically suppresses lie tellers' expressions of nervousness.

ACCURACY IN LIE DETECTION WHEN PAYING ATTENTION TO NERVOUS BEHAVIORS

If observers pay attention to nervous behaviors but lie tellers do not seem to come across as nervous, lie detection performance is expected to be poor. That was indeed found in Mann et al. (2020, Experiment 2) in which the observers reported to have paid attention to nervous behaviors whilst the smugglers made a less nervous impression on observers than the non-smugglers (see above). The accuracy rate in distinguishing between truth tellers and lie tellers in that experiment was very low, 39.2%, which was significantly below the level of chance (50%).

Two more experiments addressed the relationship between paying attention to nervous behaviors and accuracy in truth/lie detection directly, both addressing the lie detection approach advocated by Inbau et al. (2013). In their manual, Inbau et al. (2013) reported that lie tellers display a variety of nervous behaviors, including gaze aversion, unnatural posture changes, self-self-adaptors, and placing the hand over the mouth or eyes when they speak. In Mann et al. (2004) police officers watched videotaped fragments of the real-life police-suspect interviews analyzed in Mann et al. (2002) and introduced above. Before starting the lie detection task, the police officers reported what they thought were indicators of deception. Results showed that the more "Inbau-cues" they mentioned, the worse they distinguished between truths and lies. In their experiment, Kassin and Fong (1999) informed half of the observers about the visual cues that Inbau et al. (2013) discussed in their manual. These trained observers performed worse on a subsequent lie detection test than untrained observers. Both studies suggest that paying attention to nervous behaviors identified by Inbau et al. (2013) as indicative of deceit hampers distinguishing between truth tellers and lie tellers. This is not surprising. Blair and Kooi (2004) examined the extent to which these "Inbau-cues" are identified as cues to deception in DePaulo et al. (2003) meta-analysis of the scientific literature. Little evidence was found in support of the Inbau-cues.

Most lie detection studies refer to non-verbal behavior in general rather than to nervous behaviors specifically. These studies show a bleak picture regarding non-verbal lie detection. A meta-analysis examining observers' ability to detect truth and lies, showed an average accuracy rate of 52% in correctly classifying truth tellers and lie tellers when observers could only see (thus not hear) the target person, a percentage similar to chance level (Bond and DePaulo, 2006). Another meta-analysis examined the effect of training in non-verbal cues to deceit (Hauch et al., 2016). It revealed only a small positive effect.

REASONS WHY THE NOTION THAT LIE TELLERS WILL DISPLAY MORE NERVOUS BEHAVIORS EXISTS

The notion that lie tellers will display more nervous behaviors than truth tellers appears to be a misconception. Yet, this notion remains popular. We think that at least three factors contribute to its popularity. First, a moral explanation (Bond and DePaulo, 2006). The belief that lie tellers avert their gaze and increase their movements fits well with the lying-is-bad stereotype. If lying is bad, lie tellers should feel ashamed (which leads to gaze aversion) and should be afraid of getting caught (resulting in gaze aversion and an increase in movements). Second, the accusation explanation (Vrij, 2008). Accusing someone of lying could easily result in a person displaying nervous behaviors out of fear not to be believed. Although this is likely to occur in both lie tellers and truth tellers, the interviewer may subsequently misattribute the suspect's behavior to deception rather than to the accusation (Bond and Fahey, 1987). Third, the media exposure explanation (Hurley et al., 2014). There are many books [e.g., *Lie spotting* (Meyer, 2010) and *Spy the lie* (Houston et al., 2012)] and articles published in popular magazines or on the internet conveying the idea that lie tellers display non-verbal signs of nervousness. There is even a popular TV series "*Lie to Me*" about this idea. In other words, reading about deception or watching television could easily make someone think that nervous behaviors give lie tellers away.

IS THERE A FUTURE OF LIE DETECTION BASED ON NON-VERBAL BEHAVIORS?

This article presented a pessimistic picture of lie detection based on nervous behaviors, making the future of this type of lie detection in our opinion bleak. However, this does not necessarily mean that lie detection based on non-verbal behavior in general has no future. There are arguments against and in favor of lie detection based on non-verbal behavior. The argument against is that four meta-analyses have shown detecting deception based on verbal cues to be superior to non-verbal lie detection. First, a meta-analysis examining observers' ability to distinguish between truth tellers and lie tellers when observing target persons revealed an accuracy rate of 63% when observers could only hear the target person speaking, but an accuracy rate of 52% when observers could only see the target person and could not hear them speak, i.e., no verbal cues available (Bond and DePaulo, 2006). The accuracy when observers could both hear and see the target person was 56%. In addition, individual differences in the ability to distinguish between truth tellers and lie tellers seem to be minute (Bond and DePaulo, 2008). Second, a meta-analysis examining the verbal and non-verbal cues to deception revealed that verbal cues are more diagnostic indicators of deception than are non-verbal cues (DePaulo et al., 2003). Third, a meta-analysis examining the effect of training in verbal or non-verbal indicators of deception cues revealed a medium training effect for verbal lie detection training, but a small effect for non-verbal lie detection training (Hauch et al., 2016).

Two arguments can be made to continue non-verbal lie detection. First, perhaps future research will shed a more positive light on non-verbal cues to deception. Perhaps some non-verbal cues, yet unknown, will be found in the future that do reliably distinguish between truth tellers and lie tellers. Also, perhaps each lie teller gives his/her lies away in different ways (DePaulo et al., 2003; Levine, 2010; Levine et al., 2011). It will be challenging to identify the idiosyncratic pattern for each individual, but perhaps some general distinctions could show meaningful results. For example, there are individual differences in the frequency of lying (DePaulo et al., 1996; Hart et al., 2019) and perhaps frequent and infrequent lie tellers each display identifiable patterns of behavior that differ from each other. Perhaps signs of nervousness emerge in the infrequent lie tellers. Alternatively, some existing non-verbal cues may become diagnostic if they are examined differently. For example, Ekman (1985) has identified different types of smiles, including felt and false smiles. In an experiment, it was found that truth tellers displayed more felt smiles than lie tellers, whereas lie tellers displayed more false smiles than truth tellers (Ekman et al., 1988). Ekman's best known deception work relates to micro-expressions of emotions that he claims lie tellers display: facial expressions that reveal a felt emotion and are suppressed within 1/5th to 1/25th of a second (Ekman, 1985). There is no evidence that micro-expressions of emotions distinguish truth tellers from lie tellers (Burgoon, 2018) or that training in observing such micro-expressions improves lie detection (Jordan et al., 2019). Finally, it is possible that, although no diagnostic cue to deception occurs when each non-verbal cue is examined individually, a diagnostic pattern will arise when they are examined in combination with each other (DePaulo and Morris, 2004). For example, in DePaulo et al. (2003) meta-analysis the impression of being tense was more strongly (albeit, in absolute terms, still weakly) related to deception (d = 0.27) than any of the individual non-verbal cues related to nervousness. Of these individual cues, frequency of pitch was the most diagnostic cue: d = 0.21 [see DePaulo et al. (2003), Table 6 and Table 8 cues based on a larger number of estimates]. It is unclear what the concept "impression of being tense" is made of; it is even not clear whether it contains non-verbal cues, verbal cues or a mixture of both non-verbal and verbal cues. Examining which individual cues contribute toward this concept is perhaps a venue for future research.

Second, sometimes relying on non-verbal cues may be the only lie detection option available, because target persons do not speak and their physiological and brain activity cannot be measured easily. An example is spotting potential wrongdoers in public spaces (airports, train stations, sporting events, concerts etc.). Lie detection in such situation can be very important, because national security can be at stake.

National security concerns are probably the reason why the SPOT program, introduced earlier in this article, has been introduced at United States airports. However, the U.S. Government Accountability Office (2017) recommended to limit SPOT funding due to lack of scientific empirical support for the cues BDOs rely upon. In this context, Denault et al. (2020) reported that SPOT is based on pseudoscientific claims, which could be attractive because (unlike scientific knowledge) they offer immediate and easy solutions to complex challenges.

Which behaviors practitioners should pay attention to at airports or similar settings is unknown because there is no scientific research available (Vrij et al., 2018). Research regarding non-verbal cues to deception almost exclusively concentrates on non-verbal behaviors displayed by interviewees in interview settings. People's non-verbal behaviors are different when they are silent and walk rather than sit and talk. We therefore encourage researchers to examine non-verbal indicators of deception when other methods (such as speech) cannot be used. We do not expect this research to result in a clear-cut list of non-verbal cues that will identify wrongdoers. However, non-verbal lie detection may improve if the context in which the behaviors occur are taken into account, an approach advocated in verbal lie detection with promising results (Blair et al., 2010; Hartwig and Bond, 2011; Street, 2015). Another opportunity would be to increase behavioral differences between wrongdoers and others through -yet unknown- specific interventions (Hartwig and Bond, 2011).

AUTHOR CONTRIBUTIONS

RF raised the idea for the manuscript. AV drafted the manuscript. RF commented on the draft. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Centre for Research and Evidence on Security Threats (ESRC Award: ES/N009614/1).

REFERENCES

Akehurst, L., Köhnken, G., Vrij, A., and Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Appl. Cogn. Psychol.* 10, 461–471. doi: 10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413<3.0

Blair, J. P., and Kooi, B. (2004). The gap between training and research in the detection of deception. *Int. J. Police Sci. Manag.* 6, 77–83. doi: 10.1350/ijps.6.2.77.34465

Blair, J. P., Levine, T., and Shaw, A. (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 judgements. *Pers. Soc. Psychol. Rev.* 10, 214–234. doi: 10.1207/s15327957pspr1003_2

Bond, C. F., and DePaulo, B. M. (2008). Individual differences in judging deception: accuracy and bias. *Psychol. Bull.* 134, 477–492. doi: 10.1037/0033-2909.134.4.477

Bond, C. F., and Fahey, W. E. (1987). False suspicion and the misperception of deceit. *Br. J. Soc. Psychol.* 26, 41–46. doi: 10.1111/j.2044-8309.1987.tb00759.x

Buckley, J. P. (2012). Detecting of deception researchers need to collaborate with experienced practitioners. *J. Appl. Res. Mem. Cogn.* 1, 126–127. doi: 10.1016/j.jarmac.2012.04.002

Buller, D. B., and Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theor.* 6, 203–242. doi: 10.1111/j.14682885.1996.tb00127.x

Burgoon, J. K. (2018). Micro-expressions are not the best way to catch a liar. *Front. Psychol.* 9:1672. doi: 10.3389/fpsyg.2018.01672

Colwell, K., Hiscock-Anisman, C., Memon, A., Woods, D., and Michlik, P. M. (2006). Strategies of impression management among deceivers and truth tellers: how liars attempt to convince. *Am. J. Forensic Psychol.* 24, 31–38.

Denault, V., Plusquellec, P., Jupe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., et al. (2020). The analysis of nonverbal communication: the dangers of pseudoscience in security and justice contexts. *Anu. Psicol. Jurídica* 30, 1–12. doi: 10.5093/apj2019a9

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74

DePaulo, B. M., and Morris, W. L. (2004). "Discerning lies from truths: behavioural cues to deception and the indirect pathway of intuition," in *Deception Detection in Forensic Contexts*, eds P. A. Granhag and L. A. Strömwall (Cambridge: Cambridge University Press), 15-40.

Ekman, P. (1985). *Telling Lies: Clues to Deceit in the Marketplace, Politics and marriage*. New York, NY: W. W. Norton.

Ekman, P. (1997). "Deception, lying, and demeanor," in *States of Mind: American and Post-Soviet Perspectives on Contemporary Issues in Psychology*, eds D. F. Halpern and A. E. Voiskounsky (New York, NY: Oxford University Press), 93–105.

Ekman, P., and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32, 88–106. doi: 10.1080/00332747.1969.11023575

Ekman, P., and Friesen, W. V. (1972). Hand movements. *J. Commun.* 22, 353–374. doi: 10.1111/j.1460-2466.1972.tb00163.x

Ekman, P., Friesen, W. V., and O'Sullivan, M. (1988). Smiles when lying. *J. Pers. Soc. Psychol.* 54, 414–420. doi: 10.1037/0022-3514.54

Frank, M. G., and Svetieva, E. (2012). Lies worth catching involve both emotion and cognition: commentary of Vrij and Granhag (2012). *J. Appl. Res. Mem. Cogn.* 1, 131–133. doi: 10.1016/j.jarmac.2012.04.006

Gilovich, T., Savitsky, K., and Medvec, V. H. (1998). The illusion of transparency: biased assessments of others' ability to read one's emotional states. *J. Pers. Soc. Psychol.* 75, 332–346. doi: 10.1037/0022-3514.75.2.332

Gordon, N. J., and Fleisher, W. L. (2011). *Effective Interviewing and Interrogation Techniques*. Burlington, MA: Academic Press.

Granhag, P. A., Andersson, L. O., Strömwall, L. A., and Hartwig, M. (2004). Imprisoned knowledge: criminals' beliefs about deception. *Legal Criminol. Psychol.* 9, 103–119. doi: 10.1348/135532504322776889

Granhag, P. A., Strömwall, L. A., and Hartwig, M. (2007). The SUE technique: the way to interview to detect deception. *Forensic Update* 88, 25–29. doi: 10.1080/10683160701645181

Hart, C. L., Jones, J. M., Terrizzi, J. A., and Curtis, D. A. (2019). Development of the Lying in Everyday Situations Scale. *Am. J. Psychol.* 132, 343–352. doi: 10.5406/amerjpsyc.132.3.0343

Hartwig, M., and Bond, C. F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychol. Bull.* 137, 643–659. doi: 10.1037/a0023589

Hartwig, M., and Bond, C. F. (2014). Lie detection from multiple cues: a meta-analysis. *Appl. Cogn. Psychol.* 28, 661–667. doi: 10.1002/acp.3052

Hartwig, M., Granhag, P. A., Strömwall, L., and Doering, N. (2010). Impression and information management: on the strategic self-regulation of innocent and guilty suspects. *Open Criminol. J.* 3, 10–16. doi: 10.2174/1874917801003010010

Hauch, V., Sporer, S. L., Michael, S. W., and Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Commun. Res.* 43, 283–343. doi: 10.1177/0093650214534974

Hocking, J. E., and Leathers, D. G. (1980). Nonverbal indicators of deception: a new theoretical perspective. *Commun. Monogr.* 47, 119–131. doi: 10.1080/03637758009376025

Houston, P., Floyd, M., and Carnicero, S. (2012). *Spy the Lie*. New York, NY: St. Martin's Press.

Hurley, C. M., Griffin, D. J., and Stefanone, M. A. (2014). Who told you that? Uncovering the source of believed cues to deception. *Int. J. Psychol. Stud.* 6, 19–32. doi: 10.5539/ijps.v6n1p19

Inbau, F. E., Reid, J. E., Buckley, J. P., and Jayne, B. C. (2013). *Criminal Interrogation and Confessions*, 5th Edn. Burlington, MA: Jones & Bartlett Learning.

Jordan, S., Brimbal, L., Wallace, D. B., Kassin, S. M., Hartwig, M., and Street, C. N. H. (2019). A test of the micro-expressions training tool: does it improve lie detection? *J. Invest. Psychol. Offender Prof.* 16, 222–235. doi: 10.1002/jip.1532

Kassin, S. M. (2005). On the psychology of confessions: does innocence put innocents at risk? *Am. Psychol.* 60, 215–228. doi: 10.1037/0003-066X.60.3.215

Kassin, S. M., and Fong, C. T. (1999). "I'm innocent!": effects of training on judgments of truth and deception in the interrogation room. *Law Hum. Behav.* 23, 499–516. doi: 10.1023/A:1022330011811

Levine, T. R. (2010). "A few transparent liars: explaining 54% accuracy in deception detection experiments," in *Communication Yearbook*, Vol. 34, ed. C. Salmon (Newbury Park, CA: Sage).

Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., et al. (2011). Sender demeanor: individual differences in sender believability have a powerful impact on deception detection judgments. *Hum. Commun. Res.* 37, 377–403. doi: 10.1111/j.1468-2958.2011.01407.x

Macdonald, J. M., and Michaud, D. L. (1992). *Criminal Interrogation*. Denver, CL: Apache Press.

Mann, S., Deeb, H., Vrij, A., Hope, L., and Pontigia, L. (2020). Detecting smugglers: identifying strategies and behaviours in individuals in possession of illicit objects. *Appl. Cogn. Psychol.* 34, 372–386. doi: 10.1002/acp.3622

Mann, S., and Vrij, A. (2006). Police officers' judgements of veracity, tenseness, cognitive load and attempted behavioural control in real life police interviews. *Psychol. Crime Law* 12, 307–319. doi: 10.1080/10683160600558444

Mann, S., Vrij, A., and Bull, R. (2002). Suspects, lies and videotape: an analysis of authentic high-stakes liars. *Law Hum. Behav.* 26, 365–376. doi: 10.1023/A:1015332606792

Mann, S., Vrij, A., and Bull, R. (2004). Detecting true lies: police officers' ability to detect deceit. *J. Appl. Psychol.* 89, 137–149. doi: 10.1037/0021-9010.89.1.137

Masip, J., Barba, A., and Herrero, C. (2012a). Behaviour Analysis Interview and common sense. A study with novice and experienced officers. *Psychiatry Psychol. Law* 19, 21–34. doi: 10.1080/13218719.2010.543402

Masip, J., Herrero, C., Garrido, E., and Barba, A. (2012b). Is the Behaviour Analysis Interview just common sense? *Appl. Cogn. Psychol.* 25, 593–604. doi: 10.1002/acp.1728

Meyer, P. (2010). *Lie Spotting: Proven Techniques to Detect Deception*. New York, NY: St. Martin's Press.

Ofshe, R. J., and Leo, R. A. (1997). The decision to confess falsely: rational choice and irrational action. *Denver Univ. Law Rev.* 74, 979–1112.

Shallice, T., and Burgess, P. (1994). "Supervisory control of action and thought selection," in *Attention: Selection, Awareness and Control: A Tribute to Donald Broadbent*, eds L. Weiskrantz, A. Baddeley, and D. Alan (New York, NY: Clarendon Press), 171–187.

Street, C. N. H. (2015). ALIED: humans as adaptive lie detectors. *J. Appl. Res. Mem. Cogn.* 4, 335–343. doi: 10.1016/j.jarmac.2015.06.002

Strömwall, L. A., Granhag, P. A., and Hartwig, M. (2004). "Practitioners' beliefs about deception," in *Deception Detection in Forensic Contexts*, eds P. A. Granhag and L. A. Strömwall (Cambridge, England: Cambridge University Press), 229–250.

The Global Deception Team (2006). A world of lies. *J. Cross Cult. Psychol.* 37, 60–74. doi: 10.1177/0022022105282295

U.S. Government Accountability Office (2017). *Aviation Security: TSA Does not Have Evidence Supporting Most of the Revised Behavioral Indicators Used in its Behavior Detection Activities*. Retrieved from: https://www.gao.gov/products/GAO-17-608R

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd Edn. Chichester: John Wiley and Sons.

Vrij, A. (2014). Interviewing to detect deception. *Eur. Psychol.* 19, 184–195. doi: 10.1027/1016-9040/a000201

Vrij, A., Akehurst, L., and Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal Criminol. Psychol.* 11, 297–312. doi: 10.1348/135532505X60816

Vrij, A., and Granhag, P. A. (2007). "Interviewing to detect deception," in *Offenders' Memories of Violent Crimes*, ed. S. A. Christianson (Chichester: John Wiley & Sons, Ltd), 279–304. doi: 10.1002/9780470713082.ch12

Vrij, A., Hartwig, M., and Granhag, P. A. (2018). Reading lies: nonverbal communication and ceception. *Ann. Rev. Psychol.* 70, 295–317. doi: 10.1146/annurev-psych-010418-103135

Vrij, A., and Semin, G. R. (1996). Lie experts' beliefs about nonverbal indicators of deception. *J. Nonverb. Behav.* 20, 65–80. doi: 10.1007/BF02248715

Zulawski, D. E., and Wicklander, D. E. (1993). *Practical Aspects of Interview and Interrogation*. Boca Raton, FL: CRC Press.

Conflict of Interest: 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.

APPENDIX 1

Outline of Mann et al. (2020). In two experiments, smugglers and non-smugglers made a ferry-crossing. Smugglers carried an object during the crossing; non-smugglers did not, they were told the experiment would commence on the other side after the ferry-crossing. The participants were secretly videotaped during the crossing. The difference between the two experiments was that in Experiment 2, two confederates approached the (non)smugglers (without talking to them) as they were searching for someone; no intervention took place in Experiment 1. After the ferry-crossing, smugglers in both experiments reported that they felt more nervous during the crossing than non-smugglers (measured on a 7-point Likert scales). The videotapes were shown to observers (laypersons) who were asked to indicate for each participant (i) whether they thought s/he was a smuggler or non-smuggler (dichotomous scale) and (ii) the extent to which they relied on signs of nervousness to decide this (measured on a 7-point Likert scale). First, positive correlations were obtained between judging someone as a smuggler and finding the person nervous. Second, in Experiment 1 the smugglers and non-smugglers were assessed as equally nervous but in Experiment 2 the smugglers were judged as being less nervous than the non-smugglers. Third, the correct classifications of smugglers and non-smugglers was at chance level in Experiment 1 (48.0%) but below chance level in Experiment 2 (39.2%).

frontiers in Psychology

ORIGINAL RESEARCH
published: 02 September 2020
doi: 10.3389/fpsyg.2020.02134

How Our Gaze Reacts to Another Person's Tears? Experimental Insights Into Eye Tracking Technology

*Alfonso Picó[1], Raul Espert[1] and Marien Gadea[1,2]**

[1] Department of Psychobiology, Faculty of Psychology, Universitat de València, Valencia, Spain, [2] Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM)-Mental Health, Madrid, Spain

Crying is an ubiquitous human behavior through which an emotion is expressed on the face together with visible tears and constitutes a slippery riddle for researchers. To provide an answer to the question "How our gaze reacts to another person's tears?," we made use of eye tracking technology to study a series of visual stimuli. By presenting an illustrative example through an experimental setting specifically designed to study the "tearing effect," the present work aims to offer methodological insight on how to use eye-tracking technology to study non-verbal cues. A sample of 30 healthy young women with normal visual acuity performed a within-subjects task in which they evaluated images of real faces with and without tears while their eye movements were tracked. Tears were found to be a magnet for visual attention in the task of facial attribution, facilitating a greater perception of emotional intensity. Moreover, the inspection pattern changed qualitatively and quantitatively, with our participants becoming fully focused on the tears when they were visible. The mere presence of a single tear running down a cheek was associated with an increased emotional inference and greater perception of sincerity. Using normalized and validated tools (Reading the Eyes in the Mind Test and the SALAMANCA screening test for personality disorders), we measured the influence of certain characteristics of the participants on their performance of the experimental task. On the one hand, a higher level of cognitive empathy helped to classify tearful faces with higher emotional intensity and tearless faces with less emotional intensity. On the other hand, we observed that less sincerity was attributed to the tearful faces as the SALAMANCA test scores rose in clusters A (strange and extravagant) and B (immature and emotionally unstable) of our sample. The present findings highlight the advantages of using eye tracking technology to study non-verbal cues and draw attention to methodological issues that should be taken into account. Further exploration of the relationship between empathy and tear perception could be a fruitful avenue of future research using eye tracking.

Keywords: crying, eye tracking, empathy, gaze, tears

Edited by:
Marianne Schmid Mast,
University of Lausanne, Switzerland

Reviewed by:
Dennis Küster,
University of Bremen, Germany
Albert Lee,
Nanyang Technological University, Singapore

Correspondence:
Marien Gadea
marien.gadea@uv.es

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *14 February 2020*
Accepted: *30 July 2020*
Published: *02 September 2020*

Citation:
Picó A, Espert R and Gadea M (2020) How Our Gaze Reacts to Another Person's Tears? Experimental Insights Into Eye Tracking Technology. Front. Psychol. 11:2134. doi: 10.3389/fpsyg.2020.02134

INTRODUCTION

In humans, emotions are automatically transmitted through visual cues, including non-verbal behaviors such as facial expressions and body language (Kret, 2015). Among all the signals by which emotions can be expressed, visible tears – and more specifically the shedding of tears in response to an emotional state, as opposed to those in response to pain or a physical irritation of the eye – are one of the most ubiquitous displays of human emotional. Recently, the socioemotional impact of visible tears on others' perceptions and judgments is receiving growing and deserved attention as a field of empirical study (for an up-to-date non-systematic metanalysis on emotional crying, see Zickfeld et al., 2020). However, no previously published eye tracking studies have employed objective measures than self-reporting to throw light on reactions to emotional crying. We decided to apply eye tracking technology and a carefully selected series of stimuli to answer the question "How our gaze reacts to another person's tears?" The eye tracking technique has a long history (Yarbus, 1967) and allows gaze measures to be assessed with respect to the so-called "tearing effect." With the present work, we also set out to offer methodological insight and advice on how to use eye tracking technology to study non-verbal cues by providing an illustrative example of an experimental setting specifically designed to study the "tearing effect."

Gaze Measures, Before and Now

The measurement of oculomotor variables in cognitive science dates back more than 100 years (Dodge and Cline, 1901; Dodge, 1903) and constitutes a non-invasive method for evaluating a wide variety of processes, from emotional recognition to social information processing (Xiao et al., 2015). Gazing is unique among non-verbal behaviors in that the eye is a sensory organ for gathering information and, at the same time, performs the function of a signal to others (Harrigan et al., 2005). However, of the more than 1,700 articles on gaze published since 1982 and included in the review by Harrigan et al. (2005), only 13% investigated non-verbal behaviors. More recently, in a short review on the research conducted over the past 5 years, the keywords gaze and non-verbal behavior in Google Scholar, MEDLINE, Pubmed, and Scopus yielded 17,700 results. Unfortunately, this current emphasis has not always been accompanied by clear explanations about the best methodology for conducting such studies. In particular, there are very few descriptions of the methodology used to codify gaze (with the exception of some classic works, such as those by Exline and Fehr (1982) and Fehr and Exline (1987). In non-verbal research, gaze measures have traditionally been divided into (1) frequency, (2) total duration of the gaze, (3) proportion of time looking at, (4) average duration of individual glances, (5) standard deviation of glances, and (6) mutual gaze (the most investigated) (Argyle and Ingham, 1972). Other authors have determined different forms of eye movements based on their duration (Kirkland and Lewis, 1976). These traditional categorical classifications have largely been superseded by a quantitative approach that makes use of detailed records of eye movements through so-called eye tracking devices, which measure nearly 120 different metrics corresponding to basic properties of movement, position, numerosity, latency, and distance of the gaze (see **Figure 1** for an overview). Eye tracking technology has been widely used to analyze stimuli of different emotional valence in order to throw light on differences in visual behavior (Tavakoli et al., 2015; Rubo and Gamer, 2018) and how subtle differences can lead to major changes in gaze behavior. In addition, the quantitative evaluation of facial emotional expressions by eye tracking technology has provided useful insight into child and adolescent psychiatry (Rommelse et al., 2008), neurodegenerative diseases (Bek et al., 2020), mood disorders (Peckham et al., 2016; Hunter et al., 2020), and other behavioral disorders (Martin-Key et al., 2017). In this way, it is the perfect tool for our interests and presents itself as the logical next step in the investigation of emotional crying (Krivan and Thomas, 2020). Nonetheless, there are some methodological flaws that are repeated over and over again in studies employing eye tracking methodology. Some of them can be rectified using statistical techniques that take into account the special characteristics of the analyzed data (e.g., the vast majority of eye tracking metrics do not follow a Gaussian distribution). Others can be overcome by an appropriate experimental design (e.g., eye tracking metrics are idiosyncratic for most participants and stable across trials, so comparisons between groups of subjects can be problematic) or by controlling variables like gender or a misuse of the signal-to-noise ratio that could make any spurious result statistically significant in a sufficiently large sample. These and other problems are perfectly solvable if one understands the methodology of eye tracking, which will be detailed later, but they become especially problematic for researchers of non-verbal communication, as became evident when we performed a review of the literature.

Tears and Emotional Crying

Theoretical positions rooted within an evolutionary framework have suggested that tears act as biological signals, which serve to express a request for help (Roes, 1989; Kottler, 1996; Walter, 2008). Furthermore, the literature points to the functional role of emotional crying as a form of communication, for example, of the need for attention and support (Hendriks and Vingerhoets, 2006) and the need to be perceived as warmer and friendlier (van de Ven et al., 2016; Vingerhoets et al., 2016; Zickfeld et al., 2018) or more honest (Zeifman and Brown, 2011). Provine et al. (2009) claimed that emotional tears can improve emotional recognition, at least with regards to the sad feelings of the crier. These authors, and others (Provine et al., 2009; Zeifman and Brown, 2011), have proposed the "tearing effect" as a sign of improved emotional perception and processing of facial expressions in the presence of tears, which eliminates the potential emotional ambiguity toward the observed face. Importantly, it has been suggested that tears exert their intended influence provided that they are perceived as "natural;" that is, if tears are depicted running upward instead of downward from the eye (in an unnatural direction), they lose their emotional impact (Provine, 2014, 2017). This indicates that tears are a special stimulus (i.e., emotional signal) that have priority over others (Vuilleumier et al., 2002; Killgore and

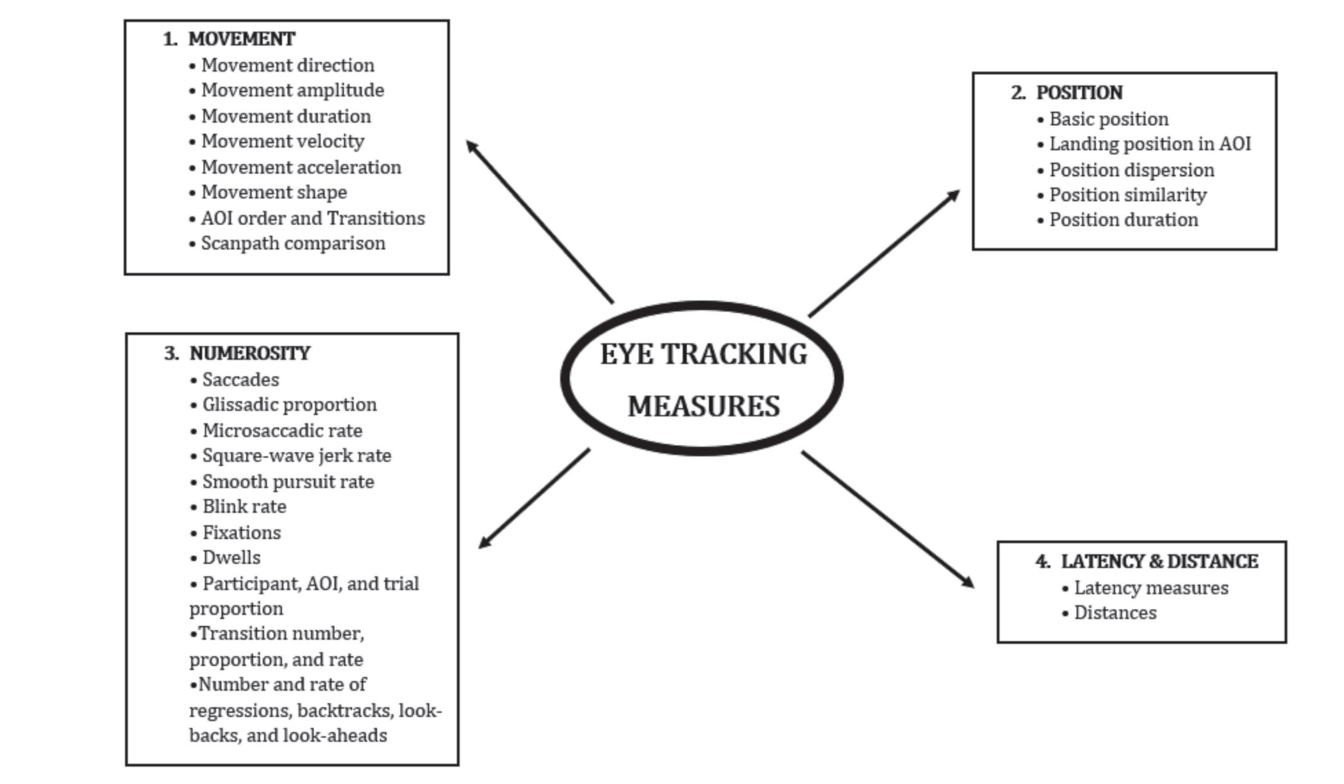


FIGURE 1 | Fundamental classification of modern eye tracking measures. This chart summarizes the taxonomy proposed by Holmqvist and Andersson (2017) for the basic classification of 120 eye tracking measurements reviewed in the literature. The number of different measurements is increasing over time due to the combination of some of the existing ones, as well as the creative use of technology that provides the opportunity to find new and interesting metrics.

Yurgelun-Todd, 2004). In summary, the literature as a whole bears testimony to the fact that the perception of emotional tears, even when operating at a preattentive level (Balsters et al., 2013; Lockwood et al., 2013), is capable of inducing important behavioral changes in the observer.

Objectives and Hypothesis

In light of the above scenario, we were interested in measuring the changing gaze behavior of observers with an objective methodology (eye tracking) and in investigating some of the putative functional roles of tears. Moreover, based on our own experience, we felt it would be useful to offer methodological advice on the use of eye tracking technology for the study of non-verbal cues beyond muscle activation in facial expressions (Kret, 2015). We argue in favor of a particular type of experimental design over others and for the selection of an appropriate sample size and eye tracking measures. Thus, we designed a study to explore some basic eye tracking measures during the observation of calm crying faces (i.e., duration of gazing and number of fixations within an area of interest – henceforth AOI – where the tear appears). We hypothesized that tearful faces would receive longer gaze time inside the AOI and that the AOI of tearful faces would receive more dwells and a greater number of fixations. With regards to the functional roles of crying, we expected that the presence of tears would facilitate the perception of the emotional intensity of the subjects' faces (related to the tearing effect), lead participants to perceive the subjects to be more sincere (related to the perception of more honesty), and elicit more sympathy from our participants toward the subjects (related to the proposed function of tears in communicating the need for help), considering sympathy as an affective experience with a prosocial motivation toward others (to help or relieve the suffering) (Walter, 2012).

An additional aim of this study was to consider the influence of factors inherent to the observer's perceptual processing of tears. The interesting review of Vingerhoets and Bylsma (2016) suggested that the study of crying was the "gateway" to achieve a better insight into important developmental processes like empathy and personality disorders. We hypothesized that people scoring high in cognitive empathy would be more prone to experiment the "tearing effect." Regarding personality features and crying, individuals with high levels of neuroticism cry relatively more (Peter et al., 2001), whereas dismissively attached people tend to cry less than others (Laan et al., 2012). Moreover, the crying of patients with borderline or narcissistic personality disorders can be perceived as manipulative and annoying by therapists in clinical settings (Alexander, 2003). Given such observations about crying with respect to personality disorders, we wondered whether the observation of other people's crying would also reveal a relation to personality disorders when measured in a non-clinical sample.

MATERIALS AND METHODS

Participants

Taking into account the experimental design, and on the basis of data from a previous pilot study (Picó et al., 2018), we performed a power analysis to justify the detection of medium effect sizes with a probability of 0.8 for a paired sample test. Subsequently, this power analysis was used to select a convenient sample size of 27 participants, but it was not employed in the correlational analyses, which occupy a secondary role in the present work. The use of power analyses that justify the sample size is essential to avoid problems of signal–noise discrimination that could cause us to incur in type I errors. To perform the analysis, we used the "pwr" package (Champely, 2018) from R software. Thirty undergraduate women aged 18–27 years (M = 22.23, SD = 2.39) were recruited from the Nursing degree at the University of Valencia (Spain) and were given a 16 GB USB memory stick as a reward for their participation in the experiment. Selection criteria included a near-perfect vision (no glasses or contact lenses), no reported history of psychiatric disorders, and no chronic pharmacological treatment. Two individuals were excluded from the eye tracking data collection due to technical issues. All the participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" and the precepts of our university's Ethics Committee, and all signed an informed consent form.

Materials

Visual Stimuli

We used a set of four photographs of neutral faces of adult persons – two women and two men – kindly provided by the photographer Marco Anelli. These photographs had been used in previous studies (van de Ven et al., 2016; Vingerhoets et al., 2016; Zickfeld et al., 2018; Stadel et al., 2019; Picó et al., 2020). The photographs were taken in the precise moment when the subject was engaged in calm crying – the particular distinction of which is the presence of visible tears with little marked emotional expression – in a spontaneous way (see details of the photographs in Picó et al., 2020). The images were manipulated to digitally remove the visible tears so that the experiment was carried out with a total of eight images: four with tears and four without, representing both genders in each case. In addition, the facial expression in each photograph was accompanied by a text consisting of an explicit affirmation (e.g., "I am not cheating on my boyfriend!") as if the phrase was being pronounced by the subject. We wrote four vignettes of text, one for each of the four subjects depicted in the photographs, and each text was paired with the two versions of the photo of the same person, once with the photo showing tears and once with the photo without tears. The order of the four photos and vignettes was completely counterbalanced. Prior to the experiment, we carried out a practice trial (not analyzed) in which the participants looked at the pictures of two women with neutral facial expressions (i.e., AF05NES and AF23NES) extracted from the Karolinska Directed Emotional Faces (Lundqvist et al., 1998), with their corresponding vignettes. The rationale for using images that depict calm crying expressions lies in the assumption that, if the effect of emotional crying is mainly due to the presence of tears, it will be detectable even in faces with little emotional display (Vingerhoets, 2013).

Eye Tracker Device

The device that we used to measure the visual variables related to attentional factors was a 150 Hz GP3 HD UX eye tracker system (Gazepoint systems, Toronto, Canada) connected to a PC with a 19″ LED Benq GL950 Senseye monitor. This eye tracker model has a wide lens, allowing relatively large head movements to be monitored during experimental tracking (∼35 cm in horizontal movement and 22 in vertical movement), without the need to restrain participants; even so, our participants were instructed to remain as still as possible, with their backs straight, up against the back of the chair. We processed the experimental data with Gazepoint Analysis UX software (Gazepoint Systems, Toronto, Canada). The most basic eye tracking data – from which the rest of the metrics can be calculated – are X- and Y-coordinates of the fixation point of gaze, measured as a fraction of the screen size at specific times (in our case, every 1/150 s). The point of gaze (POG) used is the average of the left eye and right eye POG if both are available; if not, the value of either the left or right eye is used, depending on which one is valid.

Questionnaires

The "Reading the Mind in the Eyes" test, also known as RMET (Warrier et al., 2017a), was administered as a brief social cognition test to measure cognitive empathy. Cognitive empathy is a construct closely related to Theory of Mind (ToM). Specifically, ToM refers to the ability to represent and understand, in general, the mental states of others. Cognitive empathy refers to the ability to understand and mentalize about the feelings of others, considering feelings to be a mental state among others, without necessarily implying that the empathizer is in an affective state himself (Walter, 2012). We chose cognitive empathy because, according to Warrier et al. (2017b), enhanced cognitive empathy results in a higher ability to recognize another person's mental states. In this test, a series of 36 photographs depict eye regions from different models who express a range of emotional states. Four words are presented at the same time, surrounding the photo, and each word refers to a unique mental state. Participants are asked to choose which one of the four words better suits what the person in the photograph is feeling.

The Personality Disorders Screening Test SALAMANCA questionnaire (Pérez-Urdániz et al., 2011) was administered as a brief screening tool for evaluating personality in our sample of participants. This instrument evaluates the presence of 11 personality disorders drawn from the Diagnostic and Statistical Manual of Mental Disorders (DSM) (paranoid, schizoid, schizotypal, histrionic, antisocial, narcissistic, and dependent) and the International Classification of Diseases (ICD) (emotionally unstable personality disorder-impulsive type, emotionally unstable personality disorder-borderline type, also known as limit, anankastic, and anxious). These 11 disorders are classified in three groups: Type A, strange and extravagant (paranoid, schizoid, and schizotypal); Type B, immature (histrionic, antisocial, narcissist, and both subtypes of emotional unstable disorders: impulsive and limit); and type C, avoiding (anankastic, dependent, and anxious). The SALAMANCA tool consists of a total of 22 questions; each personality trait is evaluated through two questions using a 4-point Likert scale (*false* = 0 points; *sometimes true* = 1 point; *usually true* = 2 points; *always true* = 3 points). The cutoff score is established at 3 points for every trait. This questionnaire has been validated and correlated with the Interpersonal Personality Disorder Examination and is considered an adequate test of screening, with a sensitivity of 100% and a specificity of 76.3% (Caldero-Alonso, 2009). It is important to note that this questionnaire is not intended as a diagnostic tool but rather for screening tendencies to suffering personality disorders (vulnerabilities), which should be confirmed by a psychiatrist in every case. It is a self-assessment questionnaire (< 10 min) that is easily interpreted.

Procedure

Before each participant performed the task, the eye tracking system was calibrated according to a standard protocol with nine calibration positions on the screen in order to be sufficiently personalized. The monitor was positioned 67 cm from the eyes of the participant (equally for the entire sample). Following calibration, participants carried out the task of viewing the photos of the faces with and without tears. Before each stimulus, participants were told they would be presented on the computer screen with a statement (a text vignette, for 15 s) and that they would then see the face of the person who had said the message in the text (the photograph, for 2 s). The gazing of the participants was eye tracked only while the photos were presented on the screen. Note that the photographs appeared on the screen for a very short time; this is an important methodological issue with respect to analysis of the data provided by eye tracking measures, known as "dependence between successive measurements" (Tatler and Vincent, 2008), which is rarely taken into account. The longer the stimuli is displayed on the screen, the greater are the potential bindings of the data, and classical statistical tests do not provide reliable results in this particular circumstance. One of the easiest ways to simplify the situation in eye tracking systems of < 250 Hz is to ensure that the stimulus is available on the screen for a short time (for example, for 2 s, as in the present study). Of course, this strategy is not free of problems, and the duration of the stimulus depends on the expected size of the effect to be detected and the nature of the study (Andersson et al., 2010).

Our participants were instructed to read the text and to observe the corresponding face carefully, as afterward, they would be asked to complete a questionnaire about what they had seen. In this way, immediately after the visual task, participants were given 40 s to respond to a number of questions about the stimuli on a sheet of paper (see a schematic representation of the experiment design in **Figure 2**; see details of the questionnaire in the section below). As shown in **Figure 2**, participants were presented with a first round of four text vignettes plus faces and completed the corresponding questionnaire after seeing each face. Next, the participants were told that they should relax their eyes for 10 min by sitting quietly in a comfortable chair, with their eyes closed and covered with an eye mask. Following this 10-min break, they repeated the task with the same pictures but with/without visible tears (note that the order of presentation was counterbalanced). At this point, we would like to argue in favor of within-subject experimental designs (or repeated measures designs) when using the eye tracking device due to the high variance in this measure among participants (Andrews and Coppola, 1999; Rayner, 2009; Johansson et al., 2012). Between-subject designs require a large number of participants in order to reach an acceptable power to perform a parametric evaluation of statistical differences. The main disadvantage of within-subjects designs is that the order in which the stimuli are presented can affect the validity of the causal inference process (Duchowski, 2017), and the effects of learning and fatigue are further disadvantages. However, these drawbacks can be mitigated perfectly by counterbalancing the presentation of stimuli and by employing short tasks to be carried out in less time, as we did in the present experimental setting. Finally, once we had collected all the data regarding the visual stimuli, participants were asked to complete the RMET and SALAMANCA questionnaires, after which they were given their gift and thanked for their participation.

Measures and Dependent Variables

We recorded three types of measures: measures related to the visual stimuli and the subjective reaction of the participants to them; gaze measures related to the visual stimuli and obtained by means of the eye-tracker system; and, finally, empathy and personality measures of the sample using the RMET and SALAMANCA questionnaires, respectively. Regarding the subjective measures related to the visual stimuli, each photograph of a face, with its attached text, was followed by a questionnaire, which included the following items: (1) the degree of intensity of emotionality the face seemed to show, (2) the perceived sincerity in the observed face with respect to the corresponding statement made by that person (the paired text), and (3) whether the observed face evoked sympathy (or not) in the participant. All questions were assessed on a 6-point Likert scale, where 0 indicated the complete absence of intensity, sincerity, or sympathy and 5 the highest degree of each. Regarding the gaze measures obtained through the eye tracking device, we hand-drew an area of interest (AOI) in the form of a rectangle framing both eyes and widened below the right eye to the right cheek (where tears were visualized on the crying faces), in accordance with Goldberg and Helfman (2010) advice that AOIs should be defined only on objects of interest. We measured the following dependent variables: (1) duration of the gaze inside the AOI in milliseconds, (2) fixations on the AOI (a fixation is defined as maintaining the gaze in a square of 1-degree amplitude for at least 100 ms), (3) revisits or dwells (i.e., looking at the AOI more than once), (4) number of fixations on the global stimuli (i.e., inside and outside the AOI), and (5) mean duration of said global fixations. These eye tracking metrics are available in the vast majority of current software, and we chose them to facilitate future replication of our results by other researchers. All the metrics can be calculated from eye tracking records between 60 and 2,000 Hz,

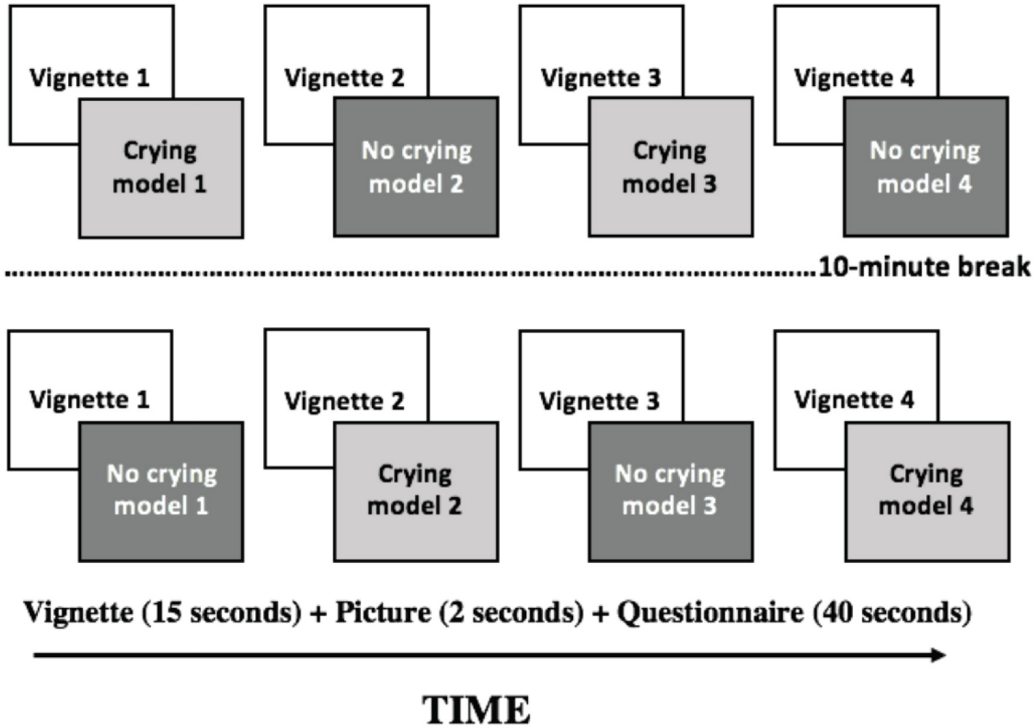


FIGURE 2 | Schematic representation of the experimental procedure for participants. Each participant was presented with the text/vignettes and the faces in a different order to avoid carryover effects. Each vignette was tailored and attached to each model. In this way, visible tears were the only difference between the two conditions.

so they are not restrictive with respect to the equipment that can be used.

Data Analysis

Data management and analysis were performed using the statistical software R version 3.6.0 (2019), R package *WRS2* (Mair and Wilcox, 2020), and *psych* (Revelle, 2018). Before applying parametric methods, we performed Shapiro–Wilks tests to check the normal distribution of the data. Since some of our variables were skewed (as expected), we selected a robust t-test for paired samples with bootstrapping (n = 1,000) to analyze differences between the tear and no tear conditions in terms of intensity of the gaze inside the AOI, fixations on the AOI, revisits of the AOI, global fixations (number), and global fixations (time). To ensure that the total duration of fixations (inside and outside the AOI) did not influence the results, we performed an analysis of covariance (ANCOVA) of the total duration time of fixations as a covariate, and the results did not reveal a significant effect of the total duration on any measure. The explanatory measure of effect size ε reported in this analysis is a robust version (Wilcox and Tian, 2011; Mair and Wilcox, 2020), which does not require equal variances and can be generalized to multiple group settings. As a reference, ε = 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes. In addition, Pearson's product–moment correlations were used to test whether personality traits and/or level of empathy of the participants were related to the experimental results. Results were significant at the $p < 0.05$ level, and p-values were corrected with Bonferroni's method for multiple comparisons. The use of robust parametric statistics (as in our case) that take into account the transgression to some of the fundamental requirements of the classical models (i.e., Gaussian distribution, independency, and homoscedasticity), or relevant transformations in dependent variables, is necessary when working with eye tracking data. We recommend a balance between the most appropriate techniques and those that are simple to interpret.

RESULTS

Intensity of Emotion, Perceived Sincerity, and Evoked Sympathy

On average, crying faces (trimmed M_{tearful} = 3.70) elicited a significantly higher mean perception of emotional intensity [$t(17)$ = 6.48, p = 0.000, $\varepsilon = 0.75$] than the faces without visible tears (*trimmed* M_{tearless} = 2.85). The crying faces were also perceived to be significantly more sincere than the same faces without tears [$t(17)$ = 3.02, $p < 0.01$, $\varepsilon = 0.34$] with trimmed M_{tearful} = 3.68 and M_{tearless} = 3.34, respectively. The crying faces (trimmed M_{tearful} = 3.21) evoked a higher mean sympathy than the tearless faces (trimmed $M_{\text{tearless}} = 3.09$), although this value

TABLE 1 | Descriptive statistics with robust paired *t*-test results.

	TEARFUL					TEARLESS					
	n	Mean	*SD*	Median	Trimmed	*n*	Mean	*SD*	Median	Trimmed	Yuen's t
Intensity	30	3.66	0.51	3.62	3.70	30	2.83	0.65	2.75	2.85	**6.48*** **
Sincerity	30	3.70	0.61	3.75	3.68	30	3.33	0.72	3.50	3.34	**3.02** **
Sympathy	30	3.17	0.69	3.25	3.21	30	3.01	1.10	3.00	3.09	0.58
Duration AOI	28	751.21	515.76	795.00	725.08	29	305.72	233.52	256.00	294.48	**3.38** **
Fixations AOI	28	1.89	0.92	2.00	1.92	29	1.28	0.92	1.00	1.24	**1.22* **
Revisits AOI	28	0.75	0.65	1.00	0.71	29	0.76	0.83	1.00	0.68	0.56
Fixations[a]	29	9.97	0.87	10.00	10.00	30	9.43	1.04	9.00	9.46	**4.40*** **
Fixations (time)[b]	28	19.68	1.97	19.47	19.53	30	20.11	2.62	19.93	20.02	−0.77

**p < 0.05, **p < 0.01, ***p < 0.001. Trimmed means were calculated with a trim level of 0.2. n reflects the final number of cases used in the analysis. [a]Number of fixations inside and outside the AOI. [b]Duration of fixation inside and outside the AOI. Crying faces were also perceived to be significantly more sincere, and with more emotional intensity. Participants spent significantly more time gazing faces with tears. Moreover, the number of fixations inside the area of interest and overall fixations were higher when inspecting crying faces.*

did not reach statistical significance [$t(17) = 0.58$, p = ns]. A summary of these results can be found in **Table 1**.

The Effect of Visible Tears on Eye Tracking Measures

With regard to eye tracking data, participants spent significantly more time gazing (duration measured in milliseconds) inside the AOI of crying (trimmed $M_{\text{tearful}} = 725.08$) vs. non-crying faces (trimmed $M_{\text{tearless}} = 294.48$), $t(17) = 3.38$, $p = 0.003$, with an explanatory effect size of εε = 0.66. The number of fixations inside the AOI was also significantly higher with respect to the crying faces (trimmed $M_{\text{tearful}} = 1.92$ and $M_{\text{tearless}} = 1.24$) [$t(17) = 1.22$, $p = 0.015$], with an effect size of 0.60 and a median difference of one fixation. The number of revisits was not statistically significant $t(17) = 0.56$, p = ns, with a trimmed mean difference of 0.11 and an $\varepsilon = 0.1$. With regards to gaze fixations and duration of the fixations on the whole stimuli (AOI plus outside the AOI), participants engaged in significantly more fixations [$t(18) = 4.40$, $p < 0.000$, $\varepsilon = 0.59$] on the crying faces (trimmed $M_{\text{tearful}} = 10$ and $M_{\text{tearless}} = 9.43$), with no significant differences in the duration of such fixations [$t(17) = -0.77$, p = ns, $\varepsilon = 0.14$] between the two faces. These results are summarized in **Table 1**.

Influence of Empathy and Personality Traits of the Sample

Regarding the scores of the RMET test for measuring cognitive empathy, we observed that the higher the RMET score was, the more emotionally intense the crying face was perceived to be ($r = 0.48$, $p < 0.01$, see the correlations regarding tearful faces in **Table 2**). Interestingly, we also observed that, as the RMET score increased, the non-crying face was perceived to be less intense ($r = -0.44$, $p < 0.01$, see correlations regarding tearless faces in **Table 3**). However, no correlations were observed between RMET levels and eye tracking measures for any of the two conditions (see **Tables 2**, **3**).

With regards to personality measured with the SALAMANCA screening test for vulnerability to personality disorders, the most relevant result was that correlations were significant when the participants were presented with the tearful faces and not when they were presented with the non-crying faces. The emotional intensity of the faces was inversely and significantly correlated to the narcissistic score ($r = -0.36$, $p < 0.05$) and positively and significantly correlated to the paranoid score ($r = 0.42$, $p < 0.05$); thus, low narcissism and higher paranoid ideation were related to the perception of a more intense emotionality in the faces with visible tears. In the case of sincerity, a higher vulnerability to personality disorders was generally related with a lower sincerity attributed to the tearful face. Specifically, a higher vulnerability to schizoid or schizotypal disorders was negatively associated with the perception of sincerity (r's = −0.50 and −0.56 with p's < 0.01). A personality with antisocial tendencies was inversely related to attributed sincerity ($r = 0.59$, $p < 0.01$). High vulnerability to narcissism was also related to low attributed sincerity ($r = -0.54$, $p < 0.01$). Lastly, vulnerability to emotional instability disorders (i.e., limit and impulsive) correlated negatively with perceived sincerity (r's = −0.39 and −0.47, with $p < 0.05$ and $p < 0.01$, respectively). It should be stressed that all the above results refer to the tearful faces and that we did not find any significant correlation among these personality measures and the attributions of emotional intensity, sincerity, or sympathy elicited by the faces without visible tears. Finally, these personality measures were not closely related to the gaze measures obtained with the eye tracker. Once again, we found no relation when judging the non-crying faces, but when faces with visible tears were viewed, we observed that the antisocial personality score rose with the duration of visual inspection outside the AOI ($r = 0.70$, $p < 0.05$). **Tables 2, 3** summarise the correlational results.

Heatmaps and Fixations: A Qualitative Inspection (Figure 3)

As an example, **Figure 3** is a graphical representation of the average visual behavior observed when a face (i.e., model

TABLE 2 | Correlations in the tearful condition.

	Intensity	Sincerity	Sympathy	Fixations	Duration of fixations	Duration (AOI)	Fixations (AOI)	Revisits (AOI)
RMET	**0.48****	−0.03	−0.01	0.50	0.23	−0.12	−0.07	−0.26
Paranoid[a]	**0.42***	−0.16	0.13	0.47	0.49	−0.31	−0.40	−0.44
Schizoid[a]	−0.15	**−0.50****	−0.08	0.01	0.05	0.21	0.08	0.19
Schizotypal[a]	0.09	**−0..56****	−0.17	0.17	0.55	−0.21	−0.35	−0.30
Histrionic[b]	−0.02	−0.14	0.08	0.12	0.40	−0.08	−0.34	−0.14
Antisocial[b]	0.06	**−0.59****	−0.20	−0.03	**0.70****	−0.21	−0.28	−0.22
Narcissistic[b]	**−0.36***	**−0.54****	−0.33	−0.42	0.52	−0.44	−0.56	−0.50
Impulsive[b]	−0.22	**−0.47****	−0.13	−0.07	0.36	0.18	−0.11	0.18
Limit[b]	0.11	**−0.39***	−0.04	0.17	0.15	0.11	−0.33	0.05
Anankastic[c]	0.17	−0.05	0.12	−0.05	0.22	0.02	−0.37	−0.13
Dependent[c]	−0.31	−0.27	0.05	0.01	0.24	0.00	−0.14	0.08
Anxious[c]	−0.06	−0.19	0.05	0.36	0.27	−0.08	−0.31	−0.50

$*p < 0.05$, $**p < 0.01$. [a]Type A personalities in the SALAMANCA screening test (strange and extravagant). [b]Type B personalities in the SALAMANCA screening test (immature and subtypes of emotional unstable disorders). [c]Type C personalities in the SALAMANCA screening test (avoiding). The higher the RMET score was, the more emotionally intense the crying face was perceived to be. Low narcissism and higher paranoid ideation were related to the perception of a more intense emotionality in the faces with visible tears, meanwhile a higher schizoid or schizotypal score was negatively associated with the perception of sincerity. The antisocial tendencies along with high vulnerability to narcissism were related to low attributed sincerity. Moreover, emotional instability disorders correlated negatively with perceived sincerity.

TABLE 3 | Correlations in the tearless condition.

	Intensity	Sincerity	Sympathy	Fixations	Duration of fixations	Duration (AOI)	Fixations (AOI)	Revisits (AOI)
RMET	**−0.44***	−0.32	−0.27	0.23	−0.17	−0.40	−0.37	−0.31
Paranoid[a]	−0.02	−0.23	−0.11	−0.02	0.14	−0.01	0.00	0.05
Schizoid[a]	0.24	−0.26	−0.13	0.14	−0.10	0.25	0.26	0.41
Schizotypal[a]	0.04	−0.31	−0.27	0.00	0.14	0.28	0.30	0.53
Histrionic[b]	0.00	−0.07	−0.01	−0.50	0.59	−0.17	0.14	0.33
Antisocial[b]	0.09	−0.20	−0.28	0.01	0.13	0.29	0.42	0.61
Narcissistic[b]	0.23	−0.12	−0.15	−0.38	0.35	0.27	0.03	0.05
Impulsive[b]	−0.17	−0.27	−0.11	−0.45	0.46	−0.27	−0.09	0.13
Limit[b]	0.04	−0.31	−0.15	−0.20	0.28	−0.10	0.13	0.34
Anankastic[c]	0.27	0.05	0.08	−0.38	0.46	0.12	0.40	0.54
Dependent[c]	0.17	−0.21	0.01	−0.44	0.34	−0.14	−0.06	0.13
Anxious[c]	−0.08	−0.23	−0.09	−0.39	0.51	−0.13	0.12	0.35

$*p < 0.05$. [a]Type A personalities in the SALAMANCA screening test (strange and extravagant). [b]Type B personalities in the SALAMANCA screening test (immature and subtypes of emotional unstable disorders). [c]Type C personalities in the SALAMANCA screening test (avoiding). As can be seen, as the RMET score increased, the non-crying face was perceived to be less intense.

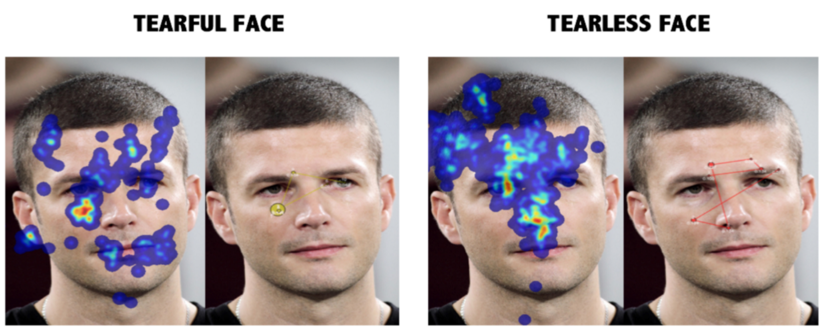


FIGURE 3 | Overall heatmap and fixations map in model 4. The overall effects represent a clear change in the pattern of gaze in model 4. Not only was the time within the zone delimited by eyes and right cheek superior in the tearful condition, but also a qualitative change was also observed.

4) was explored. We can observe the triangular geometric pattern that runs from one eye to the next and then down to the mouth and then back to the first eye (Iskra and Gabrijelcic, 2016), with extents of preference in the eyes–mouth continuum (Rogers et al., 2018) when tearless faces were judged and with brief fixation times and spreading points over the face. As the figure shows, the presence of tears alters the visual inspection pattern, breaking the triangle of fixations

concentrating them inside the AOI, as if tears were powerful visual attention magnets.

DISCUSSION

The main objectives of this study were to evaluate some of the suggested functional roles of tears and to explore the modification of gaze behavior when subjects are presented with faces with visible tears. Earlier research carried out in our laboratory showed how visible tears are capable of altering inferences regarding emotional intensity and sincerity perceived in human faces (Picó et al., 2020). In the present experiment, we replicated some of the results of our previous study with regards to faces engaged in calm crying. The participants in the present sample perceived the emotional expression of the faces to be more intense and judged it to be more sincere. Our results are also in line with previous evidence that tearful faces can facilitate the perception of emotional expression (Vingerhoets, 2013; Vingerhoets et al., 2016; Gračanin et al., 2017). Weeping is a genuine way to show emotion and is usually associated with sadness (Klineberg, 1940; van de Ven et al., 2016) but can also occur in happy situations (Vingerhoets and Cornelius, 2001). The present study shows how tears convey a message without the explicit need to identify the specific emotion that caused them. We believe this finding is especially interesting given that we have evaluated tears in calm crying faces. As Ito et al. (2019) recently pointed out (2019), visible tears seem to constitute a context in themselves that facilitates emotional inference, even in the absence of any other emotional clue. In addition, our participants judged the phrases associated with the crying faces as being more sincere. In this way, calm crying faces exerted an influence on sincerity as a state, in accordance with Zeifman and Brown (2011), who reported that the presence of visible tears increased the perception of honesty (sincerity as a trait) in subjects, and with Regan and Baker (1998), who showed that the testimonies of children who had been victims of sexual assault were perceived as more credible if they cried. According to Van Kleef (2008)'s theory of emotion as social information (2009), people use the perceived emotions of others to clarify ambiguous social situations. It is possible that an emotional sign such as visible tears makes it easier to label a specific social situation, and this might help to generate a greater sense of sincerity in the communication. If this were the case, visible tears would represent a non-verbal clue indicating sincerity, a quality that is indispensable for a fruitful collaboration in an ultrasocial species such as ours (Tomasello, 2014). Such a clue could be used by dishonest individuals in order to take advantage of their peers, and indeed, crying is also seen as one of the most conventional tactics of emotional manipulation (Buss, 1987). As for the sympathy aroused in our participants by the tearful faces, though it was greater than that provoked by the tearless versions, it did not reach statistical significance.

In contrast, the results of the eye tracking task revealed profound changes in gazing behavior provoked by crying faces. The presence of visible tears led to a greater visual inspection of the eyes and right cheek, where the most pronounced visible tear was located. The participants not only spent more time looking at this AOI in the crying faces, but they also engaged in more fixations there (i.e., they maintained their gaze on a fixed point inside the AOI more times when this area contained tears). As Loftus (1972) demonstrated, scene recognition can be expressed as a positive function of the number of fixations, and in the present study, we detected a significantly greater number of fixations in the tearful condition. The literature demonstrates that the enrichment of general stimuli leads to a greater number of fixations; in this sense, the tearing effect seems to enrich the eye area. We have not found any previous research in which this technique has been applied to the study of tears, so we are unable to compare our results. However, it is worth highlighting the work of Balsters et al. (2013), in which tears were presented as visual cues at a preattentive level and were still capable of arousing greater kindness, feelings of empathy, and connectedness. Our results are in line with these studies, as all of them point to tears functioning as a powerful visual cue that acts as a gaze magnet.

Another of our aims was to explore the relation between the cognitive empathy of the observers and the processing of tears in the calm crying faces. Interestingly, we found that a higher RMET score was significantly correlated with higher intensity of emotion only when visible tears were present, while the relationship was reversed in the absence of emotional crying. There is empirical evidence (Carr and Lutjemeier, 2005; Gery et al., 2009) that empathy is related to the ability to recognize emotions in emotional expressions, and such accurate emotional inference can be achieved during very short exposure to a facial expression. Accordingly, we found that a high level of cognitive empathy qualified people to discern and adequately label the non-crying face as being less intense and the crying face as transmitting higher emotional intensity. Interestingly, our results concerning empathy are in line with those of Harrison et al. (2007), who showed that sensitivity to the influence of pupil size, an autonomous signal related to tears in sad faces, correlated positively with the empathy score of the sample. In addition – and relevant for future research with broader samples in order to assure power – it would be interesting to examine the relation and causal direction among perceived emotional intensity, presence of tears, and cognitive empathy by means of structural equation modeling (Wang and Wang, 2020). In a recent mediation analysis (Küster, 2018), it was shown that visible tears produce an all-or-nothing effect where the intensity of crying does not appear to be a significant variable. Indeed, in the present study, we have found that the presence of a minimal signal of weeping was sufficient to provoke a measurable reaction in the observer.

Our observations regarding vulnerability to personality disorders and processing of tearful faces should be interpreted cautiously and received as suggestions to be put to the test in future studies with broader non-clinical samples and clinical populations. That said, it is noteworthy that significant correlations were detected only when the participants were judging tearful faces and that most are in line with data in the literature (clinical or otherwise). For instance, we found a positive association between higher paranoid ideation scores and the emotional intensity perceived in our calm crying faces; this is in accordance with a previously reported bias toward the perception of negative emotions in cases of clinical paranoia, with negatively biased interpretations of emotional ambiguity (Savulich et al., 2015). Regarding narcissism, which was correlated negatively with the "tearing effect," we have stated in *Introduction* that individuals with narcissistic personality disorder (NPD) cry more than others. It is perhaps plausible that a person with a higher NPD trait score will interpret tears as more "normal" and less important, given that she/he is more accustomed to crying. Our results concerning the influence of vulnerability to personality disorders on the perceived sincerity of crying faces were even more relevant; on most of the scales, higher scores were associated with lower levels of sincerity attributed to the crying model. This was especially clear in the case of the personalities grouped in clusters A (strange and extravagant) and B (immature and emotionally unstable), thus showing that these personalities interpreted the crying behavior in a slightly different way. Lastly, the isolated positive correlation between a higher score for antisocial personality and the duration of fixations on the entire tearful face (global stimulus) is of special interest. We wonder whether this kind of personality increases the visual attention given to the whole face as a way of avoiding tearful eyes. This would support the recent observation that a higher psychopathy level is a significant predictor of reduced eye contact measured with eye tracking (Gehrer et al., 2020).

STRENGTHS AND LIMITATIONS

This is the first study of an experimental line that employs an objective eye tracking protocol to evaluate emotional crying perception, and its results extend the existing behavioral data by introducing some physiological variables. To date, only one (recent) report has provided objective evidence of the tearing effect using psychophysiological measurements (Krivan et al., 2020). In our view, the present study represents a first step toward understanding crying as a visual signal of communication by means of the technology that best captures the particularities of this very special stimulus and has important social connotations. As a next step, future research should combine the psychophysical visual data obtained via eye tracking with electroencephalogram (EEG) records (e.g., event-related fixations and postsaccadic event-related potentials). Along with more traditional assessment of the socioemotional effect of tears, such research could lead to new hypotheses and new advances.

Regarding the limitations of the present work, the present design could be improved by examining the results in an additional control condition including other visual stimuli depicted in the faces of models instead of tears (e.g., a freckle, a wart, or a mole under the eye) in order to study differential gaze behavior and thus add useful physiological data to the behavioral results of Provine (2014, 2017). Moreover, we advise prudence when interpreting the correlational results: although the sample size was appropriate for the experimental study – as confirmed by the power analysis – the exploration of how personality variables are associated with facial recognition in the presence of visible tears will require a larger sample to draw solid conclusions. In addition, this study was performed with a limited number of visual stimuli (i.e., faces). We could have increased the number of stimuli to be more in line with other studies, but the selection was made with the aim of replicating and extending previous findings (van de Ven et al., 2016; Vingerhoets et al., 2016; Picó et al., 2020). Moreover, as mentioned in "Materials and Method," the subjects were selected based on their ecological validity, i.e., they were calm crying in a spontaneous way. Finally, it should be taken into account that, due to availability (high female bias), we carried out our experiments in a purely female population; therefore, until the results are replicated with male participants, our conclusions should be applied to the general population with caution. In this respect, it should be pointed out that, according to Mulac et al. (1986), female dyads make much greater visual contact during interactions than male counterparts. This trend has been observed in other cultures (Wada, 1990) and is consistent with evidence that women are more sensitive non-verbal communicators (Rosenthal and DePaulo, 1979; Rosenthal, 1979) and exhibit greater sensitivity to non-verbal cues (Keeley-Dyreson et al., 1991) than men. Therefore, we advise caution in generalizing our conclusions on eye tracking results with respect to both genders when studying non-verbal behavior.

CONCLUSION

Visible tears proved to be magnets for gaze during a face-viewing task. When they were present, the inspection pattern changed qualitatively and quantitatively, with participants becoming fully focused on the tears. The mere presence of a single teardrop running down the cheek was associated with increased emotional inference and a greater perception of sincerity. Interestingly, visible tears generated different reactions depending on the observer's personality traits, with a positive relationship observed between cognitive empathy and the perception of greater emotional intensity in tearful faces. All in all, eye tracking technology seems to be an effective tool for studying the visual aspect of emotional crying, and we hope that the present study will be the first of many empirical works that investigate the interpersonal effects of tears. Additionally, we have commented on several of the methodological aspects that should be taken into account when using eye tracking technology to study non-verbal behavior, some of which have been neglected until now. Further exploration of the relationship between empathy and tear perception using eye tracking could be a fruitful avenue for future research.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité de Ética de la Universidad de Valencia. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AP and MG conceived the study and wrote the manuscript. AP carried out the study and made the statistical analyses of the data. RE contributed to the interpretation of the study and made a critical revision of the draft. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Cristina López Celda for her help. As a last-year undergraduate student, her contribution was fundamental in collecting a substantial part of the data.

REFERENCES

Alexander, T. (2003). Narcissism and the experience of crying. *Br. J. Psychother.* 20, 27–37. doi: 10.1111/j.1752-0118.2003.tb00112.x

Andersson, R., Nystrom, M., and Holmqvist, K. (2010). Sampling frequency and eye tracking measures: how speed affects durations, latencies, and more. *J. Eye Mov. Res.* 3, 1–12.

Andrews, T., and Coppola, D. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vis. Res.* 39, 2947–2953. doi: 10.1016/s0042-6989(99)00019-x

Argyle, M., and Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica* 6, 32–49.

Balsters, M., Krahmer, E., Swerts, M., and Vingerhoets, A. (2013). Emotional tears facilitate the recognition of sadness and the perceived need for social support. *Evolut. Psychol.* 11, 148–158. doi: 10.1177/147470491301100114

Bek, J., Poliakoff, E., and Lander, K. (2020). Measuring emotion recognition by people with Parkinson's disease using eye tracking with dynamic facial expressions. *J. Neurosci. Methods* 331, 108524. doi: 10.1016/j.jneumeth.2019.108524

Buss, D. (1987). Selection, evocation, and manipulation. *J. Personal. Soc. Psychol.* 56, 1214–1221. doi: 10.1037//0022-3514.53.6.1214

Caldero-Alonso, A. (2009). *Estudio de los Resultados Obtenidos en el Cuestionario Salamanca en Población Normal*. Doctoral Dissertation Spain: University of Salamanca.

Carr, M. B., and Lutjemeier, J. A. (2005). The relation of facial affect recognition and empathy to delinquency in youth offenders. *Adolescence* 40, 601–619.

Champely, S. (2018). *pwr: Basic Functions for Power Analysis. R package version* 1.2-2.

Dodge, R. (1903). Five types of eye-movement in the horizontal meridian plane of the field of regard. *Am. J. Physiol.* 8, 307–329. doi: 10.1152/ajplegacy.1903.8.4.307

Dodge, R., and Cline, T. S. (1901). The angle velocity of eye-movements. *Psychol. Rev.* 8, 145–157. doi: 10.1037/h0076100

Duchowski, A. T. (2017). *Eye Tracking Methodology*. Cham: Springer.

Exline, R. V., and Fehr, B. J. (1982). "The assessment of gaze," in *Handbook of Methods in Nonverbal Behavior Research*, eds K. R. Scherer and P. Ekman (Cambridge: Cambridge University Press), 91–135.

Fehr, B. J., and Exline, R. V. (1987). "Social visual interactions: A conceptual and literature review," in *Nonverbal Behavior and communication,* Vol. 2, eds A. W. Siegman and S. Feldstein (Hillsdale, NJ: Lawrence Erlbaum), 225–326.

Gehrer, N. A., Duchowski, A. T., Jusyte, A., and Schönenberg, M. (2020). Eye contact during live social interaction in incarcerated psychopathic offenders. *Personal. Disord. Theory Res. Treatment*. doi: 10.1037/per0000400. Advance online publication.

Gery, I., Miljkovitch, R., Berthoz, S., and Soussignan, R. (2009). Empathy and recognition of facial expressions of emotion in sex offenders, non-sex offenders and normal controls. *Psych. Res.* 165, 252–262. doi: 10.1016/j.psychres.2007.11.006

Goldberg, J. H., and Helfman, J. I. (2010). "Comparing information graphics," in *Proceedings of the 3rd BELIV10 Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization - BELIV 10*, New York, NY doi: 10.1145/2110192.2110203

Gračanin, A., Bylsma, L. M., and Vingerhoets, A. J. J. M. (2017). "The communicative and social functions of human crying," in *Oxford Series in Social Cognition and Social Neuroscience. The Science of Facial Expression*, eds J.-M. Fernández-Dols and J. A. Russell (New York, NY: Oxford University Press), 217–233.

Harrigan, J. A., Rosenthal, R., and Scherer, K. R. (2005). *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford: Oxford University Press.

Harrison, N. A., Wilson, C. E., and Critchley, H. D. (2007). Processing of observed pupil size modulates perception of sadness and predicts empathy. *Emotion* 7, 724–729. doi: 10.1037/1528-3542.7.4.724

Hendriks, M. C. P., and Vingerhoets, A. J. J. M. (2006). Social messages of crying faces: their influence on anticipated person perception, emotions and behavioural responses. *Cogn. Emot.* 20, 878–886. doi: 10.1080/02699930500450218

Holmqvist, K., and Andersson, R. (2017). *Eye Tracking: A Comprehensive Guide to Methods, Paradigms and Measures*. Lund: Lund Eye tracking Research Institute.

Hunter, L., Roland, L., and Ferozpuri, A. (2020). Emotional expression processing and depressive symptomatology: eye tracking reveals differential importance of lower and middle facial areas of interest. *Depress. Res. Treatment* 2020, 1–7. doi: 10.1155/2020/1049851

Iskra, A., and Gabrijelcic, H. (2016). Eye tracking analysis of face observing and face recognition. *J. Graphic Eng. Design* 7, 5–11. doi: 10.24867/jged-2016-1-005

Ito, K., Ong, C. W., and Kitada, R. (2019). Emotional tears communicate sadness but not excessive emotions without other contextual knowledge. *Front. Psychol.* 10:878. doi: 10.3389/fpsyg.2019.00878

Johansson, R., Holmqvist, K., Mossberg, F., and Lindgren, M. (2012). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychol. Music* 40, 339–356. doi: 10.1177/0305735610387777

Keeley-Dyreson, M., Burgoon, J. K., and Bailey, W. (1991). The effects of stress and gender on nonverbal decoding accuracy in kinesic and vocalic channels. *Hum. Commun. Res.* 17, 584–605. doi: 10.1111/j.1468-2958.1991.tb00245.x

Killgore, W. D., and Yurgelun-Todd, D. A. (2004). Activation of the amygdala and anterior cingulate during nonconscious processing of sad versus happy faces. *NeuroImage* 21, 1215–1223. doi: 10.1016/j.neuroimage.2003.12.033

Kirkland, J., and Lewis, C. (1976). Glance, look, gaze, and stare: a vocabulary for eye-fixation research. *Percept. Mot. Skills* 43:1278. doi: 10.2466/pms.1976.43.3f.1278

Klineberg, O. (1940). *Social Psychology*. New York, NY: Holt.

Kottler, J. A. (1996). *The Language of Tears*. San Francisco: Jossey-Bass Publishers.

Kret, M. (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

Krivan, S. J., Caltabiano, N., Cottrell, D., and Thomas, N. A. (2020). I'll cry instead: mu suppression responses to tearful facial expressions. *Neuropsychologia* 2020:107490. doi: 10.1016/j.neuropsychologia.2020.107490

Krivan, S. J., and Thomas, N. A. (2020). A call for the empirical investigation of tear stimuli. *Front. Psychol.* 11:52. doi: 10.3389/fpsyg.2020.00052

Küster, D. (2018). Social effects of tears and small pupils are mediated by felt sadness: An evolutionary view. *Evol. Psychol.* 16:147470491876110. doi: 10.1177/1474704918761104

Laan, A. J., van Assen, M. A. L. M., and Vingerhoets, A. J. J. M. (2012). Individual differences in adult crying: the role of attachment styles. *Soc. Behav. Personal. Int. J.* 40, 453–472. doi: 10.2224/sbp.2012.40.3.453

Lockwood, P., Millings, A., Hepper, E., and Rowe, A. C. (2013). If i cry. do you care? *J. Individ. Differ.* 34, 41–47. doi: 10.1027/1614-0001/a000098

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cogn. Psychol.* 3, 525–551. doi: 10.1016/0010-0285(72)90021-7

Lundqvist, D., Flykt, A., and Öhman, A. (1998). *The Karolinska Directed Emotional Faces – KDEF, [CD ROM] from Department of Clinical Neuroscience, Psychology Section*. Stockholm: Karolinska Institute.

Mair, P., and Wilcox, R. R. (2020). Robust statistical methods in R using the WRS2 package. *Behav. Res. Methods* 52, 464–488. doi: 10.3758/s13428-019-01246-w

Martin-Key, N., Graf, E., Adams, W., and Fairchild, G. (2017). Facial emotion recognition and eye movement behaviour in conduct disorder. *J. Child Psychol. Psych.* 59, 247–257. doi: 10.1111/jcpp.12795

Mulac, A., Lundell, T. L., and Bradac, J. J. (1986). Male/female language differences and attributional consequences in a public speaking situation: toward an explanation of the gender-linked language effect. *Commun. Monographs* 53, 115–129. doi: 10.1080/03637758609376131

Peckham, A. D., Johnson, S. L., and Tharp, J. A. (2016). Eye tracking of attention to emotion in bipolar i disorder: links to emotion regulation and anxiety comorbidity. *Int. J. Cogn. Ther.* 9, 295–312. doi: 10.1521/ijct_2016_09_12

Pérez-Urdániz, A., Rubio Larrosa, V., and Gómez Gazol, E. (2011). *Cuestionario Salamanca de Trastornos de la Personalidad*.

Peter, M., Vingerhoets, A. J. J. M., and van Heck, G. L. (2001). Personality. gender, and crying. *Eur. J. Personal.* 15, 19–28. doi: 10.1002/per.386

Picó, A., Gadea, M., and Espert, R. (2018). Estudio del efecto lágrima con metodología de seguimiento de la mirada: un estudio piloto. *Revista De Discapacidad, Clínica y Neurociencias* 5, 1–8.

Picó, A., Gračanin, A., Gadea, M., Boeren, A., Aliño, M., and Vingerhoets, A. (2020). How visible tears affect observers' judgements and behavioral intentions: sincerity, remorse, and punishment. *J. Nonverbal Behav.* 44, 215–232. doi: 10.1007/s10919-019-00328-9

Provine, R., Krosnowski, K., and Brocato, N. (2009). Tearing: breakthrough in human emotional signaling. *Evolut. Psychol.* 7, 52–56. doi: 10.1177/147470490900700107

Provine, R. R. (2014). *Curious Behavior: Yawning, Laughing, Hiccupping, and Beyond*. Cambridge: The Belknap Press of Harvard University Press.

Provine, R. R. (2017). "Beyond the smile nontraditional facial, emotional, and social behaviors," in *The Science of Facial Expression*, eds J. A. Russell and J. M. Fernandez Dols (Oxford: Oxford Scholardship Press), 197–216. doi: 10.1093/acprof:oso/9780190613501.001.0001

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461

Regan, P. C., and Baker, S. J. (1998). The impact of child witness demeanor on perceived credibility and trial outcome in sexual abuse cases. *J. Family Violence* 13, 187–195. doi: 10.1023/A:1022845724226

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. Evanston, Il: Northwestern University.

Roes, F. L. (1989). On the origin of crying and tears. *Hum. Ethol. Newsletter* 5, 5–6.

Rogers, S. L., Speelman, C. P., Guidetti, O., and Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Sci. Rep.* 8, 4271. doi: 10.1038/s41598-018-22726-7

Rommelse, N. N., Stigchel, S. V. D., and Sergeant, J. A. (2008). A review on eye movement studies in childhood and adolescent psychiatry. *Brain Cogn.* 68, 391–414. doi: 10.1016/j.bandc.2008.08.025

Rosenthal, R. (1979). "Skill in nonverbal communication," in *Individual differences*, ed. R. Rosenthal (Cambridge, MA: Oelgeschlager, Gunn & Hain).

Rosenthal, R., and DePaulo, B. M. (1979). Sex differences in eavesdropping on nonverbal cues. *J. Personal. Soc. Psychol.* 37, 273–285. doi: 10.1037/0022-3514.37.2.273

Rubo, M., and Gamer, M. (2018). Social content and emotional valence modulate gaze fixations in dynamic scenes. *Sci. Rep.* 8:3804. doi: 10.1038/s41598-018-22127-w

Savulich, G., Freeman, D., Shergill, S., and Yiend, J. (2015). Interpretation bias in paranoia. *Behav. Ther.* 46, 110–124. doi: 10.1016/j.beth.2014.08.002

Stadel, M., Daniels, J. K., Warrens, M. J., and Jeronimus, B. F. (2019). The gender-specific impact of emotional tears. *Motiv. Emot.* 43, 696–704. doi: 10.1007/s11031-019-09771-z

Tatler, B. W., and Vincent, B. T. (2008). Systematic tendencies in scene viewing. *J. Eye Movem. Res.* 2, 1–18. doi: 10.16910/jemr.2.2.5

Tavakoli, H., Atyabi, A., Rantanen, A., Laukka, S. J., Nefti-Meziani, S., et al. (2015). predicting the valence of a scene from observers'. eye movements. *PLoS One* 10:e0138198. doi: 10.1371/journal.pone.0138198

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

van de Ven, N., Meijs, M. H. J., and Vingerhoets, A. (2016). What emotional tears convey: tearful individuals are seen as warmer, but also as less competent. *Br. J. Soc. Psychol.* 56, 146–160. doi: 10.1111/bjso.12162

Van Kleef, G. A. V. (2008). "Emotion in conflict, and negotiation: introducing. the emotions as social information (EASI) model," in *Research Companion to Emotion in Organizations*, eds N. M. Ashkanasy and C. L. Cooper (Cheltenham: Edward Elgar), doi: 10.4337/9781848443778.00034

Vingerhoets, A. (2013). *Why Only Humans Weep*. Oxford: Oxford University Press.

Vingerhoets, A., van de Ven, N., and van der Velden, Y. (2016). The social impact of emotional tears. *Motivat. Emot.* 16, 455–463. doi: 10.1007/s11031-016-9543-0

Vingerhoets, A. J. J. M., and Bylsma, L. M. (2016). The riddle of human emotional crying: a challenge for emotion researchers. *Emot. Rev.* 8, 207–217. doi: 10.1177/1754073915586226

Vingerhoets, A. J. J. M., and Cornelius, R. (2001). *Adult Crying: a Biopsychosocial Approach. Hove (UK): Brunner-*. Abingdon: Routledge.

Vuilleumier, P., Armony, J., Clarke, K., Husain, M., Driver, J., and Dolan, R. (2002). Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia* 40, 2156–2166. doi: 10.1016/s0028-3932(02)00045-3

Wada, M. (1990). The effect of interpersonal distance change on nonverbal behaviors. *Jpn. Psychol. Res.* 32, 86–96. doi: 10.4992/psycholres1954.32.86

Walter, C. (2008). *Thumbs, Toes, and Tears: and Other Traits that Make us Human*. New York, NY: Walker and Company.

Walter, H. (2012). Social cognitive neuroscience of empathy: concepts. circuits, and genes. *Emot. Rev.* 4, 9–17. doi: 10.1177/1754073911421379

Wang, J., and Wang, X. (2020). *Structural Equation Modeling: Applications Using Mplus*. Hoboken, N.J: Wiley.

Warrier, V., Bethlehem, R. A., and Baron-Cohen, S. (2017a). "The "Reading the Mind in the Eyes" Test (RMET)," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill and T. Shackelford (Cham: Springer), doi: 10.1007/978-3-319-28099-8

Warrier, V., Grasby, K., Uzefovsky, F., Toro, R., Smith, P., Chakrabarti, B., et al. (2017b). A meta-analysis of cognitive empathy: heritability and correlates of the 'Reading the mind in the Eyes' test with psychiatric conditions, psychological traits and subcortical volumes. *Eur. Neuropsychopharmacol.* 27, S503–S504. doi: 10.1016/j.euroneuro.2016.09.608

Wilcox, R. R., and Tian, T. (2011). Measuring effect size: a robust heteroscedastic approach for two or more groups. *J. Appl. Stat.* 38, 1359–1368. doi: 10.1080/02664763.2010.498507

Xiao, N., Quinn, P., Liu, S., Ge, L., Pascalis, O., and Lee, K. (2015). Eye tracking reveals a crucial role for facial motion in recognition of faces by infants. *Dev. Psychol.* 51, 744–757. doi: 10.1037/dev000001

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York, NY: Plenum Press.

Zeifman, D., and Brown, S. (2011). Age-related changes in the signal value of tears. *Evol. Psychol.* 9, 313–324. doi: 10.1177/147470491100900304

Zickfeld, J., van de Ven, N., Schubert, T. W., and Vingerhoets, A. (2018). Are tearful individuals perceived as less competent? probably not, *Compr. Results Soc. Psychol.* 3, 119–139. doi: 10.1080/23743603.2018.1514254

Zickfeld, J., Vingerhoets, A., Schubert, T. W., van de Ven, N., Pich, O., Hartanto, A., et al. (2020). *Cross Cultural Tears – A Systematic Investigation of the Interpersonal Effects of Emotional Crying Across Different Cultural Backgrounds*. Available at http://www.osf.io/fj9bd (accessed June 3, 2020).

Conflict of Interest: 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.

frontiers in Psychology

PERSPECTIVE
published: 22 October 2020
doi: 10.3389/fpsyg.2020.581701

Check for updates

More Data, Please: Machine Learning to Advance the Multidisciplinary Science of Human Sociochemistry

Jasper H. B. de Groot[1,2], Ilja Croijmans[1] and Monique A. M. Smeets[1]*

[1] Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, Netherlands, [2] Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

Communication constitutes the core of human life. A large portion of our everyday social interactions is non-verbal. Of the sensory modalities we use for non-verbal communication, olfaction (i.e., the sense of smell) is often considered the most enigmatic medium. Outside of our awareness, smells provide information about our identity, emotions, gender, mate compatibility, illness, and potentially more. Yet, body odors are astonishingly complex, with their composition being influenced by various factors. Is there a chemical basis of olfactory communication? Can we identify molecules predictive of psychological states and traits? We propose that answering these questions requires integrating two disciplines: psychology and chemistry. This new field, coined *sociochemistry*, faces new challenges emerging from the sheer amount of factors causing variability in *chemical composition* of body odorants on the one hand (e.g., diet, hygiene, skin bacteria, hormones, genes), and variability in *psychological* states and traits on the other (e.g., genes, culture, hormones, internal state, context). In past research, the reality of these high-dimensional data has been reduced in an attempt to isolate unidimensional factors in small, homogenous samples under tightly controlled settings. Here, we propose big data approaches to establish novel links between chemical and psychological data on a large scale from heterogeneous samples in ecologically valid settings. This approach would increase our grip on the way chemical signals non-verbally and subconsciously affect our social lives across contexts.

Keywords: sense of smell, machine learning, chemosignals, non-verbal communication, social and personality psychology

OPEN ACCESS

Edited by:
Marianne Schmid Mast, University of Lausanne, Switzerland

Reviewed by:
Vivian P. Ta, Lake Forest College, United States
Laiquan Zou, Southern Medical University, China

****Correspondence:***
Jasper H. B. de Groot
J.deGroot@psych.ru.nl

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *09 July 2020*
Accepted: *05 October 2020*
Published: *22 October 2020*

Citation:
de Groot JHB, Croijmans I and Smeets MAM (2020) More Data, Please: Machine Learning to Advance the Multidisciplinary Science of Human Sociochemistry. Front. Psychol. 11:581701. doi: 10.3389/fpsyg.2020.581701

INTRODUCTION

Humans are surprisingly good smellers. The pervasive myth that humans are only "tiny smellers" has been debunked by 21st century research showing a wide array of smell skills (Stevenson, 2010; de Groot et al., 2017; McGann, 2017). To name a few: humans can follow a scent-trail (like sniffer dogs; Porter et al., 2007), detect certain odorants at extremely low levels (few droplets in an Olympic size swimming pool; Yeshurun and Sobel, 2010), and identify diseases like Parkinson's before actual diagnosis (Trivedi et al., 2019). In our everyday lives, smells have a "communicative" function, informing us about the quality of food and warning us for environmental hazards (e.g., gas leaks) (Stevenson, 2010). An even less well-known function of smell is *social* communication (de Groot et al., 2017; Parma et al., 2017; Pause, 2017; Roberts et al., 2020); the topic of this article. Studies have shown that our smells provide others with cues about our identity and gender (Penn et al., 2007), age (Mitro et al., 2012), health (Olsson et al., 2014), and emotions (de Groot et al., 2015; Pause et al., 2020). This form of communication occurs without our voluntary control and generally outside of our awareness, which imbues chemical communication with mystery. Demystifying the spreading of social information through smell was listed in *Science* as one of the 125 most compelling multidisciplinary puzzles facing scientists this century (Kennedy and Norman, 2005). Our goal here is to outline *how* researchers could go about answering this query, whether there is a universal "language" of social smells. Society at large will be helped by optimally leveraging fundamental insights emerging from this view to worldwide industrial and clinical applications that could improve a person's quality of life.

Social smells are markedly complex: body odor contains thousands of molecules (de Lacy Costello et al., 2014), and massive variability is caused by factors including genotype, hormonal status, mood, skin bacteria, diet, smoking, hygiene habits, clothing, and use of fragranced products (e.g., Natsch and Emter, 2020; Roberts et al., 2020). Past studies have generally sidestepped this challenge by performing small-scale psychological experiments under carefully controlled, sterile conditions (for a meta-analysis: de Groot and Smeets, 2017; for a critical view: Wyatt, 2020). These studies formed the first stepping stones by strongly suggesting that social information can be communicated via smell under tightly controlled settings; yet, (i) the molecules transmitting the message have generally remained elusive, as well as (ii) the ecological settings in which chemical communication occurs. Not dealing with these obstacles could deadlock future research efforts to "anchor" molecules to their social source. To accelerate future research, we propose (i) multidisciplinary ways of working by integrating psychology and chemistry toward a science of human *sociochemistry* (**Box 1**), and (ii) moving outside of the sterile lab to test subjects with diverse backgrounds.

The sociochemistry we advance is a multidisciplinary, ecological approach that in view of its inherent complexity requires an ecosystem of academic institutions around the world to flourish, by working together to create speed and scale (cf. Forscher et al., 2020). We propose the building of open access databases holding information that spans across chemistry (e.g., chemical composition of sweat odor) and psychology (e.g., capturing the states and traits of those participating in the chemosignaling as well as their unique contextual information. Machine learning techniques can be applied to generate models that may accurately predict molecules' sway on our social lives across diverse contexts and samples, with technological, societal, and clinical applications following suit.

> **BOX 1 |** Definition of sociochemistry. With sociochemistry we refer to the *multidisciplinary science* examining non-verbal social communication via human body odor, particularly focusing on the chemistry between people.

In what follows, we will first outline the initial research questions, methods, and advances of past research, before we will identify current obstacles to a broader and deeper understanding of human chemical communication, and we will end with a perspective on how to overcome these hurdles in future research.

PAST RESEARCH: SIMPLIFYING A COMPLEX PROBLEM

Communication is crucial to humans. Most of our communication is non-verbal. Of all the sensory channels engaged in non-verbal communication, smells arguably pose the biggest deciphering challenge.

The past research in this field initially focused on determining *what* social information can be communicated via smell. To test this, researchers have systematically attempted to eradicate "noise" on the chemical communication channel by controlling extraneous factors (e.g., diet, hygiene, fragranced product use) and testing homogeneous samples in carefully controlled lab experiments. In these studies, sweat was collected from *senders* (who had kept to a scent-free regimen for multiple days to isolate the experimentally-induced chemical "message") and presented to *receivers* in a separate experiment. Chemical communication was inferred from recipients' behavioral, affective, physiological, neuroendocrine and/or neural responses matching the sender's state. This way, numerous double-blind experiments showed that human smells can convey information from fleeting emotions and sickness, to more enduring traits like identity, gender, reproductive status, and age (for reviews, see de Groot et al., 2017; Parma et al., 2017; Pause, 2017).

Although past research on human chemical communication has provided initial insights into the type of information human odors can bring across, we identify a number of obstacles for a better (quicker, broader, and deeper) understanding of non-verbal communication via smell.

Problem I: Small Scale, Slow Speed

The current science of non-verbal communication via smell is rooted in a longstanding tradition of strictly controlled laboratory experiments focusing on the empirical testing of hypotheses addressing cause-effect relations, using reliable and validated methods and carefully calibrated instruments (for empirical demonstrations, see e.g., Chen and Haviland-Jones, 2000; Regenbogen et al., 2017; Endevelt-Shapira et al., 2018; Quintana et al., 2019; de Groot et al., 2020b; Gomes et al., 2020; Pause et al., 2020 (for recent narrative overviews, see e.g., Loos et al., 2019; Ferdenzi et al., 2020; Havlíček et al., 2020 (for meta-analyses, see e.g., Gildersleeve et al., 2014; de Groot and Smeets, 2017). This approach, with a preference for intrinsic over extrinsic validity, has been the method of choice to build our (psychological) science for decades. Despite advantages of scientific rigor and quality, there are problems in speed and scale. With little coordination across labs around the world, different researchers may be working on similar research questions (e.g., "can humans smell fear?"; Mujica-Parodi et al., 2009; Pause et al., 2009; Prehn-Kristensen et al., 2009; Zhou and Chen, 2009), each moving through the laborious cycle of recruiting, screening, and testing senders and receivers with barely sufficient statistical power (as outlined by Wyatt, 2020). The essence to our argument is that there is a stark contrast between the *complexity* of the problem at the root of sociochemistry, which is the mystery of the correspondence between the chemical "code" and the message it carries on the one hand, and the relatively slow tactic of churning experiments one at a time.

Problem II: Generality of Findings

Second, we need to characterize the generality of findings or extrinsic validity of the traditional experiments (Simons et al., 2017). Both uniformity in subject characteristics and test settings form obstacles to a broader understanding of the potentially species-wide and real-world impact of non-verbal communication via smell. Open queries include: Is the language of smell universal? How much of this communication is modified by context, a powerful moderating factor in olfactory science (e.g., Dalton, 1999; De Araujo et al., 2005; de Groot et al., 2020a)? Can this language be "heard" beyond the thick walls of labs, in noisy field settings? Answering these questions will help chart the impact of social smells on the daily lives of many.

Because past research has been typified by (i) context-deprived lab experiments, presenting (ii) uncontaminated sweat samples, using (iii) a relatively small number of subjects with (iv) relatively uniform characteristics, we currently have no knowledge of how broadly shared human olfactory communication is. To illustrate, the male-to-female chemical communication dyad initially served to increase experimental sensitivity, with males generally having the larger and more active sweat glands, and females being the slightly better smellers (but see this meta-analysis: Sorokowski et al., 2019; and this review: Majid et al., 2017, for gender differences that are at most small and affecting only higher order smell processing). Although initially useful, this gender uniformity adds a constraint on generality, and the same goes for the almost exclusive reliance on participants that are Western, Educated, Industrialized, Rich, Democratic (WEIRD; Henrich et al., 2010) (cf. de Groot et al., 2018; Roberts et al., 2020). Generalizing research findings from WEIRD samples to other populations is a major problem in science in general, and a particularly pressing issue when one examines the breadth and scale of the non-verbal language of smells (**Box 2**).

Problem III: Unidisciplinary Research

Third, to be able to forge a link between smell molecules and behavior we need to move beyond a single-discipline research tradition. Although several psychological studies have revealed systematic patterns in the behaviors of senders and recipients (in relatively sterile, uniform settings), the *chemical message* driving this coupling has generally remained enciphered (but see Penn et al., 2007; Smeets et al., 2020). Lessons can be learned from the animal literature, where the combination of rigorous behavioral experiments (bioassay) and chemical analysis (isolating, identifying, and synthesizing the bioactive substance to recreate the bioassay-behavior) forms the golden standard to detect a common chemical "language" for a species: *pheromones* (Wyatt, 2015, 2020). But the definition of pheromones, rooted in entomological research as single molecules eliciting innate responses in a conspecific (Karlson and Lüscher, 1959), appears outdated and unsuitable for mammals like humans, as our smell perception strongly depends on learning and context, and our body emits a multitude of molecules (de Groot et al., 2017). The minimum pragmatic evidence, however, is to determine (in a collaborative, multi-lab effort) whether human chemical language is *consistent in form* (requiring a multidisciplinary approach) and *broadly shared* across the human species (requiring diverse samples and settings).

BOX 2 | Sociochemical language. A sociochemical language would imply configurations of chemical symbols that convey meaning, which meaning is acquired via learning. This notion of language would acknowledge the possibility that (i) identical chemical configurations do not mean the same to everyone, (ii) the meaning of an identical chemical configuration may vary even to a single individual depending on context, (iii) that there is (substantial) variation or "noise" around one chemical configuration, from which one single uniform meaning can still be distilled. Therefore, the language would not have to be universal.

PROSPECTIVE ADVANCES

In the wake of recent developments in psychological research and theory, chemical analytical technology, and data science (discussed below), substantial progress can be made now to unravel the symbol system of social smell. Specifically, we outline an integration of traditional psychology methods and chemistry toward a new science of human sociochemistry, studying human chemosignaling across various ecologically valid settings and samples, across all human diversity. To deal with the complexity and large, multidimensional databases that emerge from this interdisciplinary, ecologically valid endeavor, we propose applying data science approaches like machine learning. We anticipate that large scale multidisciplinary collaborations are required to get us closer to identifying the alphabet of the language of social smells and assess its real-world impact.

Multidisciplinary Approach: Deciphering the Alphabet of Social Smells

Any attempt to get closer to the answer of whether social smells convey a common language requires a multidisciplinary combination of psychological experiments and chemical analysis.

Most research on human chemical communication focused on psychological effects. The few studies that did apply chemical analysis have shown that certain characteristics and transient emotions could be identified in a sender's body odor. One pioneering study by Penn et al. (2007) showed that a person's identity and gender could be expressed in a person's body odor, with 14 molecules predicting gender with 75% accuracy. Based on remarkable anecdotal evidence that a human "super smeller" could detect Parkinson's Disease (PD) by smell, Trivedi et al. (2019) found that four compounds (eicosane, hippuric acid, octadecanal, and perillic aldehyde) were characteristic markers of PD; when smelling these compounds, the super smeller subjectively reported a strong PD smell. Other studies found chemical markers suggestive of fear (and happiness). Potential chemical markers for fear were identified by in armpit odor (Smeets et al., 2020), stress levels were also expressed in a person's breath (Preti et al., 2019; acetone, isoprene, dimethyl sulfide), and in a creative field study, (Williams et al., 2016) showed that scary and funny film events reliably changed the emission of molecules from cinema audiences. Taken together, these multidisciplinary studies show the potential for social information to be encoded in a person's smell in predictable ways, thus jumpstarting a sociochemistry approach to identify a common smell language.

Whereas on the one end of the scientific spectrum, we have this classic tradition of sequentially conducting laboratory experiments designed to address a specific causal hypothesis derived from theory, carefully controlling for measurement error and extraneous influence. On the other end there is the big data approach relying on machine learning analytical techniques performed on big databases holding what seems to be unrelated information from large populations to magically reveal unexpected correlations unencumbered by theory (Mayer-Schönberger and Cukier, 2013). Neither, on its own, will be an optimal path for unraveling human sociochemistry and the underlying language on which it is built. What we propose, instead, is a hybrid approach, a combination in which machine learning techniques are used to help us find handles on and insights into the composition of the chemical signal combinations that are the building blocks of the signal, and the related individual and external variables to further sculp this unique form of social communication. These insights will contribute to the formulation of hypotheses about cause and effect that can then be isolated and tested in controlled lab environments (cf. Wyatt, 2015, 2020).

Ecological Validity: A Broadly Shared, Widely Used Social Smell Language?

In the quest for discovering a potential universal language of smell that is also societally relevant, we argue that the highest success rate can be achieved by first examining smells whose detection generally aids survival (Schaal and Porter, 1991).

In the earliest stages of life, when vision and hearing are still underdeveloped, the smell of mothers' milk is a powerful cue that attracts a newborn to the food source (Schaal et al., 2020). Even formula-fed newborns oriented more toward the smell of an unfamiliar lactating woman than to the familiar formula smell (Porter et al., 1991); and this was not a novelty effect, as the same smell was also preferred over the breast odor of nulliparous women (Makin and Porter, 1989). There may well be universal chemical cues in the breast odor of lactating women that attract most if not all newborns under diverse ecologically valid settings, but this still requires empirical investigation from non-WEIRD samples (Schaal et al., 2020).

Humans would also benefit from picking up smells indicating danger, like fear sweat threatening physical harm, and disease sweat threatening contamination. The capacity to register these invisible, far-reaching, and long-lasting chemical warning cues would have increased our ancestors' survival chances. Indeed, the smell of fear has been shown to instigate adaptive processes: a fearful facial expression (raised eyebrows, opened nose) and increased sensory intake (eyes and nose) to better detect threat (de Groot et al., 2012); yet, typically this phenomenon has not been examined beyond WEIRD samples, with one East Asian exception (de Groot et al., 2018). Quintana and colleagues (2019) further assessed the breadth of chemical communication in a controlled yet ecologically valid Virtual Reality environment. They found that smelling fear/stress sweat induced anxiety in recipients and reduced their interpersonal trust toward a virtual character. Even outside of the lab, the smell of fear (masked in clove odor, making it undetectable) could negatively impact dental student performance (Singh et al., 2018). Indeed, odor masking (e.g., with perfume, deodorant) could not prevent recipients from making consistent and reliable smell-based social judgments at typical social distances (Gaby and Zayas, 2017). Taken together, these findings allude to fear/stress smell affecting behavior across contexts in diverse samples, but more data is needed.

The complex and resource-intensive methodology of sweat sampling and exposure has arguably held back large (field) experiments, but upscaling and including natural settings seems inevitable in an attempt to discover the commonalities in human smells and their practical application, with big data approaches providing structure within the anticipated wealth of transdisciplinary data.

Machine Learning: Solving the Big Data Challenge Ahead of Us

In vision and hearing, the wavelength of light and frequency of sound are highly predictive of color and tone; yet, predicting the smell of a molecule from its chemical structure is much harder. In the past decade, researchers have started using machine learning techniques to demonstrate links between molecular structure and odor perception (for an overview: Lötsch et al., 2018). Machine learning, a popular application of artificial intelligence, is a set of methods that can be used to automatically detect patterns in data and use these patterns to predict or classify future data (e.g., Murphy, 2012; Dhar, 2013). Although machine learning models have shown the feasibility of predicting odor perception from relatively simple, non-social smells (Khan et al., 2007; Zarzo, 2011; Snitz et al., 2013; Keller et al., 2017; Gutiérrez et al., 2018; Sanchez-Lengeling et al., 2019) a number of extra challenges emerge when machine learning is applied to uncover the language of social smells. The difference between past "non-social" models and what we propose here is that (i) past models predicted odor perception from physico-chemical properties of *single* chemical compounds, whereas body odors are mixtures of compounds, and the communicative signal also likely having a multi-component architecture (Loos et al., 2014), the composition of which requires employing chemical analytical techniques to elucidate; (ii) past model endpoints have traditionally been

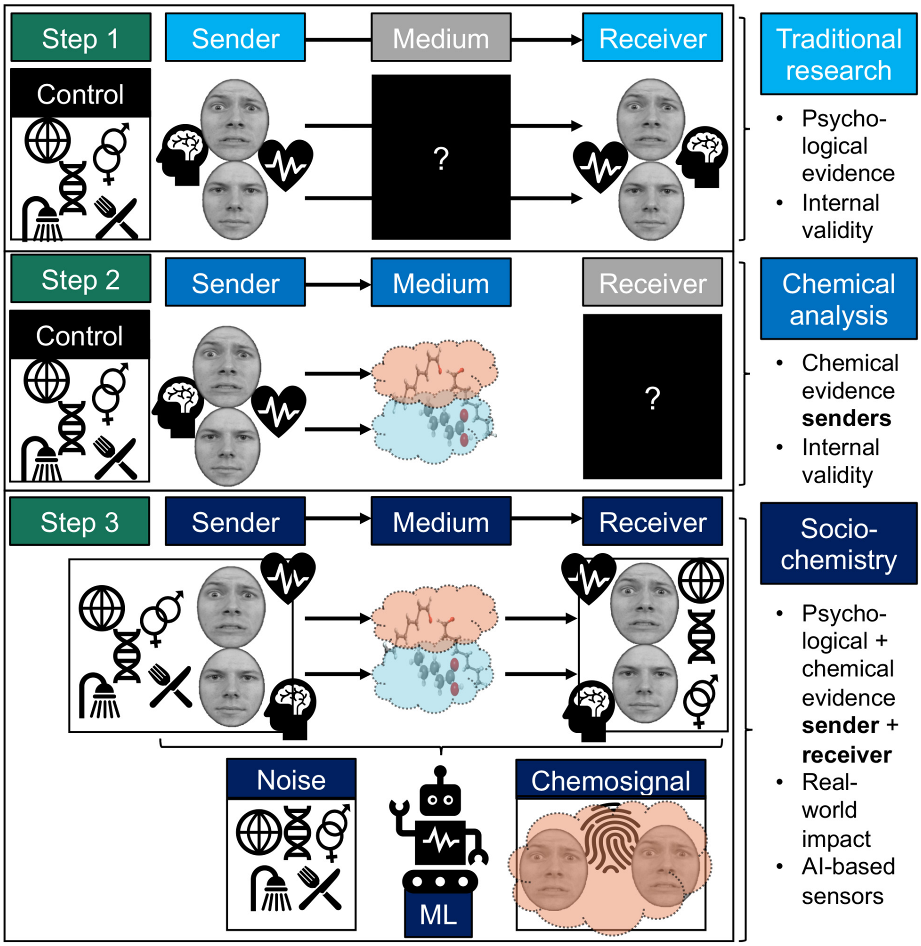


FIGURE 1 | Possible pathway to understanding sociochemistry using machine learning (ML). The different steps denote past/present (Step 1 and 2) and proposed (Step 3) approaches to elucidate human social communication via smell. The steps are increasingly data-intense and complex and go from uni- to multidisciplinary research. Although there is no strict order, Step 1 and 2 can form initial building blocks for sociochemistry (Step 3), by testing psychological correspondence between senders and receivers in traditional ways (Step 1; chemical medium "remains" black box); and by decoding psychological/clinical information from the sender's smell (Step 2; the receiver's response and therefore the *social* chemosignal remains "black box"). Controlling for various factors (e.g., genotype, culture, gender, hygiene, diet) is recommended here to initially isolate the signal and/or its psychological effect. However, true chemical communication (e.g., of emotions like fear) involves (i) studying behavioral/physiological/brain response patterns in senders *and* receivers, while (ii) identifying the molecules that *link* two humans (i.e., the social (chemo)signal in human-human interaction), (iii) under ecologically valid conditions (i.e., including "noise" factors like dietary and hygiene habits) (Step 3), to eventually develop artificial intelligence-based sensors that could be applied in the real world for senders (e.g., diagnosis) and receivers (e.g., facilitating well-being by blocking the signals from entering the nose). This is an example of fear chemosignaling (vs. neutral), using faces obtained from the Radboud Faces Database (Langner et al., 2010).

sensory endpoints (e.g., intensity, pleasantness, and qualitative descriptors like garlicky or fruity) as opposed to social-behavioral endpoints (e.g., perceivers' affect, physiology, behavior); (iii), past models have not considered various sample characteristics (excepting gene variants coding for odorants receptors) or ecologically relevant contexts that are expected to impact smell perception as well.

To identify human chemosignals within the vast amount of data that can encompass body odors (a big data challenge), we recommend moving away from using a single, traditional statistical model (e.g., logistic regression), and instead propose a sequence of different analyses, including machine learning (ML). It would seem premature to rigorously define each step in the analysis sequence, but we will sketch a possible analysis "pipeline" (**Figure 1**):

Step 1 would entail collecting sweat from senders induced to be in a particular state (e.g., fear, happiness, disgust, sickness) or having a characteristic of interest (e.g., gender, personality, genotype). A subset of these sweat samples would then be used as *stimuli* in another experiment involving human receivers, whose behavioral responses will form a benchmark for verifying effective chemical communication (requiring a sender and receiver).

In **Step 2**, the remaining sweat samples will be used for chemical analysis. After extracting the molecules using headspace, solvent, or direct extraction techniques, chemical analysis could entail two-dimensional gas chromatography-mass spectrometry (GCxGC-ToF-MS) allowing for comprehensive profiling of the volatile molecules in the sweat samples and their discriminative power between two (or more) states/traits of interest. Because there is little to no background knowledge on chemical classes associated with presumed signals in sweat odor, initial research by Smeets et al. (2020) used *untargeted* screening approaches to distinguish between fear, happiness, and a neutral state, and found a matrix of over a 1,000 chemical volatile peaks. This number could be reduced as a next step to 94 by selecting only those peak intensities that differed significantly with at least one other emotion category. Preprocessing the GC × GC-ToF-MS profiles into total-intensity-count values (TICs) is another way to yield a smaller, more manageable subset of peaks of interest (cf. Lebanov et al., 2020). What could further ease the future identification of unique chemical profiles predicting human states/traits are *templates* (reference peak profiles) that follow from overlaying all chromatograms in a set (cf. Stilo et al., 2019), or using previous datasets as templates (cf. Reichenbach et al., 2019). This requires acquiring large chemical datasets, which necessitates high-throughput approaches like automated extraction and (ultra)fast GCxGC-ToF-MS, followed by automated quantification of specific target compounds belonging to specific states and traits.

In **Step 3**, ML techniques could help identify the core chemical features of human states/traits in multiple ways. Unsupervised learning (e.g., k-means clustering) could yield potentially interesting clusters of chemicals that are involved in chemical communication not considered before. Supervised learning could be applied next by training an algorithm on a large subset of samples, and testing the trained model on the remaining set. While there are vast varieties in learning algorithms, they can broadly be divided into linear or non-linear based on the shape of the decision surface used to classify data. Linear methods, like support vector machines (SVM) with linear kernels, may be preferred because they perform at least on par with non-linear methods (e.g., Misaki et al., 2010, in the context of separating emotions with fMRI data) while remaining straightforward to interpret. The interpretability of the models from the pipeline we propose might be tested by comparing the predictive power of those models with the outcomes on receiver experiments. To illustrate, Reichenbach et al. (2019) combined GC × GC chemical profiling with SVM to predict different characteristics of wines (e.g., grape variety, origin). Although the wines had considerable overlap in their chemical composition (up to 25% overlap in grape variety), the analysis yielded a number of highly distinctive molecules that the models used to differentiate the wines with around 90% accuracy (Reichenbach et al., 2019). At the same time, the resulting models were still relatively intuitively interpretable (cf. Mori and Uchihira, 2019).

We believe that applying a chemosignal-identification pipeline as described above would also yield relatively straightforward models, with an interpretable set of chemical predictors that are highly predictive of the emotions under investigation. Feature selection in our machine learning pipeline could be based on, depending on the ML technique used, mean absolute error (MAE) of the predictor (in case of regression-based techniques) or area under the curve (AUC) measures (in case of classification-based techniques) (Molnar, 2018). Selection of the best performing predictors could be tested by application level evaluation (cf. Molnar, 2018), using follow-up lab or crowd-sourced experiments where the most likely molecule candidates are tested in appropriate molecular concentrations.

DATABASE-BUILDING: BACK TO THE FUTURE

The proposed analysis pipeline requires rather large, well-populated databases compared to current standards. At present, there is a lack of such (publicly available) databases. Ideally, data in these databases contain a vast amount of parameters from hundreds of participants (senders *and* receivers). These parameters include personal factors (e.g., gender, age, country of residence, genotype), lifestyle factors (e.g., deodorant use, hygiene habits), measures of context (e.g., sterile lab vs. field), health, personality, and emotion (e.g., subjectively reported emotions and psychophysiological measures), and thousands of additional parameters per sample resulting from chemical analysis. Hence, the complexity and vastness of the resulting database underscores the need to step away from experimenter-driven analyses techniques such as traditional regression models, and turn to automatic feature selecting analyzation algorithms instead. Using these ML techniques has another advantage – the possibility to directly apply the best performing models in artificial intelligence applications.

However, one big hurdle to take with this multivariate, machine learning approach is the need for large, ecologically valid datasets. Talking about big data, a now famous 1989 National Geographic Smell Survey managed to test and analyze data from 1.42 million respondents to examine the relation between olfaction and aging (Wysocki and Gilbert, 1989). In a related effort to build socially relevant smell databases, Snitz et al. (2019) cleverly combined online crowdsourcing with the physical distribution of "scratch-and-sniff" odorants (via regular mail), and collected data (now publicly available) from about 1,000 individuals in 100 days. In these studies, chemical communication had not been the focus. If body odors were the topic, the database should hold matrices of (i) the chemical constituents of body odors (alongside multivariate information about, e.g., the emotional state during which the body odor was produced), as well as (ii) relevant person- and situation-specific variables of senders donating the samples (e.g., diet, hygiene product use, culture, genetic variation), and (iii) person- and situation-specific variables of recipients.

Acquiring such a complex and elaborate database can impossibly be a single-lab endeavor, and in our view, coordination within a larger ecosystem of labs will be crucial. Fortunately, technical advances in communication allow large consortia of researchers to globally collaborate from the comfort of their homes, like the Global Consortium for Chemosensory Research (GCCR), which focuses on the relation between COVID-19 and chemosensory dysfunction. Within weeks, hundreds of researchers around the globe collaborated to design an online study resulting in a large (open access) database on COVID-19 and smell/taste dysfunction and a published manuscript (Parma et al., 2020). Moreover, in work that focused on smell communication (literally, being able to talk about smells), Majid et al. (2018) have examined for 20 languages whether there exists a universal hierarchy to vision being more accessible to consciousness and linguistic description than smell. These, and other examples (e.g., Iravani et al., 2020), have illustrated that global consortia can be instrumental in acquiring the necessary datasets to solve complex and urgent questions in a timely manner.

TECHNOLOGY SUPPORTING SOCIETAL AND CLINICAL IMPACT

We envision the application of machine learning to understand human non-verbal communication to yield a series of impactful consequences ranging from psychology to medicine. If machine learning techniques can pick up on statistical regularities between, for instance, emotional states and health conditions on the one hand and patterns of molecules on the other, chemical sensors can be developed to read this "smell language" in real time. Promisingly, Imam and Cleland (2020) placed an array of 72 chemosensors (based on the architecture of the mammalian olfactory bulb) in a wind tunnel, which rapidly learned and identified odor representations, despite various sources of noise. Given that body odor (itself susceptible to noise) contains information about emotions (Smeets et al., 2020) and one's health, ranging from markers for Parkinson's disease (Trivedi et al., 2019), general inflammatory reactions (Olsson et al., 2014), to possibly the presence of COVID-19 in sweat (Grandjean et al., 2020), it would be intriguing to explore whether such algorithms could be used to learn and identify the even more complex language inherent to human odors.

The near future could see a rapid growth in the diagnostic implementation of sweat odor analysis that could happen outside of a lab or clinic in a person's home, with the emergence of novel smartphone-based biosensors (Brasier and Eckstein, 2019). Through these smartphone-based on-skin biosensors, sweat compounds could become broadly available in databases as digital biomarkers. Such an in-home approach is expected to have a major influence on clinical and outpatient care, and could even prevent infectious diseases from spreading by suggesting self-quarantining. The impact of these biosensors may extend to therapeutic settings, where the smell-based detection of patients' emotions (or lack thereof) could provide an insightful role in (online) therapeutic sessions. In sum, physicians and clinicians could foresee their instrumentation being expanded in the future by sensors and machine learning to more quickly, accurately, and safely get a grip on a disease or clinical problems and their prognosis (Chen et al., 2019).

CONCLUSION

Although scientific evidence has shown that the sense of smell serves a number of crucial functions in the daily life of humans, including social communication (e.g., Stevenson, 2010; de Groot et al., 2017; McGann, 2017), the idea that humans are micro smellers has remained hardwired among scientists and laypeople. However, through smell, humans can (unwillingly) convey information about a person. These initial advances were generally obtained under the most sterile conditions, by single research groups from the perspective of a single discipline. Although initially fruitful, we caution that continuing this experimental tradition will stall scientific progress toward a broader, deeper, and quicker understanding of non-verbal communication via smell. In the quest for discovering the real-world impact of social smells in diverse samples across diverse settings, we focused on the importance of ecological testing conditions, multidisciplinary research, and open collaborations to populate high dimensional databases, with machine learning approaches "making sense" of the complicated statistical regularities between smell molecules and physical or psychological conditions (the science of sociochemistry). By informing us about food, danger, health, and hygiene, olfaction serves a crucial role in human life, and so much so, that losing our sense of smell dramatically reduces the quality of our life. Our invitation for a better fundamental and practical understanding of the language of human smells opens up a multitude of (technological) possibilities, including tailor-made or world-wide clinical and societal applications proportionate to the scale at which human odors non-verbally communicate information from a sender to a recipient, whether human or machine.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors were involved in the conceptualization of this research. JG drafted the outline of this manuscript, wrote the manuscript, and edited the manuscript. IC and MS co-wrote the manuscript and critically revised the outline of the manuscript and the manuscript itself.

FUNDING

This research was supported by a Veni Innovational Research grant (NWO-016.Veni.195.116) awarded to the JG.

ACKNOWLEDGMENTS

We thank Daniel Oberski for his contributions on machine learning approaches and Helene Loos for her contributions on chemical analytical techniques (section "Machine Learning: Solving the Big Data Challenge Ahead of Us").

REFERENCES

Brasier, N., and Eckstein, J. (2019). Sweat as a source of next-generation digital biomarkers. *Digit. Biomark.* 3, 155–165. doi: 10.1159/000504387

Chen, D., and Haviland-Jones, J. (2000). Human olfactory communication of emotion. *Percept. Mot. Skills* 91, 771–781. doi: 10.2466/pms.2000.91.3.771

Chen, Z., Chen, Z., Song, Z., Ye, W., and Fan, Z. (2019). Smart gas sensor arrays powered by artificial intelligence. *J. Semicond* 40:111601. doi: 10.1088/1674-4926/40/11/111601

Dalton, P. (1999). Cognitive influences on health symptoms from acute chemical exposure. *Heal. Psychol.* 18, 579–590.

De Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C., and Cayeux, I. (2005). Cognitive modulation of olfactory processing. *Neuron* 46, 671–679. doi: 10.1016/j.neuron.2005.04.021

de Groot, J. H. B., Beetsma, J. V., van Aerts, T. J., Le Berre, E., Gallagher, D., Aarts, H., et al. (2020a). From sterile labs to rich VR: immersive multisensory context critical for odors to induce motivated cleaning behavior. *Behav. Res. Methods* 52, 1657-1670. doi: 10.3758/s13428-019-01341-y

de Groot, J. H. B., Kirk, P. A., and Gottfried, J. A. (2020b). Titrating the smell of fear: initial evidence for dose-invariant behavioral, physiological, and neural responses. *Psychol. Sci.* (in press).

de Groot, J. H. B., Semin, G. R., and Smeets, M. A. M. (2017). On the communicative function of body odors: a theoretical integration and review. *Perspect. Psychol. Sci.* 12, 306–324. doi: 10.1177/1745691616676599

de Groot, J. H. B., and Smeets, M. A. M. (2017). Human fear chemosignaling: evidence from a meta-analysis. *Chem. Senses* 42, 663–673. doi: 10.1093/chemse/bjx049

de Groot, J. H. B., Smeets, M. A. M., Kaldewaij, A., Duijndam, M. J. A., and Semin, G. R. (2012). Chemosignals communicate human emotions. *Psychol. Sci.* 23, 1417–1424. doi: 10.1177/0956797612445317

de Groot, J. H. B., Smeets, M. A. M., Rowson, M. J., Bulsing, P. J., Blonk, C. G., Wilkinson, J. E., et al. (2015). A sniff of happiness. *Psychol. Sci.* 26, 684–700. doi: 10.1177/0956797614566318

de Groot, J. H. B., van Houtum, L. A. E. M., Gortemaker, I., Ye, Y., Chen, W., Zhou, W., et al. (2018). Beyond the west: chemosignaling of emotions transcends ethno-cultural boundaries. *Psychoneuroendocrinology* 98, 177–185. doi: 10.1016/j.psyneuen.2018.08.005

de Lacy Costello, B., Amann, A., Al-Kateb, H., Flynn, C., Filipiak, W., Khalid, T., et al. (2014). A review of the volatiles from the healthy human body. *J. Breath Res.* 8:014001. doi: 10.1088/1752-7155/8/1/014001

Dhar, V. (2013). Data science and prediction. *Commun. ACM* 56, 64–73. doi: 10.1145/2500499

Endevelt-Shapira, Y., Perl, O., Ravia, A., Amir, D., Eisen, A., Bezalel, V., et al. (2018). Altered responses to social chemosignals in autism spectrum disorder. *Nat. Neurosci.* 21, 111–119. doi: 10.1038/s41593-017-0024-x

Ferdenzi, C., Richard Ortegón, S., Delplanque, S., Baldovini, N., and Bensafi, M. (2020). Interdisciplinary challenges for elucidating human olfactory attractiveness. *Philos. T. R. Soc. B* 375:20190268. doi: 10.1098/rstb.2019.0268

Forscher, P. S., Wagenmakers, E., Coles, N. A., Silan, M. A., and IJzerman, H. (2020). A Manifesto for Team Science. *psyArxiv*[Preprint]. doi: 10.31234/osf.io/2mdxh

Gaby, J. M., and Zayas, V. (2017). Smelling is telling: human olfactory cues influence social judgments in semi-realistic interactions. *Chem. Senses* 42, 405–418. doi: 10.1093/chemse/bjx012

Gildersleeve, K., Haselton, M. G., and Fales, M. R. (2014). Meta-analyses and p-curves support robust cycle shifts in women's mate preferences: reply to Wood and Carden (2014) and Harris. Pashler, and Mickes (2014). *Psychol. Bull.* 140, 1272–1280. doi: 10.1037/a0037714

Gomes, N., Silva, F., and Semin, G. R. (2020). The lasting smell of emotions: the effects of reutilizing fear sweat samples. *Behav. Res. Meth* doi: 10.3758/s13428-020-01412-5

Grandjean, D., Sarkis, R., Tourtier, J.-P., Julien-Lecocq, C., Benard, A., Roger, V., et al. (2020). Detection dogs as a help in the detection of COVID-19 – Can the dog alert on COVID-19 positive persons by sniffing axillary sweat samples? Proof-of-concept study. *bioRxiv*[Preprint]. doi: 10.1101/2020.06.03.132134

Gutiérrez, E. D., Dhurandhar, A., Keller, A., Meyer, P., and Cecchi, G. A. (2018). Predicting natural language descriptions of mono-molecular odorants. *Nat. Commun.* 9:4979. doi: 10.1038/s41467-018-07439-9

Havlíček, J., Winternitz, J., and Roberts, S. C. (2020). Major histocompatibility complex-associated odour preferences and human mate choice: near and far horizons. *Philos. T. R. Soc. B* 375:20190260. doi: 10.1098/rstb.2019.0260

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Imam, N., and Cleland, T. A. (2020). Rapid online learning and robust recall in a neuromorphic olfactory circuit. *Nat. Mach. Intell.* 2, 181–191. doi: 10.1038/s42256-020-0159-4

Iravani, B., Arshamian, A., Ravia, A., Mishor, E., Snitz, K., Shushan, S., et al. (2020). Relationship between odor intensity estimates and COVID-19 prevalence prediction in a swedish population. *Chem. Senses* 22:bjaa034. doi: 10.1093/chemse/bjaa034

Karlson, P., and Lüscher, M. (1959). "Pheromones": a new term for a class of biologically active substances. *Nature* 183, 55–56.

Keller, A., Gerkin, R. C., Guan, Y., Dhurandhar, A., Turu, G., Szalai, B., et al. (2017). Predicting human olfactory perception from chemical features of odor molecules. *Science* 355, 820–826. doi: 10.1126/science.aal2014

Kennedy, D., and Norman, C. (2005). So much more to know. *Science* 309, 78–102.

Khan, R. M., Luk, C. H., Flinker, A., Aggarwal, A., Lapid, H., Haddad, R., et al. (2007). Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *J. Neurosci.* 27, 10015–10023. doi: 10.1523/JNEUROSCI.1158-07.2007

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

Lebanov, L., Tedone, L., Ghiasvand, A., and Paull, B. (2020). Random Forests machine learning applied to gas chromatography–Mass spectrometry derived average mass spectrum data sets for classification and characterisation of essential oils. *Talanta* 208:120471. doi: 10.1016/j.talanta.2019.120471

Loos, H. M., Doucet, S., Soussignan, R., Hartmann, C., Durand, K., Dittrich, R., et al. (2014). Responsiveness of human neonates to the odor of 5α-androst-16-en-3-one: a behavioral paradox? *Chem. Senses* 39, 693–703.

Loos, H. M., Reger, D., and Schaal, B. (2019). The odour of human milk: its chemical variability and detection by newborns. *Physiol. Behav.* 199, 88–99. doi: 10.1016/j.physbeh.2018.11.008

Lötsch, J., Kringel, D., and Hummel, T. (2018). Machine learning in human olfactory research. *Chem. Senses* 44, 11–22. doi: 10.1093/chemse/bjy067

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., et al. (2018). Differential coding of perception in the world's languages. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11369–11376. doi: 10.1073/pnas.1720419115

Majid, A., Speed, L., Croijmans, I., and Arshamian, A. (2017). What makes a better smeller? *Perception* 46, 406–430.

Makin, J. W., and Porter, R. H. (1989). Attractiveness of lactating females' breast odors to neonates. *Child Dev.* 60, 803–810.

Mayer-Schönberger, V., and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Houghton Mifflin Harcourt.

McGann, J. P. (2017). Poor human olfaction is a 19th-century myth. *Science* 356:eaam7263. doi: 10.1126/science.aam7263

Misaki, M., Kim, Y., Bandettini, P. A., and Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53, 103–118. doi: 10.1016/j.neuroimage.2010.05.051

Mitro, S., Gordon, A. R., Olsson, M. J., and Lundström, J. N. (2012). The smell of age: perception and discrimination of body odors of different ages. *PLoS One* 7:e0038110. doi: 10.1371/journal.pone.0038110

Molnar, C. (2018). *A GUide For Making Black Box Models Explainable*. Available online at: https://christophm.github.io/interpretable-ml-book (accessed September 16, 2020).

Mori, T., and Uchihira, N. (2019). Balancing the trade-off between accuracy and interpretability in software defect prediction. *Empir. Softw. Eng.* 24, 779–825. doi: 10.1007/s10664-018-9638-1

Mujica-Parodi, L. R., Strey, H. H., Frederick, B., Savoy, R., Cox, D., Botanov, Y., et al. (2009). Chemosensory cues to conspecific emotional stress activate amygdala in humans. *PLoS One* 4:e6415. doi: 10.1371/journal.pone.0006415

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

Natsch, A., and Emter, R. (2020). The specific biochemistry of human axilla odour formation viewed in an evolutionary context. *Philos. Trans. R. Soc. Lond. Ser. B* 375:20190269. doi: 10.2210/PDB6SLF/PDB

Olsson, M. J., Lundström, J. N., Kimball, B. A., Gordon, A. R., Karshikoff, B., Hosseini, N., et al. (2014). The scent of disease: human body odor contains an early chemosensory cue of sickness. *Psychol. Sci.* 25, 817–823. doi: 10.1177/0956797613515681

Parma, V., Gordon, A. R., Cecchetto, C., Cavazzana, A., and Lundström, J. N. (2017). "Processing of human body odors," in *Springer Handbook of Odor*, ed. A. Buettner, (New York, NY: Springer), 963–986. doi: 10.1007/978-3-319-26932-0

Parma, V., Ohla, K., Veldhuizen, M. G., Niv, M. Y., Kelly, C. E., Bakke, A. J., et al. (2020). More than smell – COVID-19 is associated with severe impairment of smell, taste, and chemesthesis. *Chem. Senses* 20:bjaa041. doi: 10.1093/chemse/bjaa041

Pause, B. M. (2017). "Human chemosensory communication," in *Springer Handbook of Odor*, ed. A. Buettner, (New York, NY: Springer), 987–1010.

Pause, B. M., Adolph, D., Prehn-Kristensen, A., and Ferstl, R. (2009). Startle response potentiation to chemosensory anxiety signals in socially anxious individuals. *Int. J. Psychophysiol.* 74, 88–92. doi: 10.1016/j.ijpsycho.2009.07.008

Pause, B. M., Storch, D., and Lübke, K. T. (2020). Chemosensory communication of aggression: women's fine-tuned neural processing of male aggression signals. *Philos. Trans. R. Soc. B Biol. Sci* 375:20190270. doi: 10.1098/rstb.2019.0270

Penn, D. J., Oberzaucher, E., Grammer, K., Fischer, G., Soini, H. A., Wiesler, D., et al. (2007). Individual and gender fingerprints in human body odour. *J. R. Soc. Interface* 4, 331–340. doi: 10.1098/rsif.2006.0182

Porter, J., Craven, B., Khan, R. M., Chang, S.-J., Kang, I., Judkewitz, B., et al. (2007). Mechanisms of scent-tracking in humans. *Nat. Neurosci.* 10, 27–29. doi: 10.1038/nn1819

Porter, R. H., Makin, J. W., Davis, L. B., and Christensen, K. M. (1991). An assessment of the salient olfactory environment of formula-fed infants. *Physiol. Behav.* 50, 907–911. doi: 10.1016/0031-9384(91)90413-I

Prehn-Kristensen, A., Wiesner, C., Bergmann, T. O., Wolff, S., Jansen, O., Mehdorn, H. M., et al. (2009). Induction of empathy by the smell of anxiety. *PLoS One* 4:e5987. doi: 10.1371/journal.pone.0005987

Preti, G., Dalton, P., and Maute, C. (2019). Analytical identification of stress odors from human breath. *HDIAC J.* 6, 10–15. .

Quintana, P., Nolet, K., Baus, O., and Bouchard, S. (2019). The effect of exposure to fear-related body odorants on anxiety and interpersonal trust toward a virtual character. *Chem. Senses* 44, 683–692. doi: 10.1093/gerona/gly169/5057054

Regenbogen, C., Axelsson, J., Lasselin, J., Porada, D. K., Sundelin, T., Peter, M. G., et al. (2017). Behavioral and neural correlates to multisensory detection of sick humans. *Proc. Natl. Sci. U.S.A.* 114, 6400–6405. doi: 10.1073/pnas.1617357114

Reichenbach, S. E., Zini, C. A., Nicolli, K. P., Welke, J. E., Cordero, C., and Tao, Q. (2019). Benchmarking machine learning methods for comprehensive chemical fingerprinting and pattern recognition. *J. Chromatogr. A* 1595, 158–167. doi: 10.1016/j.chroma.2019.02.027

Roberts, S. C., Havlíček, J., and Schaal, B. (2020). Human olfactory communication: current challenges and future prospects. *Philos. Trans. R. Soc. B* 375:20190258. doi: 10.1098/rstb.2019.0258

Sanchez-Lengeling, B., Wei, J. N., Lee, B. K., Gerkin, R. C., Aspuru-Guzik, A., and Wiltschko, A. B. (2019). *Machine Learning for Scent: Learning Generalizable Perceptual Representations of Small Molecules*. Available online at: http://arxiv.org/abs/1910.10685 (accessed September 16, 2020).

Schaal, B., and Porter, R. H. (1991). Microsmatic humans" revisited: the generation and perception of chemical signals. *Adv. Study Behav* 20, 135–199. doi: 10.1016/S0065-3454(08)60321-6

Schaal, B., Saxton, T. K., Loos, H., Soussignan, R., and Durand, K. (2020). Olfaction scaffolds the developing human from neonate to adolescent and beyond. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190261. doi: 10.1098/rstb.2019.0261

Simons, D. J., Shoda, Y., and Lindsay, D. S. (2017). Constraints on Generality (COG): a proposed Addition to All Empirical Papers. *Perspect. Psychol. Sci.* 12, 1123–1128. doi: 10.1177/1745691617708630

Singh, P. B., Young, A., Lind, S., Leegaard, M. C., Capuozzo, A., and Parma, V. (2018). Smelling anxiety chemosignals impairs clinical performance of dental students. *Chem. Senses* 43, 411–417. doi: 10.1093/chemse/bjy028

Smeets, M. A. M., Rosing, E. A. E., Jacobs, D. M., van Velzen, E., Koek, J. H., Blonk, C., et al. (2020). Chemical fingerprints of emotional body odor. *Metabolites* 10:84. doi: 10.3390/metabo10030084

Snitz, K., Perl, O., Honigstein, D., Secundo, L., Ravia, A., Yablonka, A., et al. (2019). Smellspace: an odor-based social network as a platform for collecting olfactory perceptual data. *Chem. Senses* 44, 267–278. doi: 10.1093/chemse/bjz014

Snitz, K., Yablonka, A., Weiss, T., Frumin, I., Khan, R. M., and Sobel, N. (2013). Predicting odor perceptual similarity from odor structure. *PLoS Comput. Biol* 9:3184. doi: 10.1371/journal.pcbi.1003184

Sorokowski, P., Karwowski, M., Misiak, M., Marczak, M. K., Dziekan, M., Hummel, T., et al. (2019). Sex differences in human olfaction: a meta-analysis. *Front. Psychol.* 10:242. doi: 10.3389/fpsyg.2019.00242

Stevenson, R. J. (2010). An initial evaluation of the functions of human olfaction. *Chem. Senses* 35, 3–20. doi: 10.1093/chemse/bjp083

Stilo, F., Liberto, E., Reichenbach, S. E., Tao, Q., Bicchi, C., and Cordero, C. (2019). Untargeted and targeted fingerprinting of extra virgin olive oil volatiles by comprehensive two-dimensional gas chromatography with mass spectrometry: challenges in long-term studies. *J. Agric. Food Chem.* 67, 5289–5302. doi: 10.1021/acs.jafc.9b01661

Trivedi, D. K., Sinclair, E., Xu, Y., Sarkar, D., Walton-Doyle, C., Liscio, C., et al. (2019). Discovery of volatile biomarkers of Parkinson's disease from sebum. *ACS Cent. Sci.* 5, 599–606. doi: 10.1021/acscentsci.8b00879

Williams, J., Stönner, C., Wicker, J., Krauter, N., Derstroff, B., Bourtsoukidis, E., et al. (2016). Cinema audiences reproducibly vary the chemical composition of air during films, by broadcasting scene specific emissions on breath. *Sci. Rep.* 6:25464. doi: 10.1038/srep25464

Wyatt, T. D. (2015). The search for human pheromones: the lost decades and the necessity of returning to first principles. *Proc. R. Soc. B Biol. Sci.* 282:20142994. doi: 10.1098/rspb.2014.2994

Wyatt, T. D. (2020). Reproducible research into human chemical communication by cues and pheromones: learning from psychology's renaissance. *Philos. Trans. R. Soc. B* 375:20190262. doi: 10.1098/rstb.2019.0262

Wysocki, C. J., and Gilbert, A. N. (1989). National geographic smell survey: effects of age are heterogeneous. *Ann. N.Y. Acad. Sci.* 561, 12–28. doi: 10.1111/j.1749-6632.1989.tb20966.x

Yeshurun, Y., and Sobel, N. (2010). An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Annu. Rev. Psychol.* 61, 219–241. doi: 10.1146/annurev.psych.60.110707.163639

Zarzo, M. (2011). Hedonic judgments of chemical compounds are correlated with molecular size. *Sensors* 11, 3667–3686. doi: 10.3390/s110403667

Zhou, W., and Chen, D. (2009). Fear-related chemosignals modulate recognition of fear in ambiguous facial expressions. *Psychol. Sci.* 20, 177–183. doi: 10.1111/j.1467-9280.2009.02263.x

Conflict of Interest: 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.

frontiers in Psychology

OPINION
published: 02 November 2020
doi: 10.3389/fpsyg.2020.573460

Check for updates

Misconceptions About Nonverbal Cues to Deception: A Covert Threat to the Justice System?

Vincent Denault [1,2,3]*

[1] *International Centre for Comparative Criminology, Montreal, QC, Canada,* [2] *Department of Communication, Université de Montréal, Montreal, QC, Canada,* [3] *Centre for Studies in Nonverbal Communication Sciences, Montreal, QC, Canada*

Keywords: courtrooms, nonverbal communication, witnesses, deception, credibility

OPEN ACCESS

Edited by:
José-Miguel Fernández-Dols, Autonomous University of Madrid, Spain

Reviewed by:
Jaume Masip, University of Salamanca, Spain

Correspondence:
Vincent Denault
vincent.denault@umontreal.ca

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *17 June 2020*
Accepted: *30 September 2020*
Published: *02 November 2020*

Citation:
Denault V (2020) Misconceptions About Nonverbal Cues to Deception: A Covert Threat to the Justice System? Front. Psychol. 11:573460. doi: 10.3389/fpsyg.2020.573460

INTRODUCTION

Nonverbal communication is studied by a worldwide community of researchers. Thousands of peer-reviewed papers have been published on the subject (Plusquellec and Denault, 2018). The same holds for deception detection. Unfortunately, misconceptions about nonverbal cues to deception are widespread. The general public holds popular beliefs (The Global Deception Research Team, 2006), unfounded or discredited claims are disseminated on social media and television, and pseudoscientific claims, that is, unfounded or discredited claims presented explicitly or implicitly as having scientific value, are promoted in manuals and seminars (Denault et al., 2015, 2020, Denault et al., submitted).

Misconceptions about nonverbal cues to deception may, at first glance, seem harmless and even entertaining. However, they can have far-reaching consequences. During police investigations, for example, they can result in coercive interrogations and, potentially, false confessions (Leo and Drizin, 2010). During trials, while less discussed within the literature, the consequences can be just as serious, perhaps even more so. Because witness credibility can be largely influenced by demeanor (Denault, 2015), when judges in bench trials (and jurors in jury trials) turn to popular beliefs about deception cues or unfounded, discredited and pseudoscientific claims, the assessment of witness credibility can be distorted. This is significant considering that "credibility is an issue that pervades most trials, and at its broadest may amount to a decision on guilt or innocence" (R. v. Handy, 2002, p. 951) and that decisions of judges are enforceable. When capital punishment is at stake, it can be an issue of life or death (Wilson and Rule, 2015, 2016).

This article aims to highlight the dangers of misconceptions about nonverbal cues to deception during trials. Their popularity among justice and legal practitioners is addressed and, subsequently, their detrimental effect on the assessment of witness credibility. This article ends with recommendations for practitioners, policy makers, and scholars to mitigate the adverse influence of unfounded, discredited and pseudoscientific claims.

THE POPULARITY OF MISCONCEPTIONS ABOUT NONVERBAL CUES TO DECEPTION

Just as for the general public, several justice and legal practitioners hold popular beliefs about deception cues (e.g., Strömwall and Granhag, 2003; Strömwall et al., 2004; Bogaard et al., 2016). Moreover, despite their considerable authority, several justice and legal practitioners are sympathetic to unfounded, discredited, and pseudoscientific claims. Within law enforcement, this is a well-known problem.

Interviewing techniques promoting nonverbal cues to deception, for example, typically start with a warning to look for clusters of gestures, changes in behaviors, and contradictions between verbal and nonverbal cues (e.g., Walters, 2003; Inbau et al., 2013). The initial warning may, implicitly, convey an impression of scientific rigor or, explicitly, refer to the work of academics. Subsequently, the importance of "body language" is typically touted well-beyond what has been conclusively demonstrated. For example, "any change in a person's constant or normal level of eye contact, which is a timely response and part of a cluster, can be sign of stress and possible deception" (Walters, 2003, p. 134). A final warning sometimes given is that deception cues depend on various factors such as "the perceived seriousness of the offense; the mental and physical condition of the subject; any underlying psychiatric or personality disorders; level of intelligence; degree of maturity; and the extent or absence of social responsibilities" (Inbau et al., 2013, p. 152).

However, research has shown that deception cues are generally faint and unreliable (DePaulo et al., 2003; Sporer and Schwandt, 2007; Luke, 2019; Vrij et al., 2019) and their use has been shown not to significantly improve lie detection accuracy (Hauch et al., 2016; see also Meissner and Kassin, 2004; Bond and DePaulo, 2006; Jordan et al., 2019). Therefore, the initial warning to look for clusters, changes and contradictions becomes trivial, all the more considering accuracy requirements (e.g., that the level of eye contact "can be" a "possible" sign of deception if it is part of a "cluster" of behaviors, a "change" from a "constant" or "normal" behavior, and a "timely" response) negate their practical value. The same holds for the initial and final warnings which, incidentally, also offer an easy response to criticism: you did not obtain the expected results because you did not adequately consider the accuracy requirements and the initial and final warnings.

While unfounded, discredited, and pseudoscientific claims about deception cues promoted to law enforcement in manuals and seminars received attention by both the media (e.g., Hager, 2017; Armstrong and Sheckler, 2019; Smith, 2020) and the academia (e.g., Lilienfeld and Landfield, 2008; Chaplin and Shaw, 2016; Denault et al., 2020), little is known about their promotion to members of the judiciary. An exception comes from Quebec where, for a few years, a number of judges from different courts received talks from proponents of synergology, an approach which, supposedly, makes it possible to "decipher body language." Proponents of synergology have claimed, among other things, that different gestures have specific meanings, which are not supported by peer-reviewed articles, and promoted concepts similar to the initial and final warnings of interviewing techniques promoting nonverbal cues to deception. Training centers in synergology are located in various countries, including Canada, France, Switzerland, and Spain (for a critical evaluation of synergology, see Denault and Jupe, 2017; Jupe and Denault, 2019; Denault et al., 2020). In other words, unsubstantiated, discredited, and pseudoscientific claims about deception cues can, in the absence of adequate policies, find their way into courtrooms.

THE DETRIMENTAL EFFECTS OF MISCONCEPTIONS ABOUT NONVERBAL CUES TO DECEPTION

In countries with adversarial justice systems (e.g., Canada, United States), rules of evidence and procedure foster, to some extent, the use of false beliefs about deception cues and unsubstantiated, discredited, and pseudoscientific claims. During bench trials, for example, judges have to establish the facts to which laws are applied. Essentially, they observe and listen to witnesses and, subsequently, assess their credibility. Based on the witnesses' credibility, judges will give more or less weight to their testimony. This is how judges will often decide what happened when witnesses have different accounts of a same event (Paciocco, 2010; Bell, 2013).

Unfortunately, in several jurisdictions, even if judges are legally authorized to use the witnesses' demeanor to assess their credibility (Mattox v. United States, 1895; Coy v. Iowa, 1988; P. (D.) v. S. (C.), 1993), evidence-based workshops or seminars to mitigate the impact of misconceptions about nonverbal cues to deception are not mandatory. In addition, expert evidence on credibility assessment is generally prohibited. As the Supreme Court of Canada points out, "the issue of credibility is an issue well within the experience of judges and juries and one in which no expert evidence is required" (R. v. Béland, 1987, p. 399). This is in keeping with the Supreme Court of the United States' ruling that jurors "are presumed to be fitted for it by their natural intelligence and their practical knowledge of men and the ways of men" (Aetna Life Ins. Co. v. Ward, 1891, p. 88; United States v. Scheffer, 1998, p. 313). Therefore, it is not uncommon for judges in bench trials (and jurors in jury trials) to turn to popular beliefs about deception cues or unfounded, discredited and pseudoscientific claims.

For example, in a 2019 decision, a judge of the Supreme Court of British Columbia gave little weight to a testimony because, among other things, "As he [the witness] gave his evidence, I observed him to cough, fidget, scratch his neck and at times appear quite nervous" (Garib v. Randhawa, 2019, p. 23). However, research has shown, unequivocally, that those behavioral cues are invalid deception cues (DePaulo et al., 2003; Sporer and Schwandt, 2007; Luke, 2019; Vrij et al., 2019). In a 2020 judgement, on the issue of voluntariness and understanding of a guilty plea, a judge of the Ontario Court of Justice wrote that "assessing body language and making eye contact can be of great assistance in deciding whether or not to accept a guilty plea, as well as weighing and making determinations about sentencing submissions" (R. v. Kerr, 2020, p. 5). However, research has not shown the existence of a body movement or a facial expression to confirm or disconfirm someone is remorseful (Bandes, 2014, 2016). These are just two among many examples (for more examples, see Denault, 2015; Denault and Dunbar, 2019).

However, while written judgments show, in practice, how judges sometimes use misconceptions about nonverbal cues to deception during trials, they are likely the "tip of the iceberg" because the influence of nonverbal communication in face to face interactions occurs much outside of conscious awareness

(Goldin-Meadow and Alibali, 2013; Todorov et al., 2015; Hall et al., 2019). And depending on the court's jurisdiction, even if judges "consciously" observe nonverbal cues to deception, they are not required to mention them in their decisions (R. v. Burns, 1994; Cojocaru v. British Columbia Women's Hospital Health Centre, 2013). Therefore, the detrimental effects of both popular beliefs about deception cues and unfounded, discredited and pseudoscientific claims is difficult to measure. In other words, misconceptions about nonverbal cues to deception are a covert threat to the justice system. This is all the more worrisome considering that, even if they are mentioned in written judgments, the assessment of witness credibility is rarely reviewed by appellate courts because, amongst other thing, they cannot "see and hear" the witnesses as judges previously did (Timony, 2000; Denault, 2015). Therefore, judges receive very little feedback and, as a consequence, could read manuals, and attend seminars promoting unfounded, discredited and pseudoscientific claims, all in good faith, throughout their career, without ever being told that, in fact, what they learned is unproven and amounts to nothing more than "junk science" (DeMatteo et al., 2019; Neal et al., 2019).

RECOMMENDATIONS FOR PRACTITIONERS, POLICY MAKERS, AND SCHOLARS

The justice system is a pillar of democracy for societies based on the rule of law. However, for the public to turn to courts when injustice occurs, public trust is fundamental. Unfortunately, misconceptions about nonverbal cues to deception are used for the assessment of witness credibility. Honest witnesses are sometimes believed to be dishonest and dishonest witnesses are sometimes believed to be honest and, as a consequence, parents in family trials can wrongfully lose their children's custody and defendants in criminal trials can wrongfully lose their liberty or their life. This can seriously jeopardize public trust in the justice system. However, a number of measures can be taken in an attempt to mitigate the adverse influence of unfounded, discredited and pseudoscientific claims.

For example, law degrees should incorporate courses on legal psychology and interpersonal communication. Waiting for lawyers to become members of the judiciary to introduce them to these subjects, expecting them to change their years old habits overnight, is irresponsible, if not delusional. Legislative changes should be made to forbid the delivery of courses promoting unfounded, discredited, and pseudoscientific claims to justice and legal practitioners. And justice and legal practitioners should be advised on how to initially assess the scientific quality of manuals and seminars of interest to them. For example, are the instructors "body language experts" or active researchers affiliated with scholarly institutions? Are the claims made during the seminars published in "international bestseller books" or in peer-reviewed publications? Are the seminars promoted using extravagant claims (e.g., "Learn to read people like a book"), appeals to authority (e.g., "We trained FBI and CIA officers"), and anecdotal evidences (e.g., "A terrorist was spotted using our approach")?

Scholars, on the other hand, should conduct deception research also with members of the judiciary, not only law enforcement, and publish articles in law journals. Changing court culture takes time, but judges regularly turn to law journals for their decisions rather than peer-review articles because unlike the latter, expert testimony is not necessarily required for the former (Hesler, 2002). Furthermore, scholars should actively promote scientific knowledge to justice and legal practitioners, respond to their questions and concerns, and stand up to unfounded, discredited and pseudoscientific claims. While science deniers sometimes turn to ad hominem attacks, and even legal threats (Dance, 2019; Jarry, 2019; Denault et al., 2020), speaking publicly about the importance of science within the justice system is of paramount importance to prevent miscarriages of justice.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author would like to thank Adam Villeneuve and the reviewer for their insightful comments and suggestions on an earlier draft of this article.

REFERENCES

Aetna Life Ins. Co. v. Ward (1891). 140 U.S. 76.

Armstrong, K., and Sheckler, C. (2019). *Why Are Cops Around the World Using this Outlandish Mind-Reading Tool? Pro Republica*. Retrieved from: https://www.propublica.org/article/why-are-cops-around-the-world-using-this-outlandish-mindreading-tool (accessed September 26, 2020).

Bandes, S. A. (2014). Remorse, demeanor, and the consequences of misinterpretation: The limits of law as a window to the soul. *J. Law Religion State* 3, 170–199. doi: 10.1163/22124810-00302004

Bandes, S. A. (2016). Remorse and criminal justice. *Emotion Rev.* 8, 14–19. doi: 10.1177/1754073915601222

Bell, E. (2013). An introduction to judicial fact-finding. *Commonwealth Law Bull.* 39, 519–552. doi: 10.1080/03050718.2013.815119

Bogaard, G., Meijer, E. H., Vrij, A., and Merckelbach, H. (2016). Strong, but wrong: lay people's and police officers' beliefs about verbal and nonverbal cues to deception. *PLoS ONE* 11:e0156615. doi: 10.1371/journal.pone.0156615

Bond, C. F. Jr., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Personal. Soc. Psychol. Rev.* 10, 214–234. doi: 10.1207/s15327957pspr1003_2

Chaplin, C., and Shaw, J. (2016). Confidently wrong: police endorsement of psycho-legal misconceptions. *J. Police Crim. Psychol.* 31, 208–216. doi: 10.1007/s11896-015-9182-5

Cojocaru v. British Columbia Women's Hospital and Health Centre (2013). SCC 30 (CanLII), [2013] 2 SCR 357.

Coy v. Iowa (1988). 487 U.S. 1012.

Dance, A. (2019). Death threats and lawsuits: John Maddox prize honours researchers who risk everything to stand up for science. *Nature.* Retrieved from:

https://www.nature.com/articles/d41586-019-03547-8 (accessed September 26, 2020)
DeMatteo, D., Fishel, S., and Tansey, A. (2019). Expert evidence: the (unfulfilled) promise of daubert. *Psychol. Sci. Public Inter.* 20, 129–134. doi: 10.1177/1529100619894336
Denault, V. (2015). *Communication non verbale et crédibilité des témoins [Nonverbal Communication and Witnesses Credibility]*. Cowansville, QC: Éditions Yvon Blais.
Denault, V., and Dunbar, N. (2019). "Credibility assessment and deception detection in courtrooms: Hazards and challenges for scholars and legal practitioners," in *The Palgrave Handbook of Deceptive Communication,* ed T. Docan-Morgan (Basingstoke: Palgrave Macmillan), 915–936. doi: 10.1007/978-3-319-96334-1_47
Denault, V., and Jupe, L. (2017). Justice at risk! *An evaluation of a pseudoscientific analysis of a witness' nonverbal behavior in the courtroom. J. Foren. Psychiatry Psychol.* 29, 221–242. doi: 10.1080/14789949.2017.1358758
Denault, V., Larivée, S., Plouffe, D., and Plusquellec, P. (2015). La synergologie, une lecture pseudoscientifique du langage corporel [Synergology, a pseudoscientific reading of body language]. *Revue de Psychoéducation* 43, 425–455. doi: 10.7202/1039262ar
Denault, V., Plusquellec, P., Jupe, L. M., St-Yves, M., Dunbar, N. E., Hartwig, M., et al. (2020). The analysis of nonverbal communication: The dangers of pseudoscience in security and justice contexts. *Anuario de Psicología Jurídica* 30, 1–12. doi: 10.5093/apj2019a9
DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74
Garib v. Randhawa (2019). BCSC 2176 (CanLII).
Goldin-Meadow, S., and Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annu. Rev. Psychol.* 64, 257–283. doi: 10.1146/annurev-psych-113011-143802
Hager, E. (2017). *The seismic change in police interrogations. The Marshall Project.* Retrieved from: https://www.themarshallproject.org/2017/03/07/the-seismic-change-in-police-interrogations (accessed September 26, 2020)
Hall, J. A., Horgan, T. G., and Murphy, N. A. (2019). Nonverbal communication. *Annu. Rev. Psychol.* 70, 271–294. doi: 10.1146/annurev-psych-010418-103145
Hauch, V., Sporer, S. L., Michael, S. W., and Meissner, C. A. (2016). Does training improve detection of deception? *A meta-analysis. Commun. Res.* 43, 283–343. doi: 10.1177/0093650214534974
Hesler, N. D. (2002). L'admissibilité des nouvelles théories scientifiques [The admissibility of new scientific theory]. *Revue du Barreau* 62, 359–385.
Inbau, F. E., Reid, J. E., Buckley, J. P., and Jayne, B. C. (2013). *Criminal Interrogation and Confessions.* Sudbury: Jones and Bartlett Publishers.
Jarry, J. (2019). *Censoring Science Communication by Screaming Defamation*. Office of Science and Society. Retrieved from: https://www.mcgill.ca/oss/article/general-science/censoring-science-communication-screaming-defamation (accessed September 26, 2020)
Jordan, S., Brimbal, L., Wallace, D. B., Kassin, S. M., Hartwig, M., and Street, C. N. H. (2019). A test of the micro-expressions training tool: does it improve lie detection? *J. Invest. Psychol. Offender Profil.* 16, 222–235. doi: 10.1002/jip.1532
Jupe, L., and Denault, V. (2019). Science or pseudoscience? *A distinction that matters for police officers, lawyers and judges. Psychiatry, Psychol. Law* 26, 753–765. doi: 10.1080/13218719.2019.1618755
Leo, R. A., and Drizin, S. A. (2010). "The three errors: Pathways to false confession and wrongful conviction," in *Police Interrogations and False Confessions: Current Research, Practice, and Policy Recommendations,* eds D. G. Lassiter and C. Meissner (Washington, DC: American Psychological Association), 9–30. doi: 10.1037/12085-001
Lilienfeld, S. O., and Landfield, K. (2008). Science and pseudoscience in law enforcement: a user-friendly prime. *Crim. Justice Behav.* 35, 1215–1230. doi: 10.1177/0093854808321526
Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Persp. Psychol. Sci.* 14, 646–671. doi: 10.1177/1745691619838258
Mattox v. United States (1895). 156 U.S. 237.
Meissner, C. A., and Kassin, S. M. (2004). "'You're guilty, so just confess!' Cognitive and behavioral confirmation biases in the interrogation room," in *Interrogations, Confessions, and Entrapment,* ed D. Lassiter (New York, NY: Kluwer Academic), 85–106. doi: 10.1007/978-0-387-38598-3_4
Neal, T. M. S., Slobogin, C., Saks, M. J., Faigman, D. L., and Geisinger, K. F. (2019). Psychological assessments in legal contexts: Are courts keeping "junk science" out of the courtroom? *Psychol. Sci. Public Interest* 20, 134–163. doi: 10.1177/1529100619888860
P. (D.) v. S. (C.) (1993). CanLII 35 (CSC), [1993] 4 RCS 141.
Paciocco, D. M. (2010). Understanding the accusatorial system. *Can Crim Law Rev.* 14, 307–325.
Plusquellec, P., and Denault, V. (2018). The 1000 most cited papers on visible nonverbal behavior: a bibliometric analysis. *J. Nonverbal Behav.* 42, 347–377. doi: 10.1007/s10919-018-0280-9
R. v. Béland (1987). CanLII 27 (SCC), [1987] 2 SCR 398.
R. v. Burns (1994). CanLII 127 (SCC), [1994] 1 SCR 656.
R. v. Handy (2002). SCC 56 (CanLII), [2002] 2 SCR 908.
R. v. Kerr (2020). ONCJ 214 (CanLII).
Smith, J. (2020). *The Junk Science Cops Use to Decide You're Lying. The Intercept.* Retrieved from: https://theintercept.com/2020/08/12/blueleaks-law-enforcement-police-lie-detection/ (accessed September 26, 2020)
Sporer, S. L., and Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychol. Public Policy Law* 13, 1–34. doi: 10.1037/1076-8971.13.1.1
Strömwall, L., Granhag, P. A., and Hartwig, M. (2004). "Practitioners' beliefs about deception," in *Deception Detection in Forensic Contexts,* eds P. A. Granhag and L. A. Strömwall (Cambridge: Cambridge University Press), 229–250. doi: 10.1017/CBO9780511490071.010
Strömwall, L. A., and Granhag, P. A. (2003). How to detect deception? *Arresting the beliefs of police officers, prosecutors and judges. Psychol. Crime Law* 9, 19–36. doi: 10.1080/10683160308138
The Global Deception Research Team (2006). A world of lies. *J. Cross Cult. Psychol.* 37, 60–74. doi: 10.1177/0022022105282295
Timony, J. P. (2000). Demeanor credibility. *Cathol. Univers. Law Rev.* 49, 903–944.
Todorov, A., Olivola, C. Y., Dotsch, R., and Mende-Siedlecki, P. (2015). Social attributions from faces: determinants, consequences, accuracy, and functional significance. *Annu. Rev. Psychol.* 66, 519–545. doi: 10.1146/annurev-psych-113011-143831
United States v. Scheffer (1998). 523 U.S. 303.
Vrij, A., Hartwig, M., and Granhag, P. A. (2019). Reading lies: Nonverbal communication and deception. *Annu. Rev. Psychol.* 70, 295–317. doi: 10.1146/annurev-psych-010418-103135
Walters, S. B. (2003). *Principles of Kinesic Interview and Interrogation.* New York, NY: CRC Press.
Wilson, J. P., and Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. *Psychol. Sci.* 26, 1325–1331. doi: 10.1177/0956797615590992
Wilson, J. P., and Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: the role of facial trustworthiness. *Soc. Psychol. Personal. Sci.* 7, 331–338. doi: 10.1177/1948550615624142

Conflict of Interest: 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.

The reviewer JM declared a past co-authorship with the author VD to the handling editor.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 December 2020
doi: 10.3389/fpsyg.2020.599703

Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport

Gary Bente[1], Eric Novotny[2]*, Daniel Roth[3] and Ahmad Al-Issa[4]

[1] Department of Communication, Michigan State University, East Lansing, MI, United States, [2] Grady College of Journalism and Mass Communication, University of Georgia, Athens, GA, United States, [3] Department of Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, Munich, Germany, [4] Department of English, American University of Sharjah, Sharjah, United Arab Emirates

The current paper addresses two methodological problems pertinent to the analysis of observer studies in nonverbal rapport and beyond. These problems concern: (1) the production of standardized stimulus materials that allow for unbiased observer ratings and (2) the objective measurement of nonverbal behaviors to identify the dyadic patterns underlying the observer impressions. We suggest motion capture and character animation as possible solutions to these problems and exemplarily apply the novel methodology to the study of gender and cultural differences in nonverbal rapport. We compared a Western, individualistic culture with an egalitarian gender-role conception (Germany) and a collectivistic culture with a more traditional gender role conceptions (Middle East, Gulf States). Motion capture data were collected for five male and five female dyadic interactions in each culture. Character animations based on the motion capture data served as stimuli in the observation study. Female and male observers from both cultures rated the perceived rapport continuously while watching the 1 min sequences and guessed gender and cultural background of the dyads after each clip. Results show that masking of gender and culture in the stimuli was successful, as hit rates for both aspects remained at chance level. Further the results revealed high levels of agreement in the rapport ratings across gender and culture, pointing to universal judgment policies. A $2 \times 2 \times 2 \times 2$ ANOVA for gender and culture of stimuli and observers showed that female dyads were rated significantly higher on rapport across the board and that the contrast between female and male dyads was more pronounced in the Arab sample as compared to the German sample. Nonverbal parameters extracted from the motion capture protocols were submitted to a series of algorithms to identify dyadic activity levels and coordination patterns relevant to the perception of rapport. The results are critically discussed with regard to the role of nonverbal coordination as a constituent of rapport.

Keywords: rapport, nonverbal behavior, motion capture, character animation, gender, culture

OPEN ACCESS

Edited by:
Miles L. Patterson, University of Missouri–St. Louis, United States

Reviewed by:
Jon Grahe, Pacific Lutheran University, United States
Masanori Kimura, Kobe College, Japan

***Correspondence:**
Eric Novotny
Eric.novotny@uga.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *27 August 2020*
Accepted: *19 November 2020*
Published: *11 December 2020*

Citation:
Bente G, Novotny E, Roth D and Al-Issa A (2020) Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport. Front. Psychol. 11:599703. doi: 10.3389/fpsyg.2020.599703

INTRODUCTION

Communication is a complex and highly demanding task. It can unfold in a harmonious and effortless way, yet sometimes also fail catastrophically. A most critical determinant of communication success is whether the partners "click" on a nonverbal level, or in other words whether they can establish rapport (Granitz et al., 2009). In general terms, rapport is characterized as being connected, tuned-in, or in-sync (Bernieri, 1988). Rapport has been shown to positively influence communication outcomes in a variety of situations, including classroom interactions (Bernieri, 1988; Murphy and Valdéz, 2005; Nguyen, 2007), conflict resolution (Drolet and Morris, 2000), child care (Burns, 1984), therapeutic interventions (Hall et al., 1995; Cooper and Tauber, 2005) business interactions (Gremler and Gwinner, 1998, 2000; Macintosh, 2009), among others. Importantly, rapport is described as an emergent social phenomenon only observable in interactions, defining the dyad as the smallest unit of analysis (Tickle-Degnen and Rosenthal, 1990; Bernieri and Gillis, 1995b). Rapport has been shown to rely on a dyad's nonverbal expressiveness (Tickle-Degnen, 2006), comprising signals of mutual attentiveness, the reciprocal exchange of positivity cues, and most importantly, the coordination of nonverbal behaviors (Tickle-Degnen and Rosenthal, 1987, 1990; Bernieri, 1988; Bernieri et al., 1996; Grahe and Bernieri, 1999). These coordination patterns include both temporal entrainment (synchrony; Lakin and Chartrand, 2003; Lakens and Stel, 2011; Ramseyer and Tschacher, 2011; Fujiwara and Daibo, 2016) and similarities in form (motor and postural mimicry; Bernieri et al., 1994; Lakin and Chartrand, 2003; Miles et al., 2009). In this sense, Grahe and Bernieri (1999) summarize: ". . .rapport is primarily a physically manifested construct; it is a construct that is visible at the surface and readily apparent" (p. 265).

In fact, observers seem to be able to assess a dyad's rapport from nonverbal interactions very swiftly and with considerable consensus (Gillis et al., 1995; Bernieri et al., 1996). However, Bernieri and Gillis (1995a) report that this consensus could not be established between observers and the interactors, which the authors discussed as a validity problem. It is questionable though, whether self-reports were an appropriate criterion for the validity of observer ratings. Evaluations of rapport provided by the interactors are based on the individual feelings. They may depend on many other factors than nonverbal behaviors and movement coordination, such as for instance others' group membership, perceived attractiveness, similarity and liking. First-person impressions might even be controversial across the interactors and thus difficult to use as a unified criterion. Third-person judgments of rapport, in contrast, are based on observations of the dyad as a whole and could provide a more neutral picture of the emergent dyadic phenomena. Yet, observation data can evidently be flawed by judgment biases and inaccuracies that may affect intuitive judgments (ratings) of perceived rapport as well as descriptive accounts (behavior coding).

The current study addresses two major methodological problems pertinent to observer studies in nonverbal rapport: (1) the production of appropriate stimulus materials that allow one to assess observers' perceptions of rapport independently from stereotypes and (2) the provision of objective measures of the nonverbal patterns underlying these perceptions independently from observers' implicit theories. We suggest motion capture technology and character animation as solutions to these problems. To demonstrate the potential of the novel methodology we present a cross-gender, cross culture study to demonstrate how the tools can be effectively used to study individual and group differences in nonverbal rapport and beyond. To achieve a maximal contrast regarding expected cultural differences in the first study of this kind, we compared a Western, individualistic culture with an egalitarian gender-role conception (Germany) and a collectivistic culture with more traditional gender role conceptions (Middle East, Gulf States). We ask (1) whether nonverbal rapport is consistently perceived by observers from different cultures solely based on the perception of dyadic movement patterns and (2) whether observer judgments reveal differences in the levels of rapport that female and male dyads as well as German and Arab dyads are able to achieve. In an exploratory analysis we finally demonstrate how to identify nonverbal interaction patterns in the motion capture protocols that account for perceived differences in rapport.

BACKGROUND

We chose gender and cultural differences for this study for two reasons: first, because both factors are under-investigated with regard to nonverbal rapport; second, because both variables are particularly relevant to stereotype activation and judgment bias (Cuddy et al., 2015; Ellemers, 2018) and thus are ideal candidates to demonstrate the advantages of the novel methodology. Only two studies came to our attention that addressed gender differences in nonverbal rapport. Puccinelli et al. (2003) reported that "female observers perceived dyad members to exhibit more rapport-facilitating behavior" (p. 211) than male observers. As the stimulus material consisted in majority of female dyads, an interaction effect between the gender of interactors and observers is likely. The authors concede that the visible gender of the interactors might have selectively primed female observers' self-stereotype (cf. Cross and Madson, 1997) and in sum led to higher rapport scores for the predominantly female stimuli. Looking at rapport, Bernieri and Gillis (1995a) reported that judgments of rapport correlated with observed "female gestures" (i.e., gestures predominantly shown within female dyads). The specific features of these gestures remain elusive as the relevant behaviors were categorized by human coders, not revealing any details. Further, one might face a potential circularity here between perceptions of rapport and the spotting of particular nonverbal cues (cf. Cappella, 1990; Bente, 2019). The latter study also included observers with different cultural background (i.e., Greek and US American observers watching American dyads). While interobserver correlations across cultures were strong, the study was inconclusive with regard to the behaviors that drove their impressions. The authors hypothesized that both groups unanimously gave ". . .insufficient weight to valid behavioral predictors of rapport (such as mutual attention, reciprocal positivity and coordination, cf. Tickle-Degnen and Rosenthal, 1990) while relying on the apparently compelling but invalid cues, smiling and expressivity" (p. 115, inserts by the authors in brackets). This interpretation remains speculative as the stimulus materials showed numerous confounds between physical appearance cues of the interactors as well as different nonverbal channels such as facial expressions, gestures, body movements and postures.

Overall, the few existing studies in this domain point to recurrent methodological problems that result from: (1) the use of video stimuli to assess observer impressions (stimulus problem) and (2) the deployment of human coders to collect behavioral data (observer problem, cf. Bente, 2019).

The Stimulus Problem

Video stimuli as predominantly used in previous observer studies not only show the nonverbal interaction, but also reveal person characteristics such as gender, race, culture, age or attractiveness that might be relevant to stereotype activation and judgment bias (Dion et al., 1990; Stangor and Crandall, 2013). For instance, gender-role stereotypes could lead to the ascription of higher rapport levels to female as compared to male dyads, just because women are expected to put more emphasis on relational harmony than men (see Cross and Madson, 1997). The same holds true for stereotypes about cultures and assumptions regarding their valuing of social harmony and relatedness (Triandis, 1995; Cuddy et al., 2009). Different techniques have been proposed to solve this problem (cf. Bernieri et al., 1994), including the use of point light displays (Johansson, 1973, 1976) or video quantization techniques (Berry et al., 1991, 1992). However, both methods display specific limitations. Quantization techniques used to degrade video images to rougher mosaic patterns in order to obscure physical appearance are not sufficient to completely eliminate clues to gender and culture (see stimulus examples in Bernieri et al., 1994). Point light displays on the other hand fail to capture postural information (see Cutting and Proffitt, 1981; Runeson and Frykholm, 1981). We here suggest the use of computer animations of standardized, neutral characters (avatars). Based on full body motion capture data collected in dyadic interactions such animations allow one to obscure gender, culture and other obvious individual characteristics of the interactors while portraying movements and postures with high fidelity (cf. Bente, 2019).

The Observer Problem

Implicit theories about relevant rapport indictors cannot only mislead observers' evaluative impressions (cf. Bernieri and Gillis, 2001), but inversely, observers' impressions of relational quality can also bias their description of the behavior. Cappella (1990) holds that for instance judgments of coordination "..., whether by participants or observers, could be confounded with judgments of positivity if judges' implicit theories of social interaction are that positive interactions are ones in which the people are in sync. If this is the case, then the judges would be assessing positivity and not synchrony, and the correlation to rapport would be an artifact" (p. 303). To avoid such circularities, descriptive movement data are needed that are independent from observers' evaluative impressions. Motion Energy Analysis (MEA) has been suggested to solve this issue (Ramseyer and Tschacher, 2011). MEA quantifies general motor activity by calculating pixel changes between pre-filtered sequential video frames. More recently, the authors introduced a method to separate body and head movement within MEA (Ramseyer and Tschacher, 2014). MEA data, however, lack information about the form of movements and postures displayed. We suggest the use of motion capture technology to overcome this constraint. In contrast to MEA, motion capture technology issues detailed protocols of body movement including rotation and translation information for all joints (cf. Poppe et al., 2014; Cornejo et al., 2018). The rich data protocols resulting from motion capturing allow to analyze a broad variety of behavioral features as possible predictors of perceived rapport. These features include aggregates of movement activity across all body parts (comparable to MEA), as well as selections of specific nonverbal subsystems, such as gestures or head and body movements and postures. Most importantly, the synchronous movement protocols of both interaction partners allow one to establish dyadic coordination patterns in terms of temporal entrainment (synchrony) as well postural similarity (mimicry).

METHOD: OBSERVER STUDY

Stimulus Material

Volunteer student participants were recruited for interaction recordings at the University of Cologne in Germany and at the American University of Sharjah in the United Arab Emirates (UAE). Recruitment in Sharjah focused on local Emiratis and students from surrounding Arab countries (Gulf States) whose mother tongue was Arabic. For the German portion of the sample, only native students were recruited in Cologne. Volunteers were randomly paired into homogenous female-female and male-male dyads. A major criterion was that the partners did not know each other before the sessions. If this was the case, the participants were reassigned to another pair. In total, 15 dyads were recorded in Germany and 15 in the UAE.

Participants were instructed that they would have a short 5–7 min conversation with another student during which they should get to know each other. Before the conversations began, participants were led into different rooms to put on the data suits necessary for motion capturing. A same sex student assistant placed the markers on the data suits and guided the interactors to the middle of the recording room where they met the experimenter. Motion capture was performed with a 12-camera Optitrack system and the capture software Arena (Optitrack, 2017). Cameras were positioned around a square area of 4 × 4 meters. Participants were then asked by the experimenter to take a T-pose (upright symmetric posture with legs closed and arms horizontally stretched out, palms down) for calibration of the tracking system. Then the participants were told that they could move freely in the square between the cameras and should use the next 5–7 min to get to know each other. Next, the experimenter left the room and the participants started the conversation. Using the capture software Arena, full body motion of both actors was captured during the conversation with a temporal resolution of 150 Hz. **Figure 1** shows a dyad wearing the data suits with the IR reflecting markers and a projection of the capture software showing both virtual characters for demonstration purposes. After completing the interaction, the participants were debriefed and received 15 Euro (Cologne), or an equivalent on-campus restaurant voucher (Sharjah) for their participation.

Movement data of all dyads were transferred to the software MotionBuilder (Autodesk, 2017) for post-production to map the animation data to the neutral computer model and to handle and animation issues, such as jitter and penetrations of body parts. The computer model appeared as a wooden mannequin (cf. Bente et al., 2010) to standardize appearance and to obscure the gender and culture of participants (see **Figure 2**).

We selected five female and five male dyads from each country with the best recording quality (fewest recording errors, i.e., jitter, erroneous joint detection or dropouts and data drop-outs because of marker occlusions) as our final stimulus material. To provide comparable stimuli with regard to length we selected 1 min segments from the middle of all 20 interaction recordings. Using MotionBuilder, the stimulus sequences were then rendered to digital videos with a 25 frames/second frequency and in a 1,024 × 768 pixel resolution.

Dependent Measures

The study focused on perceived rapport as the major variable. To account for potential variations in perceived rapport during the 1 min interactions, we used a real-time-response (RTR) measure (Bente et al., 2009) to indicate continuously the level of perceived rapport during observation. A 9-point rating scale with the extremes "+4" (very good rapport), "−4" (very bad rapport) and "0" (indifferent) was used for this purpose and was displayed as a gauge on the stimulus screen (see **Figure 2**).

As the computer characters were intended to obscure gender and culture of the stimulus dyads, we included two further questions at the end of each clip to check the effectiveness of this manipulation. We asked: (1) whether the participants assumed the respective dyad to be female or male and (2) whether they assumed the interactors to be of German or Arab origin. The respective hit rates should serve as a treatment check measure, to ensure that gender and culture were successfully obscured and perceived movement alone did not lead to the recognition of gender and culture and related stereotypes.

Participants

Student participants for the perception study were recruited at the University of Cologne (Germany) and the American University of Sharjah (UAE) via local student mailing lists and seminar announcements aiming at an equal number of male and female observers. Twenty-six female and 24 male observers participated in the study in Cologne, and 25 female observers and 21 male observers participated in Sharjah. After analyzing the biographical data, we excluded eight participants from analysis who did not meet the selection criteria (i.e., born and raised either

FIGURE 1 | Posed interaction showing the setup and the interface of the capture software as projection in the background (180° rotated).

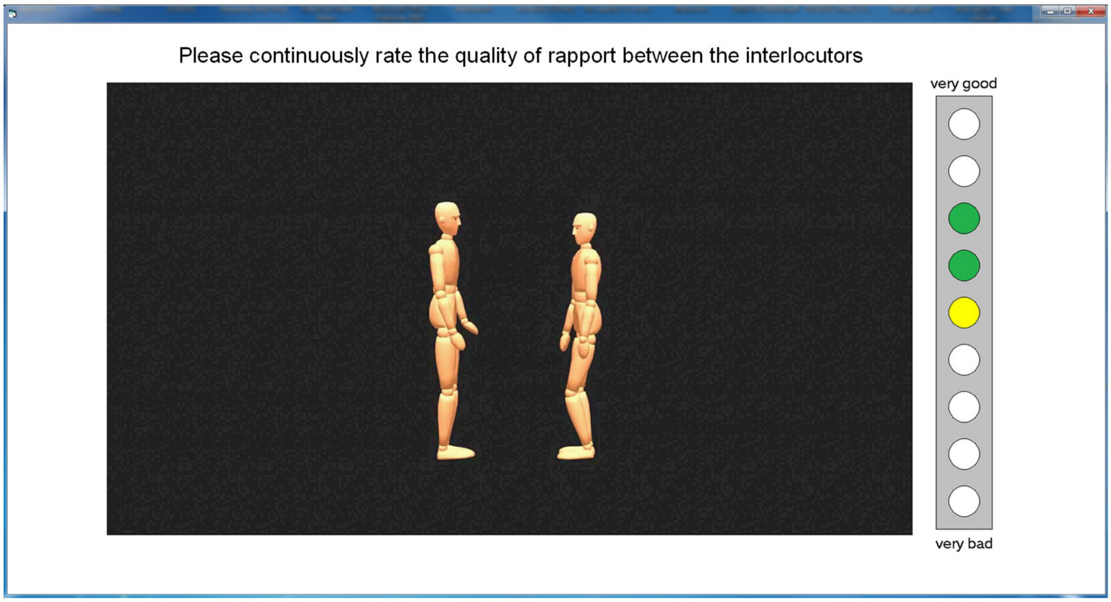


FIGURE 2 | Screenshot of the user interface for the continuous judgment of rapport (description in the text).

in Germany or an Arab country as well as having either German or Arab parents). The final data set consisted of 24 female and 23 male observers in Cologne (mean age = 24.87, *SD* = 6.66), and 22 female observers and 19 male observers in Sharjah (mean age = 20.22, *SD* = 2.33).

Procedure

Up to six participants (between three and six depending on the number showing up) were seated in a large seminar room capable of holding 30 people, leaving at least six feet of distance between each of them. Each participant faced a laptop computer with a 15 (1,980 × 1,280) widescreen monitor. They were asked not to talk to each other during the experiment. Then they were shown a screen shot printout of the user interface of the experiment software (see **Figure 2**) and received the following instructions:

> *You will now see a series of short one-minute muted videos each showing computer animations of two people in a conversation. During the video you will be asked to provide your judgment about the dyad's rapport; this means how well the two inter-actors are getting along with each other or are tuned-in during interaction. Please use the cursor up-down keys on your keyboard to continuously indicate your impression of their rapport. Your selections will be shown on the gauge on the right side of the screen (see picture in front of you). Moving the scale points on the gauge up into the green area means you have a positive impression of their rapport, moving down into the red area would indicate a negative impression. The more green or red dots light up, the better or worse is the dyads rapport, respectively. Your impression can change at any time during the interaction. Please use the cursors continuously to indicate any changes in your impression. After each clip you will be prompted, on a new screen, to indicate your opinion on whether this dyad was a female or a male dyad and whether the dyad portrayed Germans or Arabs. If you have any questions you can ask now. If you are ready, please hit the start button on the screen to launch the experiment.*

Participants then started the video sequences that were presented in random order. **Figure 2** shows the screen layout during the stimulus presentation with the RTR gauge displayed at the right of the video window. The RTR gauge could be controlled by pressing the cursor-up or cursor-down keys on the computer keyboard. At the end of the video, a new screen appeared asking for the dyads' gender (male or female), followed by a screen asking for the dyads' culture of origin (German or Arab).

RESULTS: OBSERVER STUDY

Control Check

To ensure the efficiency of our stimulus manipulation in masking stimulus gender and culture, we tested whether recognition rates for gender and culture were significantly different from chance level. Two separate one-sample *t*-tests were conducted for both variables using the chance level of 10 hits out of 20 dyads as criteria. Results indicated no significant difference from the chance level for either gender or culture. Mean hit rates were $M = 10.30$, $SD = 2.64$, $t(87) = 1.05$, $p = 0.30$ for gender, and $M = 9.97$, $SD = 2.27$, $t(87) = -0.14$, $p = 0.89$, for culture. This indicated that the participants were not able to identify reliably the gender or culture of the avatar dyads from their appearance nor from their nonverbal behavior. Accordingly, we excluded stereotype influences as accounting for variance in the rapport judgments.

Observer Agreement

To test whether perceptions of nonverbal rapport are consistent within and across the observer groups we first conducted intraclass correlations for each observer group (female German, male German; female Arab, male Arab) based on the rapport ratings toward the whole stimulus set. The correlations are presented in **Table 1**.

As **Table 1** shows, there was high agreement within observer groups in terms of rapport ratings toward all stimuli. Next, we aggregated the rapport ratings for the four groups of observers, generating a data matrix that contained the stimuli as cases and the average ratings of each group as variables. We then ran Pearson-Product-Moment-Correlations for the four observer groups across the 20 stimuli. The correlations are displayed numerically in **Table 2**, as well as visually in **Figure 3**. All groups significantly correlated in their rapport ratings, with correlations explaining between 57 and 77% of the variance. **Figure 3** illustrates that the correlations were not merely driven by the gender differences but also reflect correlation within the stimulus categories.

Dynamics of Rapport Ratings

Regarding the dynamics of the continuous rapport ratings, we conducted an exploratory graphical analysis of the RTR process measures. For this purpose, we averaged the RTR data across observers over time for each of the 20 dyads. **Figure 4A** shows the resulting time graphs. It illustrates that with only a few exceptions, average rapport ratings show little spontaneous fluctuations or any significant changes in direction during the 1 min sequences. Rather, the curves suggest that observers very early (after about 5 s) take a certain judgment direction and asymptotically approach a relatively stable level after about 30 s. This tendency becomes even more evident when averaging the dyads with high, medium, and low average rapport ratings. **Figure 4B** shows the averaged curves for the seven clips with the highest, the seven with the lowest, and the six with medium rapport ratings (lying between the seven high and low scoring stimuli). It supports the assumption that there is an initial judgment tendency driven by thin slices of behavior and that further observations just serve to consolidate the swift first impressions. It remains an open question, though, whether this process is driven by consistent stimulus characteristics, aggregating in the observers' impressions over time, or caused by selective perceptual strategies of the observers, who rapidly form their impression and then assimilate further observations.

TABLE 1 | Intraclass correlations for average rapport ratings across all stimuli.

Culture	Gender	ICC	df	*F*
German	Female	0.757***	25, 475	4.113
	Male	0.577***	23, 437	2.365
Arab	Female	0.595***	24, 456	2.471
	Male	0.566***	20, 380	2.304

****p* < 0.001.

TABLE 2 | Pearson's product moment correlations for average rapport ratings across 20 dyads.

Observers	Female German	Male German	Female Arab	Male Arab
Female German		0.877***	0.785***	0.755***
Male German			0.792***	0.772***
Female Arab				0.805***
Male Arab				

****p* < 0.001; *N* = 20 for all analyses.

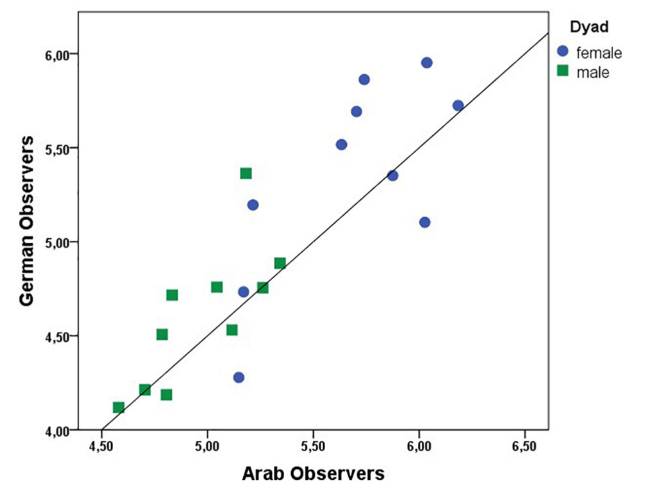


FIGURE 3 | Correlation graph for German and Arab observers, differentiated for male (green squares) and female (blue circles) stimulus dyads.

Gender and Cultural Differences

To analyze the effects of the observers' and the dyads' gender and culture on perceived rapport we conducted a 2 × 2 × 2 × 2 mixed ANOVA including culture of observers (German vs. Arab) and gender of observers (female vs. male) as between-subject factors and the culture of dyad (German vs. Arab) and gender of dyad (male vs. female) as within-subject factors. Average rapport ratings for the 1 min sequences served as the dependent variable. Between-subject factors were included despite the high inter-observer correlations to identify potential interaction effects due to in-group familiarities.

We found a main effect for the culture of observers, $F(1, 87) = 12.96$, $p < 0.001$, $\eta^2_p = 0.134$, but no main effect for gender of observers or any interaction effect between observer factors and stimulus factors. The main effect shows that Arab observers ($M = 5.29$, $SD = 0.38$) were, in general, more positive in their rapport ratings than German observers ($M = 4.99$, $SD = 0.40$). It remains unclear whether this finding reflects a general positivity bias in the social judgments of Arabs as compared to Germans, or a different sensitivity to rapport cues. As there were no interaction effects between

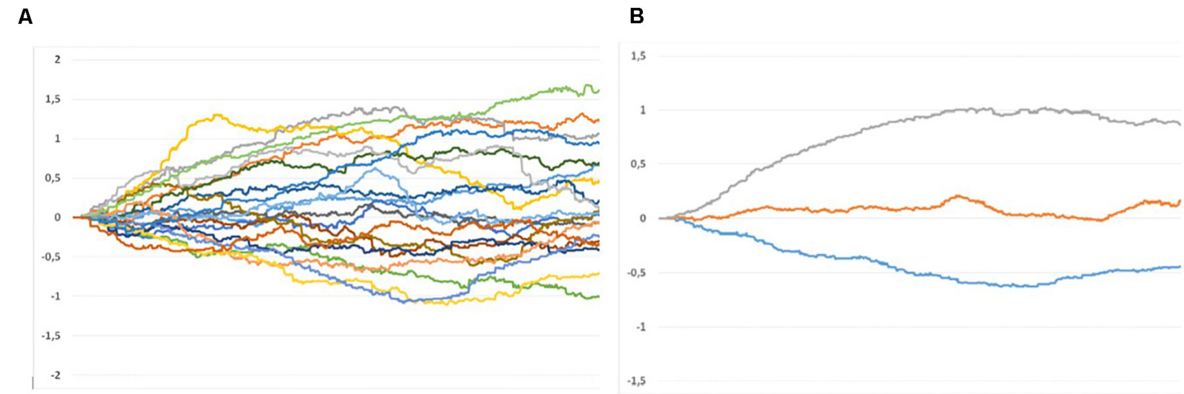


FIGURE 4 | Continuous RTR rapport ratings. **(A)** Averaged for each of the 20 stimulus clips across 88 observers. **(B)** Averaged for 7 high, 6 neutral, and 7 low rapport dyads.

the culture of observers and any other factor, we refrained from correcting the bias by grand mean standardization (Fischer, 2004).

We further found a significant main effect for the dyads' gender, $F(1, 87) = 149.99$, $p < 0.001$, $\eta^2_p = 0.641$, indicating that the female dyads ($M = 5.50$, $SD = 0.49$) were perceived as doing better in nonverbal rapport building than male dyads ($M = 4.78$, $SD = 0.52$). We also found a main effect for the culture of the dyads, $F(1, 87) = 6.62$, $p = 0.012$, $\eta^2_p = 0.073$ indicating that German dyads were perceived higher in rapport than the Arab dyads. This main effect is, however, explained by an interaction between the gender and culture of the dyads, $F(1, 87) = 4.0$, $p = 0.049$, $\eta^2_p = 0.045$. Only the male Arab dyads were rated lower in rapport than male German dyads, $t(87) = 2.87$, $p = 0.005$, $d = 0.37$, while female dyads showed no difference between the two cultures, $t(87) = -0.75$, $p = 0.391$, $d = 0.10$. **Figure 5** illustrates the interaction effect.

METHOD: BEHAVIOR ANALYSIS

Rationale

The behavioral analyses aim to showcase the information richness of the motion capture data and to demonstrate multiple ways to extract and examine the nonverbal interaction patterns that underlie the perception of dyadic rapport. We here focus on a set of behaviors that have been described in the literature as relevant to rapport. These are: *expressivity, mutual attention, reciprocal positivity,* and *coordination* (Tickle-Degnen and Rosenthal, 1990; Bernieri et al., 1996; Grahe and Bernieri, 1999; Tickle-Degnen, 2006), the latter including aspects of *synchrony* (i.e., the temporal entrainment of interactors' movement activity; Lakin and Chartrand, 2003; Lakens and Stel, 2011; Ramseyer and Tschacher, 2011; Fujiwara and Daibo, 2016) and *mimicry* (i.e., the similarity of movements and postures in form; Bernieri et al., 1994; Lakin and Chartrand, 2003; Miles et al., 2009). Importantly, we here conceive coordination as orthogonal to the individual behaviors that are subject to coordination and that have to be further processed to reveal the respective dyadic interdependencies. We exemplarily selected the following individual behaviors: (1) rotational head orientation as a proxy for visual attention (Loomis et al., 2008), (2) approach/distancing movements as an indicator of positivity/negativity (Sundstrom and Altman, 1976), and (3) overall movement activity (i.e., aggregate positional changes of all joints) as a measure of expressivity, comparable to MEA (Nelson et al., 2016). These behaviors were further submitted to algorithms quantifying the behavior of the dyads as a whole and the respective spatial–temporal coordination patterns.

Feature Extraction

A Python plugin for MotionBuilder was developed to export global translation of 15 body joints for each interaction partner from the .FBX (filmbox format) animation files to .CSV files. We also exported the coordinates for three additional virtual markers attached to the nose and the ears of the interactors. These coordinates were used to calculate head rotation instead of using the Euler angles in the .FBX files, which are difficult to interpret. Translation data were exported as absolute metric values in the shared 3D world coordinate system. Further calculations were based on these data sets (see Leuschner, 2013). During export from the capture system to MotionBuilder data, a different scale factor was used for all the Arabic dyads and one German dyad. The scale factor deviated from the real world dimensions by the factor 5/6. All sizes and distances for those data sets were therefore corrected, i.e., upscaled by the factor 1.2 before being used in parameter formation and statistical analysis.

To cover individual behaviors relevant to orientation, distance, and activity we extracted three behavior vectors for each interaction partner from the data matrices:

(1) Orientation: We calculated the angular deviation of the individual head rotation from the direct line of view, which was defined as the dynamically changing line between both computer models' nose markers (see **Figure 6**).
(2) Distance: We calculated movements toward or away from the partner as consecutive Euclidian distances between the

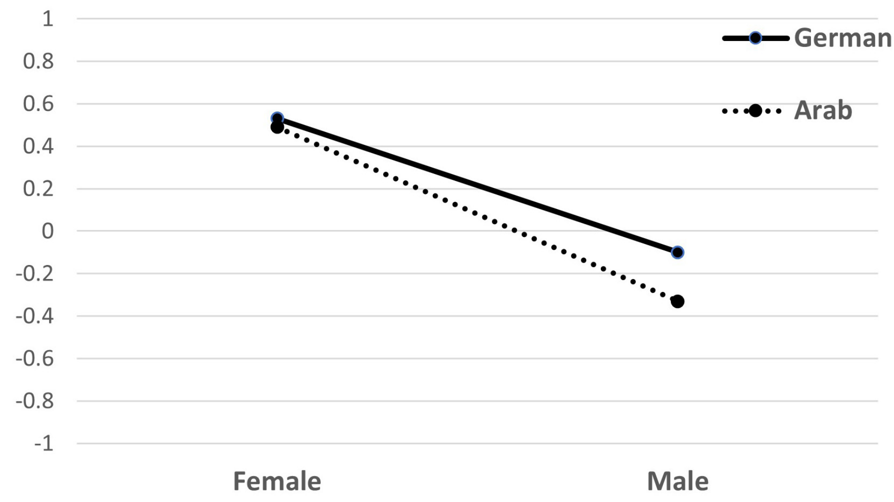


FIGURE 5 | Interaction graph for rapport ratings (estimated marginal means) depending on gender and culture of the stimulus dyads. The *y*-axis scale is set from −1 to +1 here to showcase the effect (the participants rated from −4 to +4).

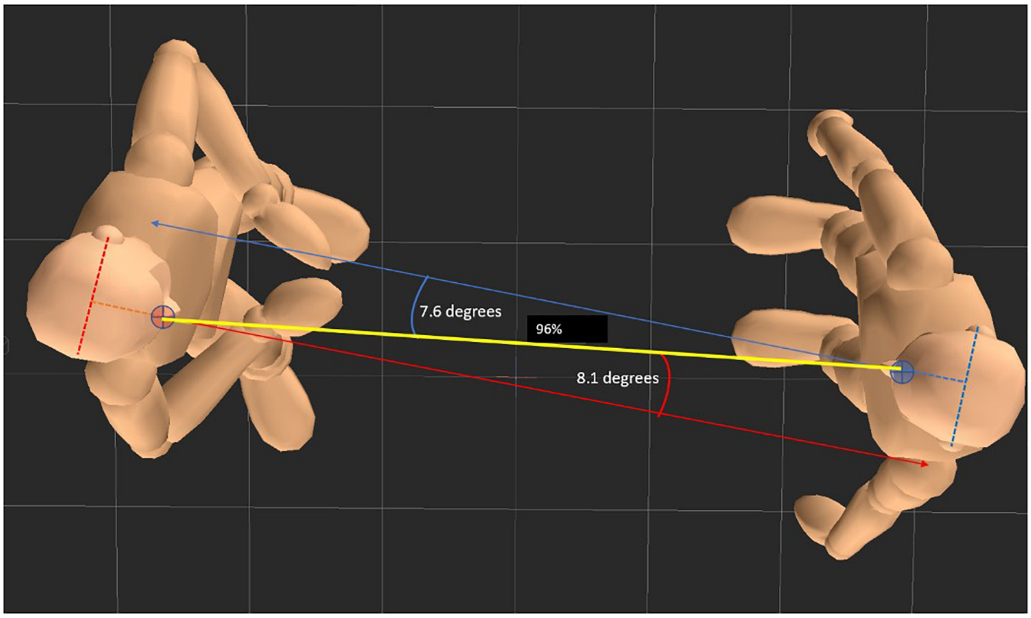


FIGURE 6 | Rationale for measuring rotational head orientation and interpersonal proximity. Invisible virtual markers between the characters' eyes were used as anchor for distance measurement and as references for the assessment of rotational deviations from the direct line of view. The angle was calculated as orthogonal to the line between two virtual ear markers projected onto the x/z plane. Distance was calibrated as percentage of both partners' arm lengths.

position of the nose marker of one partner at timepoint t and the position of the nose marker of the other person at timepoint t − 1 (see **Figure 6**).

(3) Activity: We aggregated the positional changes of all 15 joints between consecutive time points (position changes were z-transformed for each joint separately before aggregation).

To enable multidimensional comparisons of postural similarity (mimicry) we further extracted translation matrices from the animation files of both partners containing the 3D coordinates of the 15 body joints. Following the procedures suggested by Poppe et al. (2014) positional data of the joints were normalized to compensate for different body sizes and the skeletons of both partners were snapped to the shared coordinate system's origin and y-rotations were frozen. Both hips thus were always fixed in the same position and oriented toward the front of the scene. The postural similarity was then calculated for each point of time as the sum of distances between the interaction partners corresponding joints. The rationale of the procedure

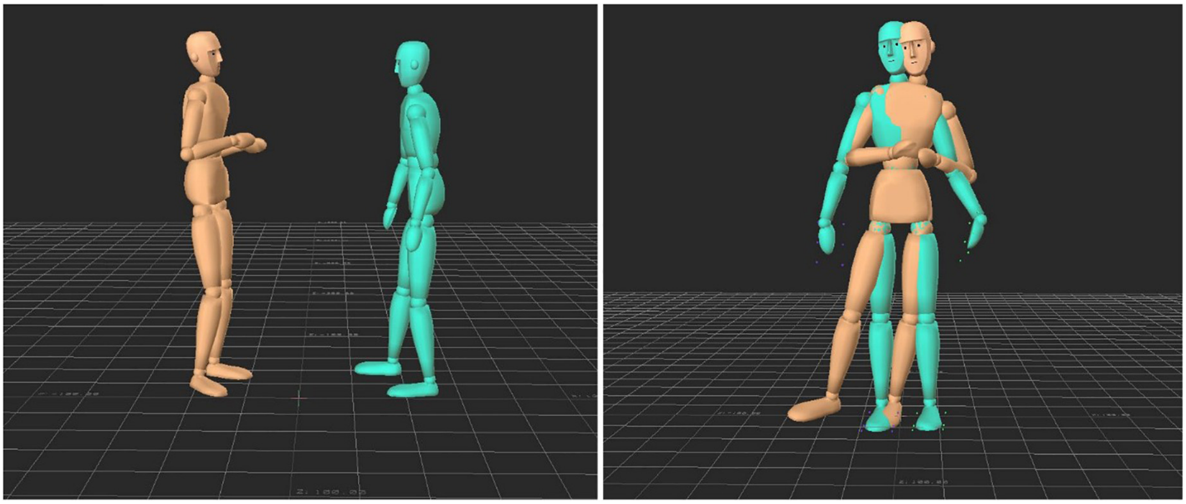


FIGURE 7 | Rationale for multidimensional comparisons: Both skeletons' hips were constrained to the position as well as the y-rotation of a static object in the coordinate system origin, thus always pointing to the front. Postural differences were calculated as Euclidian distances between the corresponding joints of both partners for each timepoint.

is illustrated in **Figure 7**. The transformations as well as the CSV-export of the transformed data were performed with the MotionBuilder Python plugin.

Parameter Formation

The extracted vectors/matrices were further processed in two ways. First, we calculated compound measures to characterize the dyads' nonverbal behavior as a whole. These measures comprised: (1) orientation: the mean absolute deviation from the direct line of view averaged for both partners and the percentage of time both partners' head deviation was less than 10° from the direct line of view (see **Figure 6**), (2) distance: the average of the Euclidian distance between the partners nose markers over time, and (3) activity: the mean of aggregated position changes across all 15 joints and both partners, and the percent of time of simultaneous activity and inactivity of both partners.

Second, the individual vectors were submitted to a suite of algorithms to quantify different aspects of coordination (see Cheong, 2019 for the algorithms and Python codes). We applied: (1) Pearson correlations, (2), Mutual Information (MI), (3) Rolling Window Time Lagged Cross-Correlation (RWTLCC), and (4) Dynamic Time Warping (DTW). DTW has only recently been introduced to the field of gesture analysis (Ten Holt et al., 2007; Chen et al., 2019). A major advantage of the method is that it takes into regard differences in form and time when comparing two signals. We also applied DTW to the multidimensional translation matrices that were extracted from the .FBX animation files after snapping the hips of the two partners (see **Figure 7**). Instead of the point-to-point differences between two vectors we here used the local sum of the point to point Euclidian distances of all 15 joints to fill the DTW data matrix. These data were also used to calculate the mean postural difference over time.

For RWTLCC we applied a step-size of 0.2 s (5 datapoints), a lag of ±2 s (±50 datapoints) and a rolling window size of 5 s (125 datapoints). To quantify the mutual interdependencies of the two signals we calculated the absolute offset of the correlation peak from the zero lag as well as the average maximal correlation found at this point. Different filters were used for the three behavioral dimensions using the "scipy.signal" library. Head rotation angles were lowpass filtered with a constant of 0.25 and standardized (divided by standard deviation) keeping the dynamic zero point as "straight orientation toward the partner." This allowed us to focus on more stable orientation patterns instead of rapid local movements (e.g., head shakes). Approach and distancing movements as well as overall motion were lowpass filtered with a filter constant of 1.0 to suppress observable jitter in the data.

Pearson "r" and the entropy measure resulting from MI analysis were used as input for statistical analysis. From RWTLCC we calculated the average maximal correlation at each point across all time lags as well the absolute offset of the correlation peak from the zero lag as general coherence and synchrony indicators. We further used the "DTW distance" measure, i.e., the minimum path cost (Cheong, 2019), to quantify the (dis)similarity between the behavioral vectors of the interactants for further analyses.

RESULTS: BEHAVIOR ANALYSIS

Pearson correlations were conducted for the $N = 20$ dyads between the extracted behavioral parameters and the average rapport ratings they received. The results for the aggregate dyadic measures are shown in **Table 3**. A significant negative correlation was found between the interpersonal distance of the partners and the rapport ratings, indicating that a closer stance was associated with higher levels of perceived rapport. Neither one of the aggregate orientation or activity parameters showed a significant correlation with rapport.

Coordination analysis of individual parameters showed a different picture. Results are summarized in **Table 4**. For the rotational head movements (i.e., turning away from or toward the partner's actual position as a proxy for directed gaze and visual attention), Pearson correlations as well as peak correlations in the ± 5 second window of RWTLCC showed significant negative correlations with rapport ratings: the lower the peak correlation between the orientation parameter, the higher were the rapport ratings. This suggests that rotational head movements in high rapport dyads occur in a complementary rather than in a symmetric fashion. **Figure 8** illustrates the result for two typical dyads. In contrast to the head rotation we found positive correlations of perceived rapport and interpersonal distance variations. The higher the peak correlation in distancing and approach behaviors within the ± 5 second window (the better coordinated the moves toward or away from each other), the higher was the perceived rapport level. Similarly, we found a significant negative correlation for the offset of the peak correlations between the two motion vectors (aggregated Euclidian distances between 15 joints) and the rapport ratings. The smaller the offset for a correlation peak in the lag window, the higher were the rapport ratings, which seems to corroborate findings from prior MEA studies (Ramseyer and Tschacher, 2011, 2014) suggesting a correlation between motor synchrony and rapport. DTW measures did not correlate in any of the parameters.

Correlations between rapport ratings and the multidimensional measures of postural similarity (mimicry), based on the aggregated Euclidian differences of the 15 joints, did not reach significance level. The similarity measure resulting from multidimensional DTW correlated with rapport ratings at $r = 0.359$ ($p = 0.120$). A similar correlation was found for the postural similarity measure, i.e., the larger the average joint distances (dissimilarity), the higher were the rapport ratings ($r = 0.367$, $p = 0.111$). It is worth noting, that the "minimal warp path length" and "average joint distance" correlated highly with one another ($r = 0.993$, $p = 0.001$). Average distance calculations are much faster to compute than DTW. It remains to be tested with larger samples whether both algorithms are functionally equivalent which would have important implications for their application in future studies.

TABLE 3 | Correlations between rapport ratings and the dyadic aggregate measures.

	Parameters	Correlation
Orientation	Average degree of head deviation	−0.093
	% of time head deviation < 10°	−0.265
Distance	Average distance between nose markers	−0.489*
Activity	Average degree of movement	0.360
	Both partners in motion	0.343
	Both partners inactive	−0.400

$p < 0.05$; N = 20 for all analyses.

DISCUSSION

The presented work addressed two methodological problems pertinent to the analysis of nonverbal rapport: (1) the production of standardized stimuli that allow for unbiased observer judgments and (2) the assessment of nonverbal interaction patterns that drive these judgments. We suggested motion capture and character animation as possible solutions (cf. Cornejo et al., 2018; Bente, 2019). We applied this methodology exemplarily to the study of gender and cultural differences in the perception of rapport as both person characteristics can be discerned from video and are prone to elicit stereotypes relevant to the judgment of rapport. Rapport is defined as a phenomenon only observable in interactions (Tickle-Degnen and Rosenthal, 1990), specifically depending on the coordination of nonverbal activity. In this sense, efforts to identify its behavioral underpinnings challenge the quality of the dyadic interaction data and the availability of algorithms to quantify the interpersonal coordination patterns (Bente and Novotny, 2020). We aimed to show that motion capture data provide unique possibilities to encounter this challenge.

Aiming at a maximal contrast between the groups in the first study of this kind, we focused on homogenous female and male dyads and selected German participants as representatives of a Western, individualistic and more emancipated culture and Middle East Arab participants (born in the United Arab Emirates or a Gulf State) to represent a more collectivistic culture with a more traditional gender role conception. Observers were also recruited from both regions. The results strongly support the novel approach. Using avatar animations instead of video we were able to successfully mask the gender and culture of the stimulus dyads. Recognition rates for both variables did not significantly deviate from chance level. On the other hand, rapport ratings correlated highly within and across observer groups (female/male, German/Arab) pointing to universal,

TABLE 4 | Correlations between rapport ratings and coordination parameters for the individual behavior vectors.

	Coordination parameters				
	Pearson r	Mutual Information	RWTLCC: peak correlation	RWTLCC: peak offset	DTW: distance
Orientation	−0.499*	0.006	−0.447*	0.340	0.218
Distance	−0.140	−0.014	0.545*	−0.432	−0.030
Activity	−0.321	−0.421	0.020	−0.533*	−0.401

$p < 0.05$; N = 20 for all analyses.

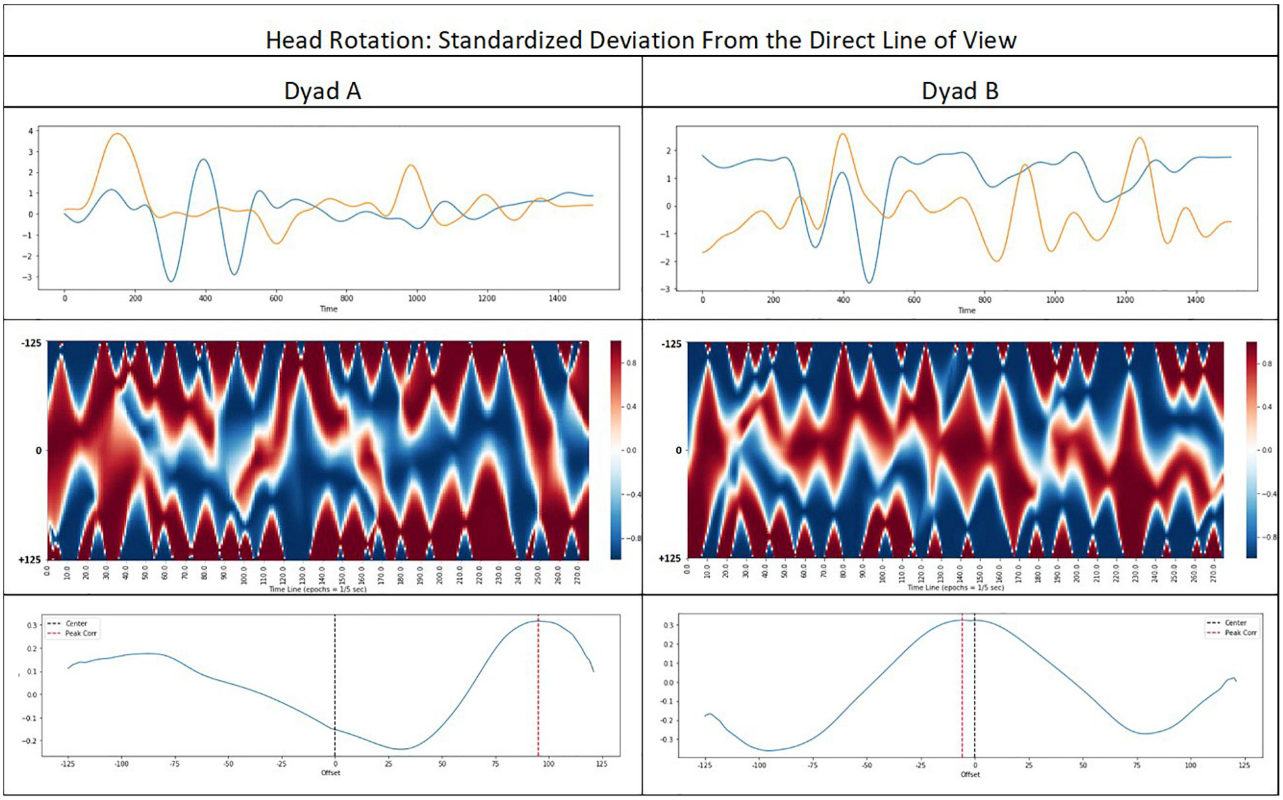


FIGURE 8 | Results of RWTLCC for the head rotation dynamics of 2 dyads differing in the peak correlation offset. The upper graphs show the low pass filtered, standardized head rotations. The mid graphs show the time lagged correlations with lags of ±5 s (±125 data points) and a moving window with a step size of 0.2 s (5 data points). The lower graphs show the mean correlations for the time lags from –125 to +125 datapoints.

gender and culture independent judgment policies (Bernieri and Gillis, 1995a). The significant inter-observer correlations also indicate that rapport can be reliably judged exclusively from the movement activity in a dyadic interaction (cf. Grahe and Bernieri, 1999) and that the nonverbal cues relevant to impression formation can be effectively portrayed by the avatar animations (Bente et al., 2010).

Beyond the high correlations in the rapport judgments, observer ratings also showed an interesting temporal dynamic. Using for the first time a continuous rating technique for rapport, we found that individual ratings already start to converge after a few seconds of observation and then asymptotically approach a final and robust level after about 20 s of the 1 min sequences. Our results corroborate earlier findings showing that rapport can be judged from a few seconds of interaction behavior (Grahe and Bernieri, 1999) and are also consistent with a "thin slices" perspective on the perception of nonverbal behavior in general (Ambady and Rosenthal, 1992). The result suggests that rapport impressions are formed swiftly and that further observations are used to consolidate the first impression rather than being sensitive to spontaneous fluctuations. In fact, this can be an effect of a perceptual bias, primed by the first impression, as well as a sign of the consistency of the specific rapport level exerted by the dyads. To test whether continuous ratings are sensitive toward local changes in nonverbal rapport one could apply the "pseudo interaction paradigm" (cf. Bernieri et al., 1988; Ramseyer and Tschacher, 2010) combining interactors from different dyads, parts of different interactions, or natural and synthetic behaviors. For instance, segments of high rapport dyads could be concatenated with segments of low rapport dyads or randomly assembled interactors from different dyads. Animation tools such as MotionBuilder provide powerful solutions to create such pseudo dyads and to smoothen the transitions by interpolating potentially distant joint postures at the seams of two segments.

An ANOVA conducted across gender and culture of stimuli and observers revealed no effect of observer gender on the general level of rapport ratings, nor did we find any interaction effects between observer gender and the other group variables (i.e., observer culture, stimulus gender, or stimulus culture). These results stand in contrast to the findings of Puccinelli et al. (2003), who reported a general positivity bias of female observers being more accommodating (i.e., generally ascribing higher rapport levels to the observed dyads). The gender effect reported by these authors, however, might be because the nearly exclusively female stimulus dyads distinctively primed rapport relevant self stereotypes in female observers. This explanation receives some support from the current results as obscuring the gender of the stimuli completely eliminated gender effects on the observer side. Future research should continue investigating this interesting interaction effect.

Regarding observer culture, we found a main effect wherein the Arab observers rated stimuli as generally higher in rapport than did the German observers. This could reflect either a general positivity bias or a different sensitivity to rapport cues. The latter assumption would be consistent with findings revealing cultural differences in nonverbal behavior between Arabs and another Western culture (i.e., US Americans; Watson, 1970). Specifically, Arabs are thought to be a higher "contact" culture, utilizing more touch and direct gaze cues, compared to their Western counterparts. As Bernieri and Gillis (1995a) note, "One might expect greater attention to such behaviors by Arabs assessing rapport than would Americans" (p. 117). It is possible that Arabs in our study were more attentive to these contact cues than Germans and thus rated stimuli higher in rapport across the board. The high correlations between Arab and German ratings, however, indicate that this difference might only concern a shift of the scale mean whereas relative perceptions of dyadic rapport are well aligned across cultures.

Beyond the intended proof of concept for the novel methodology the ANOVA results also stimulate further theoretical thoughts regarding the role of gender and culture in the establishment and maintenance of rapport. Unanimously, female dyads from both cultures were judged as significantly higher in rapport than male dyads. The perceived difference between female and male dyads was even more accentuated within the Arab dyads as revealed by a significant interaction effect. German males were rated as higher in rapport than Arab males whereas German and Arab females were rated equally in rapport. When juxtaposed with the main effect of dyad gender these findings suggest that the extent of perceived gender differs between cultures. This is consistent with the general insight that culture plays a crucial role in the establishment of gender-role expectations, respective socialization practices and resulting social behaviors (Williams and Best, 1994; Van de Vijver, 2007). As Hall and Briton (1993) posit: "Pressure to conform to stereotypes can be great. Men who are gesturally or facially expressive, for example, may be stigmatized as being weak or feminine..." (p. 283). Expressivity has been identified as a crucial behavioral feature in the perception of rapport (Bernieri and Gillis, 1995a) and as the "raw action material" in nonverbal coordination (Tickle-Degnen, 2006, p. 387). A medium level of expressivity appears to be ideal for the establishment of rapport. Nelson et al. (2016) hold that "According to Tickle-Degnen's (2006) model, optimal experiences of rapport are those where dyads feel and act in calm, yet attentive ways; suboptimal experiences foster overactive or underactive levels of action and affect. More specifically, when an actor's expressivity is overactive, information is lost between an actor and a perceiver. When expressivity is underactive, there is a shortage of nonverbal information passed between partners" (p. 3). Conceding that the Arab culture in our study has more fixed roles of masculinity and femininity than the German culture, a use of less expressive behavior by the Arab males might explain their garnering of comparably low rapport ratings.

Given the small stimulus sample size in our study, interpretations regarding gender and cultural factors in the generation of rapport should be treated with caution. Yet, our observations can be taken as a starting point for future research that investigates rapport building behavior across gender and cultures more directly. On the one hand, studies of this kind would have to be based on significantly larger interaction samples. On the other hand, they would require a theoretical framework that allows one to conceptualize the influence of gender and culture as well as their interplay in stimulating relational orientation and fostering rapport relevant behaviors. As suggested by Cross and Madson (1997) the construct of "self construal" (Markus and Kitayama, 1991) could provide such a framework. Central to the construct is the distinction between interdependent and interdependent self construals (Markus and Kitayama, 1991). This distinction refers to the way people see their self either as integral part of the group emphasizing harmony and unity or as an isolated entity striving for uniqueness and individual achievement. The construct has been primarily applied to cultural differences (Kitayama et al., 2007; Harb and Smith, 2008; Gore et al., 2009; Cross et al., 2011). Yet, a few studies also successfully applied it to the understanding of gender differences (Cross and Madson, 1997; Watkins et al., 2003). For instance, Cross and Madson (1997) found that the cultural environment even within a Western civilization still fosters independence and autonomy in men and interdependence and relatedness in women. They further hypothesize that such differences in the individuals' cognitive structure affect the micro-level of social interactions in the sense that "... individuals with an interdependent self-construal may develop skills and behaviors that facilitate the development of close relationships with others" (p. 17). While self construal research has predominantly focused on cognitive variables (e.g., Cross et al., 2002; Konrath et al., 2009), little is known about its influence on how people concretely establish rapport on the micro level of social interactions. We contend, that our understanding of the cognitive and behavioral mechanisms that foster rapport and enable or disable a smooth flow of communication could largely benefit from studies combining the assessment of self construals (cf., Singelis, 1994) with the introduced methods to analyze perceptions and behavioral correlates of rapport.

The explorative behavior analyses demonstrated the manifold possibilities to extract nonverbal features from the motion capture data and to quantify respective dyadic coordination patterns. Some of the tentative results might also inspire future studies beyond mere method demonstration. For instance, RWTLCC revealed some interesting correlations with rapport ratings. The mean peak correlation of the head rotation behavior of both partners (i.e., the highest correlation across the time lags ($\pm$2 s) found for each data point averaged over the whole 1 min sequence) correlated negatively with perceived rapport. This indicates, that the more similar the orientation dynamics of both partners within a critical time window was (i.e., both turning way from or toward the partner), the lower was the level of perceived rapport. However, the same RWTLCC parameter (mean peak correlation) for the behavioral dimension "approach and distancing behavior" of the interactors was positively correlated with rapport (i.e., the more similar the distancing and approach motions, the higher the rapport ratings). A further significant result occurred for the entrainment of the general movement activity. Here we found a negative correlation between the average offset of the peak correlation and the rapport ratings, indicating that the closer the correlation peaks to the zero-lag point (simultaneity), the higher were the rapport ratings. Overall, these results shed some critical light on previous conceptions of the role of synchrony and mimicry for rapport building (Chartrand and Bargh, 1999; Lakin and Chartrand, 2003; Miles et al., 2009; Lakens and Stel, 2011), and point to the necessity to treat the various behavioral subsystems differently with regard to the type and the level of coordination that is functional for rapport.

LIMITATIONS AND FUTURE PROSPECTS

The major limitation of the current study is the small number of interaction stimuli that only consisted of 20 dyads with 5 dyads for each gender and culture combination. It is important to reiterate, though, that the primary purpose of this study was to demonstrate the benefits of the new methodology. The possibilities of the proposed methodology reach far beyond the study of rapport and gender and cultural differences. In fact, the elimination of stereotype influence on the observer side as well as the access to objective behavioral measures provide a more universal solution for observation studies in nonverbal research, including traditional domains, such as impression formation, person perception, deception, and emotion recognition. The quality of the obtained data sets would ideally fill both sides of prominent effect models in nonverbal communication research such as Brunswik's lens model (Brunswik, 1956; Bernieri et al., 1996). Future theory driven studies following this approach certainly would require larger stimulus data sets.

This relates to a further limitation of the study. It concerns the use of a marker-based motion capture device that requires laboratory setup and larger amounts of time to equip and calibrate the participants. At this time, motion capturing still provides the most accurate method to overcome the described measurement issues in nonverbal behavior research and the fact that such systems are now rather affordable and easy to use is likely going to facilitate more widespread adoption in research and practice. However, broadly accessible machine learning tools can be expected to replace motion capture in the near future, allowing one to extract skeletal motion data from standard video (Cao et al., 2018). As the resulting data protocols (joint translations and rotations) will be compatible with motion capture data, parameters and analysis tools developed and applied to motion capture data will retain their validity.

A third limitation concerns the selective choice of behavioral parameters and algorithms to quantify patterns of coordination across different nonverbal subsystems. For demonstration purposes we here focused on a subset of the manifold possibilities to quantify behavioral interdependencies in interactions. Further algorithms and software tools can be found in Delaherche et al. (2012) and Varni et al. (2015). Promising approaches to analyze the temporal entrainment of two behavioral vectors in the frequency domain have also been introduced recently using Cross-Wavelet-Transform (Fujiwara and Daibo, 2016). It is important to note that the different algorithms are not just different ways to capture the same phenomenon, but distinctly define what is conceived as coordination or synchrony (Novotny, 2020). Therefore, further studies are needed to comparatively evaluate the different approaches with regard to their effectiveness in predicting the subjective experience or perceptions of rapport in varying contexts (cf. Bente and Novotny, 2020).

Lastly, the current study shows a limitation with regard to the measurement of perceived rapport. One might argue that it would be more important to assess the interactors' experience of rapport rather than the impressions of neutral observers. In fact, there might be fundamental differences between a first-person perspective and a third person perspective on dyadic rapport (Bernieri and Gillis, 1995a, cf. Schilbach et al., 2013). The current study focused on the third person perspective as the major objective was to demonstrate the potential of the novel methods to eliminate stereotype effects and judgment biases specifically in observer studies. It is an interesting question for future research though, why self-reports and observer judgments drift apart. Bernieri and Gillis (1995a) supposed that this might be due to the fact observers refer to socially appealing cues (such as smiles) that are less relevant for rapport and oversee more relevant ones (such as nonverbal coordination). In the current study, facial expressions were not shown in the character animation stimuli and observers could base their judgments solely on the dyadic movement patterns. Whether this might have led to a higher agreement with the interactors' self-rating remains an open question to be answered in upcoming studies. Another reason for discrepancies in rapport judgments might be that observers see the dyad as a whole whereas interactors see it from the individual perspective. Observers thus might be closer to the reality of the emergent dyadic construct of rapport than the interactors, whose evaluations can be influenced by other components of person perception such as for instance liking and control (cf. Human et al., 2013). This implies that interactors can also show discrepancies in their rapport judgments, which makes it generally questionable how these impressions can be used to validate observer judgments. Against this background it might make more sense to treat first and third person judgments of rapport as distinct, yet both relevant, information. It is important to note that impression management is not limited to preserve the individual's face but also to create a positive impression of the group as a whole in front of an "audience" (cf. Goffman, 1959; Manning, 2005; Picone, 2015). The third person perspective thus might add relevant data for our understanding of nonverbal rapport mechanisms.

DATA AVAILABILITY STATEMENT

Raw data and stimulus materials developed in this project will be made available by the authors, without undue reservation, to any qualified request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB of the American University of Sharjah (UAE) and the Department of Psychology at the University of Cologne (Germany).

AUTHOR CONTRIBUTIONS

GB and AA developed the study design and organized the data collection in the UAE and Germany. GB and DR conducted the observation studies in both countries. GB and EN analyzed the behavioral data and wrote the first manuscript draft. All authors contributed to the article and approved the submitted version.

FUNDING

Data collection was supported by a grant from the Volkswagen Foundation (VolkswagenStiftung, Germany). Method development, data analysis and publication were supported by a grant from the National Science Foundation (No. 1907807).

ACKNOWLEDGMENTS

We thank our former research assistants Thomas Dratsch, Bjoern Guenther, Lena Frischlich, and Diana Rieger for their valuable support in collecting the interaction data for stimulus creation.

REFERENCES

Ambady, N., and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol. Bull.* 111, 256–274. doi: 10.1037/0033-2909.111.2.256

Autodesk (2017). *Autodesk Motionbuilder [Computer Software]*. Available online at: http://www.autodesk.com (accessed 2017). .

Bente, G. (2019). "New tools – new insights: using emergent technologies in nonverbal communication research," in *Reflections on Interpersonal Communication*, eds S. W. Wilson and S. W. Smith (San Diego, CA: Cognella), 161–188.

Bente, G., Aelker, L., and Fürtjes, M. (2009). "Tracing ad experience. Real-time response measurement in advertising research," in *Real-Time Response Measurement in the Social Sciences: Methodological Perspectives and Applications*, eds J. Maier, M. Maier, M. Mauer, C. Reinemann, and V. Meyer (Frankfurt: Peter Lang), 117–135.

Bente, G., Leuschner, H., Al Issa, A., and Blascovich, J. J. (2010). The others: universals and cultural specificities in the perception of status and dominance from nonverbal behavior. *Conscious. Cogn.* 19, 762–777.doi: 10.1016/j.concog.2010.06.006

Bente, G., and Novotny, E. (2020). "Bodies and minds in sync: forms and functions of interpersonal synchrony in human interaction," in *The Handbook of Communication Science and Biology*, eds K. Floyd and R. Weber (New York, NY: Routledge), 416–428. doi: 10.4324/9781351235587-36

Bernieri, F., and Gillis, J. S. (1995a). The judgment of rapport: a cross-cultural comparison between Americans and Greeks. *J. Nonverbal Behav.* 19, 115–130.doi: 10.1007/BF02173170

Bernieri, F. J. (1988). Coordinated movement and rapport in teacher-student interactions. *J. Nonverbal Behav.* 12, 120–138. doi: 10.1007/BF00986930

Bernieri, F. J., Davis, J. M., Rosenthal, R., and Knee, C. R. (1994). Interactional synchrony and rapport: measuring synchrony in displays devoid of sound and facial affect. *Pers. Soc. Psychol. Bull.* 20, 303–311.doi: 10.1177/0146167294203008

Bernieri, F. J., and Gillis, J. S. (1995b). Personality correlates of accuracy in a social perception task. *Percept. Mot. Skills* 81, 168–170.doi: 10.2466/pms.1995.81.1.168

Bernieri, F. J., and Gillis, J. S. (2001). "Judging rapport: employing Brunswik's lens model to study interpersonal sensitivity," in *Interpersonal Sensitivity: Theory and Measurement*, eds J. A. Hall and F. J. Bernieri (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 67–86. doi: 10.4135/9781071800676.n7

Bernieri, F. J., Gillis, J. S., Davis, J. M., and Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: a lens model analysis. *J. Pers. Soc. Psychol.* 71, 110–129.doi: 10.1037/0022-3514.71.1.110

Bernieri, F. J., Reznick, S., and Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions. *J. Pers. Soc. Psychol.* 54, 243–253.doi: 10.1037/0022-3514.54.2.243

Berry, D. S., Kean, K. J., Misovich, S. J., and Baron, R. M. (1991). Quantized displays of human movement: a methodological alternative to the point-light display. *J. Nonverbal Behav.* 15, 81–97.doi: 10.1007/BF00998264

Berry, D. S., Misovich, S. J., Kean, K. J., and Baron, R. M. (1992). Effects of disruption of structure and motion on perceptions of social causality. *Pers. Soc. Psychol. Bull.* 18, 237–244.doi: 10.1177/0146167292182016

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley, CA: University of California Press.

Burns, M. (1984). Rapport and relationships: the basis of child care. *J. Child Care* 2, 47–57.

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., and Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *arXiv* [Preprint]. arXiv:1812.08008,

Cappella, J. N. (1990). On defining conversational coordination and rapport. *Psychol. Inquiry* 1, 303–305.doi: 10.1207/s15327965pli0104_5

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, H., Ballal, T., and Al-Naffouri, T. (2019). "A decomposition approach for complex gesture recognition using DTW and prefix tree," in *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (Osaka: IEEE), 876–877.

Cheong, J. H. (2019). *Four Ways to Quantify Synchrony Between Time Series Data. Towards Data Science*. Available online at: https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9 (accessed March 2020).

Cooper, R. A., and Tauber, A. I. (2005). New physicians for a new century. *Acad. Med.* 80, 1086–1088. doi: 10.1097/00001888-200512000-00003

Cornejo, C., Hurtado, E., Cuadros, Z., Torres, A., Paredes, J., Olivares, H., et al. (2018). Dynamics of simultaneous and imitative bodily coordination in trust and distrust. *Front. Psychol.* 9:1546. doi: 10.3389/fpsyg.2018.01546

Cross, S. E., Hardin, E. H., and Swing, B. G. (2011). The what, how, why, and where of self-construal. *Pers. Soc. Psychol. Rev.* 15, 142–179. doi: 10.1177/1088868310373752

Cross, S. E., and Madson, L. (1997). Models of the self: self-construals and gender. *Psychol. Bull.* 122, 5–37.doi: 10.1037/0033-2909.122.1.5

Cross, S. E., Morris, M. L., and Gore, J. S. (2002). Thinking about oneself and others: the relational-interdependent self-construal and social cognition. *J. Pers. Soc. Psychol.* 82, 399–418.doi: 10.1037/0022-3514.82.3.399

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., Leyens, J. P., et al. (2009). Stereotype content model across cultures: towards universal similarities and some differences. *Br. J. Soc. Psychol.* 48(Pt 1), 1–33.doi: 10.1348/014466608X314935

Cuddy, A. J., Wolf, E. B., Glick, P., Crotty, S., Chong, J., and Norton, M. I. (2015). Men as cultural ideals: cultural values moderate gender stereotype content. *J. Pers. Soc. Psychol.* 109, 622–635.doi: 10.1037/pspi0000027

Cutting, J. E., and Proffitt, D. R. (1981). "Gait perception as an example of how we may perceive events," in *Intersensory Perception and Sensory Integration*, eds R. Walk and H. Pick (Boston, MA: Springer), 249–273. doi: 10.1007/978-1-4615-9197-9_8

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., and Cohen, D. (2012). Interpersonal synchrony: a survey of evaluation methods across disciplines. *IEEE Trans. Affect. Comput.* 3, 349–365.doi: 10.1109/T-AFFC.2012.12

Dion, K. K., Pak, A. W. P., and Dion, K. L. (1990). Stereotyping physical attractiveness: a sociocultural perspective. *J. Cross Cult. Psychol.* 21, 158–179. doi: 10.1177/0022022190212002

Drolet, A. L., and Morris, M. W. (2000). Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *J. Exp. Soc. Psychol.* 36, 26–50.doi: 10.1006/jesp.1999.1395

Ellemers, N. (2018). Gender stereotypes. *Annu. Rev. Psychol.* 69, 275–298. doi: 10.1146/annurev-psych-122216-011719

Fischer, R. (2004). Standardization to account for cross-cultural response bias: a classification of score adjustment procedures and review of research in JCCP. *J. Cross Cult. Psychol.* 35, 263–282.doi: 10.1177/0022022104264122

Fujiwara, K., and Daibo, I. (2016). Evaluating interpersonal synchrony: wavelet transform toward an unstructured conversation. *Front. Psychol.* 7:516. doi: 10.3389/fpsyg.2016.00516

Gillis, J. S., Bernieri, F. J., and Wooten, E. (1995). The effects of stimulus medium and feedback on the judgment of rapport. *Organ. Behav. Hum. Decis. Process.* 63, 33–45.doi: 10.1006/obhd.1995.1059

Gore, J. S., Cross, S. E., and Kanagawa, C. (2009). Acting in our interests: relational self-construal and goal motivation across cultures. *Motiv. Emot.* 33, 75–87.doi: 10.1007/s11031-008-9113-1

Grahe, J. E., and Bernieri, F. J. (1999). The importance of nonverbal cues in judging rapport. *J. Nonverbal Behav.* 23, 253–269.

Granitz, N. A., Koernig, S. K., and Harich, K. R. (2009). Now it's personal: antecedents and outcomes of rapport between business faculty and their students. *J. Mark. Educ.* 31, 52–65.doi: 10.1177/0273475308326408

Gremler, D. D., and Gwinner, K. P. (1998). Connecting with customers: an examination of rapport in service industries. *Am. Mark. Assoc. Conf. Proc.* 9, 161–162.

Gremler, D. D., and Gwinner, K. P. (2000). Customer-employee rapport in service relationships. *J. Serv. Res.* 3, 82–104.doi: 10.1177/109467050031006

Hall, J. A., and Briton, N. (1993). "Gender, nonverbal behavior, and expectations," in *Interpersonal Expectations: Theory, Research, and Applications*, ed. P. D. Blanck (Cambridge: Cambridge University Press), 276–295. doi: 10.1017/cbo9780511527708.015

Hall, J. A., Harrigan, J. A., and Rosenthal, R. (1995). Nonverbal behavior in clinician—patient interaction. *Appl. Prev. Psychol.* 4, 21–37.doi: 10.1016/S0962-1849(05)80049-6

Harb, C., and Smith, P. B. (2008). Self-construals across cultures: beyond independence-interdependence. *J. Cross Cult. Psychol.* 39, 178–197.doi: 10.1177/0022022107313861

Human, L. J., Sandstrom, G. M., Biesanz, J. C., and Dunn, E. W. (2013). Accurate first impressions leave a lasting impression: the long-term effects of distinctive self-other agreement on relationship development. *Soc. Psychol. Personal. Sci.* 4, 395–402.doi: 10.1177/1948550612463735

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211.doi: 10.3758/BF03212378

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychol. Res.* 38, 379–393. doi: 10.1007/BF00309043

Kitayama, S., Duffy, S., and Uchida, Y. (2007). "Self as cultural mode of being," in *Handbook of Cultural Psychology*, eds S. Kitayama and D. Cohen (New York, NY: Guilford), 136–174.

Konrath, S., Bushman, B. J., and Grove, T. (2009). Seeing my world in a million little pieces: narcissism, selfconstrual, and cognitive–perceptual style. *J. Pers.* 77, 1197–1228.doi: 10.1111/j.1467-6494.2009.00579.x

Lakens, D., and Stel, M. (2011). If they move in sync, they must feel in sync: movement synchrony leads to attributions of rapport and entitativity. *Soc. Cogn.* 29, 1–14.doi: 10.1521/soco.2011.29.1.1

Lakin, J. L., and Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychol. Sci.* 14, 334–339.doi: 10.1111/1467-9280.14481

Leuschner, H. (2013). *Microanalysis of Nonverbal Communication: Development of a Nonverbal Research Method Using High-Performance 3D Character Animation*. Doctoral dissertation, Universität zu Köln, Köln.

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., and Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: implications for the dynamics of eye-gaze behavior. *Perception* 37, 1443–1457.doi: 10.1068/p5896

Macintosh, G. (2009). Examining the antecedents of trust and rapport in services: discovering new interrelationships. *J. Retail. Consum. Serv.* 16, 298–305.doi: 10.1016/j.jretconser.2009.02.001

Manning, P. (2005). "Impression management," in *Encyclopedia of Social Theory*, ed. G. Ritzer (Thousand Oaks, CA: Sage), 397–399.

Markus, H. R., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253.doi: 10.1037/0033-295X.98.2.224

Miles, L. K., Nind, L. K., and Macrae, C. N. (2009). The rhythm of rapport: interpersonal synchrony and social perception. *J. Exp. Soc. Psychol.* 45, 585–589.doi: 10.1016/j.jesp.2009.02.002

Murphy, M., and Valdéz, C. (2005). Ravaging resistance: a model for building rapport in a collaborative learning classroom. *Radic. Pedagogy* 7:130.

Nelson, A. A., Grahe, J. E., and Ramseyer, F. (2016). Interacting in flow: an analysis of rapport-based behavior as optimal experience. *SAGE Open* 6:2158244016684173. doi: 10.1177/2158244016684173

Nguyen, H. T. (2007). Rapport building in language instruction: a microanalysis of the multiple resources in teacher talk. *Lang. Educ.* 21, 284–303.doi: 10.2167/le658.0

Novotny, E. (2020). *Identifying Signatures of Perceived Interpersonal Synchrony*. Doctoral dissertation, Michigan State University, East Lansing, MI.

Optitrack (2017). *Optitrack [Motion Capture Hardware]*. Available online at: http://optitrack.com (accessed 2017).

Picone, I. (2015). "Impression management in social media," in *The International Encyclopedia of Digital Communication and Society*, eds R. Mansell and P. H. Ang (Oxford: Wiley-Blackwell), 469–476.

Poppe, R., Van Der Zee, S., Heylen, D. K., and Taylor, P. J. (2014). AMAB: automated measurement and analysis of body motion. *Behav. Res. Methods* 46, 625–633. doi: 10.3758/s13428-013-0398-y

Puccinelli, N. M., Tickle-Degnen, L., and Rosenthal, R. (2003). Effect of dyadic context on judgments of rapport: dyad task and partner presence. *J. Nonverbal Behav.* 27, 211–236.

Ramseyer, F., and Tschacher, W. (2010). "Nonverbal synchrony or random coincidence? How to tell the difference," in *Development of Multimodal Interfaces: Active Listening and Synchrony*, eds A. Esposito, N. Campbell, C. Vogel, A. Hussain, and A. Nijholt (Berlin: Springer), 182–196. doi: 10.1007/978-3-642-12397-9_15

Ramseyer, F., and Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *J. Consult. Clin. Psychol.* 79, 284–295.doi: 10.1037/a0023419

Ramseyer, F., and Tschacher, W. (2014). Nonverbal synchrony of head-and body-movement in psychotherapy: different signals have different associations with outcome. *Front. Psychol.* 5:979. doi: 10.3389/fpsyg.2014.00979

Runeson, S., and Frykholm, G. (1981). Visual perception of lifted weight. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 733–740.doi: 10.1037/0096-1523.7.4.733

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–462.doi: 10.1017/S0140525X12000660

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Pers. Soc. Psychol. Bull.* 20, 580–591.doi: 10.1177/0146167294205014

Stangor, C., and Crandall, C. S. (2013). *Stereotyping and Prejudice*. Brandon, VT: Psychology Press.

Sundstrom, E., and Altman, I. (1976). Interpersonal relationships and personal space: research review and theoretical model. *Hum. Ecol.* 4, 47–67.doi: 10.1007/BF01531456

Ten Holt, G. A., Reinders, M. J., and Hendriks, E. A. (2007). "Multi-dimensional dynamic time warping for gesture recognition," in. *Proceedings of the 13th Annual Conference of the Advanced School for Computing and Imaging*, Heijen Vol. 300, 1.

Tickle-Degnen, L. (2006). "Nonverbal behavior and its functions in the ecosystem of rapport," in *The SAGE Handbook of Nonverbal Communication*, eds V. Manusov and M. Patterson (Thousand Oaks, CA: Sage), 381–401. doi: 10.4135/9781412976152.n20

Tickle-Degnen, L., and Rosenthal, R. (1987). Group rapport and nonverbal behavior. *Rev. Pers. Social Psychol.* 9, 113–136.

Tickle-Degnen, L., and Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychol. Inquiry* 1, 285–293.doi: 10.1207/s15327965pli0104_1

Triandis, H. C. (1995). *Individualism and Collectivism*. Boulder, CO: Westview Press.

Van de Vijver, F. J. R. (2007). Cultural and gender differences in gender-role beliefs, sharing household task and child-care responsibilities, and well-being among immigrant and majority members in the Netherlands. *Sex Roles* 57, 813–824.doi: 10.1007/s11199-007-9316-z

Varni, G., Avril, M., Usta, A., and Chetouani, M. (2015). "Syncpy: a unified open-source analytic library for synchrony," in *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, Seattle,WA, 41–47.

Watkins, D., Cheng, C., Mpofu, E., Olowu, S., Singh-Sengupta, S., and Regmi, M. (2003). Gender differences in self-construal: how generalizable are Western findings? *J. Soc. Psychol.* 143, 501–519. doi: 10.1080/00224540309598459

Watson, O. M. (1970). *Proxemic Behavior: A Cross-Cultural Study*. Hague: Mouton.

Williams, J. E., and Best, D. L. (1994). "Cross-cultural views of women and men," in *Psychology and Culture*, eds W. J. Lonner and R. S. Malpass (Boston, MA: Allyn & Bacon), 191–196.

Conflict of Interest: 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.

CORRECTION
published: 15 April 2021
doi: 10.3389/fpsyg.2021.675148

Corrigendum: Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport

Gary Bente[1], Eric Novotny[2], Daniel Roth[3] and Ahmad Al-Issa[4]*

[1] Department of Communication, Michigan State University, East Lansing, MI, United States, [2] Grady College of Journalism and Mass Communication, University of Georgia, Athens, GA, United States, [3] Department of Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, Munich, Germany, [4] Department of English, American University of Sharjah, Sharjah, United Arab Emirates

Keywords: rapport, nonverbal behavior, motion capture, character animation, gender, culture

A Corrigendum on

Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport
by Bente, G., Novotny, E., Roth, D., and Al-Issa, A. (2020). Front. Psychol. 11:599703. doi: 10.3389/fpsyg.2020.599703

In the original article, there was an error.

In section *Method: Behavior Analysis, Parameter Formation, Paragraph 4,* we said: "From DTW we calculated the length of the optimal warp path as a measure of similarity of the behavior vectors when stretched or compressed in time." This is incorrect. We did not use "path length" but "DTW distance" for further statistical analyses.

A correction has been made to the section *Method: Behavior Analysis, Parameter Formation, Paragraph 4*. The corrected paragraph is shown below.

Pearson "r" and the entropy measure resulting from MI analysis were used as input for statistical analysis. From RWTLCC we calculated the average maximal correlation at each point across all time lags as well the absolute offset of the correlation peak from the zero lag as general coherence and synchrony indicators. We further used the "DTW distance" measure, i.e., the minimum path cost (Cheong, 2019), to quantify the (dis)similarity between the behavioral vectors of the interactants for further analyses.

Additionally, there was also an error in *Method: Observer Study, Stimulus Material, Paragraph 2*. In the phrase "Cameras were around a square area of 4 × 4 meters," the word "positioned" is missing.

A correction has been made to the section *Method: Observer Study, Stimulus Material, Paragraph 2*. The corrected paragraph is shown below.

Participants were instructed that they would have a short 5–7 min conversation with another student during which they should get to know each other. Before the conversations began, participants were led into different rooms to put on the datasuits necessary for motion capturing. A same sex student assistant placed the markers on the data suits and guided the interactors to the middle of the recording room where they met the experimenter. Motion capture was performed with a 12-camera Optitrack system and the capture software Arena (Optitrack, 2017). Cameras were positioned around a square area of 4 × 4 meters. Participants were then asked by the experimenter to take a T-pose (upright symmetric posture with legs closed and arms horizontally stretched out, palms down) for calibration of the tracking system. Then the participants were told that they could move freely in the square between the cameras and should use the next 5–7 min to get to know each other. Next, the experimenter left the room and the participants started the

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

****Correspondence:***
Eric Novotny
Eric.novotny@uga.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *02 March 2021*
Accepted: *03 March 2021*
Published: *15 April 2021*

Citation:
Bente G, Novotny E, Roth D and Al-Issa A (2021) Corrigendum: Beyond Stereotypes: Analyzing Gender and Cultural Differences in Nonverbal Rapport. Front. Psychol. 12:675148. doi: 10.3389/fpsyg.2021.675148

conversation. Using the capture software Arena, full body motion of both actors was captured during the conversation with a temporal resolution of 150 Hz. Figure 1 shows a dyad wearing the data-suits with the IR reflecting markers and a projection of the capture software showing both virtual characters for demonstration purposes. After completing the interaction, the participants were debriefed and received 15 Euro (Cologne), or an equivalent on-campus restaurant voucher (Sharjah) for their participation.

The authors apologize for these errors and state that they do not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Cheong, J. H. (2019). *Four Ways to Quantify Synchrony Between Time Series Data. Towards Data Science.* Available online at: https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9 (accessed March 2020).

Optitrack (2017). *Optitrack [Motion Capture Hardware].* Available online at: http://optitrack.com (accessed 2017).

frontiers in Psychology

ORIGINAL RESEARCH
published: 12 January 2021
doi: 10.3389/fpsyg.2020.612654

A Novel Test of the Duchenne Marker: Smiles After Botulinum Toxin Treatment for Crow's Feet Wrinkles

Nancy Etcoff[1†], Shannon Stock[2†], Eva G. Krumhuber[3] and Lawrence Ian Reed[4†]*

[1] *Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States,* [2] *Department of Mathematics and Computer Science, College of the Holy Cross, Worcester, MA, United States,* [3] *Department of Experimental Psychology, University College London, London, United Kingdom,* [4] *Department of Psychology, New York University, New York, NY, United States*

Smiles that vary in muscular configuration also vary in how they are perceived. Previous research suggests that "Duchenne smiles," indicated by the combined actions of the orbicularis oculi (cheek raiser) and the zygomaticus major muscles (lip corner puller), signal enjoyment. This research has compared perceptions of Duchenne smiles with non-Duchenne smiles among individuals voluntarily innervating or inhibiting the orbicularis oculi muscle. Here we used a novel set of highly controlled stimuli: photographs of patients taken before and after receiving botulinum toxin treatment for crow's feet lines that selectively paralyzed the lateral orbicularis oculi muscle and removed visible lateral eye wrinkles, to test perception of smiles. Smiles in which the orbicularis muscle was active (prior to treatment) were rated as more felt, spontaneous, intense, and happier. Post treatment patients looked younger, although not more attractive. We discuss the potential implications of these findings within the context of emotion science and clinical research on botulinum toxin.

Keywords: facial expression, Duchenne smile, botulinum toxin, emotion, attractiveness

OPEN ACCESS

Edited by:
Norah E. Dunbar,
University of California,
Santa Barbara, United States

Reviewed by:
Mollie Ruben,
University of Maine, United States
Farid Pazhoohi,
University of British Columbia,
Canada

***Correspondence:**
Lawrence Ian Reed
lr113@nyu.edu

†These authors share first authorship

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *30 September 2020*
Accepted: *08 December 2020*
Published: *12 January 2021*

Citation:
Etcoff N, Stock S, Krumhuber EG and Reed LI (2021) A Novel Test of the Duchenne Marker: Smiles After Botulinum Toxin Treatment for Crow's Feet Wrinkles. Front. Psychol. 11:612654. doi: 10.3389/fpsyg.2020.612654

INTRODUCTION

Among all communicative signals, smiles are one of the most easily recognizable facial expression, with a visual pattern that is detectable over long viewing distances (Smith and Schyns, 2009; Krumhuber et al., 2019). In morphological terms, they can be defined by the activation of the zygomaticus major muscle—a facial muscle which pulls the lip corners upwards and away from the mouth (Ekman, 1989). Despite their simplicity in appearance, smiles can occur in a range of situations and for a variety of reasons. Some smiles express positive emotions such as happiness or enjoyment; hence, they are considered spontaneous readouts of positive internal states (Ekman and Friesen, 1982). Other smiles are deliberately displayed in the absence of an underlying positive affect, for example, to signal politeness, affiliation, or feigned cooperation (Ekman, 1989; Rychlowska et al., 2017). Due to their voluntary nature, those latter expressions are typically described as posed, social, or polite smiles. What is the distinction between spontaneous and posed signals reflecting such disparate functions?

According to some researchers (Ekman, 1992; Frank and Ekman, 1993), the smile of enjoyment is reliably indicated by the contraction of the zygomaticus major muscle with the concurrent contraction of the orbicularis oculi pars lateralis, which is a circumferential muscle surrounding the eye. The latter draws skin toward the eye from the temple and the cheeks, thereby causing narrowing of the eye opening and wrinkles around the eye socket, colloquially called crow's feet. Activation of both muscles constitutes the so-called "Duchenne smile," named in homage to the neuroanatomist Duchenne de Boulogne who isolated the orbicularis oculi action and first posited its coherence with enjoyment (Boulogne, 1862/1990; Ekman et al., 1990). In the Facial Action Coding System (Ekman et al., 2002), a system for scoring visible facial movements, the appearance of the Duchenne smile is defined as Action Unit (AU) 6 (orbicularis oculi pars lateralis) coupled with AU12 (zygomaticus major).

The presence of supposedly involuntary eye constriction (AU6) has been proposed to signal happiness/enjoyment in smiles, whereas its absence termed as false, social, non-felt, or non-Duchenne smiles (Frank et al., 1993). While recent studies have revealed that Duchenne smiles occur not only as a spontaneous sign of positive affect and can be deliberately displayed (Schmidt et al., 2006; Krumhuber and Manstead, 2009; Gunnery et al., 2013), there is consistent evidence that Duchenne smiles contribute to perceptions of greater spontaneity and authenticity (Gunnery and Ruben, 2016). They are perceptually salient and perceived as more affectively intense (Malek et al., 2019; Miller et al., 2020), making the smiling person look happier, more amused, and in better humor (Scherer and Ceschi, 2000; Gosselin et al., 2002; Ambadar et al., 2009). Duchenne smiles also lead to favorable interpersonal perceptions (Harker and Keltner, 2001; Messinger et al., 2008), eliciting more positive and affiliative responses in other people (Keltner and Bonanno, 1997) and relieving the concerns of potentially cooperative partners (Reed et al., 2012). Finally, people expressing Duchenne smiles are rated as more likeable, attractive, and intelligent than those showing non-Duchenne smiles (Frank et al., 1993; Quadflieg and Rossion, 2013).

Previous research comparing perceptions of Duchenne and non-Duchenne smiles has been with subjects who had full control of the movements of their facial musculature. Some studies also employed image manipulation techniques by editing the eyes to add/remove visible signs of the Duchenne marker (Calvo et al., 2019; Namba et al., 2020). Here, we compare perceptions of Duchenne and non-Duchenne smiles of subjects before and after botulinum toxin treatment to the orbicularis oculi. Subjects were instructed to display a maximum smile at two time points. In the first, subjects had full control over their facial musculature. In the second, subjects had received botulinum toxin treatment to the orbicularis oculi in order to reduce or eliminate the appearance of crow's feet lines.

Botulinum toxin is a product of the Clostridium botulinum bacterium that disrupts vesicular exocytosis at neuromuscular junctions producing flaccid paralysis. Very small quantities of formulated pharmaceutical forms of botulinum toxin are injected directly into specific muscles to selectively inhibit activation of the injected muscle. This process is referred to as chemodenervation. In 2016, ASPS (the American Society for Plastic Surgery) reported that there were more than 7 million botulinum toxin procedures performed in the US (American Society of Plastic Surgeons, 2017). One common motivation for seeking botulinum toxin treatment is an individual's desire to look younger and more attractive. Crow's feet radiating from the lateral canthus (the outer corner of the eye where upper and lower eyelids meet) is caused by the contraction of fibers of the orbicularis oculi muscle. These wrinkles can appear during expression (dynamic lines) and can become permanent and static with age.

Studies examining the effects of panfacial aesthetic treatment (including botulinum toxin) on observers' perceptions of patients' facial characteristics have shown enhancements in physical appearance (Przylipiak et al., 2018). However, to our knowledge, there are no studies investigating these effects following the selective chemodenervation of the orbicularis oculi muscle. The aim of the current study is to examine perceptions of deliberately posed smiles displayed by patients before and after receiving botulinum toxin treatment for crow's feet lines. It was hypothesized that pre-treatment photographs with no inhibition of the Duchenne marker would be rated as more spontaneous, more intense, and happier than post-treatment photographs in which the lateral orbicularis oculi muscle has been selectively chemodenervated.

MATERIALS AND METHODS

Participants

Three hundred and ninety-three (185 female) participants from the United States were recruited via Amazon's Mechanical Turk (Buhrmeister et al., 2011) in exchange for monetary payment of $4.00. Participants' mean age was 54.25 years (*SD* = 10.31). The racial composition of the sample was 85.8% White, 8.4% African American, 4.1% Asian American, and 1.8% other. Ethical approval was granted by the Partners Human Research Committee, and subjects provided written informed consent prior to participation.

Stimulus Material

The facial stimuli featured high resolution, full color images of five adult men and 27 women who had participated in a clinical trial examining the efficacy of botulinum toxin for the treatment of crow's feet wrinkles (Carruthers et al., 2014). Each participant showed moderate or severe bilaterally symmetrical crow's feet lines as assessed with the Facial Wrinkle Scale prior to treatment. Chemodenervation was applied by injecting botulinum toxin directly into the lateral orbicularis oculi muscle tissue to prevent muscular contraction in that focal area. All individuals (referred thereafter as patients) were asked to produce a "maximum" smile before and after (approximately 1 month) receiving the botulinum toxin injection. Specifically, they were told the following: "You should smile to show your biggest natural smile; you should not force the smile. You may smile with your lips parted or with your lips together, whichever feels more natural." Each photograph was taken at an oblique (three-quarter viewing) angle using a standardized photographic apparatus with consistent lighting (see **Figures 1, 2**).

For all pre- and post-treatment images (64 images), a FACS certified coder who was blind to the treatment condition scored

FIGURE 1 | Participant before receiving botulinum toxin injection.

FIGURE 2 | Participant after (approximately 1 month) receiving botulinum toxin injection.

the presence of AU6 + 12 (the Duchenne marker) as well as AU12 intensity. Intercoder reliability was checked by a second FACS certified coder for 25% of the stimulus material (16 images). Mean agreement for the presence of AU 6 (Cohen's Kappa (κ) = 0.875) and AU 12 (κ = 1.00) was high. All patients posing "maximum" smiles displayed AU12 activity before and after treatment. A Wilcoxon signed-rank test revealed no significant difference in AU12 intensity between pre-treatment (M = 3.19, SD = 0.69) and post-treatment images (M = 3.22, SD = 0.66), Z = 0.58, p = 0.564. Hence, images were well matched for smile intensity. When analyzing AU6 activity, the majority of participants (97%) were able to voluntarily contract the orbicularis oculi muscle prior to treatment. This fell to 19% after chemodenervation, which is a significant drop in the proportion from pre- to post-treatment (McNemar's test, $p < 0.001$). As such, pre- and post-treatment photographs predominantly consisted of Duchenne (AU6 + 12) and non-Duchenne smiles (AU12), respectively.

Procedure

Participants were tested individually using Qualtrics, a web-based software. After providing informed consent they were instructed that they would view a series of images of facial expressions. Their task was to rate each expression on several dimensions. Participants then viewed the 64 images (32 pre-treatment and 32 post-treatment) in a randomized order in three blocks. Each block showed the same photograph on screen for a duration of 10 s. The first block always started with the emotion rating scales. Here, participants responded to the query: "How much of the following emotions was the person feeling?" (happiness, sadness, anger, disgust, fear, surprise, and embarrassment). Ratings were made on 8-point Likert scales, with response options ranging from 0 (not at all/none) to 8 (extremely/a great deal).

In the second block, participants evaluated the smile quality of the expression by responding to the queries: "How felt was the expression of the person?" "How spontaneous was the expression of the person?" and "How intense was the expression of the person?" Ratings were made on 10-point Likert scales, with response options ranging from 0 (not at all/none) to 10 (extremely/a great deal). In the final block, participants rated the attractiveness of the person (0-not at all/none to 10-extremely/a great deal) and guessed the person's age (from 19 and 99 years).

Participants completed the entire task in approximately 35 min. Although the experimental design allowed all 393 participants to rate each of the 32 pre-treatment and post-treatment image pairs, many stopped before viewing all images. On average, participants rated 7.86 pre-treatment and post-treatment photos, resulting in 3,090 observations at each time point.

RESULTS

To analyze the effect of treatment on each of these ratings, we used linear mixed-effects models with crossed random effects for patient and participant, and a fixed effect for treatment. Please see Baayen et al. (2008) for a detailed description of this modeling approach and Etcoff et al. (2011) for an example within the psychology literature on face perception.

The results of the linear mixed-effects models are summarized in **Table 1**. For emotion ratings, treatment was associated with a mean decrease of 0.11 in happiness ratings and a mean decrease of 0.07 in embarrassment ratings (p = 0.01 and p = 0.03 for treatment effects on happiness and embarrassment, respectively). There were no significant associations between treatment and any other measured emotions (sadness, anger, disgust, fear, or surprise).

Treatment was also associated with a statistically significant decrease in ratings of smile quality. Specifically, post-treatment photographs were associated with a mean decrease of 0.30 in felt ratings, 0.25 in spontaneity ratings, and 0.24 in intensity ratings (all treatment $p < 0.001$).

Treatment did not have a significant effect on ratings of attractiveness, although post-treatment photographs were associated with a mean age approximately 1 year younger than pre-treatment photographs ($p < 0.001$).

TABLE 1 | Mixed-effects models evaluating the effect of botulinum toxin treatment on ratings of émotions, smile, quality, attractiveness, and age.

Outcome	Covariate	Estimate	*SE*	*df*	*t*-stat	*p*
Happy	Intercept	4.441	0.14	33.6	31.51	<0.0001
	Treatment (post vs. pre)	−0.111	0.05	5,757	−2.47	0.01
Sad	Intercept	0.482	0.04	42.3	12.64	<0.0001
	Treatment (post vs. pre)	−0.035	0.03	5,764	−1.37	0.17
Anger	Intercept	0.307	0.03	43.6	10.10	<0.0001
	Treatment (post vs. pre)	0.016	0.02	5,778	0.75	0.45
Disgust	Intercept	0.357	0.04	41	9.90	<0.0001
	Treatment (post vs. pre)	0.026	0.02	5,782	1.14	0.25
Fear	Intercept	0.442	0.04	39.6	10.36	<0.0001
	Treatment (post vs. pre)	0.001	0.03	5,766	0.03	0.98
Surprise	Intercept	1.087	0.07	40.7	16.48	<0.0001
	Treatment (post vs. pre)	−0.004	0.04	5,775	−0.09	0.93
Embarrassment	Intercept	0.860	0.04	55.1	21.46	<0.0001
	Treatment (post vs. pre)	−0.074	0.03	5,751	−2.20	0.03
Felt	Intercept	5.320	0.15	34.3	36.47	<0.0001
	Treatment (post vs. pre)	−0.302	0.05	5,753	−5.79	<0.0001
Spontaneity	Intercept	4.014	0.12	37.9	34.11	<0.0001
	Treatment (post vs. pre)	−0.249	0.06	5,753	−4.06	<0.0001
Intense	Intercept	4.374	0.18	33.7	24.61	<0.0001
	Treatment (post vs. pre)	−0.239	0.06	5,757	−4.16	<0.0001
Attract	Intercept	4.149	0.15	34.5	28.15	<0.0001
	Treatment (post vs. pre)	−0.039	0.05	5,770	−0.74	0.46
Age	Intercept	42.836	0.91	32.1	46.84	<0.0001
	Treatment (post vs. pre)	−0.937	0.19	5,758	−5.03	<0.0001

None of the statistically significant treatment effects were confounded by patient age, patient sex, participant age, or participant sex (see **Supplementary Table 1**). However, we did observe significant effects of patient sex on ratings of emotions and smile quality. Specifically, compared to males, photographs of female patients were associated with mean ratings 0.22 lower for sadness, 0.44 higher for surprise, 0.95 higher for felt, 0.73 higher for spontaneity, and 1.29 higher for intensity (all $p \leq 0.02$). Ratings of female patients were also associated with a mean age approximately 3.7 years younger than males ($p = 0.03$). Among ratings of attractiveness, female participants rated images a mean of 0.23 higher than male participants ($p < 0.001$).

In a sensitivity analysis we reevaluated treatment effects after eliminating one patient without AU6 in their pre-treatment photograph and six patients with AU6 in their post-treatment photograph, which resulted in no substantive changes in our results (see **Supplementary Table 2**). Since our sample was imbalanced with respect to patient sex (27 females vs. 5 males), we also repeated our analyses among the subset of female patients. The only appreciable difference in these results was that treatment was no longer associated with a statistically significant effect on ratings of embarrassment (see **Supplementary Table 3**).

DISCUSSION

Virtually all pre-treatment photographs depicted patients displaying Duchenne smiles. These smiles were rated as being happier, more felt, more spontaneous, and more intense than those posed by the same patients under the same conditions and instructions in post-treatment photographs. The photographs were matched for smile intensity (activity of the zygomatic major muscle pulling the lip corner) suggesting that the differences were due to the inhibition of the orbicularis oculi and not by the activity of zygomatic major.

Patients were also rated as being significantly younger after treatment (by approximately 1 year) likely due to less visible crow's feet lines. There was no effect of treatment on facial attractiveness ratings. Although some have speculated that a more spontaneous smile would make a face more attractive, past research (Mehu et al., 2007b) also found no difference in ratings of attractiveness for faces displaying Duchenne vs. non-Duchenne smiles. As such, attractiveness might be more dependent on the face structure and skin health than on dynamic features.

Interestingly, we found that ratings of smile quality were dependent on sex. The smiles of female patients were rated as more felt, surprised, spontaneous, and intense as well as less sad. This is consistent with data suggesting that women are more expressive for positive valanced facial actions (McDuff et al., 2017).

These data are consistent with results of previous studies demonstrating that Duchenne smiles are perceived differently than non-Duchenne smiles (Hess and Kleck, 1994; Del Giudice and Colle, 2007; Krumhuber and Manstead, 2009;

Mehu et al., 2012; Gunnery et al., 2013). Patients in the pre-treatment photographs—consisting almost exclusively of Duchenne smiles—were perceived as feeling more genuine positive emotion in comparison to post-treatment photographs. These data are also consistent with studies reporting high frequencies of Duchenne smiles in deliberate facial action tasks (Kanade et al., 2000; Krumhuber et al., 2020). Together, these findings suggest that although the Duchenne marker can be posed in the absence of positive affect, it is still perceived by others to be indicative of genuine emotion. Future research may benefit from examining potential limitations in the production or inhibition of the Duchenne marker in facial action tasks. Such work could shed new light on how different elicitation conditions might drive the reliability of this signal (McCullough and Reed, 2016; see also Zloteanu et al., 2020 in the context of surprise expressions).

Our results have several potential implications and caveats. Our study did not support the strongest version of the Duchenne hypothesis—that inhibition of the orbicularis oculi would make the smile signal appear unfelt or weak. Non-Duchenne smiles were rated as less happy, genuine, felt, and spontaneous, though our small treatment effects suggest that the effect was subtle. However, our stimulus patients were instructed to pose "maximum smiles" (maximum zygomatic activation). It may be that more pronounced effects on smile authenticity occur with less intense smiles. In general, these small but statistically significant changes could have practical implications in natural contexts where smiles may be less intense and/or the Duchenne marker may be more conspicuous.

Previous research has shown that the Duchenne marker plays a role in communicating cooperative intent (Mehu et al., 2007a,b; Reed et al., 2012) as well as eliciting cooperation from others (Scharlemman et al., 2001; Brown and Moore, 2002). In light of the results of the current study, it is possible that when the Duchenne marker is absent (in this study through chemodenervation that inhibited the orbicularis oculi muscle and erased visible crow's feet wrinkles) signals of cooperation may be lessened. If so, augmenting other signals of positive affect such as vocal affect or body language may counter the effects. Future studies can test this idea.

The images used in this study were derived from a clinical trial evaluating the efficacy of botulinum toxin on crow's feet lines. In order to test our hypothesis, we selected a subset of patients who showed no evidence of crow's feet lines using the Facial Wrinkle Scale post treatment. The majority of patients in the trial (66%), while having clinical improvement, did not have complete elimination of their dynamic lines. Interestingly, the patients in this trial where dynamic lines were eliminated reported feeling more satisfied with their appearance after treatment than those in whom some movement was preserved (unpublished data, Allergan). This suggests a potential disconnect between the positive perception of the aesthetic outcome on the part of the patient and the subtle negative impact on emotion communication as perceived by the observers. While not within the scope of this paper, this tension warrants further exploration. As reported, perceived smile authenticity did not impact attractiveness ratings, and did make patients appear approximately a year younger.

Three specific limitations must be taken into account when interpreting our findings. First, participants rated static images as opposed to video clips. Video clips have been shown to provide richer emotional content in comparison to static images (Ambadar et al., 2005; see Krumhuber et al., 2013, for a review) and would allow for the analysis of timing characteristics of facial expressions (Ambadar et al., 2009). Second, our sample did not include pre- and post-treatment photographs of spontaneously occurring smiles. That is, we were able to test our primary hypotheses using only deliberate facial action tasks and not when patients were experiencing genuine positive emotion (see Namba et al., 2020, for a similar approach). Future research addressing these limitations could complement the present findings and broaden our understanding of the perceptual and behavioral effects of the Duchenne smile. Third, our study was done solely with participants in the United States. It may be that participants in other cultures will be more impacted by the absence of the Duchenne marker. For example, Yuki et al. (2007) found cultural difference in the use of the eyes and mouth as cues to emotion in Japan and the U.S., with participants in Japan relying more heavily on eye expression for determination of emotion, including happiness, and participants in the U.S. on the mouth.

The Duchenne smile was first reported in 1862 by Duchenne de Boulogne in his "Mechanisme de la Physiognomie Humane." Duchenne isolated facial muscle action using the novel method of electrical contraction of its muscles. These were the first physiological experiments illustrated by photography. Over 150 years later, we used a pharmacological technique to selectively chemodenervate, and therefore isolate specific facial muscles. In doing so, we shed further light on Duchenne's pioneering ideas and address current controversies. We find evidence that Duchenne smiles communicate genuine and more intense happiness and that complete inhibition of orbicularis oculi leads to subtle yet statistically significant decreases in such communication.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Research Protection Program, the Institutional Review Board for all Mass General Brigham hospitals. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

NE developed the study concept and study design. LR, NE, and SS drafted the manuscript. EK provided critical revisions. SS did the data analysis. All authors did the data interpretation and approved the final version of the manuscript for submission.

FUNDING

This study was funded by Allergan as a Phase IV Clinical Trial.

ACKNOWLEDGMENTS

We thank Conor Gallagher, Ph.D. for spearheading this project at Allergan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.612654/full#supplementary-material

REFERENCES

Ambadar, Z., Cohn, J. F., and Reed, L. I. (2009). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *J. Nonverbal Behav.* 33, 17–34. doi: 10.1007/s10919-008-0059-5

Ambadar, Z., Schooler, J., and Cohn, J. F. (2005). Deciphering the enigmatic face. *Psychol. Sci.* 16, 403–410. doi: 10.1111/j.0956-7976.2005.01548.x

American Society of Plastic Surgeons (2017). *2016 Plastic Surgery Statistics Report*. Available online at: https://www.plasticsurgery.org/documents/News/Statistics/2016/plastic-surgery-statistics-full-report-2016.pdf (accessed April 23, 2020).

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects of subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Boulogne, D. (1862/1990). *The Mechanism of Human Facial Expression*. Washington, DC: American Psychologist.

Brown, W. M., and Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: an evolutionary analysis. *Adv. Psychol. Res.* 11, 59–78.

Buhrmeister, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk: a new source of inexpenseive, yet high-quality data? *Perspect. Psychol. Sci.* 6, 3–5. doi: 10.1177/1745691610393980

Calvo, M. G., Krumhuber, E., and Fernandez-Martin, A. (2019). Visual attention mechanisms in happiness vs. trustworthiness processing of facial expressions. *Q. J. Exp. Psychol.* 72, 729–741. doi: 10.1177/1747021818763747

Carruthers, A., Bruce, S., de Connick, A., Connolly, S., Cox, S. E., Davis, P., et al. (2014). Efficacy and safety of onabotulinumtoxina for the treatment of crows feet lines: a multicenter. randomized, controlled trial. *Dermatol. Surg.* 40, 1181–1190. doi: 10.1097/dss.0000000000000128

Del Giudice, M., and Colle, L. (2007). Differences between children and adults in the recognition of enjoyment smiles. *Dev. Psychol.* 43, 769–803.

Ekman, P. (1989). "An argument and evidence about universals in facial expressions," in *Handbook of Social Psychophysiology* eds H. Wagner and A. Manstead (Oxford: John Wiley), 143–164.

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/bf00987191

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., Friesen, W. V., and Hager, J. C. (2002). *The Facial Action Coding System*. Salt Lake City: Consulting Psychologists Press.

Etcoff, N. L., Stock, S., Haley, L. E., Vickery, S. A., and House, D. M. (2011). Cosmetics as a feature of the extended human phenotype: modulation of the perception of biologically important facial signals. *PLoS One* 6:e25656. doi: 10.1371/journal.pone.0025656

Frank, M. G., and Ekman, P. (1993). Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles. *Humor Intl. J. Humor Res.* 6, 9–26.

Frank, M. G., Ekman, P., and Friesen, W. V. (1993). Behavioral markers and the recognizability of the smile of enjoyment. *J. Pers. Soc. Psychol.* 64, 83–93. doi: 10.1037/0022-3514.64.1.83

Gosselin, P., Perron, M., Legault, M., and Campanella, P. (2002). Childrens' and adults' knowledge of the distinction between enjoyment and nonenjoyment smiles. *J. Nonverbal Behav.* 26, 83–108.

Gunnery, S. D., and Ruben, M. A. (2016). Perception of duchenne and non-duchenne smiles: a meta-analysis. *Cogn. Emot.* 30, 501–515. doi: 10.1080/02699931.2015.1018817

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

Harker, L., and Keltner, D. (2001). Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *J. Pers. Soc. Psychol.* 80, 112–124. doi: 10.1037/0022-3514.80.1.112

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

Kanade, T., Cohn, J., and Tian, Y. (2000). "Comprehensive database for facial expression analysis," in *Proceedings of the Fourth IEEE International Conference*, (Grenoble: IEEE), 46–53.

Keltner, D., and Bonanno, G. A. (1997). A sutdy of laughter and dissociation: distinct correlates of laughter and smiling during bereavement. *J. Pers. Soc. Psychol.* 78, 687–702. doi: 10.1037/0022-3514.73.4.687

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., 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., Küster, D., Namba, S., and Skora, P. (2020). Human and machine validation of 14 databases of dynamic facial expressions. *Behav. Res. Methods* doi: 10.3758/s13428-020-01443-y Online ahead of print.

Krumhuber, E. G., Kuster, D., Namba, S., Shah, D., and Calvo, M. G. (2019). Emotion recognition from posed and spontaneous dynamic expressions: human observers vs. machine analysis. *Emotion* doi: 10.1037/emo0000712 Online ahead of print.

Malek, N., Messenger, D., Gao, A. Y., Krumhuber, E., Mattson, W., Joober, E., et al. (2019). Generalizing Duchenne to sad experessions with binocular rivalry and perception ratings. *Emotion* 19, 234–241. doi: 10.1037/emo0000410

McCullough, M. E., and Reed, L. I. (2016). What the face communicates: clearning the conceptual ground. *Curr. Opin. Psychol.* 7, 110–114. doi: 10.1016/j.copsyc.2015.08.023

McDuff, D., Kodra, E., Kaliouby, R. E., and LaFrance, M. (2017). A large-scale analysis of sex differences in facial expressions. *PLoS One* 12:e0173942. doi: 10.1371/journal.pone.0173942

Mehu, M., Grammer, K., and Dunbar, R. I. (2007a). Smiles when sharing. *Evol. Hum. Behav.* 28, 412–422.

Mehu, M., Little, A. C., and Dunbar, R. I. (2007b). Duchenne smiles and the perception of generosity and sociability in faces. *J. Evol. Psychol.* 5, 183–196. doi: 10.1556/jep.2007.1011

Mehu, M., Mortillaro, M., Banziger, T., and Scherer, K. (2012). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Emotion* 12, 701–715. doi: 10.1037/a0026717

Messinger, D. S., Acosta, S. I., Ambadar, Z., and Cohn, J. F. (2008). Infant smiling dynamics and perceived positive emotion. *J. Nonverbal Behav.* 32, 133–135. doi: 10.1007/s10919-008-0048-8

Miller, E. J., Krumhuber, E. G., and Dawel, A. (2020). Observers perceive the Duchenne marker as signaling only intensity for sad expressions, not genuine emotion. *Emotion* doi: 10.1037/emo0000772 Online ahead of print.

Namba, S., Rychlowska, M., Orlowska, A., Aviezer, H., and Krumhuber, E. G. (2020). Social context and culture influence judgments of non-Duchenne smiles. *J. Cult. Cogn. Sci.* 4, 309–321. doi: 10.1007/s41809-020-00066-1

Przylipiak, M., Przylipiak, J., Terlikowski, R., Lubowicka, E., Chrostek, L., and Przylipiak, A. (2018). Improvements in the perception of facial attractiveness following surgical aesthetic treatment; study based on online before and after photos. *J. Cosmet. Dermatol.* 18, 296–300. doi: 10.1111/jocd.12818

Quadflieg, S. V., and Rossion, B. (2013). Differential reliance on the Duchenne marker during smile evaluations and person judgments. *J. Nonverbal Behav.* 37, 69–77. doi: 10.1007/s10919-013-0147-z

Reed, L. I., Zeglen, K. N., and Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous prisoner's dilemma game. *Evol. Hum. Behav.* 33, 200–209. doi: 10.1016/j.evolhumbehav.2011.09.003

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

Scharlemman, J. P., Eckel, C. C., Kacelnik, A., and Wilson, R. K. (2001). The value of a smile: game theory with a human face. *J. Econ. Psychol.* 22, 617–640. doi: 10.1016/s0167-4870(01)00059-9

Scherer, K. R., 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

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

Smith, F. W., and Schyns, P. G. (2009). Smile through your fear and sadness: transmitting and identifying facial expression signals over a range of viewing distances. *Psychol. Sci.* 20, 1202–1208. doi: 10.1111/j.1467-9280.2009.02427.x

Yuki, M., Maddux, W. W., and Masuda, T. (2007). Are the windows to the soul the same as in the East and West? cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *J. Exp. Soc. Psychol.* 43, 303–311. doi: 10.1016/j.jesp.2006.02.004

Zloteanu, M., Krumhuber, E. G., and Richardson, D. C. (2020). Acting surprised: comparing perceptions of different dynamic deliberate expressions. *J. Nonverbal Behav.* 1–17.

Conflict of Interest: This study was funded by Allergan, plc. The funders had no role in data collection, data analysis, or the decision to publish this manuscript.

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.

ORIGINAL RESEARCH
published: 13 January 2021
doi: 10.3389/fpsyg.2020.579509

Anger and Sadness Expressions Situated in Both Positive and Negative Contexts: An Investigation in South Korea and the United States

Sunny Youngok Song[1,2]*, *Alexandria M. Curtis*[1] *and Oriana R. Aragón*[1]*

[1] Department of Marketing, Wilbur O. and Ann Powers College of Business, Clemson University, Clemson, SC, United States, [2] School of Marketing and International Business, Spears School of Business, Oklahoma State University, Stillwater, OK, United States

A formidable challenge to the research of non-verbal behavior can be in the assumptions that we sometimes make, and the subsequent questions that arise from those assumptions. In this article, we proceed with an investigation that would have been precluded by the assumption of a 1:1 correspondence between facial expressions and discrete emotional experiences. We investigated two expressions that in the normative sense are considered negative expressions. One expression, "anger" could be described as clenched fists, furrowed brows, tense jaws and lips, the showing of teeth, and flared nostrils, and the other "sadness" could be described as downward turned mouths, tears, drooping eyes, and wrinkled foreheads. Here, we investigated the prevalence, understanding, and use of these expressions in both positive and negative contexts in South Korea and the United States. We found evidence in both cultures, that anger and sadness displays are used to express positive emotions, a notion relevant to Dimorphous Theory. Moreover, we found that anger and sadness expressions communicated appetitive feelings of wanting to "go!" and consummatory feelings of wanting to "pause," respectively. There were moderations of our effects consistent with past work in Affect Valuation Theory and Display Rule Theory. We discuss our findings, their theoretical relevance, and how the assumptions that are made can narrow the questions that we ask in the field on non-verbal behavior.

Keywords: facial expressions, motivation and affect, Dimorphous, expression, emotion

OPEN ACCESS

Edited by:
José-Miguel Fernández-Dols, Autonomous University of Madrid, Spain

Reviewed by:
James A. Russell, Boston College, United States
Inge Schweiger Gallo, Complutense University of Madrid, Spain

****Correspondence:***
Sunny Youngok Song
sunny.song@okstate.edu
Oriana R. Aragón
oriana.aragon@gmail.com
orcid.org/0000-0002-1016-2076

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *02 July 2020*
Accepted: *26 October 2020*
Published: *13 January 2021*

Citation:
Song SY, Curtis AM and Aragón OR (2021) Anger and Sadness Expressions Situated in Both Positive and Negative Contexts: An Investigation in South Korea and the United States. Front. Psychol. 11:579509. doi: 10.3389/fpsyg.2020.579509

INTRODUCTION

A formidable challenge for the research in non-verbal behavior may lie in the assumptions that have been made concerning a correspondence between facial expressions and basic emotions e.g., the "anger face" corresponds to feelings of anger, smiles correspond to happy feelings, or tearful crying corresponds to feelings of sadness (Ekman and Friesen, 1971, 1986; Izard, 1971; Ekman, 1972). Such assumptions about correspondence between facial movements and specific emotional experiences were empirically based, and also intuitive because there *does* seem to be a tacit agreement of the normative, almost definitional understanding of expression-experience correspondence observable in our world today (Russell, 1994). For example, elementary school rooms feature charts to teach children expressions and their corresponding discrete emotions, and there was a relatively instantaneous worldwide adoption of emoji faces (Danesi, 2017), possibly because we already had an implicit consensus of what the basic emoji expressions meant.

However, such an assumption of a 1:1 correspondence between facial expressions and experience is upended when considering how these expressions are actually used in real life. Our literature (for example, Fernández-Dols and Ruiz-Belda, 1995; Carroll and Russell, 1996; Fernández-Dols et al., 1997; Aviezer et al., 2008, 2012; Fernandez-Dols and Crivelli, 2013; Aragón et al., 2015; García-Higuera et al., 2015; Durán et al., 2017; Aragón and Bargh, 2018; Aragón and Clark, 2018) and lives abound with examples of violations of this supposed correspondence, e.g., the happy tears upon the birth of one's child, the seemingly violent rage across a soccer field upon winning, a smile when embarrassed, and the tooth baring growl at a cute little baby. There might be a definitional understanding of a correspondence between expression and discreet emotions, but that appears to be separate and apart from how expression and discrete emotions correspond in real life.

In this article we proceed to report on an investigation that would have been precluded by the assumption of a 1:1 correspondence between expression and experience. This investigation put aside the definitional understanding of two expressions that, in the normative sense are considered negative expressions, but within experiments with Western samples have been found to violate normative correspondence when interpreted in context. One expression could be described as clenched fists, furrowed brows, tense jaws and lips, the showing of teeth, and flared nostrils, and the other described as downward turned mouths, tears, drooping eyes, and wrinkled foreheads. For a lack of better terminology, we will refer to these physical displays as "anger" and "sadness" expressions, respectively. Our primary aim was to understand how anger and sadness expressions are interpreted in both positive and negative contexts, and if people report to use anger and sadness expressions in both positive and negative contexts, in South Korea and the United States.

Anger and Sadness Expressions as Dimorphous Expressions of Emotion

Previous research has shown that when anger and sadness expressions are situated within positive contexts the majority of observers show consensus in their interpretation of the expressions as representing predominantly positive—not negative experiences (Fernández-Dols and Ruiz-Belda, 1995; Aviezer et al., 2012; Aragón et al., 2015; Aragón, 2016, 2017; Wenzler et al., 2016; Aragón and Bargh, 2018; Aragón and Clark, 2018). This finding that anger and sadness expressions can be associated with predominantly positive emotions has been consistent whether the expressions arose within participants themselves during emotionally evocative situations (Aragón et al., 2015; Aragón, 2017), the expressions were presented to participants to probe for reflection of their own past experiences (Aragón and Bargh, 2018), or when participants were asked to interpret what those expressions might represent (Aviezer et al., 2008; Aragón, 2016, 2017, 2020; Aragón and Bargh, 2018; Wenzler et al., 2016). These patterns were consistent whether anger and sadness expressions were pulled from photographs of real-life contexts (Aragón and Bargh, 2018), created through trained actors (Aragón, 2017; Aragón and Clark, 2018), were photographs of anger and sadness classified through facial action coding (Karolinska directed emotional faces; Lundqvist et al., 1998 as used in Aragón and Bargh, 2018), and whether experiences had been presented through static photographs, narrative accounts, or dynamic video displays.

These types of expressions, in which the normative interpretation and the contextual interpretation are opposing in valence, were termed "dimorphous" (Aragón et al., 2015). Expressions that are called dimorphous share systematic features. For one, dimorphous expressions are context dependent for accurate interpretation, i.e., an overjoyed woman who displays a downward turned mouth, flowing tears and wrinkled brow would be interpreted as experiencing a negative emotion if viewed out of context (most likely because of our implicit understanding), but is instead read as experiencing positive emotion in the context receiving her Olympic gold medal (Fernández-Dols and Ruiz-Belda, 1995; Aviezer et al., 2012; Aragón and Clark, 2018). This highlights the idea that emotional experiences and facial movements do not have a 1:1 correspondence. Dimorphous expressions are uncontrolled and spontaneous, i.e., not forced, not produced sarcastically to make a point, not in service of emotional labor, or masks to hide one's true feelings. Dimorphous expressions are displays that unfold over the course of an emotional event, e.g., "He was smiling, and he was so happy he even cried." The two expressions alternate or may combine at times during the emotional event. The term dimorphous was chosen to reflect this unfolding real-life dynamic of two expressions arising from a singularly valenced emotional experience.

Dimorphous expressions are by definition the experience of a singularly valenced emotional experience, which makes them distinct from the hypothesized simultaneous experience of positive and negative emotions as described in mixed emotions (Larsen et al., 2001), and from sequentially experienced positive and negative emotions (Carrera and Oceja, 2007; Russell, 2017). In an experiment in which participants watched a predominantly positive heart-warming story in which the hero lived a long and happy life, Aragón (2017) demonstrated that dimorphous expressions represented a singularly valenced appraisal (good things took place, and bad things did not take place) that produced a singularly valenced experience of emotion (I feel good feelings, and I do not feel bad feelings), resulting in the display of two physical expressions (I smiled and I cried) over the course of participants' own emotional experience. Those physical displays were attributed by participants to their own singularly valenced feelings (I smiled and cried because of the positive emotions I was feeling). In contrast, when participants were assigned randomly to view the same heartwarming story but in this case told that the hero had died, both positive and negative emotions were evident, as participants made two appraisals opposing in valence (both good and bad things took place), associated with two emotional experiences opposing in valence (I feel good and bad feelings), which then resulted in the display of two physical expressions over the course of an emotional episode (I smiled and I cried), that were attributed to both positive and negative feelings (I smiled and cried because of the positive and negative emotions I was feeling). These findings were consistent with self-report measures and with implicit measures of positive and negative affect.

Additionally, dimorphous expressions can be of a singular "flavor" of emotion, e.g., crying when feeling intensely relieved, but dimorphous expressions are also consistent with the idea that expressors can experience a blend of positive emotions, for example feeling both relieved and joyous when crying (Smith and Ellsworth, 1987). In dimorphous expression research, emphasis has been placed less so on which flavor of emotion is displayed, and more so upon the overall valence of the experience. For example, the expression of a positive emotion through an anger display may represent pride (Aragón and Bargh, 2018), victory (Aviezer et al., 2008), excitement (Aragón, 2016), or even overwhelming feelings of care when regarding something adorably cute (Aragón et al., 2015). Dimorphous research has focused on understanding the correspondence between the normatively understood valence of expression and the valence of the actual experience. The precise flavor of emotional experience has been of less importance for this research focus (for discussion see, Aragón and Bargh, 2018).

Researchers who have measured both emotional experiences and what we refer to as dimorphous expressions, consistently note that those who express emotion dimorphously describe their emotions as intense (e.g., Fredrickson and Levenson, 1998; Bonanno and Keltner, 2004; Ansfield, 2007; Aragón et al., 2015; Aragón, 2016, 2017; Aragón and Bargh, 2018; Aragón and Clark, 2018). Additionally, researchers note that dimorphous expressions are interpreted by raters, onlookers, or judges as highly intense (e.g., Fernández-Dols and Ruiz-Belda, 1995; Aviezer et al., 2012; Aragón and Clark, 2018), and that the situations in which these expressions arise are themselves judged to be intense in nature (e.g., Fernández-Dols et al., 2010; Wenzler et al., 2016). It is possible that the alternating display between the normatively corresponding and non-corresponding expression is a function of a fluctuating intensity of emotion. For example, as one wins an award, she may predominantly smile, however, momentarily cry when she is hit with a new wave of highly intense feelings.

When trying to understand the functional nature of these expressions, it is important to understand what they discreetly communicate. As discussed above, anger and sadness expressions have been found to be poor indicators of emotional valence because they communicate positive emotions when situated in positive contexts, and negative emotions when situated in negative contexts. Both expressions signal intense experiences, and thus do not discriminate well from each other in the aspect of intensity of experience. And, anger and sadness expressions, particularly within positive contexts can relay any of a variety of flavors of emotion, e.g., anger and sadness expressions can both signal feelings of pride, or adoration, or victory. Since anger and sadness expressions do not discriminate in the aspects outlined above, here we describe what anger and sadness expressions have been found to distinctly communicate across both the positive and negative situations.

Activation-Type Dimensions Associated With Anger and Sadness Expressions

Researchers have previously introduced useful theories about activation-type dimensions of emotion, i.e., activation and deactivation (Russell, 2003), excitement and calm (Mogilner et al., 2012), excited and peaceful happiness (Tsai et al., 2006), high and low states of action readiness to engagement (Frijda et al., 1989), high and low states of arousal (Russell, 1980; Feldman Barrett, 1998), dominance and submissiveness (Bradley and Lang, 1994), promotion and prevention focus (Higgins, 1997), and appetitive and consummatory aspects of pleasure (Berridge and Robinson, 2003). In a quest to understand what angry and sadness expressions might consistently communicate, Aragón and Bargh (2018) tested these overlapping constructs in a series of experiments to see which, if any, would be associated distinctly with anger or sadness expressions, whether situated in positive or negative contexts.

Over a series of studies, the constructs of high arousal (feelings associated with words such as excited, active, and alert) and low arousal (words such as calm, depleted, and sleepy) were differentiated for anger and sadness i.e., anger expressions were viewed as more so high arousal, and sadness expressions were viewed as more so lower arousal, but this was only true when those expressions were situated in negative contexts. When asked if a winning athlete who showed an anger expression was excited, participants indicated yes, she was. But also, when asked if a winning athlete who cried was excited, again participants would indicate yes. And when given an open text box to describe how they interpreted the sadness expression in a positive context, participants would describe, "She's excited. She just needs to stop for a minute," or "She was just overwhelmed and needed to pause a minute." A similar pattern emerged for the aspects of dominance and submissiveness, which again only discriminated between anger and sadness expressions in negative contexts. These findings were true whether the paradigm tested for participants' own experiences with anger and sadness expressions, or participants' interpretations of what others were feeling when those others displayed anger or sadness expressions.

Over many iterations only one conceptualization within the overlapping activation-type constructs showed a consistent discrimination between the anger and sadness expressions across both negative and positive contexts, those were the fundamental motivational orientations of feelings of "wanting to go" and "wanting to stop." Anger expressions that arose in positive or negative contexts represented and communicated positive and negative emotional experiences, respectively, that were imbued with antsy feelings of wanting to go, move or accelerate as put forth by Berridge and Robinson (2003) in the concept of appetitive pursuit. In contrast, sad expressions that arose in both positive and negative contexts, represented and communicated positive and negative emotional experiences, respectively, that were imbued with spent feelings of wanting to pause, stop, or be still, akin to the concept of consummatory states in which one pauses from pursuit (Berridge and Robinson, 2003).

Even more important to the possible reason why dimorphous expression might exist at all, i.e., "Why don't smiles seem to suffice in these intensely positive moments?" dimorphous expressions appear to provide information about the expressers' appetitive and consummatory orientations with a specificity that smiles did not. This is supported by the fact that positive emotions expressed through anger and sadness displays have also been found to impact inferences about expressers' experiences and future product preferences. For example, people who expressed they were so happy as to "yell 'YES!'" were deemed to be more likely to prefer an action vacation package and people who expressed that they were so happy they cried were deemed more likely to prefer a relaxation vacation package. These orientations even imbued participants' inferences about the types of products that were in use when the expressions had arisen, i.e., drivers with anger expressions were driving zippy sports cars, and drivers who displayed joyous tears were driving luxury sedans (Aragón, 2016). In American samples to date, anger and sadness expressions have provided robust signals to onlookers about the expressers' feelings of wanting to go or wanting to pause, respectively. Considering the choreography of social interactions, such information would seem vital, particularly in situations in which emotions are intense.

Overview

This study included respondents from both the United States and South Korea. In the experimental portion of the study, participants were assigned randomly to view people donning anger, smiling, or sadness expressions situated within short vignettes. There were five vignette themes designed to elicit different flavors of emotion. Each of the five vignette themes was tailored to feature a positive and negative version within the same theme, e.g., someone fulfills a lifelong dream, versus someone fails to fulfill a lifelong dream. Every participant saw a total of 10 vignettes (5 themes $\times$ 2 versions) that displayed models who varied in gender, background (Asian, Black, Latino/a, and White), age—some being apparently college-aged, and some appearing to be in their early to mid-thirties, and who showed no apparent signals of authority, e.g., in plain clothing against plain backgrounds. Our dependent variables of interest were participants' interpretations of the expressers' emotional intensity, experience (positive and negative), and the expressers' appetitive and consummatory motivational orientations. Additionally, to understand prevalence of these expressions cross-culturally, we also asked participants if they had seen, known someone, or had themselves used the depicted expression in a similar context. Separated in time from the experimental portion but in the same research session we also collected individual difference measures of participants' tendencies to express emotion dimorphously (Dimorphous Expression Questionnaire, Aragón et al., 2015).

Central Study Aims

Our central question pertained to the existence of dimorphous expressions across these two cultures and that was to be addressed through the experimental design, and the individual difference measure. We reasoned that if dimorphous expressions existed in South Korea as they have been observed in the United States, that in the experimental portion South Korean participants would interpret anger and sadness expressions in positive contexts much as Americans do, as intense positive, but not negative emotional experiences. And when asked if they had seen, known, or themselves used such an expression of anger or sadness in a positive context, agreement that they had would be further evidence for dimorphous expressions. Additionally, we expected that the individual difference measures of dimorphous expressions could capture the existence of dimorphous expressions.

> H1: We predicted that when anger, smiling, and sadness expressions were situated in positive contexts they would be interpreted as representing predominantly positive emotional experiences, and when situated in negative contexts they would be interpreted as representing predominantly negative emotional experiences.
> H2: We predicted that anger and sadness expressions would communicate appetitive and consummatory orientations, respectively, and smiling expressions would not clearly communicate either appetitive or consummatory orientations.
> H3: We predicted that anger and sadness expressions would communicate more intense emotional experiences than smiling expressions.
> H4: We predicted that participants in both the South Korea and the United States would self-report the use of dimorphous expression.

Three Cross-Cultural Considerations for Our Hypotheses

One consideration was that people from Western, individualistic contexts strive to maximize positive and minimize negative emotions, whereas those from Eastern, collectivist contexts instead value experiencing both positive and negative emotions. Thus whether conceptualized as a mixed experience of emotions (Larsen et al., 2001) or sequentially experienced emotions (Carrera and Oceja, 2007), people from Eastern contexts have more experiences with a combination of both positive and negative emotions than people from Western contexts (Sims et al., 2015). These cultural differences in the presence of both positive and negative emotions apply more so in positive than negative settings (Kim et al., 2014), presumably because people from Eastern contexts see a mix of both positive and negative emotions as creating a tempered balance or harmony during positive events, and those from Western contexts strive to feel more purely positive. Therefore, we considered the following possibility:

> H5: We left open the possibility that participants from South Korea might not express emotions dimorphously, and as such might not interpret anger and sadness

expressions situated in positive contexts as representing predominantly positive emotions.

A second consideration involved previous work in affect valuation theory that has found that North American (United States and Canada) students idealize high arousal positive affect more so than low arousal, and East Asian (Chinese, Japanese, and Korean) students more so idealize low arousal positive affect over high arousal (Tsai et al., 2006). Idealized affect in turn predicts typically experienced affect ostensibly because the experiences one might choose to engage in could differ by how one would like to feel during the experience. These experience selections are reflected in the products (Chim et al., 2018; Park et al., 2020), professional services such as doctors (Sims et al., 2018), and leaders (Tsai et al., 2016) that are preferred cross-culturally. Higher arousal versions are preferred in Western cultures and lower arousal versions preferred in Eastern cultures. Therefore, one might expect similar patterns in the prevalence of the expression of positive emotion as anger and sadness expressions, respectively, because anger expressions are related to appetitive states that conceptually overlap with high arousal, and sadness expressions are related to consummatory states that conceptually overlap with low arousal.

H6: We predicted that South Korean participants would report using more sadness than anger expressions and United States participants would report using more anger than sadness expressions to communicate positive emotions.

A third consideration was in regard to how display rules might differ between these cultures. Individualistic cultures are more so focused on the development of the self (Markus and Kitayama, 1991), personal goals (Yamaguchi, 1994), and the expression of emotion (Butler et al., 2007; Matsumoto et al., 2008b). In contrast, collectivist cultures are more so focused on the development of the in-group, the in-group's goals, place a lesser value on the expression of an individual's emotions (Matsumoto et al., 2008a), particularly anger and sadness expressions (Matsumoto, 1990; Safdar et al., 2009), and encourage the suppression, i.e., holding back (Gross, 2001) of emotion, or the masking of negative emotions with smiles (Friesen, 1972; Ekman and Friesen, 1982; Matsumoto and Kupperbusch, 2001; Rychlowska et al., 2017).

H7: We predict that our South Korean sample would report fewer expressions of emotion than our American sample, with the exception of smiling to mask negative experiences.

MATERIALS AND METHODS

Participants

Students from an American State University and South Korean Universities were recruited through classroom announcements and through university electronic bulletin boards to participate in this approximate 20-min study for either course credit or approximately $5 US dollars during the fall of 2019. Participants who participated for cash compensation were paid via an online application. Students participating for course credit were compensated through a department subject pool.

Sample size was estimated with the experimental design's between factors in mind: 2 (country) × 2 (gender) × 3 (expression) = 12 cells at 50 participants per cell estimated in Aragón and Bargh (2018). This calculation rendered a goal to recruit 600 participants. Our sample for the dimorphous expression questionnaire ($N = 659$) included data from all non-international (i.e., native to each respective country) participants who passed attention checks and completed through the individual difference measure of dimorphous expressions located near the beginning of the survey: South Korean ($N = 305$, 132 men, $M_{age} = 20.16$, $SD = 2.55$, SK age- corrected for cultural differences in numerical assignment) and American ($N = 354$, 169 men, $M_{age} = 20.03$, $SD = 2.05$). There was attrition ($n = 75$) during this 20-min survey. Our sample for the experimental portion which occurred about 10 min into the survey included all who completed through that portion ($N = 584$).

Materials and Procedure

Data were collected as part of a larger investigation. See **Supplementary Appendix A** for study details. All materials were developed in English, translated into South Korean (by Song), and back translated into English by paid interpreters. Originals and backtranslations were then compared for meaning (by Aragón). When discrepancies arose, adjustments to the translated version were made (Brislin, 1970). This process took three iterations. We ran a pilot of this study in the summer of 2019 with 66 participants from South Korea. Only a few minor wording changes were made after the pilot study. The findings from the pilot and this study are nearly identical.

Experimental Paradigm

Participants were assigned randomly to consider either anger, smiling, or sadness expressions, displayed within five vignette themes, that were designed with a both a positive and negative version (10 trials in total). The vignette themes were selected to reflect the instances in which participants have previously reported to express dimorphously (Aragón et al., 2015; Aragón, 2016; Aragón and Bargh, 2018; Aragón and Clark, 2018), such as when (1) a person has the opportunity to fulfill a lifelong dream, (2) a person views a beautiful nature scene, (3) a person accomplishes a big life goal after a long struggle to succeed, (4) a person is reunited with family after a long absence, and (5) a fan is able to see a favorite celebrity. Negative versions also were created for each of these themes, please see **Supplementary Appendix B** for all scenarios.

We chose five themes so that our effects would not be bound to a single "flavor" of emotion, because we would be able to demonstrate for example, that sadness expressions can communicate positive emotions that would come about in a beautiful nature scene as well as would come about in accomplishing a big life goal. To have confidence that the positive events were considered predominantly positive, and the negative events were considered predominantly negative, an independent sample of participants ($N = 61$ online; 38%

women, M_{age} = 24.41) validated that the "positive" vignettes were considered predominantly positive, and not negative, and the "negative" vignettes were considered predominantly negative, and not positive. There were no interactions of vignette, meaning all of the "positive" vignettes were interpreted as equally positive, and equally not negative, and all of the "negative" vignettes were interpreted as equally negative, and equally not positive. See **Supplementary Appendix B** for means.

Inserted just below the wording of each vignette was a photograph of a man or woman (randomly selected) of apparent Asian, White, Black, or Hispanic/Latinx background (counterbalanced). All vignette, expression, gender, and model pairing combinations were presented an equal number of times. This design made it less likely that our observed effects would be due idiosyncratic factors of a model's gender, the vignette type, a model's ethnicity, or a model's particular anger, smiling, or sadness expression. See **Figure 1**.

The photographed models were purchased (Shutterstock, 2019) and independently validated (N = 105, online sample; 37% women, M_{age} = 36.70) to give us confidence that what we considered normative anger, smiling, and sadness expressions were actually understood normatively as representing anger, happiness, and sadness, respectively. In a forced-choice paradigm, participants viewed each of our modeled expressions and selected from labels of happy, angry, sad, surprised, fearful, or disgusted. This methodology has been used by basic emotion researchers, and we considered it appropriate to capture a normative understanding of the expression - experience correspondence (as described in Russell, 1994). In our stimuli validation study 73.1% of respondents considered our "angry" models to be angry, 93.7% considered our "smiling" models to be happy, and 76.5% considered our "sadness" models to be sad. Asian models were overrepresented in our design (4 Asian, 2 Black, 2 Latino/a, and 2 White models) because we took into account that our South Korean participants would come from a less diverse context (Fearon, 2003). Without this adjustment, South Korean participants would have evaluated a greater percentage of models who appeared to be outgroup members than would have American participants. See **Figure 2** for examples of the expressions used.

Following each of the 10 vignettes, participants indicated their inferences about the expressers' emotions with, "He (or she) is feeling ________" positive valence: "strong positive emotions" and "good emotions," and for negative valence: "strong negative emotions" and "bad emotions." We also asked about the motivational orientations of the expresser with, "He (or she) is feeling like he (she) wants to ________" appetitive items: "go, go, go!" and "get moving," and consummatory items "stop for a moment" and "be still for a moment." We also asked the intensity of the emotion perceived "He (she) is feeling intense emotions." Then, to assess the prevalence of anger, smiling and sadness expressions for these vignettes for each culture, three items asked participants if they had seen, known someone, or had themselves used "this type of expression in a positive (or negative) event." Response options were 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Somewhat Disagree*, 4 = *Somewhat Agree*, 5 = *Agree*, and 6 = *Strongly Agree*. See **Table 1**, for descriptive statistics.

Individual Difference Measure

Respondents answered a revised version of the Dimorphous Expression Questionnaire (Aragón et al., 2015). The questionnaire was modified from the original by the addition of a preface to increase participants' understanding of what we are asking, and questions were added to reflect the greater diversity of contexts in which dimorphous expression have been reported to occur (Aragón, 2016; Aragón and Bargh, 2018). In the preface, participants viewed five photographs of men and women displaying anger and sadness expressions. On the next page the photographs had labeling that indicated the situation in which each expression had arose. Some were positive situations, and some were negative. We made clear to participants that we were interested in when normatively negative facial expressions represented positive emotions, and when normatively positive facial expressions represented negative emotions. Participants then responded to five items (α = 0.88) that captured anger expressions when feeling positive emotions, and six items (α = 0.72) that captured participants tendency to cry or appear sad when feeling positive emotions. Response options were 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Somewhat Disagree*, 4 = *Somewhat Agree*, 5 = *Agree*, and 6 = *Strongly Agree*. See **Table 2** descriptive statistics.

RESULTS

When deciding upon an analytical strategy for the experimental paradigm, we first ran a nested mixed linear model to test for effects of the counterbalance of the vignette type, counterbalance of vignette valence, the gender of model, and the background/ethnicity of the models, all of which varied on each trial. There were no significant main effects of these variables. Results were similar with and without counterbalance and model characteristic variables added as controls. We chose to run our analysis with generalized linear models because they provided effect sizes and power statistics, which are not available for nested mixed linear models. Results are in the same direction and of similar magnitude and significance with either the mixed linear or the generalized linear models. *Post hoc* comparisons have been Bonferroni corrected. We report key findings here and details in tables. Tables have been created with detailed descriptive statistics, including percentages of participants who agreed/disagreed with our prompts to provide our readers a full sense of how participants responded to our questions.

Experimental Paradigm

Inferred Intensity of Emotion

In a repeated measure, general linear model, we predicted intensity ratings with repeated effects of context (positive event and negative event) and vignette (5 types), with between subject effects of expression (anger, smiling, and sadness) and country (South Korea and United States). We entered all main effects and possible interactions. Consistent with past research, anger (M = 5.12, SE = 0.05) and sadness (M = 5.01, SE = 0.05) expressions communicated more intense emotional experiences than did smiles (M = 4.30, SE = 0.05), $F(2, 569) = 74.76$, $p < 0.001$,

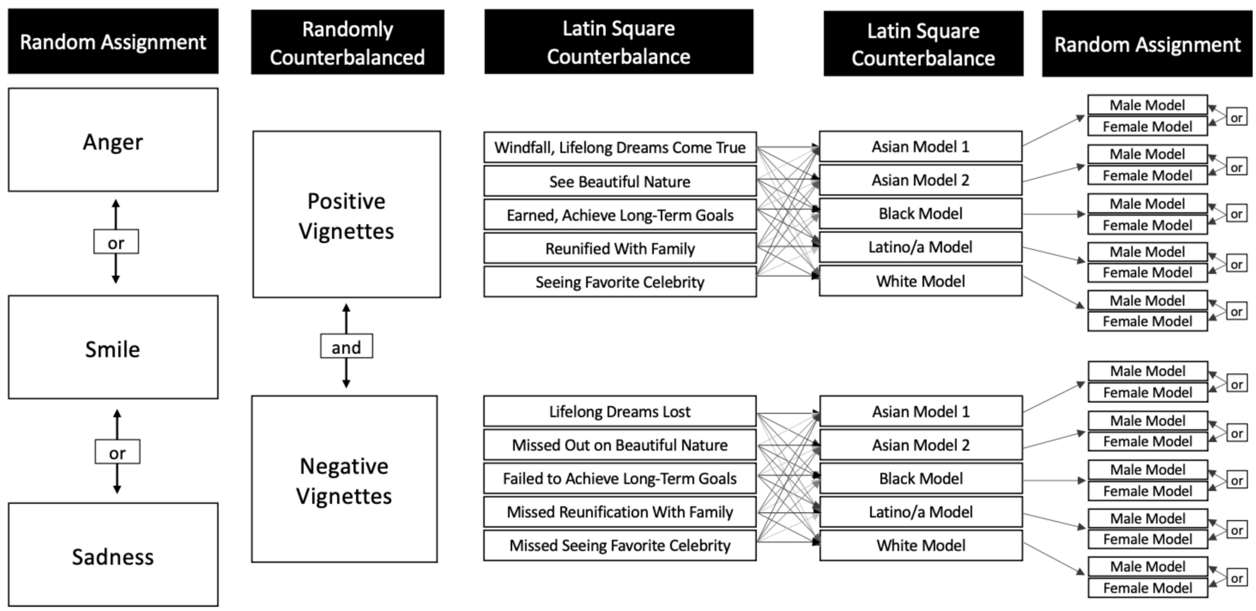


FIGURE 1 | In the experimental portion, participants were assigned randomly to view models with either an anger, smiling, or sadness expression. Counterbalancing ensured that all participants responded to five positive and five negative vignettes, with Asian, Black, Latino/a, and White models, that were either male or female.

FIGURE 2 | Pictured are examples from the 30 photographs [10 models each depicting anger **(A)**, smiling **(B)**, and sadness **(C)**] used for our experiment. Standard content license was purchased with signed model release on file at Shutterstock.

$\eta_p^2 = 0.21$, observed power = 1.00. There was also a significant country × context × expression interaction, $F(2, 569) = 16.26$, $p < 0.001$, $\eta_p^2 = 0.02$, observed power = 0.86, that revealed South Korean participants in the smile condition interpreted positive vignettes as more intense than negative vignettes. See **Table 3** and **Figure 3**.

Inferred Emotional Experience

In a repeated measure, general linear model, we predicted affective valence with repeated effects of context (positive event and negative event), vignette (5 types), and valence (positive and negative), with between subject effects of expression (anger, smiling, and sadness) and country (South Korea and United States). We entered all main effects and possible interactions. One interaction accounted for the majority of the variance explained by the model, that was the interaction between context and valence, $F(1, 578) = 1508.47$, $p < 0.001$, $\eta_p^2 = 0.72$, observed power = 1.00 (see **Figure 4A**). Emotions in positive contexts were interpreted as predominantly positive ($M = 4.61$, $SE = 0.04$), not negative ($M = 2.33$, $SE = 0.04$). In contrast,

TABLE 1 | Descriptive statistics for the seen, known, and used the expression items.

	Vignette	Expression condition	Positive vignettes				Negative vignettes			
			South Korea		**United States**		**South Korea**		**United States**	
			% Agreed	**Mean (SD)**	**% Agreed**	**Mean (SD)**	**% Agreed**	**Mean (*SD*)**	**% Agreed**	**Mean *(SD)***
Seen % who agreed or strongly agreed (5 or 6) and mean scores (*SD*)	Fortune	Anger-like	36.7	3.54 (1.69)	51.9	4.21 (1.49)	16.0	2.71 (1.57)	63.9	4.52 (1.25)
		Smile	55.4	4.43 (1.27)	72.9	4.92 (0.80)	36.3	3.86 (1.33)	29.5	3.64 (1.36)
		Sadness-like	30.2	3.52 (1.57)	57.5	4.38 (1.37)	11.9	2.69 (1.49)	72.8	4.84 (1.01)
	Awe	Anger-like	22.9	3.05 (1.60)	31.8	3.46 (1.48)	16.5	2.97 (1.48)	57.4	4.42 (1.18)
		Smile	54.7	4.49 (1.10)	70.5	4.80 (0.97)	39.0	4.04 (1.29)	23.6	3.46 (1.31)
		Sadness-like	20.5	3.25 (1.41)	40.4	3.87 (1.48)	16.9	2.76 (1.57)	65.2	4.64 (1.14)
	Pride	Anger-like	36.7	3.74 (1.67)	61.7	4.57 (1.28)	13.4	2.61 (1.55)	68.2	4.84 (1.06)
		Smile	60.8	4.65 (0.91)	70.3	4.88 (0.95)	33.8	3.77 (1.38)	31.0	3.60 (1.39)
		Sadness-like	32.6	3.45 (1.65)	53.5	4.35 (1.39)	17.9	2.79 (1.62)	73.0	4.81 (1.08)
	Interpersonal	Anger-like	28.4	3.36 (1.60)	45.8	4.07 (1.38)	16.8	2.68 (1.53)	67.6	4.70 (1.10)
		Smile	58.4	4.57 (1.04)	67.4	4.70 (0.95)	34.2	3.65 (1.47)	29.7	3.58 (1.36)
		Sadness-like	34.9	3.86 (1.51)	67.5	4.78 (1.20)	11.8	2.49 (1.60)	68.7	4.68 (1.20)
	Ecstatic	Anger-like	37.4	3.62 (1.71)	42.6	3.94 (1.59)	12.4	2.66 (1.48)	66.7	4.56 (1.15)
		Smile	55.4	4.47 (1.11)	70.5	4.83 (0.89)	38.0	3.80 (1.35)	34.6	3.80 (1.32)
		Sadness-like	32.2	3.69 (1.46)	56.1	4.43 (1.30)	20.2	2.70 (1.61)	65.2	4.63 (1.20)
	Total, % agreed in at least 1 vignettes, mean (*SD*)	Anger-like	63.1	3.43 (1.24)	83.3	4.05 (0.94)	41.6	2.76 (1.10)	89.8	4.60 (0.83)
		Smile	84.6	4.51 (0.75)	89.9	4.82 (0.69)	70.4	3.84 (1.00)	58.5	3.63 (1.01)
		Sadness-like	65.2	3.54 (1.05)	84.2	4.36 (1.00)	46.0	2.70 (1.12)	89.6	4.72 (0.86)
Known % who agreed or strongly agreed (5 or 6) and mean scores (*SD*)	Fortune	Anger-like	38.8	3.52 (1.69)	45.4	4.04 (1.50)	16.2	2.74 (1.54)	54.6	4.40 (1.24)
		Smile	52.0	4.37 (1.32)	68.2	4.75 (0.94)	35.0	3.88 (1.34)	25.6	3.55 (1.32)
		Sadness-like	29.1	3.50 (1.63)	53.1	4.30 (1.38)	13.1	2.71 (1.57)	68.4	4.72 (1.04)
	Awe	Anger-like	19.8	2.97 (1.54)	27.8	3.43 (1.44)	17.3	2.89 (1.48)	56.5	4.33 (1.20)
		Smile	52.0	4.39 (1.06)	72.9	4.81 (0.91)	33.3	4.00 (1.28)	24.4	3.47 (1.31)
		Sadness-like	18.0	3.19 (1.40)	39.5	3.88 (1.43)	9.6	2.52 (1.43)	60.0	4.57 (1.08)
	Pride	Anger-like	41.8	3.84 (1.65)	58.3	4.43 (1.29)	16.5	2.66 (1.57)	65.1	4.70 (1.10)
		Smile	56.8	4.57 (1.09)	71.9	4.87 (0.89)	33.8	3.82 (1.33)	24.8	3.57 (1.30)
		Sadness-like	27.9	3.49 (1.64)	55.3	4.42 (1.30)	17.9	2.79 (1.63)	71.3	4.71 (1.08)
	Interpersonal	Anger-like	29.8	3.23 (1.66)	42.6	3.89 (1.43)	18.9	2.67 (1.54)	61.1	4.62 (1.11)
		Smile	46.8	4.32 (1.15)	66.4	4.66 (1.03)	31.6	3.53 (1.43)	25.0	3.49 (1.36)
		Sadness-like	37.2	3.90 (1.49)	67.5	4.73 (1.22)	11.8	2.52 (1.58)	68.7	4.65 (1.16)

(Continued)

TABLE 1 | Continued

	Vignette	Expression condition	Positive vignettes				Negative vignettes			
			South Korea		United States		South Korea		United States	
			% Agreed	**Mean (SD)**	**% Agreed**	**Mean (SD)**	**% Agreed**	**Mean (*SD*)**	**% Agreed**	**Mean (*SD*)**
	Ecstatic	Anger-like	31.3	3.37 (1.24)	42.6	3.94 (0.93)	16.7	2.63 (1.56)	60.2	4.49 (1.16)
		Smile	52.7	4.40 (0.79)	62.8	4.76 (0.64)	32.5	3.64 (1.35)	25.2	3.69 (1.26)
		Sadness-like	34.5	3.54 (1.06)	59.6	4.37 (0.97)	18.8	2.62 (1.64)	62.6	4.53 (1.19)
	Total, % agreed in at least 1 vignettes, mean (*SD*)	Anger-like	64.1	3.42 (1.70)	79.6	3.92 (1.60)	46.5	2.77 (1.11)	87.0	4.50 (0.87)
		Smile	82.1	4.43 (1.17)	93.8	4.71 (0.95)	70.4	3.78 (0.88)	59.2	3.57 (0.93)
		Sadness-like	65.3	3.70 (1.53)	82.5	4.50 (1.14)	43.7	2.63 (1.09)	88.7	4.64 (0.82)
Used % who agreed or strongly agreed (5 or 6) and mean scores (*SD*)	Fortune	Anger-like	17.3	2.85 (1.52)	38.0	3.72 (1.59)	12.0	2.38 (1.48)	49.1	4.11 (1.36)
		Smile	45.3	4.04 (1.50)	64.3	4.67 (0.94)	32.5	3.59 (1.46)	20.2	3.18 (1.38)
		Sadness-like	14.0	2.86 (1.47)	27.2	3.40 (1.41)	8.3	2.32 (1.38)	57.0	4.37 (1.22)
	Awe	Anger-like	10.4	2.45 (1.38)	22.2	3.00 (1.49)	5.1	2.27 (1.30)	45.4	4.05 (1.27)
		Smile	41.3	4.05 (1.37)	59.7	4.55 (1.03)	26.9	3.64 (1.31)	15.0	3.07 (1.29)
		Sadness-like	10.1	2.61 (1.38)	21.9	3.07 (1.49)	7.2	2.31 (1.38)	41.7	3.98 (1.41)
	Pride	Anger-like	28.6	3.27 (1.67)	43.5	4.06 (1.47)	7.2	2.16 (1.38)	47.7	4.27 (1.32)
		Smile	45.9	4.26 (1.19)	60.2	4.65 (1.02)	27.3	3.48 (1.39)	18.6	3.19 (1.35)
		Sadness-like	17.4	2.97 (1.58)	36.0	3.68 (1.51)	9.5	2.30 (1.39)	53.9	4.34 (1.26)
	Interpersonal	Anger-like	18.9	2.82 (1.58)	31.5	3.59 (1.50)	5.3	2.21 (1.26)	46.3	4.19 (1.23)
		Smile	40.3	4.17 (1.13)	58.9	4.52 (1.12)	31.6	3.52 (1.48)	18.0	3.06 (1.36)
		Sadness-like	18.6	3.13 (1.44)	44.7	3.95 (1.53)	10.6	2.38 (1.54)	53.0	4.25 (1.28)
	Ecstatic	Anger-like	15.2	2.84 (1.63)	27.8	3.36 (1.60)	3.1	2.15 (1.29)	50.0	4.18 (1.36)
		Smile	51.4	4.22 (1.35)	55.0	4.43 (1.18)	28.7	3.43 (1.39)	22.8	3.39 (1.36)
		Sadness-like	17.2	2.98 (1.45)	26.3	3.32 (1.42)	7.1	2.33 (1.39)	45.2	3.98 (1.39)
	Total, % agreed in at least 1 vignettes, mean (*SD*)	Anger-like	26.9	2.82 (1.22)	74.1	3.55 (0.97)	23.8	2.26 (1.00)	79.6	4.16 (1.03)
		Smile	74.4	4.14 (0.95)	87.6	4.56 (0.73)	59.1	3.56 (0.95)	45.4	3.20 (0.99)
		Sadness-like	41.6	2.90 (1.09)	61.4	3.48 (1.12)	24.1	2.34 (1.08)	79.1	4.18 (1.01)

TABLE 2 | Descriptive statistics for Dimorphous expression, anger and sadness expression items.

Dim.Exp	Item description	South Korea	United States	Total	*F* statistic		Percentage who agreed or strongly agreed (scores of 5 or 6)		
		Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	*df* = 1, 658	*p*-value	South Korea	United States	Total
Normative anger expressions	I can look angry (e.g., clenched jaw and pumping fists) when I feel intense accomplishment (for example when getting a great grade on an important exam, and shouting "YES!").	3.55 (1.64)	4.17 (1.45)	3.88 (1.57)	26.93	<0.001	34.8	48.0	41.9
	I could make an expression that looks angry (e.g., clenched jaw and pumping fists), if I experienced a large windfall (for example winning $10 million dollar lottery).	3.54 (1.58)	3.72 (1.49)	3.64 (1.54)	2.32	0.128	31.8	34.7	33.4
	I can look angry (e.g., clenched jaw and pumping fists) when I feel intense excitement (for example when at a rock concert shouting "YEAHHH!").	3.17 (1.57)	4.00 (1.45)	3.61 (1.56)	50.36	<0.001	23.0	42.9	33.7
	I can look angry (e.g., clenched jaw and pumping fists) when I feel intense anticipation (for example when heading into an athletic competition shouting "YEAH!").	3.49 (1.63)	4.22 (1.33)	3.88 (1.52)	39.77	<0.001	32.5	47.7	40.7
	I can have physical expressions that might look like anger (e.g., clenched jaw, gritted teeth, pumping fists, or pinching and squeezing), when I am actually overwhelmed with positive feelings.	3.09 (1.51)	3.38 (1.40)	3.24 (1.46)	6.57	0.011	19.0	21.8	20.5
								% agreed to at least 1 item	
	All items	3.37 (1.30)	3.90 (1.16)	3.65 (1.25)	31.11	<0.001	54.4	68.1	61.8
Normative sadness expressions	I cry when I see loved ones emotionally reunite (for example when a person returns home after a long absence).	3.91 (1.11)	3.93 (1.36)	3.92 (1.25)	0.03	0.855	32.5	36.7	34.7
	I would cry if I experienced a large windfall (for example winning $10 million dollar lottery).	3.66 (1.49)	3.86 (1.51)	3.77 (1.50)	2.90	0.089	30.5	36.7	33.8
	I cry when I see a person give unselfishly to another (for example when someone donates a home to a needy family).	3.89 (1.32)	3.50 (1.27)	3.68 (1.30)	15.09	<0.001	35.7	20.9	27.8
	I cry when I achieve something that I worked long and hard to obtain (for example at graduation, or when receiving an award).	4.67 (1.23)	3.51 (1.41)	4.05 (1.45)	125.16	<0.001	63.0	25.7	42.9
	I cry when in awe of nature (for instance when looking out at a beautiful tropical island).	3.17 (1.48)	2.51 (1.29)	2.81 (1.42)	37.57	<0.001	21.6	8.5	14.6
	I cry when I feel very close to a loved one (for instance when feeling mutual love with another person).	3.73 (1.35)	3.59 (1.38)	3.65 (1.37)	1.67	0.197	31.5	26.6	28.8
								% agreed to at least 1 item	
	All items	3.84 (0.70)	3.48 (1.00)	3.65 (0.89)	27.30	<0.001	89.5	65.5	76.6

TABLE 3 | General linear repeated measures models testing inferred: intensity and valence of emotion.

Description	Omnibus					Descriptive statistics								Pairwise comparisons		
	df	*F*	*p*-value	η_p^2	Observed power	Interaction variable	Interaction variable	Level 1	*M* (*SE*)	Level 2	*M* (*SE*)	Level 3	*M* (*SE*)	*p*-value 1 and 2	*p*-value 1 and 3	*p*-value 2 and 3
Intensity of experience analysis																
Context (pos.context, neg.context)	(1, 569)	35.65	<0.001	0.059	1.00			pos. context	4.90 (0.03)	neg. context	4.72 (0.03)	–	–	$p < 0.001$	–	–
Country (South Korea, United States)	(1, 569)	1.12	=0.291	0.002	0.18			–	–	–	–	–	–	–	–	–
Expression (anger, smile, and sadness)	(2, 569)	74.76	<0.001	0.208	1.00			Anger	5.12 (0.05)	Smile	4.30 (0.05)	Sadness	5.01 (0.05)	<0.001	=0.394	<0.001
Context × Country	(1, 569)	16.26	<0.001	0.028	0.98		*w/in pos. context*	South Korea	4.92 (0.04)	United States	4.87 (0.04)	–	–	>1.00	–	–
							w/in neg. context	South Korea	4.63 (0.04)	United States	4.81 (0.04)	–	–	=0.023	–	–
Context × Expression	(2, 569)	15.96	<0.001	0.053	1.00		*w/in pos. context*	Anger	5.11 (0.06)	Smile	4.49 (0.06)	Sadness	5.09 (0.06)	<0.001	>1.00	<0.001
							w/in neg. context	Anger	5.13 (0.06)	Smile	4.10 (0.06)	Sadness	4.93 (0.06)	<0.001	=0.053	<0.001
Country × Expression	(2, 569)	8.18	=0.184	0.006	0.36			–	–	–	–	–	–	–	–	–
Context × Country × Expression	(2, 569)	5.63	=0.004	0.019	0.859	*w/in South Korea*	*w/in pos. context*	Anger	5.16 (0.08)	Smile	4.63 (0.09)	Sadness	4.99 (0.09)	<0.001	=0.462	=0.012
							w/in neg. context	Anger	5.09 (0.09)	Smile	3.99 (0.10)	Sadness	4.81 (0.09)	<0.001	=0.078	<0.001
						w/in United States	*w/in pos. context*	Anger	5.19 (0.07)	Smile	4.36 (0.07)	Sadness	5.19 (0.07)	<0.001	=0.687	<0.001
							w/in neg. context	Anger	5.16 (0.08)	Smile	4.22 (0.07)	Sadness	5.05 (0.08)	<0.001	=0.867	<0.001
Affect valence analysis																
Context (pos. context, neg. context)	(1, 578)	84.80	<0.001	0.128	1.00			pos. context	3.47 (0.01)	neg. context	3.43 (0.01)	–	–	$p < 0.001$	–	–
Country (South Korea, United States)	(1, 578)	28.74	<0.001	0.047	1.00			South Korea	3.35 (0.02)	United States	3.46 (0.01)	–	–	$p < 0.001$	–	–
Expression (anger, smile, and sadness)	(2, 578)	1.54	=0.214	0.005	0.33			–	–	–	–	–	–	–	–	–

(Continued)

TABLE 3 | Continued

Description	Omnibus					Descriptive statistics								Pairwise comparisons		
	df	*F*	*p*-value	η_p^2	Observed power	Interaction variable	Interaction variable	Level 1	*M* (*SE*)	Level 2	*M* (*SE*)	Level 3	*M* (*SE*)	*p*-value 1 and 2	*p*-value 1 and 3	*p*-value 2 and 3
Valence (pos.emotion, neg.emotion)	(1, 578)	0.57	=0.450	0.001	0.12			–	–	–	–	–	–	–	–	–
Context × Valence	(1, 578)	1508.47	<0.001	0.723	1.00		*w/in pos. context*	pos.emo.	4.61 (0.04)	neg.emo.	2.17 (0.04)	–	–	<0.001	–	–
							w/in neg. context	pos.emo.	2.33 (0.04)	neg.emo.	4.52 (0.04)	–	–	<0.001	–	–
Expression × Valence	(2, 578)	151.10	<0.001	0.343	1.00		*w/in anger*	pos.emo.	3.06 (0.04)	neg.emo.	3.80 (0.04)	–	–	<0.001	–	–
							w/in smile	pos.emo.	3.89 (0.04)	neg.emo.	2.88 (0.04)	–	–	<0.001	–	–
							w/in sadness	pos.emo.	3.23 (0.04)	neg.emo.	3.59 (0.04)	–	–	<0.001		
Country × Valence	(1, 578)	0.410	=0.522	0.001	0.01			–	–	–	–	–	–	–	–	–
Context × Expression × Valence	(2, 578)	8.94	<0.001	0.030	0.97	*w/in pos. context*	*w/in pos.emo*	Anger	4.25 (0.07)	Smile	4.99 (0.07)	Sadness	4.60 (0.07)	<0.001	=0.001	=0.001
							w/in neg.emo	Anger	2.68 (0.07)	Smile	1.98 (0.07)	Sadness	2.33 (0.07)	<0.001	=0.001	<0.001
						w/in neg. context	*w/in pos.emo*	Anger	1.87 (0.06)	Smile	2.79 (0.07)	Sadness	1.86 (0.06)	<0.001	> 1.00	<0.001
							w/in neg.emo	Anger	4.92 (0.06)	Smile	3.79 (0.06)	Sadness	4.84 (0.06)	<0.001	> 1.00	<0.001
Context × Country × Valence	(1, 578)	6.19	=0.013	0.011	0.70	*w/in pos. context*	*w/in pos.emo*	South Korea	4.52 (0.06)	United States	4.71 (0.05)	–	–	=0.020	–	–
							w/in neg.emo	South Korea	2.36 (0.06)	United States	2.31 (0.05)	–	–	=0.545	–	–
						w/in neg. context	*w/in pos.emo*	South Korea	2.19 (0.06)	United States	2.16 (0.05)	–	–	=0.690	–	–
							w/in neg.emo	South Korea	4.36 (0.06)	United States	4.68 (0.05)	–	–	<0.001	–	–
Expression × Valence × Country	(2, 578)	1.16	=0.315	0.004	0.25			–	–	–	–	–	–	–	–	–
Country	(2, 578)	0.452	=0.637	0.002	0.12			–	–	–	–	–	–	–	–	–

All post hoc pairwise comparison p-values have been Bonferroni corrected for multiple comparisons.

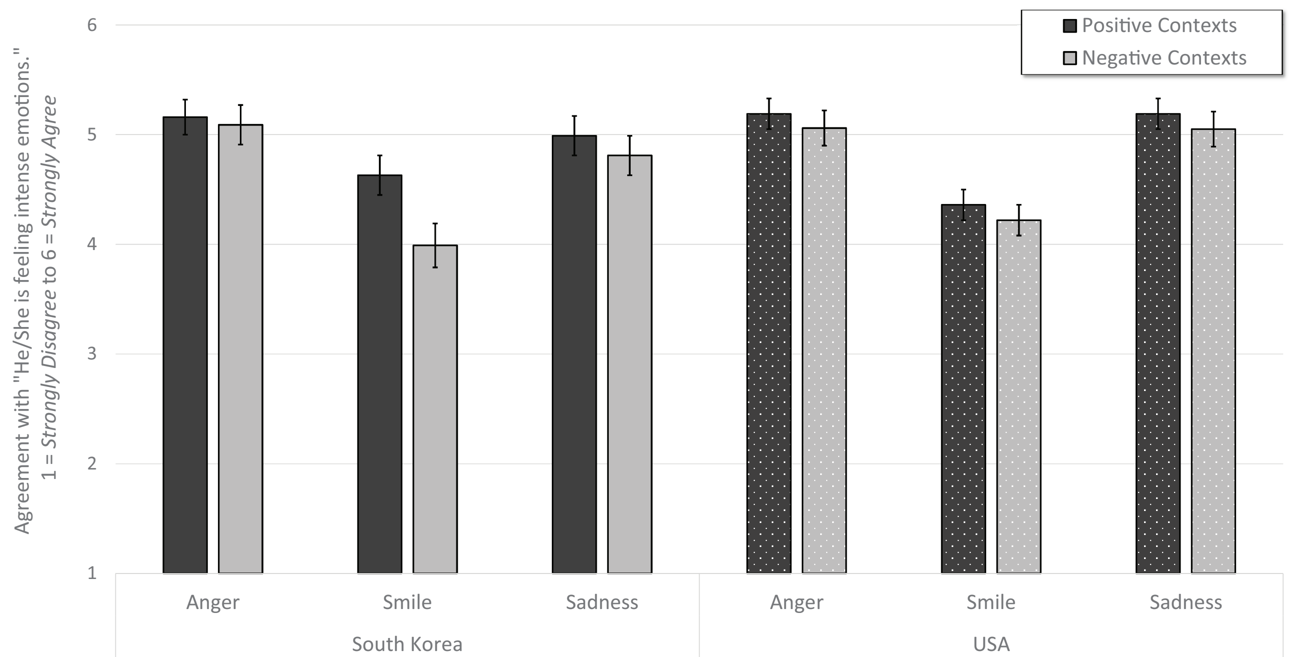


FIGURE 3 | Participants who viewed anger and sadness expressions interpreted that expressers had more intense experiences than those participants who read the same 10 vignettes in the smile condition. Smiles in positive contexts were seen as more intense than in negative contexts for South Korean participants. Error bars indicate ±2 standard errors.

emotions in negative contexts were interpreted as predominantly negative (M = 4.52, SE = 0.04), not positive (M = 2.17, SE = 0.04).

A far less robust, yet significant interaction was context × valence × country, $F(1, 578) = 6.19$, $p = 0.013$, $\eta^2_p = 0.01$, observed power = 0.70 (see **Figure 4B**). American participants (M = 4.71, SE = 0.05) interpreted more positivity in positive contexts than did South Korean participants (M = 4.52, SE = 0.06), $p = 0.020$, and American participants (M = 4.68, SE = 0.05) interpreted more negativity in negative contexts than did South Korean participants (M = 4.36, SE = 0.06), $p < 0.001$. The interaction between context and valence was also moderated by expression, $F(2, 578) = 8.94$, $p < 0.001$, $\eta^2_p = 0.03$, observed power = 0.97. This moderation was mainly driven by the smile condition. Smiles featured in negative vignettes were interpreted as representing less negative emotion ($M_{neg.}$ = 3.79, SE = 0.06) than were anger ($M_{neg.}$ = 4.92, SE = 0.06) and sadness ($M_{neg.}$ = 4.84, SE = 0.06) expressions. See **Table 3** and **Figure 4C**.

Inferred Motivational Orientations

Using the same statistical strategy, we tested participants' inferences about our models' appetitive and consummatory motivations. See **Table 4**. Central to this investigation and as hypothesized, there was a significant interaction between expression and motivational orientation, $F(2, 578) = 86.14$, $p < 0.001$, $\eta^2_p = 0.23$, observed power = 1.00. When participants were assigned randomly to view anger expressions, they interpreted that the expressers had higher appetitive (M = 3.65, SE = 0.05) than consummatory (M = 3.24, SE = 0.05) orientations, $p < 0.001$. In contrast, those participants assigned randomly to view sadness expressions inferred higher consummatory (M = 3.74, SE = 0.05) than appetitive (M = 2.83, SE = 0.05) orientations, $p < 0.001$. Participants who viewed smiles, did not distinguish between appetitive (M = 3.34, SE = 0.05) or consummatory motivations (M = 3.34, SE = 0.05), $p = 1.00$. Replicating past work (Aragón and Bargh, 2018), these effects did not depend upon whether the expressions arose in positive or negative contexts, as the expression × motivation × context interaction was not significant, $F(2, 578) = 0.139$, $p = 0.870$. See **Figure 5A**.

There was also a less robust yet significant interaction between expression × motivation × country $F(2, 578) = 10.76$, $p < 0.001$, $\eta^2_p = 0.04$, observed power = 0.99. American participants (M = 3.78, SE = 0.07) reported slightly higher agreement that anger expressions were appetitive than did South Korean participants (M = 3.52, SE = 0.07), $p = 0.009$. Americans (M = 3.94, SE = 0.06) reported slightly higher agreement that sadness expressions were consummatory than did South Korean participants (M = 3.54, SE = 0.07), $p < 0.001$. And South Korean participants (M = 3.13, SE = 0.08) did infer that smiles were less consummatory than did American participants (M = 3.55, SE = 0.06), $p < 0.001$. There was no significant interaction between context × expression × motivation × country, $p = 0.504$. See **Figure 5B**.

Seen, Known, Used: Anger, Sadness, and Smiling Expressions

Data were again analyzed in a repeated measure, general linear model, with repeated effects of item (seen, known, and

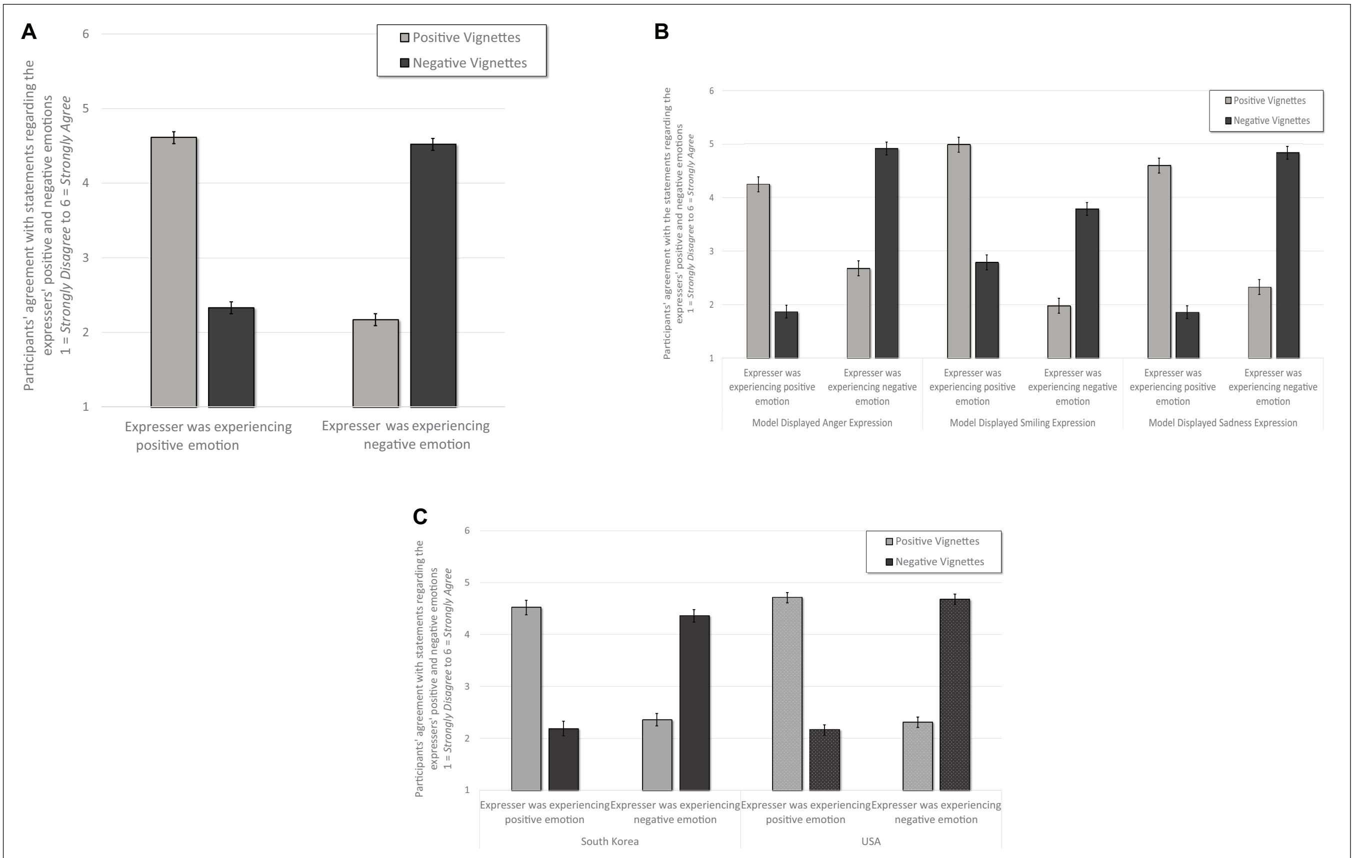


FIGURE 4 | Overall, positive contexts were deemed to produce positive- not negative emotions, and negative contexts were deemed to produce negative – not positive emotions, as in every case means and deviations clearly fell within agreement and disagreement categories. Panel **(A)** illustrates that participants' inferences about the valence of experience for the expresser was chiefly determined by the positive or negative framing of the context. Panel **(B)** shows a moderation by expression, which was chiefly driven by smiles seen as representing lower negative and higher positive emotions in negative situations. Panel **(C)** illustrates that American participants rated higher positive emotion in positive vignettes, and higher negative emotion in negative vignettes than did South Korean participants. Error bars indicate ±2 standard errors.

used) × valence (positive and negative), and vignette (5 types). Condition (anger, smiling, and sadness) and country were entered as fixed factors. As one would expect, participants reported strongest agreement for having seen an expression ($M = 3.92$, $SE = 0.04$), next highest for having known someone who expresses in such a manor ($M = 3.86$, $SE = 0.03$), and lowest scores for having expressed in such a way themselves ($M = 3.42$, $SE = 0.04$), $F(1,566) = 293.75$, $p < 0.001$, $\eta_p^2 = 0.34$, observed power = 1.00, all pairwise p's < 0.001. Item and vignette type did not interact with country, valence and country, or valence, expression and country, all p's > 0.05. Results indicated that all three questions showed a consistency that occurred across all vignette types and will be reported out here as prevalence of these expressions. Vignette and item-specific details are offered in **Table 1**.

As hypothesized, there was a large main effect of country $F(1,566) = 178.16$, $p < 0.001$, $\eta_p^2 = 0.24$, observed power = 1.00, with South Korean participants reporting a lower prevalence of the depicted expressions ($M = 3.30$, $SE = 0.05$) than American participants ($M = 4.17$, $SE = 0.04$). There was a robust interaction between valence × expression × country, $F(2,566) = 57.32$, $p < 0.001$, $\eta_p^2 = 0.17$, observed power = 1.00. American participants reported a higher prevalence of anger and sadness expressions in negative contexts ($M_{ang.} = 4.41$, $SE = 0.09$; $M_{sad.} = 4.51$, $SE = 0.09$) than in positive contexts ($M_{ang.} = 3.85$, $SE = 0.09$; $M_{sad.} = 4.08$, $SE = 0.08$). In contrast, South Korean participants reported a higher prevalence of anger and sadness expressions in positive contexts ($M_{ang.} = 3.24$, $SE = 0.10$; $M_{sad.} = 3.37$, $SE = 0.10$) than in negative contexts ($M_{ang.} = 2.58$, $SE = 0.10$; $M_{sad.} = 2.56$, $SE = 0.10$) contexts, all pairwise p's < 0.001. Additionally, although South Korean and American participants reported higher prevalence of smiles in positive contexts ($M_{SK} = 4.35$, $SE = 0.11$; $M_{USA} = 4.72$, $SE = 0.10$) than in negative contexts ($M_{SK} = 3.68$, $SE = 0.11$; $M_{USA} = 3.44$, $SE = 0.08$), both p's < 0.00, American participants reported a higher prevalence of smiles in positive contexts than did South Korean participants, $p = 0.006$. South Korean participants reported marginally higher prevalence of smiles in negative contexts than did American participants, $p = 0.079$. See **Table 4** and **Figure 6**.

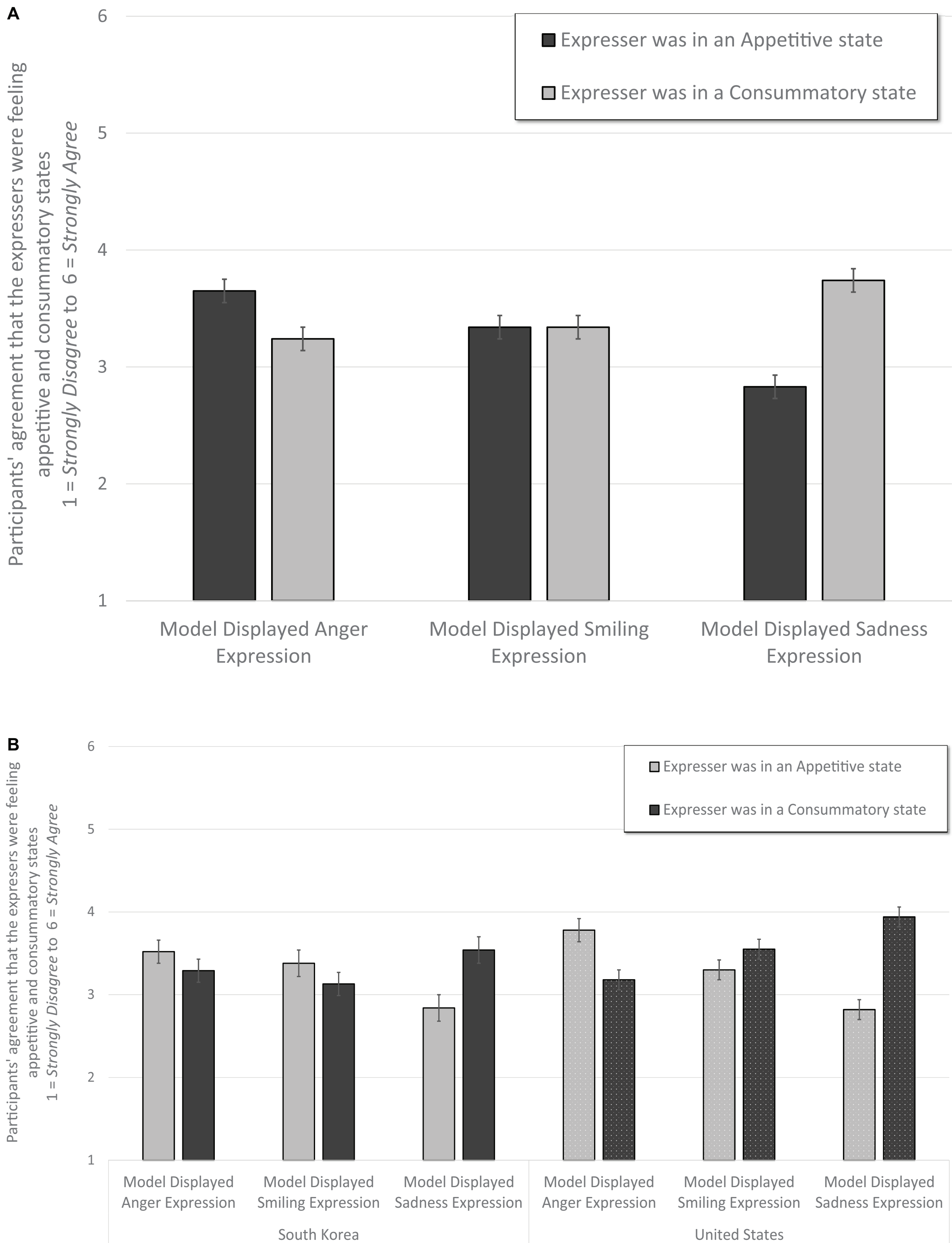


FIGURE 5 | Panel **(A)** illustrates that across these five different "flavors" of emotion, and both positive and negative events, anger expressions communicated more appetitive, less consummatory experiences, and sadness expressions communicated more consummatory, less appetitive experiences. Smiles did not differentiate between these two motivational aspects. Panel **(B)** illustrates that this largely held true cross culturally, but South Korean participants rated lower appetitive for anger and lower consummatory for sadness and smile expressions than did American participants. Error bars indicate ±standard errors.

Individual Difference Measure of Positive Emotions Expressed Through Anger and Sadness

The same analytic strategy was used to test the prevalence of positive emotion expressed through anger and sadness displays. Overall, South Korean participants (M = 3.60, SE = 0.05) did not differ from American participants (M = 3.69, SE = 0.04) in their reports of displaying normatively negative expressions when feeling highly positive emotions. There was a significant interaction between country and type of expression (anger and sadness), $F(1, 567) = 61.17$, $p < 0.001$, $\eta_p^2 = 0.09$, observed power = 1.00. South Korean participants (M = 3.84, SE = 0.05) reported higher usage of sadness expressions within positive contexts than did American participants (M = 3.48, SE = 0.05), $p < 0.001$. In contrast, American participants (M = 3.90, SE = 0.07) reported a higher usage of anger expressions in positive situations than did South Korean participants (M = 3.37, SE = 0.07), $p < 0.001$. See **Figure 7**.

In the dimorphous expression questionnaire there was an interaction of higher use of sadness than anger expressions by South Korean participants, and the higher use of anger than sadness expressions by American participants. In the item that asked participants if they had used expressions, in the mean scores the interaction was not apparent. However, when we created binary scores (those who agreed or strongly agreed that they had expressed) for our descriptive statistics, the interaction was again evident, in that for our South Korean participants 41.6% agreed that they had used the sadness expression and 26.9% agreed to have used the anger expression in at least one of the positive vignettes. In contrast, in our American participants 74.1% agreed that they had used the anger expressions, and 61.4% agreed to have used the sadness expressions in at least one of the positive vignettes.

DISCUSSION

Our central question pertained to the existence of dimorphous expressions across these two cultures in South Korea and the United States. It appears that dimorphous expressions do exist in South Korea as they have been observed in the United States, in that there are instances in which individuals report to use normatively negative expressions to express positive emotions. As well, participants from South Korea generally interpreted anger and sadness expressions within positive contexts as representing predominantly positive- not negative, or positive and negative mixed or sequentially experienced emotions. Also, consistent with dimorphous theory, anger and sadness expressions situated within positive contexts were interpreted as representing intense emotional experiences. When participants were queried if they had seen, known, or themselves used anger and sadness expressions within positive contexts, again, there was evidence for the existence of dimorphous expressions in both cultures.

An interesting pattern emerged when participants reported on their own use of anger and sadness displays when feeling positive emotions. When asked through the dimorphous expression questionnaire, the overall prevalence of dimorphous expressions did not differ by country. However, when asked within specific vignettes, and provided with specific exemplars of expression, there was a large main effect of country with South Korean participants reporting overall a lower agreement in using anger and sadness displays within positive contexts (we discuss negative contexts below). It could very well be that the vignettes were not equally compatible for both samples, i.e., if the specific vignettes did not tap into South Korean experiences as well as they had American, we might have inadvertently created the main effect of culture. Differences between South Korea and the United States did not appear for the dimorphous expression questionnaire, and those questions asked more generally about expression with a greater breadth of instances in which dimorphous expressions occur. This investigation provided evidence that dimorphous expressions exist in both South Korea and the United States, and future research will be needed to determine the extent to which these dimorphous expressions are used.

In regard to the communication of motivational orientations through expressions, the experimental portion conceptually replicated Aragón and Bargh (2018). Anger and sadness expressions communicated appetitive and consummatory motivations, respectively, in both positive and negative contexts in both the United States and South Korea. Consistent with past research, smiling expressions did not provide consistent signals about appetitive or consummatory orientations. This pattern of results speaks to a possible functional reason for why smiles are not the only expressions that arise for positive feelings. When experiencing highly intense positive feelings paired with an antsy feeling of wanting to go or a consuming feeling of wanting to stop, anger and sadness displays, respectively, communicate those feelings better than do smiles. It seems that this would be important social information to be able to communicate when emotions are running high for the coordination, cooperation, and compensatory behaviors that facilitate social interactions.

Another intriguing pattern in our results was that in many of the analyses South Korean and American participants were in agreement as to what a certain expression did not represent, but when it came to stating what the expression did represent it seemed that the South Korean participants were less adamant about what they were viewing. For example, both South Korean and American participants agreed that anger and sadness expressions in negative contexts were not representing positive emotions, but South Koreans appeared less adamant that they were negative emotions. The same was true for anger and sadness expressions in positive contexts, there was cross-cultural agreement that they were not negative experiences, but South Korean participants were less extreme in rating how positive they were. The same pattern emerged when evaluating anger and sadness expressions for motivational orientations. There was cross-cultural agreement that anger expressions were not consummatory and sadness expressions were not appetitive, but again, South Korean participants were not as emphatic that anger expressions were appetitive and sadness expressions were consummatory as American participants. This suggests a possible reporting bias because in each case, South Korean participants were on par with Americans in declaring what an expression did not communicate, but they were less confident to say what it did communicate.

TABLE 4 | General linear repeated measures models testing inferred motivational orientations, and seen, known, used items.

Description	Omnibus					Descriptive statistics								Pairwise comparisons		
	df	***F***	***p*-value**	η_p^2	**Observed power**	**Interaction variable**	**Interaction variable**	**Level 1**	***M* (*SE*)**	**Level 2**	***M* (*SE*)**	**Level 3**	***M* (*SE*)**	**Levels 1 and 2**	**Levels 1 and 3**	**Levels 2 and 3**
Motivational orientations analysis																
Context (pos. context, neg. context)	(1, 578)	42.77	<0.001	0.069	1.00			pos. context	3.43 (0.02)	neg. context	3.28 (0.02)	–	–	<0.001	–	–
Country	(1, 578)	14.85	<0.001	0.025	0.970			South Korea	3.28 (0.03)	United States	3.43 (0.02)	–	–	<0.001	–	–
Expression	(2, 578)	6.42	=0.002	0.022	0.903			Anger	3.44 (0.03)	Smile	3.34	Sadness	3.28 (0.03)	<0.001	=0.635	=0.077
Motivation (appetitive., consumm.)	(1, 578)	15.03	<0.001	0.025	0.972			Appetitive.	3.27 (0.03)	consumm.	3.44 (0.03)	–	–	<0.001	–	–
Context × Motivation	(1, 578)	211.92	<0.001	0.268	1.00		*pos. context*	Appetitive.	3.64 (0.04)	consumm.	3.21 (0.04)	–	–	<0.001	–	–
							neg. context	Appetitive.	2.90 (0.04)	consumm.	3.67 (0.04)	–	–	<0.001	–	–
Expression × Motivation	(2, 578)	86.14	<0.001	0.230	1.00		*w/in anger*	Appetitive.	3.65 (0.05)	consumm.	3.24 (0.05)	–	–	<0.001	–	–
							w/in smile	Appetitive.	3.34 (0.05)	consumm.	3.34 (0.05)	–	–	=0.997	–	–
							w/in sadness	Appetitive.	2.83 (0.05)	consumm.	3.74 (0.05)	–	–	<0.001	–	–
Country × Motivation	(1, 578)	4.614	=0.032	0.008	0.573	*w/in appetitive.*		South Korea	3.25 (0.04)	United States	3.30 (0.04)	–	–	=0.352	–	–
						w/in consumm.		South Korea	3.32 (0.04)	United States	3.56 (0.04)	–	–	<0.001	–	–
Context × Expression × Motivation	(2, 578)	0.139	=0.870	0.000	0.07			–	–	–	–	–	–	–	–	–
Context × Country × Motivation	(1, 578)	341.07	<0.001	0.371	1.00	*w/in pos. context*	*w/in app.mot.*	South Korea	3.95 (0.06)	United States	3.34 (0.05)	–	–	<0.001	–	–
							w/in cons.mot.	South Korea	2.66 (0.05)	United States	3.75 (0.05)	–	–	<0.001	–	–
						w/in neg. context	*w/in app.mot.*	South Korea	2.55 (0.06)	United States	3.26 (0.05)	–	–	<0.001	–	–
							w/in cons.mot.	South Korea	3.98 (0.06)	United States	3.36 (0.05)	–	–	<0.001	–	–

(Continued)

TABLE 4 | Continued

Description	Omnibus					Descriptive statistics								Pairwise comparisons		
	df	***F***	***p*-value**	η_p^2	**Observed power**	**Interaction variable**	**Interaction variable**	**Level 1**	***M (SE)***	**Level 2**	***M (SE)***	**Level 3**	***M (SE)***	**Levels 1 and 2**	**Levels 1 and 3**	**Levels 2 and 3**
Expression × Motivation × Country	(2, 578)	10.76	0.000	0.036	0.99	*w/in anger*	*w/in app.mot.*	South Korea	3.52 (0.07)	United States	3.78 (0.07)	–	–	=0.009	–	–
							w/in cons.mot.	South Korea	3.29 (0.07)	United States	3.18 (0.06)	–	–	=0.244	–	–
						w/in smile	*w/in app.mot.*	South Korea	3.38 (0.08)	United States	3.30 (0.06)	–	–	=0.478	–	–
							w/in cons.mot.	South Korea	3.13 (0.08)	United States	3.55 (0.06)	–	–	<0.001	–	–
						w/in sadness	*w/in app.mot.*	South Korea	2.84 (0.08)	United States	2.82 (0.06)	–	–	=0.802	–	–
							w/in cons.mot.	South Korea	3.54 (0.07)	United States	3.94 (0.06)	–	–	<0.001	–	–
Context × Exp. × Motiv. × Country	(2, 578)	0.686	=0.504	0.002	0.17			–	–	–	–	–	–	–	–	–
Seen, known, and used the expressions analysis																
Context (pos. context, neg. context)	(1, 566)	101.41	<0.001	0.152	1.00			pos. context	3.94 (0.04)	neg. context	3.53 (0.04)	–	–	<0.001	–	–
Country (South Korea, United States)	(1, 566)	178.16	<0.001	0.239	1.00			South Korea	3.30 (0.05)	United States	4.17 (0.04)	–	–	<0.001	–	–
Expression (anger, sadness, and smile)	(2, 566)	23.64	<0.001	0.077	1.00			Anger	3.52 (0.06)	Smile	4.05 (0.06)	Sadness	3.63 (0.06)	<0.001	=0.173	<0.001
Item (seen, known, and used)	(2, 566)	293.75	<0.001	0.342	1.00			Seen	3.92 (0.04)	Known	3.86 (0.03)	Used	3.42 (0.04)	=0.001	<0.001	<0.001
Context × Country	(1, 566)	58.53	<0.001	0.094	1.00	*w/in pos. context*		South Korea	3.65 (0.06)	United States	4.22 (0.05)	–	–	<0.001	–	–
						w/in neg. context		South Korea	2.94 (0.06)	United States	4.12 (0.05)	–	–	<0.001	–	–

(Continued)

TABLE 4 | Continued

Description	Omnibus					Descriptive statistics								Pairwise comparisons		
	df	*F*	*p*-value	η_p^2	Observed power	Interaction variable	Interaction variable	Level 1	*M* (*SE*)	Level 2	*M* (*SE*)	Level 3	*M* (*SE*)	Levels 1 and 2	Levels 1 and 3	Levels 2 and 3
Context × Expression	(2, 566)	50.10	<0.001	0.150	1.00	*w/in anger*		pos. context	3.55 (0.07)	neg. context	3.50 (0.07)	–	–	=0.446	–	–
						w/in smile		pos. context	4.53 (0.07)	neg. context	3.56 (0.07)	–	–	<0.001	–	–
						w/in sadness		pos. context	3.73 (0.07)	neg. context	3.54 (0.07)	–	–	=0.006	–	–
Country × Expression	(2, 566)	37.80	<0.001	0.118	1.00	*w/in anger*		South Korea	2.91 (0.08)	United States	4.13 (0.08)	–	–	<0.001	–	–
						w/in smile		South Korea	4.01 (0.09)	United States	4.08 (0.07)	–	–	=0.568	–	–
						w/in sadness		South Korea	2.96 (0.09)	United States	4.30 (0.07)	–	–	<0.001	–	–
Context × Expression × Country	(2, 566)	57.32	<0.001	0.168	1.00	*w/in pos. context*	*w/in anger*	South Korea	3.24 (0.10)	United States	3.85 (0.09)	–	–	<0.001	–	–
							w/in smile	South Korea	4.35 (0.11)	United States	4.72 (0.08)	–	–	=0.006	–	–
							w/in sadness	South Korea	3.37 (0.10)	United States	4.08 (0.08)	–	–	<0.001	–	–
						w/in neg. context	*w/in anger*	South Korea	2.58 (0.10)	United States	4.41 (0.09)	–	–	<0.001	–	–
							w/in smile	South Korea	3.68 (0.11)	United States	3.44 (0.08)	–	–	=0.079	–	–
							w/in sadness	South Korea	2.56 (0.10)	United States	4.51 (0.09)	–	–	<0.001	–	–

All post hoc pairwise comparison p-values have been Bonferroni corrected for multiple comparisons.

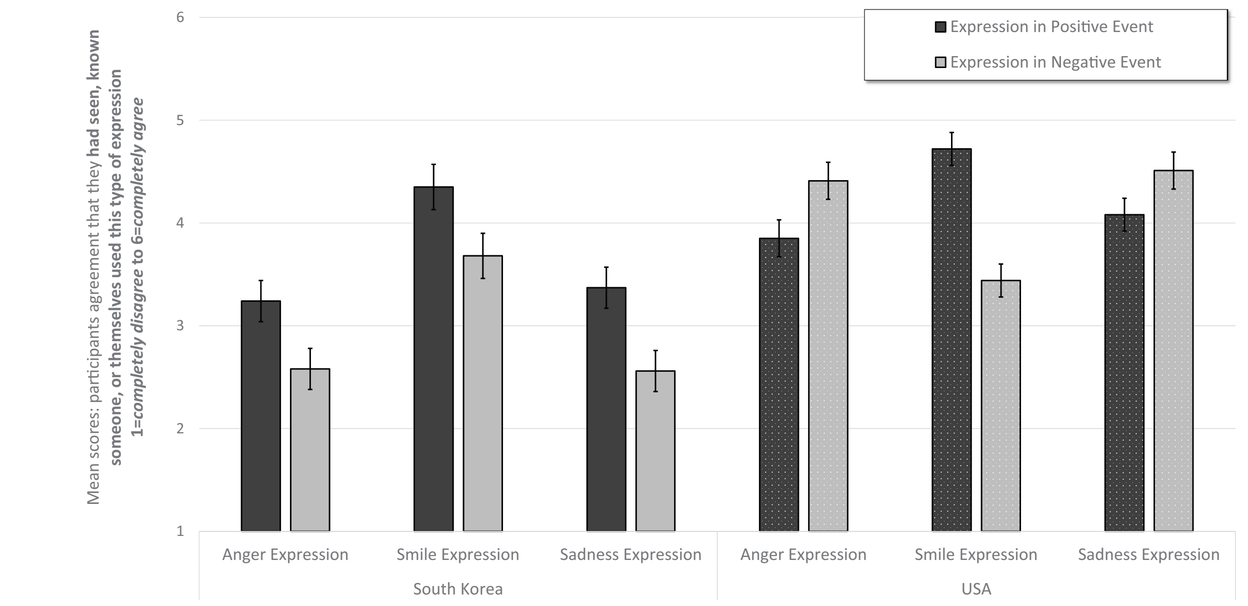


FIGURE 6 | Figure illustrates the prevalence of anger, sadness or smiling expression in positive and negative events. Error bars indicate ±2 standard errors.

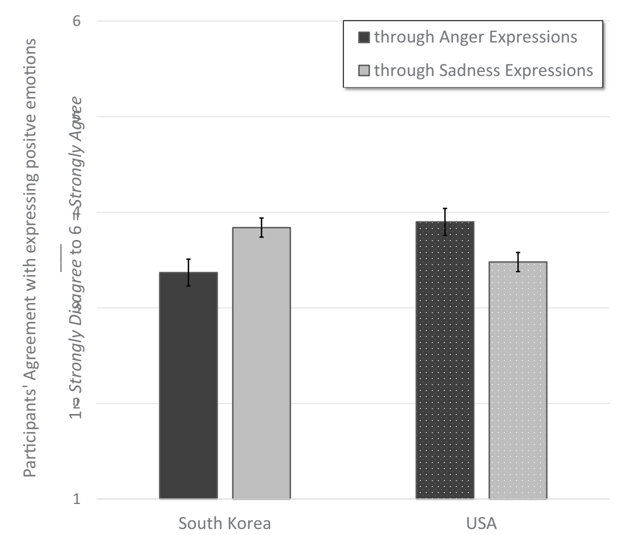


FIGURE 7 | Panel illustrates that South Korean participants reported expressing positive emotions with more sadness and fewer anger expressions than did American participants. Error bars indicate ±2 standard errors.

The addition of the smiling condition was intended as type of control condition because smiles are presumed to be the normative expression for positive emotions, and our central aim was to understand anger and sadness displays in positive contexts in these two cultures. However, the smile condition provided the most complex, least straightforward results of our investigation. Here we attempt to address these complexities. Concerning what was communicated by smiles, generally, when smiles were presented in positive contexts they were interpreted as representing positive experiences, and when smiles were presented in negative contexts they were interpreted as representing negative experiences. As previous research has demonstrated smiles did not clearly communicate appetitive or consummatory motivations, but South Korean participants appeared to read smiles as being more so appetitive and less consummatory than did American participants (see **Figure 5B**). Work in Affect Valuation Theory has found that more subtle smiles are more so the norm in Eastern relative to Western contexts (Tsai et al., 2019). The types of smiles in our stimuli contained 7 smiles that exposed teeth, and 4 which were open-mouthed. Thus, it is possible that the larger, toothier grins may have seemed to communicate a higher-arousal emotion to our South Korean participants.

As previous research had suggested South Korean participants did endorse that they had seen, known and used smiles more so in negative contexts than did American participants. This phenomenon is thought to represent masking of negative emotions. We note though that smiles in negative contexts could represent masking, but they could also represent dimorphous expressions of negative emotions, reappraisal of the negative experience, or mixed or sequentially experienced positive and negative emotions. In a similar experiment that provided participants text boxed with which to comment about how they would feel if smiling in context of losing an important sporting event (Aragón, 2020), participants noted masking "(I would) try to not show how I am having very negative feelings about losing," dimorphous expressions "I would feel so frustrated and upset that I would laugh. Sometimes when I am frustrated or fed up with something, I laugh (but not because I am happy)," reappraisal "I was pleased with my effort and gave it everything I had," and mixed or sequentially experienced positive and negative emotions "I would feel discouraged but also proud." Therefore, it is possible that any of these factors, i.e., masking, dimorphous expression of negative emotions, the tendency to reappraise (Butler et al., 2007; English et al., 2017), or the prevalence of mixed or sequential emotions could account for the differences we observed in the interpretation and prevalence of smiles across both positive and negative contexts in South Korean and American participants. These findings will be interesting to probe in future research.

Additionally, previously reported display rules research has shown that individuals from eastern contexts are less likely to express negative emotions than are those from western contexts. However, the results from the self-report dimorphous expression questionnaire suggest that those display rules may be less tied to physical displays and more tied to rules about which emotional experiences are appropriate to communicate, because overall normatively negative displays that communicated positive emotional experiences were equally prevalent in both American and South Korean participants. In the items that asked if participants had seen, known or used expressions, expression and emotion appeared to contribute independently to the prevalence of such expressions, because anger and sadness displays were less prevalent in South Korean participants (effect of expression), but particularly so when they were communicating negative emotions (moderated by the effect of the communicated valence of emotion).

Most astounding was the interaction between country, context, and expression in regard to anger and sadness displays in the items that asked participants about having seen, known, or used expressions in the given situations. South Korean participants reported a lower prevalence in the use of anger and sadness expressions overall, but if they were to use those expressions, they reported using those expressions to express positive more so than negative emotions. The reverse was true for our American participants, who were significantly more likely to use anger and sadness expressions for the display of negative than positive emotions. Future work might explore if people from Eastern contexts might try to maintain social harmony through the use of dimorphous expressions, particularly the sadness expression. In an Eastern context that is sensitive to power distance, particularly in cases in which one is opened up to envious attacks such as when an individual has won an award, or experienced a great windfall, it might be prudent to express positive emotions through sadness or crying because such displays been found to reduce aggressive sentiments and upregulate caring responses toward the expresser (Hendriks et al., 2008; Aragón and Clark, 2018).

Limitations of this investigation include those issues highlighted above, i.e., the use of more pronounced smiles and the use of vignettes that might not have been equally compatible for both cultures. This study is also limited in that it was entirely self-report, and as such is vulnerable to issues of self-knowledge and self-presentation. Likewise, the study was conducted online, which always leaves open the possibility waning attention and effort provided by the study's participants. Another limitation is that our stimuli is not equivalent to real-life instances. We attempted to ameliorate this shortcoming by using many different types of exemplars of expression, with both male and female models, across different types of scenarios.

Returning to the idea presented in the opening of this article, the assumption of a 1:1 correspondence between expressive displays and discrete emotional experiences would have precluded this investigation. Participants in this study and others have demonstrated and reported that expressions normatively considered negative, in this case anger and sadness expressions, can represent and communicate either positive or negative states, and can appear in different types of contexts that should elicit different what we call "flavors" of emotion. This investigation found the first evidence in both South Korea and the United States that anger and sadness displays communicate appetitive and consummatory motivational orientations, respectively. It *could* be that anger and sadness displays simply are expressions of intensity and motivation. Of course, future work will need to be done to know if that might be true. Clearly future investigations that explore manipulations of in-group/out-group status, gender, or social status of the expresser could prove interesting. Also, future research might consider if a singular expresser versus multiple expressers interact with culture. An overarching conclusion is that sometimes questioning the foundation on which our work in non-verbal behavior has been built can lead to questions that may inform subsequent work in ways that had not been considered. We hope to have made such a contribution with this work.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Clemson IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SS and OA designed and implemented the experiments, with input from AC. OA analyzed the data. SS, AC, and OA wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the Clemson University Creative Inquiry Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.579509/full#supplementary-material

REFERENCES

Ansfield, M. E. (2007). Smiling when distressed: when a smile is a frown turned upside down. *Pers. Soc. Psychol. Bull.* 33, 763–775. doi: 10.1177/0146167206297398

Aragón, O. R. (2016). "I couldn't help but to cry!" "I couldn't help but to yell "Yes!"" Dimorphous expressions inform consumers of users' motivational orientations," in *Advances in Consumer Research*, Vol. 44, eds P. Moreau and S. Puntoni (Duluth, MN: Association for Consumer Research), 384–385.

Aragón, O. R. (2017). "Tears of joy" and "tears and joy?" personal accounts of dimorphous and mixed expressions of emotion. *Motiv. Emot.* 41, 370–392. doi: 10.1007/s11031-017-9606-x

Aragón, O. R. (2020). "Using consumer neuroscience to explain product choice," in *Proceedings of the Association for Consumer Research, North American Conference*, Paris.

Aragón, O. R., and Bargh, J. A. (2018). "So happy I could shout!" and "So happy I could cry!" Dimorphous expressions represent and communicate motivational aspects of positive emotions. *Cogn. Emot.* 32, 286–302. doi: 10.1080/02699931.2017.1301388

Aragón, O. R., and Clark, M. S. (2018). "Tears of joy" and "smiles of joy" prompt distinct patterns of interpersonal emotion regulation. *Cogn. Emot.* 32, 913–940. doi: 10.1080/02699931.2017.1360253

Aragón, O. R., Clark, M. S., Dyer, R. L., and Bargh, J. A. (2015). Dimorphous expressions of positive emotion: displays of both care and aggression in response to cute stimuli. *Psychol. Sci.* 26, 259–273. doi: 10.1177/0956797614561044

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. *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

Berridge, K. C., and Robinson, T. E. (2003). Parsing reward. *Trends Neurosci.* 26, 507–513. doi: 10.1016/s0166-2236(03)00233-9

Bonanno, G. A., and Keltner, D. (2004). The coherence of emotion systems: comparing "on-line" measures of appraisal and facial expressions, and self-report. *Cogn. Emot.* 18, 431–444. doi: 10.1080/02699930341000149

Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9

Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301

Butler, E. A., Lee, T. L., and Gross, J. J. (2007). Emotion regulation and culture: are the social consequences of emotion suppression culture-specific? *Emotion* 7, 30–48. doi: 10.1037/1528-3542.7.1.30

Carrera, P., and Oceja, L. (2007). Drawing mixed emotions: sequential or simultaneous experiences? *Cogn. Emot.* 21, 422–441. doi: 10.1080/02699930600557904

Carroll, J. M., and Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *J. Pers. Soc. Psychol.* 70, 205–218. doi: 10.1037/0022-3514.70.2.205

Chim, L., Hogan, C. L., Fung, H. H. H., and Tsai, J. L. (2018). Valuing calm enhances enjoyment of calming (vs. exciting) amusement park rides and exercise. *Emotion* 18, 805–818. doi: 10.1037/emo0000348

Danesi, M. (2017). *The Semiotics of Emoji: The Rise of Visual Language in The Age of the Internet*. London, UK: Bloomsbury Publishing.

Durán, J., Reisenzein, R., and Fernández-Dols, J. M. (2017). "Coherence between emotions and facial expressions," in *The Science of Facial Expressions*, eds J.-M. Fernández-Dols and J. A. Russell (Oxford: Oxford University Press), 107–129.

Ekman, P. (1972). "Universal and cultural differences in facial expression of emotion," in *Nebraska Symposium on Motivation*, ed. J. R. Cole (Lincoln: Nebraska University Press), 207–283.

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

Ekman, P., and Friesen, W. V. (1982). Felt, false, and miserable smiles. *J. Nonverbal Behav.* 6, 238–252. doi: 10.1007/bf00987191

Ekman, P., and Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motiv. Emot.* 10, 159–168. doi: 10.1007/bf00992253

English, T., Lee, I. A., John, O. P., and Gross, J. J. (2017). Emotion regulation strategy selection in daily life: the role of social context and goals. *Motiv. Emot.* 41, 230–242. doi: 10.1007/s11031-016-9597-z

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *J. Econ. Growth* 18, 195–222.

Feldman Barrett, L. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cogn. Emot.* 12, 579–599. doi: 10.1080/026999398379574

Fernández-Dols, J. M., Carrera, P., and Crivelli, C. (2010). Facial behavior while experiencing sexual excitement. *J. Nonverbal Behav.* 35, 63–71. doi: 10.1007/s10919-010-0097-7

Fernandez-Dols, J. M., and Crivelli, C. (2013). Emotion and expression: naturalistic studies. *Emot. Rev.* 5, 24–29. doi: 10.1177/1754073912457229

Fernández-Dols, J. M., and Ruiz-Belda, M. A. (1995). Are smiles a sign of happiness? Gold medal winners at the olympic games. *J. Pers. Soc. Psychol.* 69, 1113–1119. doi: 10.1037/0022-3514.69.6.1113

Fernández-Dols, J. M., Sánchez, F., Carrera, P., and Ruiz-Belda, M. A. (1997). Are spontaneous expressions and emotions linked? An experimental test of coherence. *J. Nonverbal Behav.* 21, 163–177.

Fredrickson, B. L., and Levenson, R. W. (1998). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cogn. Emot.* 12, 191–220. doi: 10.1080/026999398379718

Friesen, W. V. (1972). *Cultural Differences in Facial Expressions in A Social Situation: An Experimental Test of the Concept of Display Rules*. Unpublished doctoral dissertation, University of California, San Francisco.

Frijda, N. H., Kuipers, P., and Terschure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *J. Pers. Soc. Psychol.* 57, 212–228. doi: 10.1037//0022-3514.57.2.212

García-Higuera, J.-A., Crivelli, C., and Fernández-Dols, J.-M. (2015). Facial expressions during an extremely intense emotional situation: toreros' lip funnel. *Soc. Sci. Inform.* 54, 439–454. doi: 10.1177/0539018415596381

Gross, J. J. (2001). Emotion regulation in adulthood: timing is everything. *Pers. Soc. Psychol. Bull.* 34, 396–406.

Gross, J. J., and John, O. P. (1995). Facets of emotional expressivity: three self-report factors and their correlates. *Pers. Individ. Diff.* 19, 555–568. doi: 10.1016/0191-8869(95)00055-b

Hendriks, M. C., Croon, M. A., and Vingerhoets, A. J. (2008). Social reactions to adult crying: the help-soliciting function of tears. *J. Soc. Psychol.* 148, 22–41. doi: 10.3200/SOCP.148.1.22-42

Higgins, T. (1997). Beyond pleasre and pain. *Am. Psychol.* 52, 1281–1300.

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

Kim, J., Seo, M., Yu, H., and Neuendorf, K. (2014). Cultural differences in preference for entertainment messages that induce mixed responses of joy and sorrow. *Hum. Commun. Res.* 40, 530–552. doi: 10.1111/hcre.12037

Larsen, J. T., McGraw, A. P., and Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *J. Pers. Soc. Psychol.* 81, 684–696. doi: 10.1037/0022-3514.81.4.684

Lundqvist, D., Flykt, A., and Ohman, A. (1998). *The Karolinska Directed Emotional Faces- KDEF*. Solna: Karolinska Institutet.

Markus, H. R., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295x.98.2.224

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motiv. Emot.* 14, 195–214. doi: 10.1007/bf00995569

Matsumoto, D., and Kupperbusch, C. (2001). Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. *Asian J. Soc. Psychol.* 4, 113–131. doi: 10.1111/j.1467-839X.2001.00080.x

Matsumoto, D., Seung Hee, Y., and Fontaine, J. (2008a). Mapping expressive differences around the world. *J. Cross Cult. Psychol.* 39, 55–74. doi: 10.1177/0022022107311854

Matsumoto, D., Yoo, S. H., Nakagawa, S., and Members of the Multinational Study of Cultural Display (2008b). Culture, emotion regulation, and adjustment. *J. Pers. Soc. Psychol.* 94, 925–937. doi: 10.1037/0022-3514.94.6.925

Mogilner, C., Aaker, J. L., and Kamvar, S. D. (2012). How happiness affects choice. *J. Consum. Res.* 39, 429–443. doi: 10.1086/663774

Park, B., Genevsky, A., Knutson, B., and Tsai, J. (2020). Culturally valued facial expressions enhance loan request success. *Emotion* 20, 1137–1153. doi: 10.1037/emo0000642

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

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

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295x.110.1.145

Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emot. Rev.* 9, 111–117. doi: 10.1177/1754073916639658

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

Safdar, S., Friedlmeier, W., Matsumoto, D., Yoo, S. H., Kwantes, C. T., Kakai, H., et al. (2009). Variations of emotional display rules within and across cultures: a comparison between Canada, USA, and Japan. *Can. J. Behav. Sci.* 41, 1–10. doi: 10.1037/a0014387

Shutterstock (2019). *Images Purchased with Standard License*. London: Shutterstock.

Sims, T., Koopmann-Holm, B., Young, H. R., Jiang, D., Fung, H., and Tsai, J. L. (2018). Asian Americans respond less favorably to excitement (vs. calm)-focused physicians compared to European Americans. *Cult. Divers. Ethn. Minority Psychol.* 24, 1–14. doi: 10.1037/cdp0000171

Sims, T., Tsai, J. L., Jiang, D., Wang, Y., Fung, H. H., and Zhang, X. (2015). Wanting to maximize the positive and minimize the negative: implications for mixed affective experience in American and Chinese contexts. *J. Pers. Soc. Psychol.* 109, 292–315. doi: 10.1037/a0039276

Smith, C. A., and Ellsworth, P. C. (1987). Patterns of appraisal and emotion related to taking an exam. *J. Pers. Soc. Psychol.* 52, 475–488. doi: 10.1037/0022-3514.52.3.475

Tsai, J. L., Ang, J. Y. Z., Blevins, E., Goernandt, J., Fung, H. H., Jiang, D., et al. (2016). Leaders' smiles reflect cultural differences in ideal affect. *Emotion* 16, 183–195. doi: 10.1037/emo0000133

Tsai, J. L., Blevins, E., Bencharit, L. Z., Chim, L., Fung, H. H., and Yeung, D. Y. (2019). Cultural variation in social judgments of smiles: the role of ideal affect. *J. Pers. Soc. Psychol.* 116, 966–988. doi: 10.1037/pspp0000192

Tsai, J. L., Knutson, B., and Fung, H. H. (2006). Cultural variation in affect valuation. *J. Pers. Soc. Psychol.* 90, 288–307. doi: 10.1037/0022-3514.90.2.288

Wenzler, S., Levine, S., van Dick, R., Oertel-Knochel, V., and Aviezer, H. (2016). Beyond pleasure and pain: facial expression ambiguity in adults and children during intense situations. *Emotion* 16, 807–814. doi: 10.1037/emo0000185

Yamaguchi, S. (1994). "Collectism among the Japanese: a perspective from the self," in *Individualism and Collectivism: Theory, Method, and Applications*, Vol. 18, eds U. Kim and H. C. Triandis (Thousand Oaks, CA: Sage), 175–188.

Conflict of Interest: 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.

frontiers in Psychology

PERSPECTIVE
published: 14 January 2021
doi: 10.3389/fpsyg.2020.611248

Expression Authenticity: The Role of Genuine and Deliberate Displays in Emotion Perception

Mircea Zloteanu[1,2] and Eva G. Krumhuber[3]*

[1] Department of Criminology and Sociology, Kingston University London, Kingston, United Kingdom, [2] Department of Psychology, Kingston University London, Kingston, United Kingdom, [3] Department of Experimental Psychology, University College London, London, United Kingdom

OPEN ACCESS

Edited by:
Marianne Schmid Mast,
University of Lausanne, Switzerland

Reviewed by:
Walter Gerbino,
University of Trieste, Italy
Steven Grant Greening,
University of Manitoba, Canada

****Correspondence:***
Mircea Zloteanu
m.zloteanu@kingston.ac.uk;
m.zloteanu@ucl.ac.uk

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *28 September 2020*
Accepted: *21 December 2020*
Published: *14 January 2021*

Citation:
Zloteanu M and Krumhuber EG (2021) Expression Authenticity: The Role of Genuine and Deliberate Displays in Emotion Perception. Front. Psychol. 11:611248. doi: 10.3389/fpsyg.2020.611248

People dedicate significant attention to others' facial expressions and to deciphering their meaning. Hence, knowing whether such expressions are genuine or deliberate is important. Early research proposed that authenticity could be discerned based on reliable facial muscle activations unique to genuine emotional experiences that are impossible to produce voluntarily. With an increasing body of research, such claims may no longer hold up to empirical scrutiny. In this article, expression authenticity is considered within the context of senders' ability to produce convincing facial displays that resemble genuine affect and human decoders' judgments of expression authenticity. This includes a discussion of spontaneous vs. posed expressions, as well as appearance- vs. elicitation-based approaches for defining emotion recognition accuracy. We further expand on the functional role of facial displays as neurophysiological states and communicative signals, thereby drawing upon the *encoding-decoding* and *affect-induction* perspectives of emotion expressions. Theoretical and methodological issues are addressed with the aim to instigate greater conceptual and operational clarity in future investigations of expression authenticity.

Keywords: emotion, facial expressions, genuine, posed and spontaneous, authenticity discrimination

INTRODUCTION

The accurate recognition of emotions plays a crucial role in communication and social interaction. Knowing what another person is feeling is relevant for predicting their psychological state, likely future behavior, and interaction outcome (Hall et al., 2009). However, the advantage of such knowledge hinges on the emotional displays matching the person's true underlying affect.

Humans have great control over their facial behavior (Zuckerman et al., 1986; Smith, 2004), with the ability to produce complex expressions. This implies that not all displays genuinely reflect a person's underlying emotional state (Barrett, 2006). Deliberate expressions reflect the strategic intent of the individual in the absence of felt emotions (Ekman and Rosenberg, 2005). During social interaction, individuals may consciously regulate and suppress their emotions and portray expressions of unfelt emotions. This raises the issue of *expression authenticity*.

While people seem capable of recognizing emotions from specific facial configurations with high accuracy (Calvo and Nummenmaa, 2015),[1] the ability to distinguish the authenticity of such expressions is much poorer (Ekman and O'Sullivan, 1991; Hess and Kleck, 1994; McLellan et al., 2010). Interestingly, the reason(s) for this has not been fully determined yet, with difficulties in explanations partly stemming from disagreements about the nature of emotions and the function of facial expressions. Recent propositions have attempted to elucidate some of the inconsistencies of past research, considering facial expressions as both *innate cues* and *communicative signals* (Crivelli and Fridlund, 2018; Barrett et al., 2019). Here, we build on this work, thereby focusing on human expression authenticity judgments: assessing whether the emotional expression displayed by another person is a genuine reflection of their underlying affect. This operational definition is representative of the task participants typically perform and the instructions they receive; however, as will be discussed, how one conceptualizes emotions and operationalizes facial expressions will ultimately determine what an authenticity judgment indicates and the inferences that can be drawn from it.[2]

THE FUNCTION OF FACIAL EXPRESSIONS: ENCODING-DECODING VS. AFFECT-INDUCTION

Conceptually, there are two main perspectives regarding the function of facial expressions. These view facial displays either as (a) innate cues reflecting genuine affect or (b) communicative signals of affect and intent.

According to the *encoding-decoding* perspective (Ekman, 2003), observers (called *decoders*) "decode" the meaning behind emotional displays of others (called *senders*). Facial expressions are considered to be innate, neurologically activated, fixed facial muscle patterns that occur in response to emotion-eliciting stimuli (Tomkins, 1962). Their appearance is an evolved response to specific events that are difficult (if not impossible) to suppress (Hurley and Frank, 2011), resulting in facial *leakage* (i.e., involuntary displays of felt emotions; Ekman, 1997). As such, they are functionally not a source of emotional information but became so as an exaptation (Darwin, 1872). Under this account, voluntary modulations of expressions come in the form of *display rules*, which are socio-cultural norms regulating the expression of displays (Ekman and Friesen, 1971). For expression authenticity, this perspective places emphasis on presumed *reliable muscles*, which are facial markers said to activate only during felt affect and being impossible to voluntarily control (Ekman, 2003). Under this view, differences between genuine and deliberate displays exist, and expression authenticity is a function of the decoder's perceptual ability and knowledge for making accurate inferences. While being popular, this view has been criticized due to its vague conceptualization and lack of empirical support (Barrett et al., 2019).

According to the *affect-induction* perspective (Crivelli and Fridlund, 2018, 2019),[3] facial displays function as a signaling mechanism to communicate one's emotional states, motivating corresponding states in the observer (called *receivers*). Evolutionary there is no reason why facial expressions and emotion perception could not have co-evolved as part of a social signaling system (Izard, 1994; Dezecache et al., 2013). Indeed, a growing body of evidence suggests that humans are adept at producing facial expressions for communicative reasons (Smith, 2004). Under this view, the function of emotional displays is to signal emotional information and intent (i.e., they are not cryptic "cues" needed to be decoded; Crivelli and Fridlund, 2019). This perspective is not without its limitations. For instance, there are clear examples of behavior, such as blushing (Crozier, 2010), which can be used to infer the emotional states the sender may wish to suppress but is unable to do so. Also, the perspective does not adequately account for emotional leakage or solitary reactions (e.g., smiling when alone; but see Crivelli and Fridlund, 2018).

When synthesized both perspectives are useful for understanding human expression authenticity judgments. For instance, the encoding-decoding perspective provides the foundation for considering genuine (i.e., innate, involuntary responses) and deliberate (i.e., voluntary, communicative signals) expressions. It is important to note, though, that the argument for clear differences in expression authenticity (Ekman et al., 1988) is neither consistent with empirical investigations (Barrett et al., 2019) nor reflected in human judgments of facial expressions (Zloteanu et al., 2020, 2018). First, senders seem to possess the ability to produce genuine-looking displays of emotion (Surakka and Hietanen, 1998; Gosselin et al., 2010; Gunnery et al., 2013). Second, when considering facial expressions as social signals, as done in the affect-induction perspective, it is possible to understand why expression authenticity judgments are relatively poor. In deceptive scenarios, deliberate emotional cues serve an obvious communicative purpose: they convey an affective state to an observer which strategically benefits the sender (maliciously or otherwise).

SPONTANEOUS VS. POSED: A SUFFICIENTLY NUANCED DICHOTOMY?

Irrespective of the perspective adopted, researchers typically employ an experimental design that separates facial expressions

[1] High accuracy rates may also result from prototypical and posed expressions typically being used in research (see Krumhuber et al., 2019, 2020; Dupré et al., 2020).

[2] If one conceives affect as a knowable and measurable phenomenon, then expression authenticity judgments reflect the ability to detect the emotion being expressed (i.e., it is an objective task, with a correct answer as defined by the researcher). Such a paradigm measures accuracy, i.e., the proximity of the judgment to the target emotion, and precision, i.e., the variability between and within judges and expressions. Alternatively, if one believes that underlying affect is unverifiable, then expression authenticity judgments reflect the perception of different types of emotional displays (i.e., it is a subjective task, considering judgment formation, and variability). Such a paradigm measures only precision.

[3] Readers familiar with emotion theories may view encoding-decoding as reflecting the Basic Emotion Theory (BET) and affect-induction as reflecting the Behavioral Ecology View (BECV). The present terminology restricts our reliance on these accounts to certain elements concerned with the conceptualization of facial cues.

into *spontaneous* and *posed*. These are generated in various ways, ranging from emotion-induced exemplars to directed facial muscle activations (Coan and Allen, 2007; Quigley et al., 2013; Siedlecka and Denson, 2019). The conceptualization, however, has been criticized for not reflecting the nuances in expression elicitation (Shackman and Wager, 2019).

Proponents of the encoding-decoding perspective treat spontaneous and posed displays as categorical, with "spontaneous" reflecting felt emotional displays and "posed" reflecting unfelt deliberate displays. The origin of this dichotomy stems from neuroanatomical research alleging separate neural pathways for the production of involuntary and voluntary facial expressions (Rinn, 1984). The two systems are argued to produce visible differences in facial muscle activation, intensity, facial symmetry, and dynamics (Ekman, 2003). Yet, these have been challenged in recent work. For example, research finds no strong support for reliable muscles in either laboratory (Krumhuber and Manstead, 2009) or naturalistic studies (Fernández-Dols and Crivelli, 2013). Also, intensity relates more to the production method than to veracity (Zloteanu et al., 2020, 2018; Miller et al., 2020), and differences in the dynamic components are found to be subtle and varied between emotions (Cohn and Schmidt, 2004; Namba et al., 2016).

Proponents of the affect-induction perspective treat spontaneous and posed as one dimension of emotional displays. The use of actors, for instance, has been proposed to be a valid approach for studying expression authenticity (Gur et al., 2002). Proponents of actor portrayals argue that unmodulated and authentic expressions absent of socio-cultural influence are rare and difficult (if not impossible) to elicit under laboratory conditions (Scherer and Bänziger, 2010). The use of actors permits the creation of well-controlled, reliable, and recognizable displays to investigate the "shared code of emotional signalling" (Scherer and Bänziger, 2010, p. 166); although, the specific acting technique may play a similarly important role (Orlowska et al., 2018; Krumhuber et al., 2020). Nonetheless, such research has often been criticized due to the intentional communicative nature of portrayals, arguing that the reliance on actors for both spontaneous and posed displays invalidates the concept of authenticity (Cowie et al., 2005; Sauter and Fischer, 2018). We conjecture that the use of actors raises an interesting theoretical point. If actors can reproduce elements of genuine, spontaneous, felt displays (e.g., Carroll and Russell, 1997), it calls into question whether authenticity discrimination as a perceptual ability is possible *per se*.

Ultimately, terms such as "genuine" and "spontaneous" should be treated with caution as—theoretically and methodologically—they are debatable concepts. While some researchers treat them as synonyms, others consider them as different dimensions (i.e., genuine-deceptive and spontaneous-posed). For encoding-decoding scholars, genuine and spontaneous reflect similar concepts, namely, the absence of modulation and intentionality in the emotional display (Ekman, 2003). Yet, for affect-induction scholars the genuine-deceptive dimension reflects the intent of the sender (Crivelli and Fridlund, 2018), while spontaneous/posed are labels given to displays with specific facial characteristics. It is important to note that emotional congruence and sender control are complex issues. For instance, an expression may match the person's emotional state but be deliberately produced, such as exaggerated displays (e.g., laughing more strongly in the presence of others; Fridlund, 1991). Based on emotional congruence this would be considered as genuine (and potentially even spontaneous); yet, based on control it can be labeled as deceptive (and posed). Careful considerations should also be given to the type of expression as there are many ways of eliciting either spontaneous or posed displays (Zloteanu et al., 2018, 2020; Krumhuber et al., 2019).

HOW DO WE MEASURE "ACCURACY": APPEARANCE-BASED OR ELICITATION-BASED?

Emotional experiences are often difficult to measure, with some scholars even arguing that they are empirically unverifiable (i.e., we can never truly know what someone is experiencing). Most investigations rely on proxies such as self-report or bodily measures (for recent commentaries see Barrett et al., 2019; Crivelli and Fridlund, 2019). This begs the question: what do we mean by "accurate emotion recognition?"

A review of the literature reveals multiple processes with similar yet not equivalent terms and definitions, such as emotion *identification*, *categorization*, *discrimination*, *inference,* and *recognition*. These are used interchangeably or separately, and sometimes the same term has different definitions (see Gonçalves et al., 2018), making it difficult to know if two scholars pertain to the same phenomenon. For instance, in our research (Zloteanu et al., 2020, 2018) we define *emotion classification accuracy* as the ability to infer specific emotions from facial displays, and *emotion authenticity discrimination* as the ability to differentiate between spontaneous (genuine) and posed (deliberate) displays. By contrast, Buck et al. (2017) use the exact opposite definitions which they label *emotion categorization ability* and *emotion communication accuracy*. Such interchangeable use in terminology may lead to confusion or misleading conclusions and interfere with attempts to synthesize research (see Fiske, 2020). This is a symptom of a larger issue within psychology relating to the use of operational definitions to explain phenomena (see Lilienfeld et al., 2015).

Much of the expression authenticity research has employed an *appearance-based* approach, thereby focusing on stimulus features, such as the Duchenne marker for the distinction between genuine and deliberate smiles. Appearance-based approaches make strong assumptions for the presence/absence of specific facial markers and dynamic features (Ekman, 2003) and impose constraints as to which exemplars are representative of authenticity (thereby excluding facial responses if they fail to meet relevant criteria). Under this approach, judges engage in a categorization task that prompts them to classify facial exemplars based on pre-selected criteria (e.g., Ekman et al., 1983). While such procedure allows for clear and reliable assessments, it may not be sufficient for measuring expression authenticity, as investigations can be conducted with stimuli produced to "look" authentic even though the sender did not experience genuine affect (as often the case with actor portrayals; Scherer and Bänziger, 2010). As such, it only assesses the perceptual or categorization ability of the observer.

The *elicitation-based pathway* is an alternative approach that places the focus on the methods used to produce facial expressions. Here, expression authenticity is operationalized on the basis of sender veracity (or intent), where genuine expressions reflect responses to an emotion-evoking event and non-genuine expressions are voluntarily produced displays in the absence of such an event. It makes no assumptions as to what constitutes a veridical emotional display and merely refers to the congruence between the eliciting event and the external behavior. Natural variations in displays between senders are considered relevant for the judgment process by decoders. Exemplars are selected based on the elicitation technique, allowing researchers to explore how differently produced displays affect people's judgments. Under this approach, labels such as "genuine" and "deliberate" apply only to the inferences made by judges. While elicitation-based approaches introduce more variability in judgments, they capture the diversity of facial displays and mirror the emotional inferences made in real life. This is in stark contrast to the appearance-based approach in which facial exemplars must adhere to a strict morphological or dynamic criterion regardless of the production method being used.

Both approaches have merits yet answer different questions. The appearance-based approach permits investigations of universal representations of expressions (i.e., prototypical displays), in decoders' ability to detect specific facial configurations, and how alterations of such patterns impact perception and judgment. The elicitation-based approach permits investigations of the variability in human responses to emotional events, how such behaviors are affected by context or experimental manipulation, and how people infer meaning from such displays. Noteworthy, measures of accuracy have a different meaning under the two approaches. According to the latter, judgment accuracy is more akin to *congruency* (as in Dawel et al., 2017), where a judgment is correct if the expression is judged as "genuine" and the sender was experiencing an emotion. By contrast, appearance-based accuracy reflects the correct identification and grouping of expressions with similar facial patterns irrespective of the sender's affective experience.

RECOMMENDATIONS FOR FUTURE RESEARCH

For future investigations of expression authenticity, we recommend the use of advanced statistical analyses, such as Bayesian mixed-effects models, to account for individual differences in senders and judges as sources for variability (see Sorensen et al., 2016). For example, a study comparing genuine and deliberate expressions may find no overall difference in genuineness ratings, yet inspection of the stimuli reveals that some expressions in the deliberate condition were rated overly genuine, thereby influencing the aggregate score. Omitting those expressions as "bad" exemplars may be unjustifiable as one would need to assume the existence of "good" exemplars; instead, the respective senders may have just been excellent actors who produced convincing portrayals. In a similar vein, some observers may systematically underrate genuine expressions, minimizing potential differences between conditions. Using mixed-effects models such variability can easily be accounted for without the need to remove data or make assumptions regarding its validity, thereby allowing for more robust analyses (Brysbaert and Stevens, 2018).

Separate from the sender-judge variability within and between studies, considerations should be given to biases in authenticity judgments. For sender-specific biases, the *demeanor bias* – the finding that some senders produce general impressions of (dis)honesty irrespective of their veracity (Zuckerman et al., 1981; Levine et al., 2011) – plays an important role. A person's demeanor may result in their display being judged as non-genuine, irrespective of appearance, or intent. Merely examining facial features (i.e., Duchenne marker) will not reveal such perceptual biases toward particular senders. For judge-specific biases, response tendencies such as the *truth-bias* may impact expression judgments. Overestimating others' truthfulness results in inflated accuracy scores which do not reflect true detection ability but a response preference (see Zuckerman et al., 1981). People may be biased toward disproportionally assuming that facial expressions are genuine (i.e., *authenticity bias*; Gosselin et al., 1995; Zloteanu, 2020). Hence, it is crucial to separate response biases from signal detection when measuring accuracy (Stanislaw and Todorov, 1999).

Future research should also embrace the wide range in which facial expressions occur. Studies concerned with authenticity typically employ one set of spontaneous and posed stimuli, pre-selected from many exemplars and based on specific criteria (see Krumhuber et al., 2017). Rarely do investigations target multiple types of displays (e.g., Soppe, 1988). Given that "spontaneous" and "posed" serve as umbrella terms (see Sauter and Fischer, 2018; Siedlecka and Denson, 2019), judgments under one operationalization may not generalize to another, and aggregating findings will result in incorrect and misleading inferences. In Zloteanu et al. (2018, 2020), we illustrated how producing deliberate expressions using different methods results in judgment differences for each expression type. Under a classical one genuine vs. one deliberate design, these results would not be easily interpretable as each comparison produces different insights into expression authenticity judgments.

Finally, it would be desirable to aim for greater transparency and consistency in the use of operational definitions, urging researchers to be explicit, comprehensive, and transparent in their methodology. While some scholars may be aware of the nuances and shortcomings of specific terminology (Barrett et al., 2019), over-labeling measures and phenomena increase the risk of confusion within and across a domain (Lilienfeld et al., 2015). Labels should serve as mere conveniences for scientific communication, but do not represent unchallengeable and unfalsifiable constructs. Given that emotion scholars still debate the exact definition of emotions (Ortony and Turner, 1990; Izard, 2007; Kagan, 2007), their taxonomy (Fiske, 2020), and whether they are discrete (Siegel et al., 2018) and universal

(Barrett, 2006), it may be premature to taut certainty in a field debating the fundamentals.

CONCLUSION

Deciphering what another person is feeling is a complex task. Here, we address the role of facial expressions as innate cues and communicative signals, proposing a shift from accuracy measures to judgments in expression authenticity. This includes a comparison of encoding-decoding and affective-induction perspectives to offer insights into the process of emotion expression recognition. We conceive of senders as strategic performers who utilize their full expressive capabilities in social interaction and judges as attempting to infer meaning and intent from emotional displays. To help accelerate progress in the field we encourage researchers to carefully consider theory and methodology in how they operationalize facial expressions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MZ: conceptualization. MZ and EK: writing and revision. Both authors contributed to the article and approved the submitted version.

FUNDING

We acknowledge that the publication fees for this article were fully paid by Kingston University London.

REFERENCES

Barrett, L. F. (2006). Are emotions natural kinds? *Perspect. Psychol. Sci.* 1, 28–58.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol. Sci. Public Interest* 20, 1–68. doi: 10.1177/1529100619832930

Brysbaert, M., and Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *J. Cogn.* 1:9.

Buck, R., Powers, S. R., and Hull, K. S. (2017). Measuring emotional and cognitive empathy using dynamic, naturalistic, and spontaneous emotion displays. *Emotion* 17, 1120–1136. doi: 10.1037/emo0000285

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

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

Coan, J. A., and Allen, J. J. B. (eds). (2007). *Handbook of Emotion Elicitation and Assessment*. New York, NY: Oxford University Press.

Cohn, J. F., and Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *Int. J. Wavelets Multiresolution Inf. Process.* 02, 121–132. doi: 10.1142/s021969130400041x

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

Crivelli, C., and Fridlund, A. J. (2018). Facial displays are tools for social influence. *Trends Cogn. Sci.* 22, 388–399. doi: 10.1016/j.tics.2018.02.006

Crivelli, C., and Fridlund, A. J. (2019). Inside-out: from basic emotions theory to the behavioral ecology view. *J. Nonverbal Behav.* 43, 161–194. doi: 10.1007/s10919-019-00294-2

Crozier, W. R. (2010). The puzzle of blushing. *Psychologist* 23, 390–393.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.

Dawel, A., Wright, L., Irons, J., Dumbleton, 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. Methods* 49, 1539–1562. doi: 10.3758/s13428-016-0813-2

Dezecache, G., Mercier, H., and Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. *J. Pragmat.* 59, 221–233. doi: 10.1016/j.pragma.2013.06.007

Dupré, D., Krumhuber, E. G., Küster, D., and McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLoS One* 15:e0231968. doi: 10.1371/journal.pone.0231968

Ekman, P. (1997). "Expression or communication about emotion," in *Uniting Psychology and Biology: Integrative Perspectives on Human Development*, eds N. L. Segal, G. E. Weisfeld, and C. C. Weisfeld (Washington, DC: American Psychological Association), 315–338. doi: 10.1037/10242-008

Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York, NY: Owl Books.

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

Ekman, P., Friesen, W. V., and O'Sullivan, M. (1988). Smiles when lying. *J. Pers. Soc. Psychol.* 54:414. doi: 10.1037/0022-3514.54.3.414

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 O'Sullivan, M. (1991). Who can catch a liar? *Am. Psychol.* 46, 913–920. doi: 10.1037/0003-066x.46.9.913

Ekman, P., and Rosenberg, E. L. (2005). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford: Oxford University Press.

Fernández-Dols, J.-M., and Crivelli, C. (2013). Emotion and expression: naturalistic studies. *Emot. Rev.* 5, 24–29. doi: 10.1177/1754073912457229

Fiske, A. P. (2020). The lexical fallacy in emotion research: mistaking vernacular words for psychological entities. *Psychol. Rev.* 127, 95–113. doi: 10.1037/rev0000174

Fridlund, A. J. (1991). Sociality of solitary smiling: potentiation by an implicit audience. *J. Pers. Soc. Psychol.* 60, 229–240. doi: 10.1037/0022-3514.60.2.229

Gonçalves, A. R., Fernandes, C., Pasion, R., Ferreira-Santos, F., Barbosa, F., and Marques-Teixeira, J. (2018). Effects of age on the identification of emotions in facial expressions: a meta-analysis. *PeerJ* 6:e5278. doi: 10.7717/peerj.5278

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

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

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., et al. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *J. Neurosci. Methods* 115, 137–143. doi: 10.1016/s0165-0270(02)00006-7

Hall, J. A., Andrzejewski, S. A., and Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: a meta-analysis. *J. Nonverbal Behav.* 33, 149–180. doi: 10.1007/s10919-009-0070-5

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

Hurley, C. M., and Frank, M. G. (2011). Executing facial control during deception situations. *J. Nonverbal Behav.* 35, 119–131. doi: 10.1007/s10919-010-0102-1

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychol. Bull.* 115, 288–299. doi: 10.1037/0033-2909.115.2.288

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspect. Psychol. Sci.* 2, 260–280. doi: 10.1111/j.1745-6916.2007.00044.x

Kagan, J. (2007). *What is Emotion? History, Measures, and Meanings*. New Haven, CT: Yale University Press.

Krumhuber, E. G., Küster, D., Namba, S., Shah, D., and Calvo, M. (2019). Emotion recognition from posed and spontaneous dynamic expressions: human observers versus machine analysis. *Emotion* doi: 10.1037/emo0000712 [Epub ahead of print].

Krumhuber, E. G., Küster, D., Namba, S., and Skora, L. (2020). Human and machine validation of 14 databases of dynamic facial expressions. *Behav. Res. Methods* doi: 10.3758/s13428-020-01443-y [Epub ahead of print].

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

Levine, T. R., Serota, K. B., Shulman, H., Clare, D. D., Park, H. S., Shaw, A. S., et al. (2011). Sender demeanor: individual differences in sender believability have a powerful impact on deception detection judgments. *Hum. Commun. Res.* 37, 377–403. doi: 10.1111/j.1468-2958.2011.01407.x

Lilienfeld, S. O., Sauvigné, K. C., Lynn, S. J., Cautin, R. L., Latzman, R. D., and Waldman, I. D. (2015). Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. *Front. Psychol.* 6:1100. doi: 10.3389/fpsyg.2015.01100

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

Miller, E. J., Krumhuber, E. G., and Dawel, A. (2020). Observers perceive the Duchenne marker as signaling only intensity for sad expressions, not genuine emotion. *Emotion* doi: 10.1037/emo0000772 [Epub ahead of print].

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

Orlowska, A. B., Krumhuber, E. G., Rychlowska, M., and Szarota, P. (2018). Dynamics matter: recognition of reward, affiliative, and dominance smiles from dynamic vs. static displays. *Front. Psychol.* 9:938. doi: 10.3389/fpsyg.2018.00938

Ortony, A., and Turner, T. J. (1990). What's basic about basic emotions? *Psychol. Rev.* 97:315. doi: 10.1037/0033-295x.97.3.315

Quigley, K. S., Lindquist, K. A., and Barrett, L. F. (2013). "Inducing and measuring emotion and affect: tips, tricks, and secrets," in *Handbook of Research Methods in Social and Personality Psychology*, eds H. T. Reis and C. M. Judd (New York, NY: Cambridge University Press), 220–252. doi: 10.1017/cbo9780511996481.014

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

Sauter, D. A., and Fischer, A. H. (2018). 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.

Shackman, A. J., and Wager, T. D. (2019). The emotional brain: fundamental questions and strategies for future research. *Neurosci. Lett.* 693, 68–74. doi: 10.1016/j.neulet.2018.10.012

Siedlecka, E., and Denson, T. F. (2019). Experimental methods for inducing basic emotions: a qualitative review. *Emot. Rev.* 11, 87–97. doi: 10.1177/1754073917749016

Siegel, E. H., Sands, M. K., Van den Noortgate, W., Condon, P., Chang, Y., Dy, J., et al. (2018). Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychol. Bull.* 144, 343–393. doi: 10.1037/bul0000128

Smith, D. L. (2004). *Why We Lie: The Evolutionary Roots of Deception and the Unconscious Mind*. New York, NY: St. Martin's Griffin press.

Soppe, H. J. G. (1988). Age differences in the decoding of affect authenticity and intensity. *J. Nonverbal Behav.* 12, 107–119. doi: 10.1007/bf00986929

Sorensen, T., Hohenstein, S., and Vasishth, S. (2016). Bayesian linear mixed models using Stan: a tutorial for psychologists, linguists, and cognitive scientists. *Quant. Methods Psychol.* 12, 175–200. doi: 10.20982/tqmp.12.3.p175

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/bf03207704

Surakka, V., and Hietanen, J. K. (1998). Facial and emotional reactions to Duchenne and non-Duchenne smiles. *Int. J. Psychophysiol.* 29, 23–33. doi: 10.1016/s0167-8760(97)00088-3

Tomkins, S. S. (1962). *Affect Imagery Consciousness: Volume I: The Positive Affects*. New York, NY: Springer publishing company.

Zloteanu, M. (2020). "Reconsidering facial expressions and deception detection," in *Handbook of Facial Expression of Emotion*, eds A. Freitas-Magalhaes and J. Borod (Porto: FEELab Science Books & Leya), 238–284.

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

Zloteanu, M., Krumhuber, E. G., and Richardson, D. C. (2020). Acting surprised: comparing perceptions of different dynamic deliberate expressions. *J. Nonverbal Behav.* doi: 10.1007/s10919-020-00349-9 [Epub ahead of print].

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, TX: Pennsylvania State University Press), 13–35.

Zuckerman, M., Larrance, D. T., Spiegel, N. H., and Klorman, R. (1981). Controlling nonverbal displays: facial expressions and tone of voice. *J. Exp. Soc. Psychol.* 17, 506–524. doi: 10.1016/0022-1031(81)90037-8

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 January 2021
doi: 10.3389/fpsyg.2020.611670

Is Technology Enhancing or Hindering Interpersonal Communication? A Framework and Preliminary Results to Examine the Relationship Between Technology Use and Nonverbal Decoding Skill

Mollie A. Ruben[1,2], Morgan D. Stosic[1], Jessica Correale[1] and Danielle Blanch-Hartigan[3]*

[1] Department of Psychology, University of Maine, Orono, ME, United States, [2] Center for Healthcare Organization and Implementation Research, VA Boston Healthcare System, Boston, MA, United States, [3] Department of Natural and Applied Sciences, Bentley University, Waltham, MA, United States

OPEN ACCESS

Edited by:
Norah E. Dunbar, University of California, Santa Barbara, United States

Reviewed by:
Yann Coello, Université Lille Nord de France, France
Anders Flykt, Mid Sweden University, Sweden

***Correspondence:**
Mollie A. Ruben
mollie.ruben@maine.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *29 September 2020*
Accepted: *08 December 2020*
Published: *15 January 2021*

Citation:
Ruben MA, Stosic MD, Correale J and Blanch-Hartigan D (2021) Is Technology Enhancing or Hindering Interpersonal Communication? A Framework and Preliminary Results to Examine the Relationship Between Technology Use and Nonverbal Decoding Skill. Front. Psychol. 11:611670. doi: 10.3389/fpsyg.2020.611670

Digital technology has facilitated additional means for human communication, allowing social connections across communities, cultures, and continents. However, little is known about the effect these communication technologies have on the ability to accurately recognize and utilize nonverbal behavior cues. We present two competing theories, which suggest (1) the potential for technology use to *enhance* nonverbal decoding skill or, (2) the potential for technology use to *hinder* nonverbal decoding skill. We present preliminary results from two studies to test these hypotheses. Study 1 (N = 410) found that global screen time was unrelated to nonverbal decoding skill. However, how participants spent their time using technology mattered. Participants who reported more active technology use (i.e., posting content) self-reported that their nonverbal decoding skill (as measured by the Emotional Sensitivity subscale of the Social Skills Inventory) was superior but performed worse on objective measures of decoding skill (using standardized tests including the Diagnostic Analysis of Nonverbal Accuracy-Adult Faces and the Workplace Interpersonal Perception Skill). By contrast, passive users performed significantly better on objective measures of nonverbal decoding skill; although they did not self-report any difference in their skill compared to less passive users. Study 2 (N = 190), and a mini-meta analysis of both studies, replicated this pattern. These effects suggest a roadmap for understanding the theoretical relationship between technology use and nonverbal communication skills. We also provide recommendations for future research, including the use of experimental designs to determine causal pathways and to advance our conceptual understanding of the relationship between technology use and nonverbal decoding skill.

Keywords: technology, nonverbal communication, decoding ability, interpersonal accuracy, communication skills

INTRODUCTION

A young-professional is woken up to the sound of a buzzing alarm, and grudgingly rolls over to grab their phone. Perhaps this individual begins their morning by passively scrolling through their Facebook feed in order to determine their colleague's reaction to the heated presidential debate the night before. Or maybe they snap a quick picture of their #OOTD (i.e., Outfit of the Day) to send to their close friend. After returning home from a long day of work-based videoconference calls, this individual may spend the next few hours sucked into the whereabouts of their favorite social media influencer, or casually swiping through some dating profiles. Before retiring to bed, however, they make sure to post a quick inspiring quote to their Twitter profile.

This scenario, while fictitious, illustrates the increasing relationship many individuals have with technology from the instant they wake up, to the instant they go to bed. Technology serves various functions, from increasing office productivity, facilitating big data collection, enhancing record keeping, and above all else, providing a distinctly digital way for humans to *communicate* with one another. Indeed, the rate of communicative instances via technology per day in 2020 is astounding: 350 million photos uploaded to Facebook, 500 million tweets, 3 billion snapchats, and over 26 billion texts by Americans alone (Aslam, 2020a,b; Sayce, 2020; Tocci, 2020).

While the digital revolution has certainly changed the way individuals can communicate, little empirical results exists regarding the effect of technology on an individual's communication skills. Specifically, because technology markedly changes the available information individual's use to decode the communicative intents of others (e.g., determining a friend's emotional state via short text message instead of their facial expression), are those who spend large quantities of time communicating online better or worse decoders of nonverbal information? Not only is nonverbal decoding a crucial component of general social and communication skills, but it has been tied to better interpersonal outcomes (e.g., Hall et al., 2009), can be easily assessed with validated, reliable, and standardized objective measures, and can be improved with practice and feedback trainings (e.g., Schlegel et al., 2017b). Therefore, the question of whether technology may affect nonverbal decoding, or how accurately a perceiver can recognize and interpret the nonverbal behaviors of another person, is important to empirically address.

Supplementing or even fully replacing face-to-face communication with technology-mediated communication affects both the *number* of nonverbal cues, as well as the *types* of nonverbal cues that individuals use to decode communicative meaning (Vinciarelli, 2017). For example, text messages may not allow access to important vocal cues (e.g., pitch, tone, inflections), but may have distinct timing and spacing cues to draw from Döring and Pöschl (2008). By contrast, video conferencing technologies may allow access to vocal cues, but may limit the ability to engage in mutual eye gaze or perceive body movements and gestures (Ferrán-Urdaneta and Storck, 1997; Neureiter et al., 2013). If individuals rely more heavily on technology-mediated, as opposed to face-to-face, interactions as a primary means of communication, it seems likely that the nonverbal decoding skill individuals ordinarily employ in face-to-face communication would be impacted (e.g., worsened, or perhaps enhanced).

This paper applies communication skills theories and conceptual accounts of technology use to examine the role of technology use on an individual's ability to accurately perceive the nonverbal behavior displayed by others (i.e., nonverbal decoding skill). For the purposes of this paper, we define technology use as any technology or application on a smart phone that contributes to *communication* online (e.g., use of social media sites, texting, emailing). Cell phone use is the predominant method of technology use by young adults in the United States today with 96% of 18–26 years-old young adults reporting ownership of a smart phone (Pew Research Center., 2019). Therefore, for the remainder of the paper, when discussing technology use, we are referring specifically to smart phone use.

We start by reviewing two competing hypotheses, that technology use either enhances or hinders communication skills. We then present results from two cross-sectional studies and a mini meta-analysis of these studies on the relationship between technology use and nonverbal decoding skill to inform our understanding of which of the competing hypotheses is more likely supported. Finally, we make recommendations for future research aimed at disentangling the causal relationship between technology use and nonverbal decoding skill.

TECHNOLOGY USE MAY ENHANCE COMMUNICATION SKILLS

The most effective way to improve nonverbal decoding skill is by practicing decoding nonverbal cues and receiving feedback on the accuracy of one's perceptions (Blanch-Hartigan et al., 2012; Schlegel et al., 2017a). Regarding the relationship between technology use and nonverbal decoding skill, some theorists have argued that technology-mediated communication may enhance communication skills by providing a safe environment to practice sending and receiving nonverbal cues, and allowing for feedback regarding the accuracy of one's perceptions (e.g., Stritzke et al., 2004; Ellison et al., 2007; Valkenburg and Peter, 2009). Because it is unusual in face-to-face interactions to receive feedback about one's decoding ability, it may be that spending more time using technology to interact with others may facilitate face-to-face interactions by providing this type of practice and feedback to users on a regular basis.

Liberated Relationship Perspective

One hypothesis which falls into this "enhancement" framework is the Liberated Relationships Perspective (Hu et al., 2004). This theory argues that increased internet usage has allowed individuals who may not typically engage in conversation the opportunity to engage with one another through technology-mediated communication. Some of the constraints may be psychological, such as in cases of shyness and social anxiety (Stritzke et al., 2004), or physical, such as in cases of distant geographical locations (Ellison et al., 2007). According to this framework, internet usage may afford an increase in the *number* of interactions an individual is able to engage in. If the internet supplements, instead of detracts from, face-to-face interactions, individuals may have increased opportunities to practice nonverbal decoding with a greater number and variety of communication partners.

Internet Enhanced Self-Disclosure Hypothesis

While not directly related to communication skill, the Internet Enhanced Self-Disclosure Hypothesis also provides support for improved nonverbal decoding skill with increased technology use (Valkenburg and Peter, 2009). This theory posits that greater technology use may enhance social connectedness and wellbeing by enhancing *online self-disclosure*. The authors define online self-disclosure as "online communication about personal topics that are typically not easily disclosed, such as one's feelings, worries, and vulnerabilities" (p. 2). Because online platforms allow for the sharing of intimate information to a significantly greater degree than do face-to-face interactions, it is likely that individuals are afforded more opportunities to practice decoding and receive feedback regarding affective information. Individuals who engage in technology-mediated communication more frequently may become more skilled decoders of nonverbal information, perhaps for affective information in particular.

TECHNOLOGY USE MAY HINDER COMMUNICATION SKILLS

While these two "enhancement" theories describe the ways in which increased technology usage may allow individuals more opportunities to practice decoding nonverbal communication, others have argued a competing perspective. Specifically, researchers have argued that technology may hinder specific communication skills. Spending time communicating via technology may result in less face-to-face interactions and therefore less practice decoding nonverbal information in whole, as well as from specific cue channels (e.g., vocal tone) which are reduced or absent in many technology platforms (Kraut et al., 1998; Nie, 2001; Patterson, 2019). In this way, the type of communication skills learned or practiced in technology-mediated communication are not equivalent to, and may even hinder, the skills required to decode nonverbal behavior in face-to-face interactions.

Reduction Hypothesis

In the early 1990s, several researchers theorized that the internet had detrimental effects on adolescent wellbeing and social connectedness (Kraut et al., 1998; Nie, 2001). It was assumed that because the internet motivates adolescents to form superficial online relationships with strangers that are less beneficial than their real-world relationships, time spent online occurs at the expense of time spent with existing relationships. The Reduction Hypothesis posits that it is the lack of or decrease in face-to-face interacting that leads to detrimental communicative consequences rather than technology itself (Valkenburg and Peter, 2009).

Valkenburg and Peter (2009) propose two important updates to this theory based on changes in how individuals use the internet to communicate since the Reduction Hypothesis was first introduced. First, in the second half of the 1990s, it was hard to maintain a pre-existing social network on the internet because not a lot of people had access to it, often resulting in online friends separate from offline friends. Today, with more widespread access and utilization of the internet and social media, individuals spend more time online connecting with people they also spend time with in face-to-face interactions as opposed to forming online-only relationships with strangers (Valkenburg and Peter, 2009). However, the communication skills, such as nonverbal decoding, that individuals develop through online interactions may not translate to actual face-to-face interactions. As such, time spent online may stunt the development of nonverbal decoding necessary for face-to-face interactions. Therefore, although our internet habits have changed, the Reduction Hypothesis is still relevant to theorizing regarding the effects of technology use on nonverbal decoding ability.

Cues-Filtered–Out Theory

In addition to reducing the amount of time individuals spend interacting face-to-face, theorists have also noted that many technology-mediated communication platforms greatly reduce both the number as well as the kinds of nonverbal cues technology users are exposed to. Cues absent from some technology-mediated communication (e.g., social media, texting, emailing) can include physical appearance, tone of voice, facial expression, gaze, posture, touch, space, and gestures (Kiesler et al., 1984; Siegel et al., 1986). These nonverbal cues are important in expressing relative status, affect, relationship roles, and many other interpersonal dimensions. This Cues-Filtered-Out Theory (Culnan and Markus, 1987; Sproull and Kiesler, 1986) suggests that without these cues available, especially for low bandwidth technology (i.e., communication systems with access to only one or two channels such as vocal, kinesics, or proxemics), certain communicative functions are lost. Although higher bandwidth systems may allow for certain nonverbal cues, these cues are often more obvious and lack complexity, which may cause individuals to lose the ability to decode more subtle nonverbal cues (e.g., facial expressions are more complex than emoji's, vocal intensity is more complex than CAPITALIZING words). Therefore, this theory suggests that the filtering out of important nonverbal cues (e.g., especially for individuals who use low bandwidth technology systems) impacts an individual's ability to receive practice and feedback on the accuracy of their nonverbal decoding attempts, thereby hindering nonverbal decoding skill (Walther and Parks, 2002).

CURRENT RESEARCH AND HYPOTHESES

The primary objective of the current research is to empirically examine the relationship between technology use and nonverbal decoding skill via two studies and a mini meta-analysis combining results from these two studies. Because individuals may use technology the same amount but differ in *how* they spend their time online, we measured users' online communication activity via objective global screen time use taken from iPhone users, as well as the degree of self-reported active technology use (posting selfies and photographs, responding to others' posts) and the degree of self-reported passive technology use (scrolling through photographs and others' posts but not responding or posting themselves). In addition, we also sought to be thorough in our assessment of nonverbal decoding skill, as researchers have demonstrated that there are different *kinds* of decoding skills subsumed by a higher-order global decoding skill (Schlegel et al., 2017a). Therefore, we employed three distinct measures of nonverbal decoding, two objective assessments of skill using a standardized, validated, and reliable test of emotion recognition [i.e., Diagnostic Analysis of Nonverbal Accuracy-Adult Faces (DANVA-2AF; Nowicki and Duke, 1994)] and a newly developed test that assesses relevant decoding ability in the workplace such as inferring behavioral intentions, personality traits, status, interpersonal attitudes (dominance/cooperativeness and motivations), behavioral outcomes, and thoughts and feelings [i.e., the Workplace Interpersonal Perception Skill (WIPS; Dael et al., in preparation)], and one self-report measure [the Emotional Sensitivity subscale of the Social Skills Inventory (SSI; Riggio, 2005)]. Together, we utilized these various measures of technology and nonverbal decoding skill in order to test the preceding competing hypotheses: (1) more technology use is related to better nonverbal decoding skill vs. (2) more technology use is related to poorer nonverbal decoding skill.

MATERIALS AND METHODS

Study 1

Participants

Data were collected from 410 participants in the University of Maine introductory participant pool for a study on perceiving nonverbal signals in others. Of these, 51% were male and 48% were female. A total of 377 (92%) participants identified as white, 15 (4%) as Asian, 14 (3%) as American Indian or Alaska Native, 12 (3%) as Black, 2 (0.5%) as Native Hawaiian or Pacific Islander, and 33 (8%) as Other. Their ages ranged from 18 to 29 (M = 19.09, SD = 1.56). A power analysis conducted using G*Power (Faul et al., 2007) assuming a small to medium effect (r = 0.15) of technology use on nonverbal decoding skill indicated that 343 participants would be needed to achieve 80% power using an alpha level of 0.05 (two-tailed). The final sample of participants exceeds this threshold, indicating that the present study is sufficiently powered to detect small to medium effects.

Measures

Technology Use

Three separate measures of technology use were collected from participants. For iPhone users, participants were instructed to navigate to their phone settings and extract their average daily screen time over the last 7 days in minutes (N = 263). This screen time metric is a real-time report of how much time a participant spends with their phone screen *turned on* in an average week (i.e., listening to music with one's screen off is not included). To ensure participants did not alter their responses in order to appear more socially desirable, we also required that they upload a screenshot of this information. In addition to this objective measure of technology use, participants were asked to self-report on a scale of 0–10 from "does not describe me at all" to "describes me very well" how well the following statements described their technology use, "I tend to be an active user, posting frequently" and "I tend to be a passive user, scrolling through posts and photos." These two questions comprised our self-report measures of technology use: the degree to which a participant endorsed themselves as an active user separately from the degree to which a participant endorsed themselves as a passive user. Because active user endorsement and passive user endorsement were single item questions rather than a single bipolar item, participants could report any combination of active and passive technology use. That is, a participant could endorse a high degree of active use and a high degree of passive use, they could report a low degree of both, or a high degree of one and not the other. For all analyses, we entered both continuous variables to examine how the independent contribution of active and passive use predicted our outcomes of interest.

Nonverbal Decoding Measures

The newly developed WIPS test (Workplace Interpersonal Perception Skill; Dael et al., in preparation; a = 0.67) assesses multiple aspects of decoding skill using 41 brief video segments with and without sound from three types of role-played workplace interactions: a recruiter-applicant negotiation, a helpdesk trouble-shooting scenario, and a company team meeting. Each segment is paired with a multiple-choice question for which the correct answer was based on actual behavior (what happened in the interaction during or after the video segment), instructions that the actors received (e.g., to be competitive), actors' self-reported personality, or post-interaction evaluations (e.g. perceptions of the other as competitive) and response options varied from 2 options to 6 options depending on the item. In this way, participants must decode multiple simultaneous nonverbal cues (e.g., tone of voice, facial expression) in order to accurately assess the interpersonal characteristics of any given situation. For some items, the video consisted of multiple short segments (e.g., You will see the same person in two different negotiations signing a contract. In which negotiation did the person negotiate the better deal for herself?) while other videos were based off of just one video (e.g., In the following video, you will see 6 people enter the room for a team meeting. Who is the team leader?). Accuracy is calculated as the proportion correct responses compared against a criterion or correct response for each segment.

Participants also completed the Diagnostic Analysis of Nonverbal Accuracy-Adult Faces (DANVA-2AF; Nowicki and Duke, 1994; a = 0.60), a test of emotion recognition ability using static and posed photographs. This measure presents 24 photographs of adult faces with high and low intensity portrayals of the four basic emotions of happiness, anger, sadness, and fear. Accuracy was calculated as the proportion correct.

Finally, participants completed the Emotional Sensitivity (ES; a = 0.80) subscale of the Social Skills Inventory (SSI; Riggio, 2005). The ES subscale consists of 15 self-report items, with a 5-point response scale ranging from "Not at all like me" to "Exactly like me." The ES subscale specifically assesses self-reported skill for decoding emotional and other nonverbal messages (e.g., *I always seem to know what people's true feelings are no matter how hard they try to conceal them)*. For analysis purposes, a sum was calculated across items.

Study 2

Our second study was an exact replication of Study 1 launched approximately 3 months after Study 1 with data from 190 participants from the University of Maine introductory participant pool. Because we had not hypothesized *a priori* the effect of active and passive technology use on nonverbal decoding skill, we wished to collect a second sample of participants in order to investigate whether the pattern of results we describe in Study 1 would replicate. The demographics of this second sample were comparable to those from our first study, with 91 male participants (48%) and 99 females (52%). Of these, 179 (94%) identified as white, 9 (5%) as Asian, 5 (3%) as Black, 2 (1%) as American Indian or Alaska Native, 1 (0.5%) as Native Hawaiian or Pacific Islander, and 6 (3%) as Other. Participant's ages ranged from 18 to 31 (M = 19.43, SD = 1.57). A power analysis conducted using G*Power (Faul et al., 2007) assuming a small to medium effect derived from Study 1 (r = 0.20) indicated that 191 participants would be needed to achieve 80% power using an alpha level of 0.05 (two-tailed).

Analyses

To test our competing hypotheses about the relationship between technology use and nonverbal decoding skill, we first examined bivariate correlations between our study variables. Next, we ran a series of linear regressions on the whole sample in Study 1 and Study 2 controlling for participant gender to examine the independent contribution of active and passive technology use on each of our nonverbal decoding skill measures (accuracy scores on the WIPS test, accuracy scores on the DANVA, and self-reported emotional sensitivity).

To combine results from Study 1 and Study 2, a mini meta-analysis (Goh et al., 2016) was performed for each technology use variable and each nonverbal decoding variable. We used fixed effects in which the mean effect size (i.e., mean correlation) was weighted by sample size. All correlations were Fisher's z transformed for analyses and converted back to Pearson correlations for presentation.

RESULTS

Study 1

Means, standard deviations, and bivariate correlations are presented in **Table 1**. Contrary to what would be predicted

TABLE 1 | Study 1 and study 2 means, standard deviations, and bivariate correlations between technology use, nonverbal decoding skill, and gender.

Variable	M (SD)		2		3		4		5		6		7	
	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2	Study 1	Study 2
DANVA 2-AF	0.75 (0.11)	0.74 (0.13)	0.30***	0.42***	0.05	0.11	0.09	0.11	0.03	0.01	0.09	0.10	0.16***	0.30***
WIPS test	0.75 (0.11)	0.74 (0.13)			0.03	0.21**	0.00	−0.03	−0.17***	-0.16*	0.14**	0.27***	0.15**	0.22**
Emotional sensitivity subscale	85.56 (16.93)	87.93 (17.49)					0.02	0.17*	0.20***	0.25***	0.04	−0.03	0.15**	0.35***
Screen time (minutes)	297.88 (136.24)	363.40 (176.50)							0.11†	0.24**	0.01	−0.04	0.08	0.12
Active use	4.28 (2.81)	4.00 (2.55)									−0.15**	−0.36***	0.26***	0.23**
Passive use	8.25 (3.05)	8.50 (3.07)											0.02	-0.08
Gender	Male N = 210 Female N = 196	Male N = 92 Female N = 98												

*†$p < 0.10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Study 1 N's range from 263 to 410 and Study 2 N's range from 129 to 190 depending on screen time use measure as not all participants had iPhones or screen time turned on. DANVA 2-AF is the Diagnostic Analysis of Nonverbal Accuracy-Adult Faces; Nowicki and Duke (1994). WIPS test is the Workplace Interpersonal Perception Skill test; Dael et al., in preparation. Emotional Sensitivity is a self-reported subscale from the Social Skills Inventory (Riggio, 2005). Active and passive use were measured on 0–10 Likert scale from "does not describe me well" to "describes me very well." Gender coded as (0 = male, 1 = female).*

by either theoretical framework, screen time use was unrelated to every measure of nonverbal decoding skill we employed. However, when examining the ways in which participants self-reported spending their time online, a more complex pattern emerged. Specifically, more active technology use was related to higher self-reported nonverbal decoding skill ($r = 0.20$, $p < 0.001$) but lower accuracy score on the WIPS ($r = -0.17$, $p < 0.001$). That is, participants who identified as more active users (i.e., posting frequently) believed that they were *better* judges of others' nonverbal communication, but performed significantly *worse* on an objective test of nonverbal decoding skill (i.e., the WIPS test). On the other hand, participants who reported being more passive users (i.e., reading through posts and looking at other people's photographs) were significantly *more accurate* in decoding nonverbal behavior, as measured by the WIPS ($r = 0.14$, $p = 0.005$), although they did not self-report any differences in their nonverbal decoding skills from less passive users as highlighted by the correlation between passive user endorsement and self-reported skill on the ES subscale of the SSI ($r = 0.04$, $p = 0.484$). Neither self-reported passive nor active technology use was significantly related to an individual's ability to decode facial expressions of emotions, measured via the DANVA-2AF (p's > 0.07).

TABLE 2 | Regression results from study 1 and study 2 examining the independent contribution of technology use variables on nonverbal decoding skill.

Study 1

Predictors	**Objective**		**Self-report**
	DV: WIPS test β_{std} ***t*** **(*p*-value)**	**DV: DANVA-2AF** β_{std} ***t*** **(*p*-value)**	**DV: Emotional sensitivity** β_{std} ***t*** **(*p*-value)**
Active use	**−0.21** −4.17 ($p < 0.001$)	−0.01 −0.16 ($p = 0.871$)	**0.18** 3.51 ($p < 0.001$)
Passive use	**0.11** 2.31 ($p = 0.021$)	0.09 1.77 ($p = 0.077$)	0.06 1.12 ($p = 0.264$)
Gender	**0.21** 4.14 ($p < 0.001$)	**0.17** 3.24 ($p = 0.001$)	0.10 1.95 ($p = 0.052$)
R^2	$R^2 = 0.084$; $F(3, 401) = 12.17$, $p < 0.001$	$R^2 = 0.035$; $F(3, 401) = 4.81$, $p = 0.003$	$R^2 = 0.051$; $F(3, 401) = 7.17$, $p < 0.001$

Study 2

Predictors	**DV: WIPS test** β_{std} ***t*** **(*p*-value)**	**DV: DANVA-2AF** β_{std} ***t*** **(*p*-value)**	**DV: Emotional sensitivity** β_{std} ***t*** **(*p*-value)**
Active use	−0.13 −1.73 ($p = 0.085$)	−0.02 −0.23 ($p = 0.815$)	**0.21** 2.76 ($p = 0.006$)
Passive use	**0.25** 3.42 (p = 0.001)	0.12 1.59 ($p = 0.114$)	0.06 0.88 ($p = 0.382$)
Gender	**0.27** 3.93 ($p < 0.001$)	**0.32** 4.44 ($p < 0.001$)	**0.31** 4.42 ($p < 0.001$)
R^2	$R^2 = 0.15$; $F(3, 188) = 10.87$, $p < 0.001$	$R^2 = 0.11$; $F(3, 188) = 7.46$, $p < 0.001$	$R^2 = 0.16$; $F(3, 188) = 11.41$, $p < 0.001$

β_{std} is standardized Beta. Bolded values reflect significance at the $p < 0.05$ level.

Gender, Technology Use, and Nonverbal Decoding Skill

Because active and passive technology use were not mutually exclusive (i.e., an individual could report being high on active and passive use), and because gender is related to both technology use (Jackson et al., 2008) as well as nonverbal decoding skill (Hall and Gunnery, 2013), we wished to determine the independent effects of active and passive technology use on nonverbal decoding skill while controlling for gender. Therefore, we first entered active use, passive use, and gender into a regression predicting accuracy scores on the WIPS. Active use remained a significant negative predictor ($\beta_{std} = -0.21$, $p < 0.001$; **Table 2**), suggesting that those who are more active users were worse at decoding nonverbal behavior. Passive use also remained a significant positive predictor ($\beta_{std} = 0.11$, $p = 0.02$), where those who reported spending their time looking at others' posts and pictures were more accurate in decoding nonverbal behavior. Further, these two effects were significant even after controlling for gender, which also significantly predicted higher scores on the WIPS test ($\beta_{std} = 0.21$, $p < 0.001$; female coded as 1, male coded as 0). Approximately 8% of the variance in WIPS test scores was accounted for when active use, passive use, and gender were entered as predictors.

We next entered active use, passive use, and gender into a regression predicting accuracy scores on the DANVA-2AF. None of these variables, apart from gender ($\beta_{std} = 0.17$, $p = 0.001$), significantly predicted scores on the DANVA-2AF (**Table 2**). Approximately 4% of the variance in DANVA-2AF scores was accounted for by these predictor variables.

When active use, passive use, and gender were entered into a regression predicting self-reported nonverbal decoding skill, active use remained a significant positive predictor ($\beta_{std} = 0.18$, $p < 0.001$), such that those who were more active users self-reported that they were better at decoding nonverbal information from others (**Table 2**). While more passive use was unrelated to self-reported nonverbal decoding skill, gender remained a marginally significant positive predictor ($\beta_{std} = 0.10$, $p = 0.052$) indicating that females reported being more skilled nonverbal decoders than males. Approximately 5% of the variance in self-reported nonverbal decoding skill was accounted for when active use, passive use, and gender were entered as predictors.

Study 2

While results from Study 1 were neither supportive of an enhancing or suppressing effect of global technology usage on nonverbal decoding skill, we did find that the ways individuals used technology mattered (i.e., actively versus passively). Because this active/passive relationship was not hypothesized *a priori*, we examined these effects in a separate sample of participants. Therefore, akin to Study 1, we first examined the bivariate correlations between our measures of technology use and nonverbal decoding skill. We once again found that screen time use was unrelated to objective measures of nonverbal decoding skill—i.e., the DANVA and WIPS (p's > 0.20). However, in Study 2 objective screen time use was significantly and positively related to self-reported nonverbal decoding skill (r = 0.17, p = 0.050) (**Table 1**).

Replicating Study 1's findings, active technology use was also related to higher self-reported nonverbal decoding skill (r = 0.25, p = 0.001), but lower objective nonverbal decoding skill as measured by the WIPS (r = −0.16, p = 0.028). Individuals who identified as more passive users were once again significantly more accurate in decoding nonverbal behavior, as measured by the WIPS (r = 0.27, p < 0.001), although they did not self-report any differences in their nonverbal decoding skills from less passive users (r = −0.03, p = 0.653). Neither self-reported passive nor active technology use was significantly related to an individual's ability to decode facial expressions of emotions, measured via the DANVA-2AF (p's > 0.167).

We deconstructed these effects by entering active use, passive use, and gender into three separate linear regressions predicting the WIPS, DANVA-2AF, and self-reported nonverbal decoding skill. We regressed our three predictor variables on scores from the WIPS. Replicating regression results from Study 1, active technology use was a marginally significant negative predictor of nonverbal decoding skill (β_{std} = −0.13, p = 0.085), passive use remained a significant positive predictor of nonverbal decoding skill (β_{std} = 0.25, p = 0.001), and gender was a significant predictor, with females scoring higher on the WIPS test compared to males (β_{std} = 0.27, p < 0.001). This model accounted for 15% of the variance in WIPS scores.

Next, we regressed active use, passive use, and gender on scores from the DANVA-2AF. Once again, gender was the only significant positive predictor (β_{std} = 0.32, p < 0.001), with females scoring significantly higher than males. Approximately 11% of the variance in DANVA-2AF scores was accounted for by these three predictors.

When active use, passive use, and gender were entered into a regression predicting self-reported nonverbal decoding skill, active use was a significant positive predictor, similar to Study 1, (β_{std} = 0.21, p = 0.006), such that those who were more active technology users self-reported having more skill in decoding nonverbal information. Reporting more passive technology use was unrelated to self-reported nonverbal decoding skill. Gender remained a significant positive predictor (β_{std} = 0.31, p < 0.001) indicating that females self-reported more nonverbal decoding skill than males. Approximately 16% of the variance in self-reported nonverbal decoding skill was accounted for when active use, passive use, and gender were entered as predictors.

Mini Meta-Analysis

Finally, we conducted a mini meta-analysis (Goh et al., 2016) in order to provide a consistent account regarding the relationship between technology use and objective and self-reported measures of nonverbal decoding skill across these two studies. After combining these effects across both studies, we found that individuals who self-reported more active technology use self-reported higher nonverbal decoding skill (Mr = 0.22, p < 0.001), but scored lower on one objective index of nonverbal decoding skill (i.e., the WIPS test: Mr = −0.17, p < 0.001). Moreover, individuals who self-reported more passive use scored significantly higher on both objective indices of nonverbal decoding (i.e., the WIPS test: Mr = 0.18, p < 0.001 and the DANVA2-AF: Mr = 0.09, p = 0.023), but did not self-report higher levels of nonverbal decoding skill (Mr = 0.02, p = 0.667; **Table 3**).

DISCUSSION

While many have theorized about the potential positive or negative effects that technology may have on communication skills, no studies to date have empirically examined the relationship between technology use and nonverbal decoding skill. In order to begin to understand the ways in which technology use and nonverbal decoding skill are related, we measured multiple facets of each construct to more thoroughly examine their empirical relationships with one another.

While overall screen time was unrelated to any measure of nonverbal decoding skill, interesting and consistent patterns emerged when looking at the way individuals spent their time using technology. Specifically, individuals who reported actively posting and engaging with technology-mediated communication self-reported that they were more accurate at decoding the nonverbal behaviors of others. However, these more active users were more likely to score lower on objective measures of nonverbal decoding skill. Conversely, individuals who reported spending their time online passively viewing others' posts and photos scored higher on objective nonverbal decoding skill but did not self-report that their skills were any better.

These findings lend support to the role of practice and feedback as an effective way to increase nonverbal decoding skill (Blanch-Hartigan et al., 2012). Passive users of communication technology likely receive practice in decoding nonverbal cues simply by being exposed to other users' content (e.g., pictures, posts, videos) and thus a greater frequency of nonverbal cues. Indeed, the average screen time reported across both studies was about 5 h a day, meaning that passive users may spend up to 5 h each day practicing decoding nonverbal cues. In contrast to "other-focused" passive users, active users likely lose out on a plethora of communication cues as they report spending their time online engaging in "self-focused" activities. That is, although active users likely receive a great deal of practice *encoding* their own thoughts, feelings, attitudes, etc., they do not receive this same practice when it comes to *decoding* the thoughts, feelings, attitudes, etc. of others.

Therefore, these results support both the hypothesis that technology use enhances nonverbal decoding skill, and the hypothesis that technology use worsens nonverbal decoding skill. The key lies in *how* one spends their time using technological platforms. Those who use technology to practice making judgments of others may benefit from time online and learn skills to enhance their face-to-face interactions. However, greater technology use may have the opposite effect for those who choose to spend their time online creating and posting their own content, instead of interacting with the content of others. In these cases, technology may have adverse effects on an individual's nonverbal decoding skill in face-to-face interactions.

The current research is not without limitations. First, we are limited by our homogenous sample of college participants in one US state. More research is needed to see if the relationship between active and passive technology use and nonverbal

TABLE 3 | Mini meta-analysis results from study 1 and study 2 examining combined correlations between measures of technology use and nonverbal decoding skill.

	Objective		Self-report			
	WIPS test		DANVA-2AF		Emotional sensitivity	
	M*r* (SE)	Combined *Z* [95% CI]	M*r* (SE)	Combined *Z* [95% CI]	M*r* (SE)	Combined *Z* [95% CI]
Screen time (minutes)	−0.01 (0.05)	−0.19 [-0.11, 0.09]	$0.10^{\dagger}$ (0.05)	1.90 [0.00, 0.19]	0.02 (0.05)	0.34 [−0.08, 0.12]
Active use	-0.17^{***} (0.04)	−4.09 [−0.24, −0.09]	0.02 (0.04)	0.57 [−0.06, 0.10]	0.22^{***} (0.04)	5.33 [0.14, 0.30]
Passive use	0.18^{***} (0.04)	4.47 [0.10, 0.26]	0.09^{*} (0.04)	2.27 [0.01, 0.17]	0.02 (0.04)	0.43 [−0.06, 0.10]

Mr = weighted mean correlation (converted from rz to r). SE is standard error of mean r. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{***}p < 0.001$, *all two-tailed.*

decoding skill will generalize more broadly. In addition, while the WIPS test has many advantages to other tests of nonverbal decoding ability (e.g., good reliability and validity, real-world workplace context, dynamic stimuli, many domains of nonverbal sensitivity), it is not yet a published, validated test of decoding ability. Additionally, although self-reporting active and passive technology use provides valid information regarding the way participant's view their online activity, or the way they are motivated to be, future studies should confirm these self-reports with objective measures in order to assess the accuracy of individual's self-perceptions. We also examined one aspect of technology use on smartphone devices and the questions focused on self-reported social media use. The role of other technology-mediated communication platforms, such as teleconferencing or interactive video gaming, deserve future study. In our regression models, only 4–16% of the variance in decoding skills was explained by our predictors; therefore, there are many other factors that impact decoding skill ability which should be explored in future work. While the WIPS test is not validated yet (i.e., in prep), it is more ecologically valid than many other available standardized tests of decoding ability because it includes many workplace scenarios and dynamic video rather than focusing on one domain (e.g., emotion recognition like the DANVA-2AF) or using just static photographs where participants often show a ceiling effect on accuracy. In addition, and explained extensively below, we cannot make causal claims about the direction of the relationships given that our data was cross-sectional.

Suggestions to Further Theories of Technology Use and Nonverbal Decoding Skill

Although our data suggest that the way in which an individual communicates with technology may impact nonverbal decoding skills globally (i.e., as measured by the WIPS test), we only observed a marginally significant effect to suggest that technology use was related to an individual's ability to decode facial expressions of emotion measured via the DANVA-2AF. While it may be that technology truly does not impact this facet of nonverbal decoding skill, it is also possible that we did not measure technology use at a detailed enough level to reveal any meaningful relationships. Although participants reported technology use generally, different social media and technology communication platforms are vastly different in their bandwidth and each emphasize distinct cue channels. For example, while some platforms emphasize visual cues (e.g., Instagram, Snapchat) others may underscore more verbal cues (e.g., Facebook, Twitter). Collapsing technology use across all platforms may dilute interesting relationships between particular social media apps, cue channels, and nonverbal decoding skill. For instance, it may be that individuals who passively use applications which highlight posting pictures or videos receive more practice in decoding facial expressions, and therefore may score higher on emotion decoding tests such as the DANVA-2AF. Therefore, we urge future researchers to be thoughtful in selecting the most relevant nonverbal decoding skill measure for their particular study Stosic and Bernieri (in prep) taking into account domain (e.g., emotion recognition or general workplace decoding skills) as decoding ability does not appear to be a single skill (Schlegel et al., 2017a), and to further explore the ways in which specific technology-mediated platforms, opposed to global technology use, impact vital communication skills.

In addition to delineating more precise constructs, the areas of technology and nonverbal communication research would benefit from an increase in experimental designs. While we have interpreted our data as technology use potentially influencing nonverbal decoding skills, it is highly plausible that the causal relationship is reversed. Individuals who are more accurate perceivers of others' nonverbal behavior may be more likely to use technology in a passive way because they are more practiced, more comfortable, or more engaged with others. Those who are less accurate perceivers of others' nonverbal behavior may use technology more actively because they are more self-focused or find perceiving others to be more challenging or less rewarding. The correlational nature of the current studies does not allow us to untangle the direction of these effects. Therefore, we urge future work to consider experimental designs to examine the causal relationship between technology use and communication ability, particularly nonverbal decoding skill.

While experimental designs on this topic are rare, we are aware of one study that employed a quasi-experimental design to manipulate technology use. Age-matched cohorts of preteens attended a summer camp in a staggered order such that one group went earlier than the other group (Uhls et al., 2014). While at camp, electronics including television, computers, and mobile phones were not allowed. The first group to attend camp was the experimental group ($N = 51$) and the group that stayed at school while the first group was at camp was considered the control group ($N = 54$). After just 5 days of interacting face-to-face without the use of any technology, preteens' recognition of nonverbal emotion cues from photographs and videos (using the

DANVA-2 Child and Adult Faces and the Child and Adolescent Social Perception Measure) was significantly greater compared to the control group. From this, we can gather that the short-term effects of increased opportunities for face-to-face interaction, combined with time away from screen-based media and digital communication, improved preteens' understanding of and ability to decode nonverbal emotion cues.

Completely removing technology can be difficult in a real-world context; however, there are a variety of methods we propose to untangle the relationship between technology use and nonverbal decoding skill. There are applications and settings on most smartphones that display an alert when the user has reached a screen time maximum for the day. Researchers could consider a dose-response experiment in which they randomly assign different allowed hours of screen time to users each day for a series of days. One could then understand if different doses of screen time lead to higher or lower levels of nonverbal decoding skill.

In another potential research design, researchers could randomly assign the way technology is used by participants. Researchers could assign individuals as "passive users" who are not allowed to post but must read through others' posts and/or photographs. Some questions to consider are whether or not this would facilitate practice, contribute to learning, and improve nonverbal decoding skill. Another quasi-experimental design could follow emerging adolescents with or without phones and assess differences in their nonverbal decoding skills, accounting for covariates and confounders such as gender, socioeconomic status, parents' educational levels, and baseline communication skills.

In addition to experimentally manipulating technology use, research could examine and potentially rule out the reverse causality claim that nonverbal decoding skill is driving technology use. To do this, researchers could train participants on nonverbal decoding skill using validated trainings, such as the Geneva Emotion Recognition Test training (GERT; Schlegel et al., 2017b), and then assess whether technology use changes over time or if training nonverbal decoding skill makes technology-mediated communication smoother or more rewarding.

CONCLUSION

As the use of technology-mediated communication continues to expand, it is crucial for psychological research to address the positive and negative consequences of technology use on communication skills, in particular nonverbal communication. The current research suggests that it may not be the technology use itself, but rather how actively or passively users engage with technology, that facilitates or hinders nonverbal decoding skill. We ultimately found support for all hypotheses (i.e., Liberated Relationship Perspective, Internet Enhanced Self Disclosure Hypothesis, Reduction Hypothesis, and Cues Filtered Out Theory) but the ways in which the hypotheses were supported depended on how users interacted with technology. Our results showed that those who use technology in a more passive way (reading and look at others' posts) had higher nonverbal decoding accuracy. That is, more passive users may benefit from time online and learn skills to enhance their face-to-face communication (supporting the Liberated Relationship Perspective and Internet Enhanced Self Disclosure Hypothesis). For those who reported more active use (creating and posting their own content), they had lower nonverbal decoding accuracy. For these more active users, technology may have adverse effects on their ability to read and respond to others in face-to-face communication (supporting the Reduction Hypothesis and Cues Filtered Out Theory).

We believe these results to be encouraging, as some of the fears regarding the negative impact of technology on an individual's communication skills may not come to fruition if technology is used in a more passive, observational manner rather than an active, self-focused manner. Beyond these results, we also provide researchers with suggestions to further the field of technology use and communication skills. Due to the growing diversity in technology-mediated communication platforms, we urge researchers to account for the different functions theses platforms afford users. In addition, and perhaps most importantly, we urge researchers to explore experimental designs to determine causal pathways in the complex relationship between technology and communication skills. Researchers are beginning to understand how the technological revolution is changing the ways in which humans navigate social interactions. A deeper appreciation for this complexity can lead to the development of interventions to enhance and not hinder our communication skills with the increasing presence and benefits of technology in our lives.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Maine IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MR, MS, and JC contributed to conception, design of the study, and wrote the first draft of the manuscript. MR organized the database and performed the statistical analysis. DB-H wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

We thank research assistant, Vasiliqi Turlla, for her help in data collection and data cleaning and Herbert Ruben for always asking what technology was doing to our communication skills.

REFERENCES

Aslam, S. (2020a). *Facebook by the Numbers (2020): Stats, Demographics & Fun Facts*. Available online at: https://www.omnicoreagency.com/facebook-statistics/ (accessed September 28, 2020)

Aslam, S. (2020b). *Snapchat by the Numbers (2020): Stats, Demographics & Fun Facts*. Available online at: https://www.omnicoreagency.com/snapchat-statistics/ (accessed September 28, 2020)

Blanch-Hartigan, D., Andrzejewski, S. A., and Hill, K. M. (2012). The effectiveness of training to improve person perception accuracy: a meta-analysis. *Basic Appl. Soc. Psychol.* 34, 483–498. doi: 10.1080/01973533.2012.728122

Culnan, M. J., and Markus, M. L. (1987). "Information technologies," in *Handbook of organizational Communication: an Interdisciplinary Perspective*, eds F. M. Jablin, L. L. Putnam, K. H. Roberts, and L. W. Porter (Thousand Oaks, CA: Sage Publications, Inc), 420–443.

Döring, N., and Pöschl, S. (2008). "Nonverbal cues in mobile phone text messages: the effects of chronemics and proxemics," in *the Reconstruction of Space and Time: Mobile Communication Practices*. eds R. Ling and S. W. Campbell (Piscataway, NJ: Transaction Publishers), 109–136. doi10.4324/9781315134499-6

Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook "friends:" social capital and college students' use of online social network sites. *J. Comput. Mediat. Commun.* 12, 1143–1168. doi: 10.1111/j.1083-6101.2007.00367.x

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

Ferrán-Urdaneta, C., and Storck, J. (1997). "Truth or deception: the impact of videoconferencing for job interviews," in *Proceedings of the Eighteenth International Conference on Information Systems (ICIS 97)*, (Atlanta, GA).

Goh, J. X., Hall, J. A., and Rosenthal, R. (2016). Mini meta-analysis of your own studies: some arguments on why and a primer on how. *Soc. Pers. Psychol. Compass* 10, 535–549. doi: 10.1111/spc3.12267

Hall, J. A., and Gunnery, S. D. (2013). "Gender differences in nonverbal communication," in *Handbooks of Communication Science. Nonverbal Communication*, eds J. A. Hall and M. L. Knapp (Berlin: De Gruyter Mouton), 639–669.

Hall, J. A., Andrzejewski, S. A., and Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: a meta-analysis. *J. Nonverbal Behav.* 33, 149–180. doi: 10.1007/s10919-009-0070-5

Hu, Y., Wood, J. F., Smith, V., and Westbrook, N. (2004). Friendships through IM: examining the relationship between instant messaging and intimacy. *J. Comput. Mediat. Commun.* 10, 38–48.

Jackson, L. A., Zhao, Y., Kolenic, A. III., Fitzgerald, H. E., Harold, R., and Von Eye, A. (2008). Race, gender, and information technology use: the new digital divide. *Cyber Psychol. Behav.* 11, 437–442. doi: 10.1089/cpb.2007.0157

Kiesler, S., Siegel, J., and McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *Am. Psychol.* 39, 1123–1134. doi: 10.1037/0003-066x.39.10.1123

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., and Scherlis, W. (1998). Internet paradox: a social technology that reduces social involvement and psychological well-being? *Am. Psychol.* 53, 1017–1031. doi: 10.1037/0003-066x.53.9.1017

Neureiter, K., Fuchsberger, V., Murer, M., and Tscheligi, M. (2013). "Hands and eyes: how eye contact is linked to gestures in video conferencing," in *Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems*, eds W. E. Mackay, S. A. Brewster, and S. Bødker (New York, NY: Association for Computing Machinery), 127–132.

Nie, N. H. (2001). Sociability, interpersonal relations and the internet: reconciling conflicting findings. *Am. Behav. Sci.* 45, 420–435. doi: 10.1177/00027640121957277

Nowicki, S., and Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: the diagnostic analysis of nonverbal accuracy scale. *J. Nonverbal Behav.* 18, 9–35. doi: 10.1007/bf02169077

Patterson, M. L. (2019). A systems model of dyadic nonverbal interaction. *J. Nonverbal Behav.* 43, 111–132. doi: 10.1007/s10919-018-00292-w

Pew Research Center. (2019). *Mobile Fact Sheet*. Available online at: www.pewresearch.org/internet/fact-sheet/mobile/ (accessed November 15, 2020)

Riggio, R. E. (2005). "The social skills inventory (SSI): measuring nonverbal and social skills," in *the Sourcebook of Nonverbal Measures Going Beyond Words*, ed V. L. Manusov. New Jersey, NJ: Lawrence Erlbaum Associates Publishers, 25–33.

Sayce, D. (2020). *The Number of Tweets Per Day in 2020*. Available online at: https://www.dsayce.com/social-media/tweets-day/ (accessed September 28, 2020)

Schlegel, K., Boone, R. T., and Hall, J. A. (2017a). Individual differences in interpersonal accuracy: a multi-level meta-analysis to assess whether judging other people is one skill or many. *J. Nonverbal Behav.* 41, 103–137. doi: 10.1007/s10919-017-0249-0

Schlegel, K., Vicaria, I. M., Isaacowitz, D. M., and Hall, J. A. (2017b). Effectiveness of a short audiovisual emotion recognition training program in adults. *Motiv. Emot.* 41, 646–660. doi: 10.1007/s11031-017-9631-9

Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. W. (1986). Group processes in computer-mediated communication. *Organ. Behav. Hum. Decis. Process.* 37, 157–187. doi: 10.1016/0749-5978(86)90050-6

Sproull, L., and Kiesler, S. (1986). Reducing social context cues: electronic mail in organizational communication. *Management Science* 32, 1492–1512. doi: 10.1287/mnsc.32.11.1492

Stritzke, W. G., Nguyen, A., and Durkin, K. (2004). Shyness and computer-mediated communication: a self-presentational theory perspective. *Media Psychol.* 6, 1–22. doi: 10.1207/s1532785xmep0601_1

Tocci, M. (2020). *Text Marketing Statistics You Need To Know: SimpleTexting*. Available online at: https://simpletexting.com/all-the-text-marketing-statistics-you-need-to-know/ (accessed September 28, 2020).

Uhls, Y. T., Michikyan, M., Morris, J., Garcia, D., Small, G. W., Zgourou, E., et al. (2014). Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues. *Comput. Hum. Behav.* 39, 387–392. doi: 10.1016/j.chb.2014.05.036

Valkenburg, P. M., and Peter, J. (2009). Social consequences of the internet for adolescents: a decade of research. *Curr. Dir. Psychol. Sci.* 18, 1–5. doi: 10.1111/j.1467-8721.2009.01595.x

Vinciarelli, A. (2017). "Body language without a body: nonverbal communication in technology mediated settings," in *Proceedings of the International Multimedia Conference*, (Mountain View, CA).

Walther, J. B., and Parks, M. R. (2002). "Cues filtered out, cues filtered". *Handbook of Interpersonal Communication*, eds M. L. Knapp and G. R. Miller (Thousand Oaks, CA: Sage Publication)

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 January 2021
doi: 10.3389/fpsyg.2021.624177

Check for updates

Nonverbal Behaviors "Speak" Relational Messages of Dominance, Trust, and Composure

Judee K. Burgoon[1], Xinran Wang[1], Xunyu Chen[1], Steven J. Pentland[2] and Norah E. Dunbar[3]*

[1] Center for the Management of Information, The University of Arizona, Tucson, AZ, United States, [2] Information Technology and Supply Chain Management, Boise State University, Boise, ID, United States, [3] Department of Communication, UC Santa Barbara, Santa Barbara, CA, United States

Nonverbal signals color the meanings of interpersonal relationships. Humans rely on facial, head, postural, and vocal signals to express relational messages along continua. Three of relevance are dominance-submission, composure-nervousness and trust-distrust. Machine learning and new automated analysis tools are making possible a deeper understanding of the dynamics of relational communication. These are explored in the context of group interactions during a game entailing deception. The "messiness" of studying communication under naturalistic conditions creates many measurement and design obstacles that are discussed here. Possibilities for their mitigation are considered.

Keywords: nonverbal communication, relational messages, dominance, nervousness, trust, deception

Edited by:
Leonardo Carlucci,
Università degli Studi "G. d'Annunzio" Chieti – Pescara, Italy

Reviewed by:
Gary D. Bond,
Eastern New Mexico University, United States
Kevin Shockley,
University of Cincinnati, United States

****Correspondence:***
Judee K. Burgoon
jburgoon@cmi.arizona.edu;
judee@arizona.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *30 October 2020*
Accepted: *04 January 2021*
Published: *26 January 2021*

Citation:
Burgoon JK, Wang X, Chen X, Pentland SJ and Dunbar NE (2021) Nonverbal Behaviors "Speak" Relational Messages of Dominance, Trust, and Composure. Front. Psychol. 12:624177. doi: 10.3389/fpsyg.2021.624177

INTRODUCTION

A mainstay of interpersonal communication is the concept of relational communication, constituted through a constellation of dimensions along which actors express implicit messages about how they regard one another and their interpersonal relationship. These messages are expressed predominantly through nonverbal rather than verbal signals. Although Burgoon and Hale (1984) have identified up to 12 non-orthogonal themes or dimensions along which relational messages can be exchanged, three of the most prominent ones are dominance, trust, and composure. Until recently, the subtlety with which these messages are sent and received has challenged the ability of scientists to capture and describe them. Human observational skills are subjective and operate at a macroscopic level that constrains the measurement of such messages. Moreover, the laborious nature of manual behavioral coding has been a limiting factor on their use in discerning complex social dynamics. Now, with the benefit of new technologies and methods, the nonverbal means by which humans "speak" relational messages can be uncovered objectively, microscopically and dynamically, sometimes to the point of measurement outstripping our clear understanding but at least prompting intriguing possibilities.

Laboratory studies of human behavior are often critiqued for being artificial and highly scripted, with confederates following strict interview protocols and engaging in unnaturally brief interactions (see, e.g., Frank et al., 2008; Fuller et al., 2011; Frank and Svetieva, 2012). In this paper, we report the results of an experiment in which interactions unfold naturally rather than being scripted, the experimental induction introduces enough range in sentiment for participants to develop favorable and unfavorable judgments of one another, interactions are lengthy enough to produce changes in sentiments, and relational messages are measured at multiple intervals so that their dynamics can be captured over time. Moreover, the methods afford measurement of a wealth of nonverbal signals from the head, face, torso and voice as predictors of participants' own understanding of the relational messages they are receiving from fellow participants. This permits us to identify the nonverbal signals most likely to express three relational dimensions of interest here–dominance, composure and trust–as interactions progress.

BACKGROUND

The concept of relational messages can be traced to the term "metacommunication," coined by Bateson (1951, 1958) to describe signals that distinguish between the "report" and "command" functions of communication and create a frame for understanding it. The report level refers to the content, whereas the command level directs the recipient, the signaler, or both, how to interpret the verbal content. Usually, the metacommunication is considered the nonverbal signals that accompany the verbal content and serves to clarify, amplify or even contradict the verbal content. This distinction was applied in the clinical context, where Watzlawick et al. (1967) used it to refer to observations of how patients interact with their therapists. Their body language in particular expressed implicit messages of how the patients regarded the therapists. These implicit messages, known as relational communication, became a mainstay of interpersonal communication.

Early work applied the construct in such other contexts as a theory of personality (Leary, 1957), the dimensions of meaning in language (Osgood et al., 1957), interpersonal needs (Schutz, 1966), source credibility (McCroskey, 1966), group decision-making (Bales, 1968), immediacy (Mehrabian, 1971), categories of social relationships (Mehrabian and Ksionzky, 1972), intraspecific displays (Andrew, 1972), transactional social relationships (Millar and Rogers, 1976), and interpersonal interaction (Rogers and Farace, 1975; Parks, 1977) and relationship terms (Knapp et al., 1980). Based on a review of these various literatures, Burgoon and Hale (1984) expanded on the concept to 12 topoi, or generic themes, of relational communication *continua*. The dimensions that emerged as most central and recurrent were dominance-submission and affection-hostility. Additional dimensions included trust and composure. Given their relevance to interpersonal and group communication, these dimensions were chosen to reflect participant judgments in an experiment on group communication. The investigation had as a central focus how deception is enacted in group deliberations, making the topoi of dominance, trust and composure particularly germane. Because exploring the dynamics of relational messaging was additionally one of the objectives of the investigation, and it was thought that affection-hostility (liking) would be unlikely to change over an hour's discussion, affection-hostility was only measured at the end of the discussion.

OVERVIEW

The experiment examined relational communication and deception over multiple phases during group interaction. The sample was multicultural. The exploration of group interaction across multiple, diverse cultures represents a rare approach in several respects. It examines actual nonverbal behavior as opposed to imagined behavior or self-reports of recollected behavior. It allows lengthy rather than brief interchanges and group rather than dyadic interactions. As well, its inclusion of samples from multiple, diverse cultures is also an improvement over studies that make comparisons between two countries chosen for convenience' sake, or comparisons by countries rather than self-defined by cultural orientations (see, e.g., Giles et al., in press). The inclusion of samples from eight different locations and six different countries with diverse self-reported cultural orientations adds significant range to the cultures that are represented. All of these characteristics—actual interactions, lengthy interactions, group deliberations, and cultural comparisons across multiple cultural orientations—represent advances in deception and relational communication research. Here we present that portion of the research concerned with the nonverbal features associated with relational communication.

Seldom have the nonverbal behaviors associated with relational communication dimensions been studied in depth because of the laborious nature of manually coding nonverbal behavior (for an exception, see Burgoon and Le Poire, 1999, which was a 3-year undertaking). The current project represents a significant advance into the behavioral particulars and dynamics inherent in nonverbal relational message exchange. The nonverbal behaviors were measured using automated tools and analyses incorporating artificial intelligence. Not only did these measurement and analysis tools make it possible to measure far more behaviors in far less time than with manual coding but made it possible to measure microscopic behavior that is neither measurable by human observers nor observable with the naked eye. It was also possible to record analyses over a longer period of time so as to capture the dynamics of those nonverbal behaviors that are not static. In the current case, we recorded group interactions 1 h in length.

HYPOTHESES

Dominance

Dominance-submission is one of the most fundamental and widely recognized dimensions of human relations (Massey-Abernathy and Haseltine, 2019). Though dominance can be defined from different disciplinary perspectives, we adopt the definition proposed by Burgoon et al. (1998) that interpersonal dominance is "a relational, behavioral, and interactional state that reflects the actual achievement of influence or control over another via communication actions." This definition indicates that unlike power, which entails potentialities for exerting influence on others, dominance is accomplished through actual dyadic interaction. It is achieved behaviorally through particular interaction strategies, such as threat, elevation or initiation.

Nonverbal behaviors associated with perceived dominance are multi-faceted and vary according to the context. For example, silence can be a symbol of threat and dominance (Bruneau, 1973) in one case while an embodiment of submission in another case. Previous studies have reported that facial expressions, such as lowered brows or a non-smiling mouth, are associated with perceived dominance (Keating et al., 1981; Witkower et al., 2020). On the opposite side, body collapse and gaze avoidance correlate with submissiveness (Weeks et al., 2011). On the vocal side, lower pitch (Cheng et al., 2016), loudness (Tusing and Dillard, 2000), vocal variability, rapid speech rate (Hall et al., 2005), jitter, shimmer, and pleasing voice quality (Hughes et al., 2014) have also been reported to correlate with perceived dominance. Tusing and Dillard (2000) also suggested a gender differentiation in vocal indicators of dominance. Bente et al. (2010) reported *trans*-cultural universalities in the recognition of interpersonal dominance.

Burgoon and Dunbar (2006) categorized three overarching principles associated with the nonverbal expressions of dominance: physical potency, resource control and interaction control, each of which has certain nonverbal manifestations. This set of principles delineates dominance-establishing strategies on a higher level, and based on this taxonomy, we can hypothesize the nonverbal behaviors that might influence perceived dominance. Many of these strategies such as threat or elevation would be inappropriate for ostensibly cooperative group settings. However, others, such as signals of potency through intimidating facial expression, dynamic facial expression and loud voices, indicators of size through gesture and posture, and control of interaction through turn initiation, speaker interruptions and control of gaze patterns would (see Burgoon et al., 2002). Thus, we hypothesize that dominant group members, compared to nondominant members, will have:

- More expansive and upright postures and head positions,
- More gaze and receipt of gaze,
- Less smiling but more expressive facial expressions,
- More initiations of turns-at-talk and longer turns at talk,
- Louder voices, and
- More interruptions of others.

A fuller delineation of dominance signals appears in **Table 1**, hypothesized relationships between perceived dominance and nonverbal behaviors. Their opposites would signify non-dominance and submission. So, for example, non-dominant group members would have fewer and shorter speaking turns, quieter voices, few interruptions, more smiling, more constricted postures, more head tilt, more rigid facial expressions and more eye gaze while listening than speaking.

To the extent that they can be measured automatically, each of the behaviors in **Table 1**: Hypothesized relationships between perceived dominance and nonverbal behaviors, can be hypothesized as indicators of dominance. Ones such as elevation, choice of seating position and self-nomination that would not be involved are omitted. The measurements are described in the "Materials and Methods" Section.

TABLE 1 | Hypothesized relationships between perceived dominance and nonverbal behaviors.

Principles	Strategies	Hypothesized nonverbal signals of dominance
Physical potency	Threat	• More glare and stare
	Size or strength	• Louder voice • Deep-pitched voice • Clear articulation (higher voice quality) • Non-smiling face • Upright head and posture
	Expressivity	• More facial expression • More variation in pitch • More head/body movement • More rapid speaking tempo
Resource control	Command of space	• More open body position • More expansive posture
	Precedence	• Initiation of more turns at talk • Longer turns at talk
	Prerogative	• Choice of seating position
	Possession of valued commodities	• More turns-at-talk • Longer turns at talk
Interactional control	Centrality	• More looking while speaking, less looking while listening • Interruption of others' speaking turns
	Elevation	• Standing or seating above others
	Initiation	• Initiating a conversation
	Non-reciprocation	• Non-matching of others' behavior
	Task performance cues	• Self-nomination

Composure

In the context of relational communication, composure is the degree of tension or relaxation experienced within a relationship. Generally, increased levels of composure during interactions leads to more positive outcomes. For instance, manager composure leads to increased employee satisfaction, motivation, and organizational commitment (Mikkelson et al., 2017). Further, professionals may attempt to present themselves as composed in order to instill trust and confidence (Finch, 2005). In a comprehensive analysis of the nonverbal behaviors that influence perceptions of composure, Burgoon and Le Poire (1999) found indicators associated with pleasantness or positivity, expressivity, involvement and immediacy, relaxation, and conversational management. This suggests that composed individuals are active and engaged communicative partners, while also creating an atmosphere of pleasant relaxation and accessibility for the interaction. Not all of these factors like conversational management are as easily exhibited in groups as in dyads, and proxemic immediacy toward one group member might mean non-immediacy with another, but most are relevant in the group context. Thus, the composed group member should show pleasant facial and vocal emotion, expressive (varied) faces and voices, high amount of talk time, relaxed posture, relaxed face and head, relaxed voice, deep pitch, relaxed laughter, moderate loudness and moderate tempo. On the other hand, lower duration of eye contact, non-smiling mouth movement and more jittery hand movement may be signals of anxiety (Waxer, 1977).

Regarding vocal activity, Daly et al. (1977) found that frequency and duration of interactions was positively related to composure, although ratings of composure leveled off and declined as individuals' vocal activity surpassed 50–60 percent of total time in groups. Presumably, composed individuals must balance between being active and engaged but also preserving a degree of comfort that requires some holding back. Excessive vocalization may undermine the impression of composure as individuals walk the line between being engaged but not overwhelmingly so.

At the other end of the composure continuum, nervousness and communication avoidance or apprehensiveness may also lessen composure evaluations (Burgoon and Koper, 1984). Whereas composure can instill confidence and success, reticent communication is seen as communicating disinterest or social anxiety (Burgoon and Hale, 1983). Reticent communicators are described as being disengaged from a conversation, withdrawing or attempting to avoid interaction altogether. This can occur for a variety of reasons including chronic communication apprehension or situational attempts to suppress information exchange. Reticence can lead to suspicion and stalemate.

Perceptions of reticent communicators arise from nonverbal behaviors associated with negative arousal, non-immediacy, tension, and anxiousness (Burgoon and Koper, 1984; Mann et al., 2020). Under circumstances that are moderately anxiety-provoking, the reticent individual may exhibit stress-related indicators such as increased fidgeting, adaptor gestures, elevated pitch and strident voice quality; under circumstances that are more anxiety-provoking, such as an interrogation, the communicator may go into "lock-down," exhibiting the rigidity pattern associated with tension—flat affect, reduced facial and head expressivity, little vocal variety and the like.

Given the need to suppress information during deception, deception researchers have investigated behavioral cues associated with reticent communication. People often assume that nervousness is a sign of deceptiveness even if objectively, that is untrue (Vrij and Fisher, 2020). Although the proposition that deception can be revealed through shifty eyes has all but been disproven, cues associated with deception do overlap with ones associated with anxiety and nervousness. In particular, rigidity is a potential indicator of deception (Twyman et al., 2014; Pentland et al., 2017). Although this research does not necessarily align rigidity with nervousness, nervous behaviors are associated with rigidity, including kinesic cues (Gregersen, 2005) and vocal tension (Laukka et al., 2008). This is partly because deception is thought to increase cognitive load, and higher cognitive load reduces overall activation (Vrij and Fisher, 2020). However, previous studies showed mixed results in the associations between vocal variations and nervousness. Whereas Laukka et al. (2008) found no variation differences when specifically analyzing nervousness, Hagenaars and van Minnen (2005) found lower variation in pitch during episodes of fear versus happiness. Nervousness is potentially a more salient trait than composure to perceive in a group setting. In this context, nervousness is viewed as the bipolar opposite of composure. Given that the behaviors of a calm and collected group member may go unnoticed, participants in the current study were asked to rate the nervousness of group members. We hypothesize that perceptions of nervousness (either caused by social pressures or attempts to conceal information) will be exhibited in more rigid and tense behaviors.

In sum, we expect to replicate Burgoon and Le Poire (1999) and Pentland et al. (2017) in finding nonverbal indicators of nervousness and tension, as compared to composure, that include:

- Rigid and tense behaviors such as reduced head and face movements,
- Less immediacy (less gaze and indirect facing),
- Less vocal and kinesic pleasantness (e.g., less relaxed laughter and vocal resonance),
- Softer vocal amplitude, less vocal fluency,
- Higher pitch,
- Fewer and shorter turns-at-talk.

Left as a research question are the other vocal variations perceived as nervousness. **Table 2**: Hypothesized relationships between perceived nervousness and nonverbal behaviors, operationalizes the relationship between these principles and hypothesized nonverbal signals.

Trust

Scholars from a variety of fields have studied trust through differing disciplinary lenses. Psychologists, for example, might emphasize the attributes of the individual that foster perceptions held by another, whereas sociologists might examine the relationships present within groups that lead to trust

TABLE 2 | Hypothesized relationships between perceived nervousness and nonverbal behaviors.

Principles	Strategies	Hypothesized nonverbal signals of nervousness
Withdrawn from engagement	Low expressivity	• Softer amplitude • More rigidity in facial animation • More rigidity in head movements
	Low conversational management	• Shorter talk time • Fewer turns-at-talk
Unpleasantness	Negativity	• Fewer pleasant facial expressions • Less nodding
	Non-immediacy	• Direct facing (not measurable in a group setting)
	Tension	• Higher pitch • Less vocal variation • Less vocal fluency • More rigidity in facial animation • More rigidity in head movements • Fidgety with hands and feet (not measured)

(Rousseau et al., 1998). At its core, trust is an *expectancy* about future behavior since one must assume that a person, group, or organization will behave in a particular way. If we trust a person, we are taking a risk and making ourselves vulnerable to them, so we want some assurance about what will happen when we do. Gottman (2011) argues this comes down to two factors: transparency and positive moral certainty. Transparency is the opposite of deceptiveness and allows us to rely on the other person when necessary. Positive moral certainty means we believe our partner is an ethical, moral person and will "treat us and others with high moral standards, integrity, honesty, kindness, love, and goodwill" (Gottman, 2011, p. 177). DeSteno (2014) argues that trust has two facets: integrity and competence. We are willing to trust someone who has both a high moral character and the expertise we need.

Trust always entails some level of risk, uncertainty, or willingness to be vulnerable through reliance and disclosure (van der Werff and Buckley, 2017). Simpson's (2007) dyadic model of trust determines whether or not trust will result from an interaction between two interdependent actors or groups. The participants must be willing to take a risk and make themselves vulnerable for the sake of a mutually beneficial outcome. Each partner makes an independent assessment of whether the other is making decisions contrary to their own self-interest in favor of the best interests of the partner or the relationship (called "transformation of motivation"). Burgoon et al. (in press) recently posited an integrated adaptative "spiral model of trust." The spiral model suggests that positive violations of expectancies (ones that conform to or surpass expectations) are more welcome than confirmations and therefore are more likely to foster trust than negative violations (ones that fail to meet expectations). These expectancy violations can take both verbal and nonverbal forms but our focus here is on the nonverbal ones. Boone and Buck (2003) proposed that emotional expressivity, the degree to which individuals accurately communicate their feeling states, helps to establish trustworthiness. Although existing research has found correlations between perceived trustworthiness and both verbal and nonverbal behavior (Wood, 2006; Lucas et al., 2016; Chen et al., 2020), people place overreliance on facial expressions when assessing credibility (Tsankova et al., 2012; Lucas et al., 2016).

What does trust look like, nonverbally? It can be expressed dyadically in the form of reciprocity, convergence, synchrony, or involvement that two partners share, or it can be examined individually in terms of the amount of uncertainty, tension, and suspicion that is expressed. Burgoon (in press) proposed that suspicion's relationship to trust is curvilinear in that it is associated with the highest degree of uncertainty. As uncertainty is reduced, suspicion either morphs into distrust (greater certainty about the other's untrustworthiness) or trust (greater certainty about the other's trustworthiness) (Toma and Hancock, 2012). Once the nature of another person's motives become known, whether or not they are trustworthy becomes known. Furthermore, we expect trust to correlate with dominance positively, because dominant individuals tend to convey confidence and appear competent (Anderson and Kilduff, 2009), and competence instills trust. Additionally, trust is expected to be associated with less nervousness, because individuals may employ nervousness as a heuristic when judging veracity, although nervousness may not imply lying (Feeley and deTurck, 1995; Vrij and Fisher, 2020). Consequently, we can hypothesize:

- Nonverbal indicators of dominance are positively associated with trust.
- Nonverbal indicators of tension and nervousness are negatively associated with trust.
- Uncertainty is negatively associated with trust.
- Nonverbal indicators of involvement and immediacy are positively associated with trust, specifically,

(1) High amounts of gaze, (2) direct facing, (3) forward lean, (4) rapid speech, (5) short response latencies, (6) fluent speech, and (7) long turns-at-talk.

MATERIALS AND METHODS

Participants

We conducted 95 experimental sessions at eight universities around the world. Due to video recording failures, we used 56 games and 379 players (166 males and 213 females) for this study. Specifically, this sample includes 9 games (61 players) from the Western United States, 6 games (41 players) from the Southwestern United States, 6 games (42 players) from the Northwestern United States, 3 games (20 players) from Israel, 8 games (58 players) from Fiji, 4 games (21 players) from Zambia, 10 games (66 players) from Singapore, and 10 games (70 players) from Hong Kong, China. Furthermore, participants recruited at the same site were culturally diverse because of recruitment of college students with international experiences and various tribal backgrounds. Age averaged 21.90 years (*sd* = 3.46 years), with 12 participants not reporting their age. Among the 366 participants who reported their ethnicities, 48.1% were Asian, 18.9% were white, 13.7% were Fijian, 6.6% were black, while Latin/Hispanic, multiracial and other individuals accounted for 5.5, 4.4, and 3%, respectively. Additionally, among the 368 participants who reported their native languages, 44.3% were native English speakers.

Experiment Procedures

The experiment consisted of group interactions using a scenario modified from popular board games, Mafia (designed by Dimitry Davidoff) and the Resistance (designed by Don Eskridge). Groups of six to eight participants were seated equidistant from one another in a circle at desks, each with a laptop computer. Participants first took turns to introduce themselves and answer a follow-up question from another participant as an icebreaker activity. Next, two to three of them were randomly assigned the role of a spy, while the rest of the group were assigned the role of villagers. Villagers were to conduct missions to eliminate spies, who were attempting to infiltrate the village. Spies were to try to sabotage the missions. Villagers would win a point for each mission that succeeded; spies would win a point for each mission that failed. At various junctures, the players rated one another on the relational dimensions of dominance, nervousness, and trustworthiness. Because the spies were working against the interest of villagers, variance was introduced in how players judged one another. Only the spies knew who the other spies were, creating uncertainty among villagers as to who to trust. In this sample, 229 players were assigned to be villagers, and 150 were spies; 53.6% of them had played a game similar to our experiment scenario before. Villagers won 28 games.

The game consisted of up to eight rounds and was capped at 1 h. For each round, participants were asked to complete missions in a hypothetical town. First, they elected a leader. The leader then chose a team of three to five (depending on the size of the group and how many rounds had been played) that had to be approved by a majority vote from the group. Next, teams "completed" the mission through anonymous votes. Villagers were always expected to vote for the missions to succeed. Spies were expected to vote to fail the missions, although they might vote strategically for a mission to succeed. One or two failed votes caused the mission to fail, depending on the size of the group and how many rounds had been played. The final winners were those who won more rounds, which earned team members monetary rewards. In addition, elected leaders would earn extra money. Audio-visual signals from each player were recorded during the entire game, including the icebreaker activity (see Dorn et al., in press, for a complete description of the experimental protocol).

Independent Variables – Nonverbal Behaviors

The nonverbal behavioral features covered in this study include facial, head pose, and vocalic features. To extract facial features, videos were captured from front-facing cameras built into the computer tablets, an overhead 360-degree camera, and a webcam that recorded the entire group interaction. We fed the video recordings of every player into OpenFace, an open-source deep-neural-network (DNN) based facial recognition tool (Baltrusaitis et al., 2018), which output an intensity score (from zero to five) of 17 facial action units (FAUs) for each frame. We calculated the mean and standard deviation of these 17 FAUs for every player and game phase. **Table 3**: Facial action units (AU) output by OpenFace lists the names of the FAUs (Ekman and Friesen, 1978). OpenFace also output three features that represent the 3D location of the head with respect to camera and three features of head rotation (i.e., pitch, yaw, and roll). The time to take within-game surveys was excised from each video. OpenFace developers conducted extensive experiments to demonstrate the tool's state-of-the-art performances on FAU detection and head pose estimation (Baltrusaitis et al., 2018), thus our analysis results are reliable to the extent that OpenFace is a valid tool.

To extract vocalic features, we first developed a pipeline, a procedure for speech detection and audio alignment to segment audio files of players' turns-at-talk. Specifically, we started with detecting speech in audio recordings of each player. Because audio waves picked up by each microphone were slightly different due to different distances between microphones and speakers, we employed an audio alignment algorithm named dynamic time warping to align the audio waves of the same utterance. Then, we identified speakers based on the highest loudness, because the loudest recording should be picked by the microphone assigned to the speaker. Finally, we segmented audio files into players' turns-at-talk. **Figure 1** summarizes the pipeline.

TABLE 3 | Facial action units (AU) output by OpenFace.

AU number	Description
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
12	Lip corner puller
14	Dimpler
15	Lip corner depressor
17	Chin raiser
20	Lip stretcher
23	Lip tightener
25	Lips part
26	Jaw drop
45	Blink

The audio segments were processed by OpenSmile, a software tool for automatic feature extraction from audio signals (Eyben and Schuller, 2015). Based on the demonstrations of the validity of OpenSmile for multimedia recognition tasks by Eyben et al. (2013), we judged OpenSmile to be a valid tool for our analysis.

Table 4: Acoustic measures and descriptions lists the turn-at-talk vocalic features output from OpenSmile used in our multilevel regression analyses. These features were averaged for each game phase. Additionally, we standardized speaking tempo by dividing the word count by speech time for each player in every game phase. Because of the high costs associated with obtaining accurate transcripts, we used a subset of 28 games whose transcripts were available for further analysis of speaking tempo. Lastly, we standardized all the nonverbal behavioral features within the same game session and the same game phase.

Dependent Variables – Ratings

To measure the relational dimensions, participants rated each other on dominance, composure, and trustworthiness on five-point scales. The survey items are described below. The questions incorporated several of the typical bipolar adjectives used to measure each dimension. To avoid fatigue, these were combined into single item measures.

- Please rate how dominant each player was during this round. Were they active and forceful or passive and quiet? A rating of 5 would mean you thought they were assertive, active, talkative, and persuasive. A score of 1 would mean

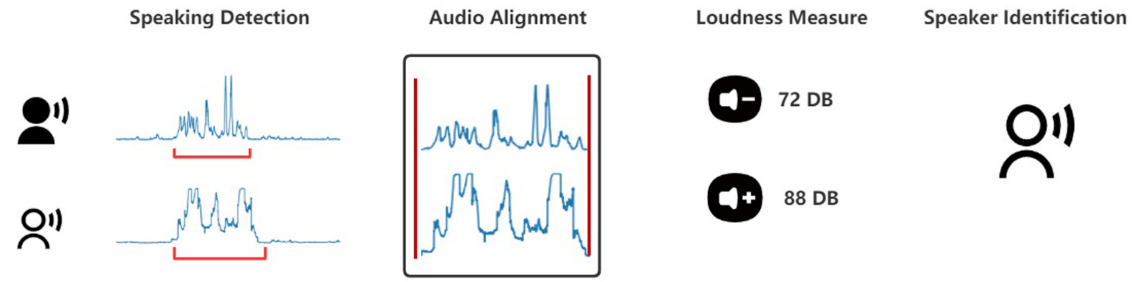


FIGURE 1 | Pipeline for segmenting audio files of players' turns-at-talk.

you thought they were unassertive, passive, quiet and not influential. Please mark any number from 1 to 5.

- Please rate how nervous each player was during this round. A rating of 5 would mean you thought they were anxious, uncomposed, and tense. A rating of 1 would mean you thought they were calm, composed, and relaxed. Please mark any number from 1 to 5.
- Please rate how much you trusted each player during this round. Were they trustworthy or suspicious? A rating of 5 would mean they were honest, reliable and truthful and 1 would mean you thought they were dishonest, unreliable and deceitful.

Ratings were collected after the icebreaker and every two rounds. This allowed measurement of dynamics in the relational messages. Because games varied in the number of rounds, all games were segmented into three game phases, namely, the icebreaker (phase 1), Round 1 and 2 (phase 2) combined, and Round 3 till the end of the game, combined (phase 3). For each game phase, villagers' ratings of each player were averaged separately on dominance, nervousness, and trustworthiness. Ratings from spies were excluded because of contamination by their knowledge of others' game roles. Additionally, self-ratings were excluded.

RESULTS

The Mixed-Effects Regression Model

To test the relationships among dominance, nervousness, trust and nonverbal behaviors, multivariate mixed-effects regression models were specified for each of the dependent variables of dominance, nervousness and trust perceptions. Control variables included game phase (phase 1, phase 2, or phase 3), game role (spy or villager), gender (male or female), previous game experience (yes or no), English as a second language (yes or no), and game score difference between spies and villagers by game phase. The interaction effect between game phase and game role was included because perceptions of players with different roles may have had different trends as the games progressed. As shown in the **Supplementary Material**, spies were perceived as less dominant (**Supplementary Tables 1–3**) and less trustworthy (**Supplementary Tables 7–9**) over time. Individual nonverbal behaviors were set as independent variables and their unique game name was specified as a random effect. Equation 1 specifies the regression equation.

Equation 1: Mixed-Effects Regression Model

$$\begin{aligned} &\textit{Relational Message Score} \\ &\quad = \textit{Game Phase} + \textit{Game Role} + \textit{Game Phase} \times \textit{Game Role} \\ &\qquad + \textit{Gender} + \textit{Game Experience} + \textit{Native Language} \\ &\qquad + \textit{Game Status} + \textit{Nonverbal Behavior} + (1|\textit{Game}) + \epsilon \end{aligned}$$

Table 1: Hypothesized relationships between perceived dominance and nonverbal behaviors and **Table 2**: Hypothesized relationships between perceived nervousness and nonverbal behaviors represent a theoretical delineation of dominance and nervousness principles and hypothesized nonverbal signals. Trust signals are presumed to draw from the dominance and nervousness indicators. The tables present an expansive look at nonverbal signals available from full frontal videos from the shoulders up. Due to data collection constraints or behavioral coding limitations, some behaviors are not included in the current data analysis. Specifically, features related to eye behavior, trunk and limb movement, and interactional dynamics (non-reciprocation and self-nomination) are excluded from the analysis.

Table 5: Mixed-effect regression results for nonverbal behaviors related to dominance presents test results for each nonverbal behavior analyzed with respect to perceptions of dominance. For simplicity, the results of control variables are omitted.

Results indicate that perceptions of dominance are associated with a louder voice, more expressive facial behavior, more head movement, and more and longer turns-at-talk. Vocal pitch, head

TABLE 4 | Acoustic measures and descriptions.

Measure name	Description
F_0 (pitch) Mean F_0 (pitch) Std	The low to high level of a tone perceived by humans as pitch.
Loudness-Mean Loudness-Std	The amplitude of sound pressure perceived as loudness.
Turn-at-talk Duration	Duration of a turn-at-talk in seconds

TABLE 5 | Mixed-effect regression results for nonverbal behaviors related to dominance.

Principle	Strategies	Hypothesized	Measurement	β(*SE*)
Physical potency	Size of strength	Louder voice	Mean vocal loudness	0.157 (0.03)***
		Deep-pitched voice	Mean vocal pitch	−0.04 (0.03)
		Upright head and posture	Mean head pitch	−0.008 (0.03)
			Mean head yaw	−0.017 (0.03)
			Mean head roll	0.017 (0.03)
	Expressivity	More facial expression	SD AU01	0.054 (0.03)*
			SD AU02	0.064 (0.03)*
			SD AU04	0.006 (0.03)
			SD AU05	0.059 (0.03)*
			SD AU06	0.075 (0.03)**
			SD AU07	0.029 (0.03)
			SD AU09	0.067 (0.03)**
			SD AU10	0.072 (0.03)**
			SD AU12	0.043 (0.03)+
			SD AU14	0.123 (0.03)***
			SD AU15	0.106 (0.03)***
			SD AU17	0.057 (0.03)*
			SD AU20	0.059 (0.03)*
			SD AU23	0.108 (0.03)***
			SD AU25	0.126 (0.03)***
			SD AU26	0.085 (0.03)***
			SD AU45	0.075 (0.03)**
		More variation in pitch	SD vocal pitch	−0.047 (0.03)+
		More head movement	SD head pitch	0.12 (0.03)***
			SD head yaw	0.069 (0.03)**
			SD head roll	0.074 (0.03)**
		More rapid speaking tempo	Word count/ speaking time	0.029 (0.039)
Resource control	Possession of valued commodities	More turns-at-talk	Count of turn-at-talk	0.34 (0.024)***
		Longer turns at talk	Mean turn-at-talk duration	0.153 (0.025)***

*+p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001.*
*Only beta weights and standard errors for each nonverbal behavior are presented. See **Supplementary Tables 1–3** for full model results.*

position, and speaking tempo were not found to be significantly related to perceptions of dominance.

Next, we tested relationships between selected nonverbal behaviors and nervousness. **Table 6**: Mixed-effect regression results for nonverbal behaviors related to nervousness presents these results. Generally, we expect to see withdrawn and tense behaviors.

The results show that nervousness is associated with more rigid head movements, and fewer and shorter turns-at-talk.

TABLE 6 | Mixed-effect regression results for nonverbal behaviors related to nervousness.

Principle	Strategies	Hypothesized	Measurement	β(*SE*)
Withdrawn from engagement	Low expressivity	Softer amplitude	Mean vocal loudness	−0.038 (0.02)+
		More rigidity in facial animation	SD AU01	−0.013 (0.02)
			SD AU02	−0.06 (0.02)**
			SD AU04	0.033 (0.02)
			SD AU05	−0.013 (0.02)
			SD AU06	−0.006 (0.02)
			SD AU07	−0.014 (0.02)
			SD AU09	0.002 (0.02)
			SD AU10	−0.022 (0.02)
			SD AU12	0 (0.02)
			SD AU14	−0.032 (0.02)
			SD AU15	−0.038 (0.02)
			SD AU17	−0.005 (0.02)
			SD AU20	−0.024 (0.02)
			SD AU23	−0.031 (0.02)
			SD AU25	−0.032 (0.02)
			SD AU26	−0.037 (0.02)
			SD AU45	−0.031 (0.02)
		More rigidity in head movement	SD head pitch	−0.05 (0.02)*
			SD head yaw	−0.044 (0.02)+
			SD head roll	−0.061 (0.02)**
	Low conversational management	Small amount of talk time	Mean turn-at-talk duration	−0.052 (0.02)*
		Fewer turns-at-talk	Count of turn-at-talk	−0.063 (0.02)**
Unpleasantness	Tension	Higher pitch	Mean vocal pitch	0.038 (0.03)
		Less vocal variation	SD vocal pitch	0.032 (0.02)

*+p < 0.1, *p < 0.05, **p < 0.01.*
*Only beta weights and standard errors for each nonverbal behavior are presented. See **Supplementary Tables 4–6** for full model results.*

Measures associated with loudness, pitch, vocal variation, and facial animation were not found to be significant.

Lastly, the correlation between trust and dominance is 0.31 ($p < 0.001$), and the correlation between trust and nervousness is −0.082 ($p < 0.01$). Both correlations are in the same direction as expected, but the relationship between trust and nervousness accounts for virtually no variance, indicating these measures are independent. **Table 7**: Mixed-effect regression results for nonverbal behaviors related to trust presents the relationship between trust and the hypothesized nonverbal behaviors. We expect to see indicators of trust similar to those of dominance and opposite to those of nervousness.

Results show that more, longer, and slower turns-at-talk are positively associated with perceptions of trust, while loudness, pitch, upright head and posture, and less rigidity in

TABLE 7 | Mixed-effect regression results for nonverbal behaviors related to trust.

Dominance indicators	Nervousness indicators	Hypothesized for trust	Measurement	β(*SE*)
Louder voice	Softer amplitude	Louder voice	Mean vocal loudness	0.026 (0.02)
Deep-pitched voice	Higher pitch	Deep-pitched voice	Mean vocal pitch	−0.011 (0.03)
Upright head and posture		Upright head and posture	Mean head pitch	−0.021 (0.02)
			Mean head yaw	−0.010 (0.02)
			Mean head roll	−0.020 (0.02)
More facial expression	More rigidity in facial animation	More facial expression (i.e., less rigidity in facial animation)	SD AU01	−0.026 (0.02)
			SD AU02	−0.010 (0.02)
			SD AU04	0.003 (0.02)
			SD AU05	−0.010 (0.02)
			SD AU06	−0.003 (0.02)
			SD AU07	−0.007 (0.02)
			SD AU09	0.004(0.02)
			SD AU10	0.041 (0.02)$^{+}$
			SD AU12	0.003 (0.03)
			SD AU14	0.008 (0.02)
			SD AU15	0.004 (0.03)
			SD AU17	−0.022 (0.02)
			SD AU20	0.011 (0.02)
			SD AU23	−0.014(0.02)
			SD AU25	0.028 (0.02)
			SD AU26	0.020 (0.02)
			SD AU45	−0.039 (0.02)
More variation in pitch	Less vocal variation	More variation in pitch	SD vocal pitch	−0.002 (0.03)
More head movement	More rigidity in head movement	More head movement (i.e., less rigidity in head movement)	SD head pitch	−0.016 (0.03)
			SD head yaw	0.013 (0.02)
			SD head roll	0.000 (0.02)
More rapid speaking tempo		More rapid speaking tempo	Word count/speaking time	−0.090 (0.04)*
More turns-at-talk	Fewer turns-at-talk	More turns-at-talk	Count of turn-at-talk	0.051 (0.03)*
Longer turns at talk	Small amount of talk time	Longer turns at talk	Mean Turn-at-talk duration	0.054 (0.025)*

$^{+}p < 0.1$, $^{*}p < 0.05$.

Only beta weights and standard errors for each nonverbal behavioral are presented. See ***Supplementary Tables 7–9*** *for full model results.*

facial expressions and head movement did not correlate with perceptions of trust.

Exploratory Analysis

Computational extraction of behavioral features provides insight into nonverbal behaviors that manual coding cannot. The current video corpus was processed with automated voice and face behavioral analysis software which produced 75 features. After calculating the mean and standard deviation of these features, our dataset resulted in 150 nonverbal behavior features. Although conducting separate statistical tests on this many features increases the probability of Type 1 errors, we do so with the intention of exploring macro-level patterns, not hypothesis testing. **Table 8**: Proportion of significant nonverbal features, Means and **Table 9**: Proportion of significant nonverbal features, Standard Deviations below provide insights into the relationships among dominance, nervousness, trust, and each category of features that had *p*-values below 0.05. The values in the table provide a count of the number of statistically significant features by dependent variable (dominance, nervousness, and trust), channel (face, head, or voice) and summary statistic (mean or standard deviation). The tables also count the number of positive or negative coefficients.

Dominance was associated with the greatest number of significant features with 101. Nervousness had far fewer with 27 significant features, and trust, with only 15 significant features. Among the 143 significant relationships, 104 were standard deviations and 39 were means, which indicates that the variation, rather than the average level of nonverbal signals, influenced the perceptions of relational dimensions more.

Interestingly, when looking at the count of positive or negative coefficients associated with measured standard deviation (**Table 9**: Proportion of significant nonverbal features, Standard Deviations), we see far more significant features with positive coefficients for dominance and almost all with negative coefficients for nervousness. In the case of standard deviations, positive coefficients correspond to more dynamic behaviors and negative coefficients correspond to muted or rigid behaviors. Almost all (68/70) of the significant standard deviation measures were positive for dominance, and all (21/21) were

TABLE 8 | Proportion of significant nonverbal features, means.

	Summary statistic = Mean		
	# Features $p < 0.05$/# Features	# Positive Coef./# Features $p < 0.05$	# Negative Coef./# Features $p < 0.05$
Dominance	**31/75**	**17/31**	**14/31**
Face	12/17	12/12	0/12
Head	1/6	0/1	1/1
Voice	18/52	5/18	13/18
Nervousness	**6/75**	**4/6**	**2/6**
Face	3/17	1/3	2/3
Head	0/6	0	0
Voice	3/52	3/3	0/3
Trust	**2/75**	**1/2**	**1/2**
Face	0/17	0	0
Head	0/6	0	0
Voice	2/52	1/2	1/2

The positive (negative) coefficients indicate that a higher mean of the nonverbal feature is associated with a higher (lower) level of perception of dominance, nervousness, or trustworthiness.

TABLE 9 | Proportion of significant nonverbal features, standard deviations.

	Summary statistic = Standard deviation		
	# Features $p < 0.05$/# Features	# Positive Coef./# Features $p < 0.05$	# Negative Coef./# Features $p < 0.05$
Dominance	**70/75**	**68/70**	**2/70**
Face	14/17	14/14	0/14
Head	6/6	6/6	0/6
Voice	50/52	48/50	2/50
Nervousness	**21/75**	**0/21**	**21/21**
Face	1/17	0/1	1/1
Head	2/6	0/2	2/2
Voice	18/52	0/18	18/18
Trust	**13/75**	**13/13**	**0/13**
Face	0/17	0	0
Head	0/6	0	0
Voice	13/52	13/13	0/13

The positive coefficients indicate that a higher value for the standard deviation of the nonverbal feature (i.e., a higher degree of variability) is associated with a higher level of perceived dominance, nervousness, or trustworthiness. The negative coefficients indicate that the variable has a negative sign in the final equation such that less variability is associated with a higher degree of perceived dominance, nervousness or trustworthiness.

negative for nervousness. Again, although we expect that some Type 1 errors are likely, this finding supports our general hypothesis that dominance is associated with more energetic behaviors (more variability) and nervousness is associated with tension (less variability). Furthermore, a majority (13/15) of significant features for trust were standard deviations of voice features, and they all had a positive coefficient, indicating that trust tends to be associated with more variability in voice.

Classification Results

Our analysis revealed the significant effects of the nonverbal signals on relational dimensions. However, statistical significance does not necessarily imply practical significance. In this section, we aim to predict the relational scores with behavioral measures. Given the reported significance of many nonverbal signals, we assume that such variables will help to predict the relational dimensions.

To formulate the prediction of relation dimensions as a classification problem and to mitigate individual level rating bias, we binarized the aggregated score of the three relational dimensions. The median of each player's response was used as the cutoff to dichotomize the original scores. In this way, labels of "High/Low dominance," "High/Low nervousness," and "High/Low trustworthiness" were assigned to each player, and the generated categories were roughly balanced.

The same nonverbal variables in the previous analysis were used as predictors. Six popular machine learning algorithms, Logistic Regression, Random Forest, Naïve Bayes, Support Vector Machine (SVM), and two ensemble learning methods, Bagging and Boosting, were used to predict the binary categories of dominance, nervousness and trustworthiness. Ensemble methods, which combine multiple learning algorithms to achieve better prediction, have gained wide popularity due to their superior performance. The two ensemble methods that we used combine multiple decision trees to make the final decision. An 80/20 split was applied to construct the training set from which the model was then applied to the test set. To obtain a more reliable estimate of the model's prediction ability, we adopted a "repetitive random split" strategy, that is, we randomly split the full data set 100 times. For each train-test split, the accuracy and F1 score of this model was recorded. The F1 score averages accuracy in predicting dominance and non-dominance, nervousness and composure, or trust and distrust. The mean of accuracy and F1 score over 100 splits was used to evaluate the model's prediction performance. A modified stepwise variable selection (MSVS) method was used to search through the gigantic model space (Draper and Smith, 1981).

Table 10: Prediction accuracy and F1score of the machine learning models (RF, Random Forest; LR, Logistic Regression; NB, Naïve Bayes; BAG, Bagging; XGB, Boosting) summarizes the best accuracy and F1 score that our models achieved in the three prediction tasks. In our results, the highest accuracies on predicting dominance, nervousness and trustworthiness were all achieved by the bagging models, which shows its superiority as an ensemble method. The naïve Bayes models were outperformed by the bagging models only by a narrow margin, and their F1 scores were even higher when predicting dominance and nervousness. One consistent observation across different machine learning algorithms is that higher accuracies and F1 scores were attained when predicting dominance than when predicting nervousness and trustworthiness, which implies that perceived dominance is the most predictable relation dimension with nonverbal signals. On the other hand, the manifestations of nervousness and trustworthiness in nonverbal behaviors were

TABLE 10 | Prediction accuracy and F1 score of the machine learning models (RF, random forest; LR, logistic regression; NB, Naïve Bayes; BAG, bagging; XGB, boosting).

Relational dimension	RFF1	RFACC	LRF1	LRACC	SVMF1	SVMACC	NBF1	NBACC	BAGF1	BAGACC	XGBF1	XGBACC
Dominance	0.61	0.70	0.63	0.72	0.63	0.72	0.69	0.72	0.69	0.72	0.68	0.70
Nervousness	0.44	0.62	0.46	0.66	0.45	0.66	0.58	0.66	0.56	**0.66**	0.58	0.65
Trustworthiness	0.43	0.63	0.43	0.66	0.43	0.66	0.59	0.65	0.59	**0.66**	0.58	0.65

The accuracies of the XGB (bagging) model (in bold) were highest among the six machine learning classifiers.

more subtle and dynamic and cannot be accurately reflected in our aggregated predictors.

Since variable importance can be output by some decision tree-based models, we further examined which variables mattered most to the prediction of relational dimensions by calculating the means of variable importance reported by the best random forest models. The results are presented in **Figure 2**. The best random forest model of predicting dominance contains 7 nonverbal signals, and 6 of them are vocalic variables. The leftmost bar in **Figure 2A** represents the most prominent predictor, which is the summation of turns-at-talk of a player. The variable importance analyses of trustworthiness prediction (**Figure 2B**) and nervousness prediction (**Figure 2C**) exhibit a similar pattern: the majority of the most important variables are vocalic features. Due to the models' limited predictive power, the variable importance scores were all very low. However, the importance of vocalic signals in the impressions of relational dimensions can be inferred from these figures.

Table 11: Most important variables in the best Random Forest model reports the most important variables when predicting the perceived dominance, trustworthiness and nervousness, respectively. The rank was based on each variable's importance given by the random forest models.

A few model-free observations can be drawn. First, the majority of the important predictive variables comes from the vocalic signals. The only two non-vocalic variables are the mean of AU 15 (lip corner depressor) and the standard variation of AU 20 (lip stretcher), both of which, interestingly, are lip movements. Secondly, when predicting trustworthiness and nervousness, most of the important variables are standard deviations of an original measurement, which is consistent with the findings of the linear models. A few variables (bold in **Table 11**: Most important variables in the best Random Forest model) appeared twice in **Table 11**, for example, MFCC Channel 1 Standard Deviation, which represents the standard deviation of the first spectral envelop of MFCC, and Loudness Derivative Standard Deviation, which reflects the standard deviation of vocal loudness. The co-occurrence of such variables not only demonstrates the robustness of our analysis, but also calls for efforts to interpret these nonverbal signals.

DISCUSSION

Relational communication is a fundamental aspect of interpersonal communication and nonverbal signals are a centerpiece of understanding that endeavor. How people regard one another and their relationship can be expressed in ways that speak what words cannot. Nonverbal relational messages express meanings and sentiments that people may refrain from saying out loud, such as declaring a romantic interest or expressing schadenfreude, or can ease the burden of delivering hurtful messages such as the end of a relationship, a death in the family or a terminal cancer diagnosis.

Relational communication can take myriad forms, as enumerated by Burgoon and Hale (1984, 1987; Hall et al., 2005). Only three are the focus here, three that coincide with our investigation of cross-cultural group deception, but it should be understood that the diverse themes of relational messages all entail nonverbal signals to varying degrees and so are subject to the same opportunities and obstacles that we discuss here.

A major impetus for featuring relational communication in this special issue is that the development of new automated tools for measuring nonverbal signals and new machine learning methods for analyzing them has made it possible to delve into the heretofore elusive and ephemeral topic of relational messaging. We often know when a significant message has been exchanged between two people, we just cannot always put our finger on what its basis is. The current investigation begins to put "flesh" on the skeleton of relational communication, to discover the possibilities of examining nonverbal communication in a more microscopic way than in the past and to discover what obstacles we encounter along the way. At the same time, this synergistic effort brings together computer science, psychology and communication methods in social signal processing. This direction is at the forefront of work on affective computing and neurocognitive psychology. The outgrowth of this program of research is that technical fields will benefit from theories originating from psychology and communication, and the social sciences will benefit from the technological advances of computer science fields. For example, predictive models using input features suggested by social sciences may facilitate understanding relational messages in real time and designing interactive systems that recognize and interpret human affects. Meanwhile, using machine-learning based automated tools for measurement is more scalable and less costly than manual coding, and these tools are helpful for testing and refining social science theories. Additional side benefits of the current project are that it situated nonverbal interaction in a group setting rather than the usual dyadic one and explored such interaction in natural, ongoing discussion rather than scripted or brief interchanges. This has brought with it the messiness that accompanies naturalistic

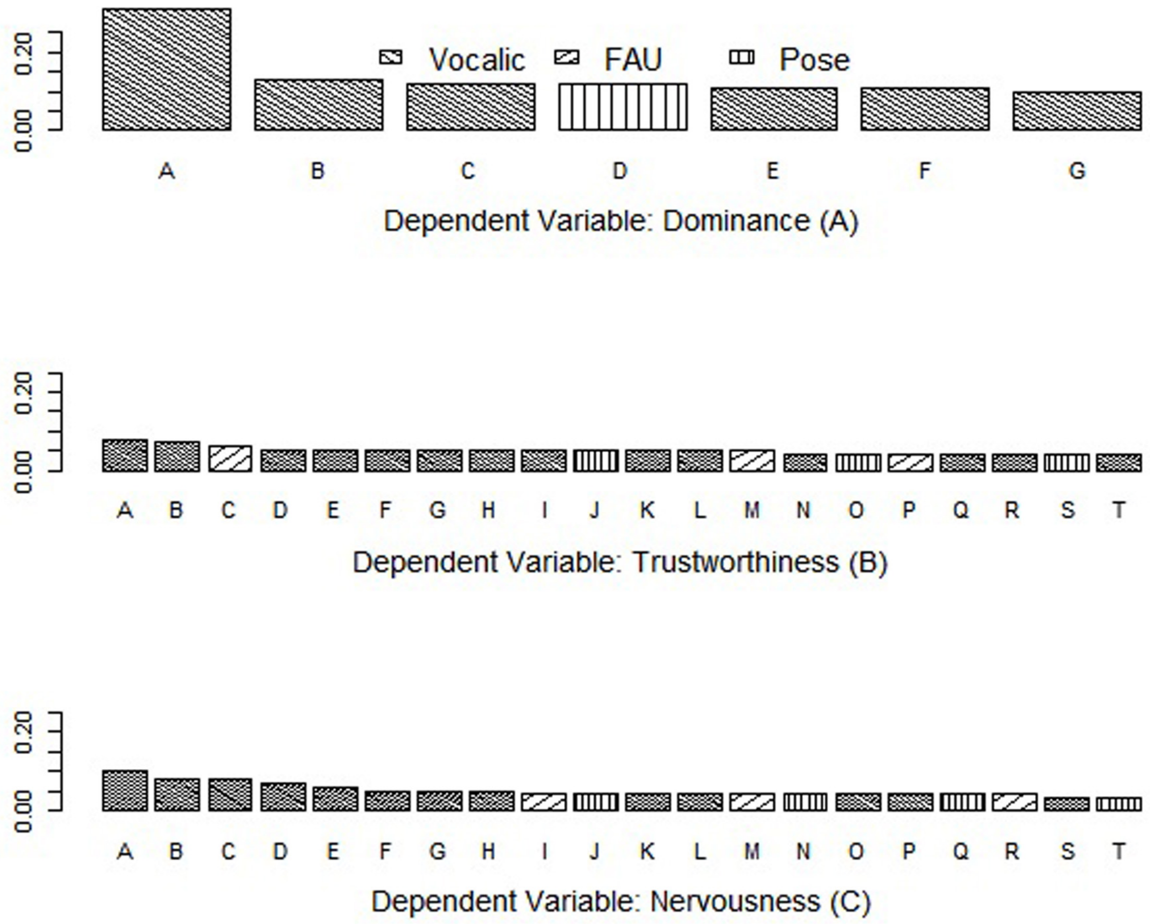


FIGURE 2 | Variable importance reported by the best random forest model on predicting Dominance **(A)**, Trustworthiness **(B)**, and Nervousness **(C)**.

human interaction and is the basis for many of the obstacles, and mitigating strategies, we will discuss.

Nonverbal Signals of Dominance and Non-dominance

Of the three relational message themes examined in this paper, dominance was associated with the widest variety and greatest number of cues. It was associated with 101 of the 150 variables tested, whereas the other two relational themes, composure and trust, were associated with far fewer. The results indicate that perceptions of dominance were associated with a louder voice, more expressive facial behavior, more head movement (in terms of pitch, yaw, and roll), and more and longer turns-at-talk. Additionally, several of the FAUs were associated with dominance as well; of the 17 we measured, 15 of them were associated with perceptions of dominance. Dunbar (2004) argued that while power is a perception based on a relationship and the resources to which one has access, dominance is based in the particular contextual behaviors that one enacts during a relationship and interaction. That was certainly true in this case where the objectively measured behaviors that have been commonly associated with dominance in the research literature (see e.g., Dunbar and Burgoon, 2005) were related to the perceptions that one was behaving dominantly.

Nonverbal Signals of Nervousness and Composure

Nonverbal signals of nervousness were associated with softer vocal amplitude, less head movement, and fewer and shorter turns-at-talk. The hypothesis that nervousness is reflected in rigid facial animation, higher pitch, and less vocal variation was not confirmed in the present study. However, overall results largely support that nervousness leads to tense and rigid nonverbal behaviors. The exploratory analysis showed that all significant face, head, and voice features were associated with reduced variation. Most notably, all 18 voice measures with a p-value below 0.05 had a negative coefficient, which indicates that perceptions of nervousness are amplified as vocal animation decreases. In the current study, the vocal channel seems to have been the predominant signal of nervousness. This is likely due to kinesic controllability and how emotional indicators in the voice are difficult to conceal. Additionally, markers of nervousness in the face and head are likely better assessed at critical moments

TABLE 11 | Most important variables in the best random forest model.

Dependent variable	Importance rank	Variable name	Variable category	Variable description
Dominance	**A**	Turn-at-talk Duration Summation	Vocalic	The total duration of turns-at-talk for a player
	B	**MFCC Channel 1 Standard Deviation**	Vocalic	The standard deviation of the first spectral envelope of MFCC
	C	LSP Channel 5 Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of fifth line spectral pair frequency
	D	AU15 Intensity Standard Deviation	FAU	The standard derivation of the intensity of the fifteenth facial action unit
	E	Fundamental Frequency Mean	Vocalic	The mean of fundamental frequency
	F	MFCC Channel 1 Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the first spectral envelope of MFCC
	G	LSP Channel 1 Mean	Vocalic	The mean of the first line spectral pair frequency
Trustworthiness	**A**	**Loudness Derivative Standard Deviation**	Vocalic	The standard deviation of the derivative of the normalized loudness
	B	ZCR Standard Deviation	Vocalic	The standard deviation of the zero-crossing rate of time signal
	C	AU 20 Intensity Mean	FAU	The mean of the intensity of the 20th facial action unit
	D	MFCC Channel 12 Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the 12th spectral envelope of MFCC
	E	**MFCC Channel 11 Derivative Standard Deviation**	Vocalic	The standard deviation of the derivative of the 11th spectral envelope of MFCC
	F	Fundamental Frequency Derivative Mean	Vocalic	The mean of the derivative of fundamental frequency
	G	Fundamental Frequency Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of fundamental frequency
Nervousness	**A**	MFCC Channel 1 Standard Deviation	Vocalic	The standard deviation of the first spectral envelope of MFCC
	B	MFCC Channel 1 Derivative Mean	Vocalic	The mean of the derivative of the first spectral envelope of MFCC
	C	Loudness Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the normalized loudness
	D	LSP Channel 1 Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the first line spectral pair frequency
	E	MFCC Channel 11 Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the 11th spectral envelope of MFCC
	F	Voice Probability Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the voicing probability
	G	MFCC Channel 2 Derivative Standard Deviation	Vocalic	The standard deviation of the derivative of the 2nd spectral envelope of MFCC

"sma" indicates that the variable is smoothed by a moving average filter with window length 3. (2) A bolded variable name indicates that this variable appeared twice in Table.

during interactions such as immediately before or after turns-at-talk. Analysis over a wide timespan, such as the case in this study, may obscure subtle indicators of nervousness.

Surprisingly, vocal pitch was not a significant indicator of nervousness in our study. A possible explanation is that the manifestation of nervousness in voice may be confounded by one's phonatory attributes, which correlate with gender, age, native language and even cultural background. Though these factors have been controlled in our model, the chance of nervousness's effect being weakened by human's highly varied vocal timbre still exists. It might be more meaningful to conduct within-subject vocalic analysis (e.g., compare utterances from the same individuals when they are nervous and those when they are not) to reveal the effect of nervousness. Another explanation is that nervousness would be more evident in dyadic interaction where an individual is the sole target of scrutiny and suspicion. In a group, it is easy to deflect attention to self by passively avoiding turns at talk or focusing attention on others, thus reducing one's cognitive load and anxiety.

Nonverbal Signals of Trust and Suspicion

Of the three relational message themes studied here, trust and its converse of distrust or suspicion, had the fewest nonverbal signals that predicted it. Only two single mean vocal features were associated with trust and 13 vocalic standard deviations were, all pointing to more variation in the voice contributing to the perception of trustworthiness. No facial or head features predicted trust, and none of the nonverbal features predicted distrust. This coincides with finding few features in the machine learning models as important for classifying high or low trust. Of the machine learning models, the bag-of-words methods achieved the highest, but paltry, 66% accuracy that was virtually the same as all the other methods. The F1 score was highest at 59%, indicating that averaging the accuracy scores for trust and distrust reduced overall accuracy. Except for those features of trust that overlapped with the significant predictors of dominance or composure, then, trustworthiness was clearly not easily predictable from nonverbal signals.

Why might that be? Is it the case that nonverbal features are not intrinsic ingredients in gauging who might be viewed as trusted? We do not think so. There are many possible explanations for the minimal appearance of nonverbal signals in the alchemy of trust. First is the fact that because participants had little basis for making judgments, group members tended to rate everyone similarly at the start. This would have created a restriction in range statistically. Secondly, because group members did not know one another, aggregating across time periods and moment-to-moment changes in behavior may have blurred any important signals into one average and meaningless soup. Were we to design a new study, we would seek to develop more nuanced measures of trust reflective of specific actions of what might have engendered trust or piqued suspicion, much like studies of close relationships seek to identify turning points or significant events in ongoing communication.

Third, and contrariwise, our moment-to-moment measurements were related to specific blocks of rounds, yet trust may grow out of the accretion of actions across time, something our measurement did not capture well. For example, by the end of the game, a villager could recall which time another player had been on a winning or failed mission and so make reasonable guesses about who were villagers and who, spies. This historical information had nothing to do with that player's nonverbal actions at that point in time. This possibility points to the importance of selecting time slices that best reflect the granularity of the question of interest and deciding whether measurements should be geared to microscopic moments or a broader sweep of time. They are also a reminder that our understandings are best achieved by combining verbal and nonverbal information as well as other contextual information in our models.

Measurement error and the relatively simple variable construction may also account for the biases in our modeling results. The measurement error may come from multiple sources. If the subject's face was not captured by the camera (for example, due to large amplitude of body movement) or multiple faces appeared in the same frame, the OpenFace software would output invalid measurements. Tablets can be slightly moved by a pressing finger, which results in drastic fluctuations in the head pose measurements. The performance of the SOTA audio diarization algorithm was also far from being perfect when being applied to our data set. As a result, the audio files generated in our pipeline might not reflect the subject-level utterances exactly. Considering such measurement errors, the effect size and significance level of some nonverbal signals may not reflect the reality. On the other hand, the bi-round-level aggregated mean and standard deviation may not represent the subtlety of nonverbal signals well either. Intuitively, the nonverbal messages matter most to the receivers' perception when the sender is having the group's attention. As a result, nonverbal analysis focusing on certain "critical moments" may be more meaningful, and this can be accomplished by narrowing analysis to specific segments of interaction.

Although automated tools could be employed to measure nonverbal behaviors, manual methods were still necessary for obtaining a few features accurately given the current technologies. For example, we required an accurate and precise count of words to calculate speaking tempo. Machine-generated transcripts were of low quality because of crosstalk, background noise, and accented speech, so we resorted to tedious manual correction. Furthermore, hesitations and interruptions could be identified easily by human coders, but those marked by automated transcription services did not fully correspond to what human coders would perceive as disfluencies. Because of the large amount of data, we did not manually code hesitations and interruptions for our analysis. However, these are important nonverbal features that could affect perceptions of dominance, nervousness, and trustworthiness. Future research could profitably explore reliable automated tools for measuring these nonverbal features.

SUMMARY

In summary, nonverbal signals color the meanings of interpersonal relationships. Humans rely on facial, head, postural and vocal signals to express relational messages of dominance or non-dominance. They rely on vocal signals to convey nervousness or composure. And to a lesser extent, some of these signals contribute to meanings of trust and distrust. Emerging automated analysis tools and machine learning methods have made possible a deeper understanding of the dynamics of relational communication and have exposed much of the messiness of studying communication under naturalistic conditions. Improvements in measurement and experimental design may mitigate some of these complications in the future.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are not readily available because privacy restrictions exist and data is not available until the end of grant. Requests to access the datasets should be directed to JB, judee@arizona.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board (IRB) at UC Santa Barbara. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

FUNDING

Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-16-1-0342. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the United States Government. The United States Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.624177/full#supplementary-material

REFERENCES

Anderson, C., and Kilduff, G. J. (2009). The pursuit of status in social groups. *Curr. Direct. Psychol. Sci.* 18, 295–298. doi: 10.1111/j.1467-8721.2009.01655.x

Andrew, R. J. (1972). "The information potentially available in mammal displays," in *Nonverbal Communication*, ed. R. A. Hinde (Cambridge: Cambridge University Press), 179–204.

Bales, R. F. (1968). "Interaction process analysis," in *International Encyclopedia of the Social Sciences*, ed. D. C. Sills (New York, NY: MacMillan and Free Press), 465–471.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. P. (2018). "Openface 2.0: facial behavior analysis toolkit," in *Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, (Lake Placid, NY: IEEE), 59–66. doi: 10.1109/FG.2018.00019

Bateson, G. (1951). "Information and codification," in *Communication: The social Matrix of Psychiatry*, eds J. Ruesch and G. Bateson (New York, NY: W.W. Norton), 168–211.

Bateson, G. (1958). *Naven*, 2nd Edn. Palo Alto, CA: Stanford University Press.

Bente, G., Leuschner, H., Al Issa, A., and Blascovich, J. J. (2010). The others: universals and cultural specificities in the perception of status and dominance from nonverbal behavior. *Conscious. Cogn.* 19, 762–777. doi: 10.1016/j.concog.2010.06.006

Boone, R. T., and Buck, R. (2003). Emotional expressivity and trustworthiness: the role of nonverbal behavior in the evolution of cooperation. *J. Nonverb. Behav.* 27, 163–182. doi: 10.1023/A:1025341931128

Bruneau, T. J. (1973). Communicative silences: forms and functions. *J. Commun.* 23, 17–46. doi: 10.1111/j.1460-2466.1973.tb00929.x

Burgoon, J. K., and Dunbar, N. E. (2006). "Dominance, power and influence," in *The SAGE Handbook of Nonverbal Communication*, eds V. Manusov and M. Patterson (Thousand Oaks, CA: Sage), 279–298.

Burgoon, J. K., Dunbar, N. E., and Jensen, M. (in press). "An integrated spiral model of trust," in *Detecting Trust and Deception in Group Interaction*, eds V. S. Subramanian, J. K. Burgoon, and N. E. Dunbar (Berlin: Springer).

Burgoon, J. K., and Hale, J. L. (1983). A research note on the dimensions of communication reticence. *Commun. Q.* 31, 238–248. doi: 10.1080/01463378309369510

Burgoon, J. K., and Hale, J. L. (1984). The fundamental topoi of relational communication. *Commun. Monogr.* 51, 193–214. doi: 10.1080/03637758409390195

Burgoon, J. K., and Hale, J. L. (1987). Validation and measurement of the fundamental themes of relational communication. *Commun. Monogr.* 54, 19–41. doi: 10.1080/03637758709390214

Burgoon, J. K., Johnson, M. L., and Koch, P. T. (1998). The nature and measurement of interpersonal dominance. *Commun. Monogr.* 65, 309–335. doi: 10.1080/03637759809376456

Burgoon, J. K., and Koper, R. J. (1984). Nonverbal and relational communication associated with reticence. *Hum. Commun. Res.* 10, 601–626. doi: 10.1111/j.1468-2958.1984.tb00034.x

Burgoon, J. K., and Le Poire, B. A. (1999). Nonverbal cues and interpersonal judgments: participant and observer perceptions of intimacy, dominance, composure, and formality. *Commun. Monogr.* 66, 105–124. doi: 10.1080/03637759909376467

Burgoon, J. K., Segrin, C., and Dunbar, N. E. (2002). "Nonverbal communication and social influence," in *The Persuasion Handbook: Developments in Theory and Practice*, eds J. P. Dillard and M. Pfau (New York, NY: Sage), 445–473.

Chen, X., Ita Levitan, S., Levine, M., Mandic, M., and Hirschberg, J. (2020). Acoustic-prosodic and lexical cues to deception and trust: deciphering how people detect lies. *Trans. Assoc. Comput. Linguist.* 8, 199–214. doi: 10.1162/tacl_a_00311

Cheng, J. T., Tracy, J. L., Ho, S., and Henrich, J. (2016). Listen, follow me: dynamic vocal signals of dominance predict emergent social rank in humans. *J. Exp. Psychol. Gen.* 145, 536–547. doi: 10.1037/xge0000166

Daly, J. A., McCroskey, J. C., and Richmond, V. P. (1977). Relationships between vocal activity and perception of communicators in small group interaction. *West. J. Commun.* 41, 175–187. doi: 10.1080/10570317709389610

DeSteno, D. (2014). *The Truth About Trust: How it Determines Success in Life, Love, Learning, and More*. London: Penguin Books.

Dorn, B., et al. (in press). "A system for multi-person, multi-modal data collection in behavioral information systems," in *Detecting Trust and Deception in Group Interaction*, eds V. S. Subramanian, J. K. Burgoon, and N. E. Dunbar (Berlin: Springer).

Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis*. Hoboken, NJ: John Wiley and Sons.

Dunbar, N. E. (2004). Dyadic power theory: constructing a communication-based theory of relational power. *J. Fam. Commun.* 4, 235–248. doi: 10.1080/15267431.2004.9670133

Dunbar, N. E., and Burgoon, J. K. (2005). "Nonverbal measurement of dominance," in *The Sourcebook of Nonverbal Measures: Going Beyond Words*, ed. V. Manusov (Mahwah, NJ: Erlbaum), 361–374.

Ekman, P., and Friesen, W. V. (1978). *Manual of the Facial Action Coding System (FACS)*. California, CA: Consulting Psychologists Press. doi: 10.1037/t27734-000

Eyben, F., and Schuller, B. (2015). OpenSMILE:) The Munich open-source large-scale multimedia feature extractor. *ACM SIGMultimedia Rec.* 6, 4–13. doi: 10.1145/2729095.2729097

Eyben, F., Weninger, F., Gross, F., and Schuller, B. (2013). "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*, Barcelona, 835–838. doi: 10.1145/2502081.2502224

Feeley, T. H., and deTurck, M. A. (1995). Global cue usage in behavioral lie detection. *Commun. Q.* 43, 420–430. doi: 10.1080/01463379509369989

Finch, L. P. (2005). Nurses' communication with patients: examining relational communication dimensions and relationship satisfaction. *Int. J. Hum. Car.* 9, 14–23. doi: 10.20467/1091-5710.9.4.14

Frank, M. G., Menasco, M. A., and O'Sullivan, M. (2008). "Human behavior and deception detection," in *Handbook of Science and Technology for Homeland Security*, ed. J. G. Voeller (Hoboken, NJ: Wiley), 1–13. doi: 10.1016/j.chb.2014.02.027

Frank, M. G., and Svetieva, E. (2012). Lies worth catching involve both emotion and cognition. *J. Appl. Res. Mem. Cogn.* 1, 131–133. doi: 10.1016/j.jarmac.2012.04.006

Fuller, C. M., Biros, D. P., and Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. *Expert Syst. Appl.* 38, 8392–8398. doi: 10.1016/j.eswa.2011.01.032

Giles, M., Hansia, H., Metzger, M., and Dunbar, N. (in press). "The impact of culture in deception and deception detection," in *Detecting Trust and Deception in Group Interaction*, eds V. S. Subramanian, J. K. Burgoon, and N. E. Dunbar (Berlin: Springer).

Gottman, J. M. (2011). *The Science of Trust: Emotional Attunement for Couples*. New York, NY: WW Norton & Company.

Gregersen, T. S. (2005). Nonverbal cues: clues to the detection of foreign language anxiety. *For. Lang. Ann.* 38, 388–400. doi: 10.1111/j.1944-9720.2005.tb02225.x

Hagenaars, M. A., and van Minnen, A. (2005). The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *J. Anxiety Disord.* 19, 521–537. doi: 10.1016/j.janxdis.2004.04.008

Hall, J. A., Coats, E. J., Smith, and LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* 131, 898–924. doi: 10.1037/0033-2909.131.6.898

Hughes, S. M., Mogilski, J. K., and Harrison, M. A. (2014). The perception and parameters of intentional voice manipulation. *J. Nonverb. Behav.* 38, 107–127. doi: 10.1007/s10919-013-0163-z

Keating, C. F., Mazur, A., Segall, M. H., Cysneiros, P. G., Kilbride, J. E., Leahy, P., et al. (1981). Culture and the perception of social dominance from facial expression. *J. Pers. Soc. Psychol.* 40, 615–626. doi: 10.1037/0022-3514.40.4.615

Knapp, M. L., Ellis, D. G., and Williams, B. A. (1980). Perceptions of communication behavior associated with relationship terms. *Commun. Monogr.* 47, 262–278. doi: 10.1080/03637758009376036

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö, Faria, V., et al. (2008). In a nervous voice: acoustic analysis and perception of anxiety in social phobics' speech. *J. Nonverb. Behav.* 32, 195–214. doi: 10.1007/s10919-008-0055-9

Leary, T. (1957). *Interpersonal Diagnoses of Personality*. New Yokr, NY: Ronald press.

Lucas, G., Stratou, G., Lieblich, S., and Gratch, J. (2016). "Trust me: multimodal signals of trustworthiness," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, (New York, NY: ACM), 5–12. doi: 10.1145/2993148.2993178

Mann, S., Deeb, H., Vrij, A., Hope, L., and Pontigia, L. (2020). Detecting smugglers: identifying strategies and behaviours in individuals in possession of illicit objects. *Appl. Cogn. Psychol.* 34, 372–386. doi: 10.1002/acp.3622

Massey-Abernathy, A. R., and Haseltine, E. (2019). Power talk: communication styles, vocalization rates and dominance. *J. Psycholinguist. Res.* 48, 107–116. doi: 10.1007/s10936-018-9592-5

McCroskey, J. C. (1966). Scales for the measurement of ethos. *Speech Monogr.* 33, 65–72. doi: 10.1080/03637756609375482

Mehrabian, A. (1971). *Silent Messages*. Belmont, CA: Wadsworth Pub. Co.

Mehrabian, A., and Ksionzky, S. (1972). Categories of social behavior. *Comp. Group Stud.* 3, 425–436. doi: 10.1177/104649647200300404

Mikkelson, A. C., Hesse, C., and Sloan, D. (2017). Relational communication messages and employee outcomes in supervisor/employee relationships. *Commun. Rep.* 30, 142–156. doi: 10.1080/08934215.2017.1300677

Millar, F. E., and Rogers, L. E. (1976). "A relational approach to interpersonal communication," in *Explorations in Interpersonal Communication*, ed. G. R. Miller (Thousand Oaks, CA: Sage), 87–104.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Champaign, IL: University of Illinois Press.

Parks, M. R. (1977). Relational communication: theory and research. *Hum. Commun. Res.* 3, 372–381. doi: 10.1111/j.1468-2958.1977.tb00541.x

Pentland, S. J., Twyman, N. W., Burgoon, J. K., Nunamaker, J. F. Jr., and Diller, C. B. (2017). A video-based screening system for automated risk assessment using nuanced facial features. *J. Manag. Inform. Syst.* 34, 970–993. doi: 10.1080/07421222.2017.1393304

Rogers, L. E., and Farace, R. V. (1975). Analysis of relational communication in dyads: new measurement procedures. *Hum. Commun. Res.* 1, 222–239. doi: 10.1111/j.1468-2958.1975.tb00270.x

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. (1998). Not so different after all: a cross-discipline view of trust. *Acad. Manag. Rev.* 23, 393–404. doi: 10.5465/amr.1998.926617

Schutz, W. C. (1966). *The Interpersonal Underworld*. Science and Behavior Books.

Simpson, J. A. (2007). Psychological foundations of trust. *Curr. Direct. Psychol. Sci.* 16, 264–268. doi: 10.1111/j.1467-8721.2007.00517.x

Toma, C. L., and Hancock, J. T. (2012). What lies beneath: the linguistic traces of deception in online dating profiles. *J. Commun.* 62, 78–97. doi: 10.1111/j.1460-2466.2011.01619.x

Tsankova, E., Aubrey, A. J., Krumhuber, E., Möllering, G., Kappas, A., Marshall, D., et al. (2012). "Facial and vocal cues in perceptions of trustworthiness," in *Asian Conference on Computer Vision*, eds J. I. Park and J. Kim (Berlin: Springer), 308–319. doi: 10.1007/978-3-642-37484-5_26

Tusing, K. J., and Dillard, J. P. (2000). The sounds of dominance. Vocal precursors of perceived dominance during interpersonal influence. *Hum. Commun. Res.* 26, 148–171. doi: 10.1111/j.1468-2958.2000.tb00754.x

Twyman, N. W., Elkins, A. C., Burgoon, J. K., and Nunamaker, J. F. (2014). A rigidity detection system for automated credibility assessment. *J. Manag. Inform. Syst.* 31, 173–202. doi: 10.2753/mis0742-1222310108

van der Werff, L., and Buckley, F. (2017). Getting to know you: a longitudinal examination of trust cues and trust development during socialization. *J. Manag.* 43, 742–770. doi: 10.1177/0149206314543475

Vrij, A., and Fisher, R. P. (2020). Unraveling the misconception about deception and nervous behavior. *Front. Psychol.* 11:1377. doi: 10.3389/fpsyg.2020.01377

Watzlawick, P., Beavin, J., and Jackson, D. (1967). *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes*. New York, NY: W.W. Norton.

Waxer, P. H. (1977). Nonverbal cues for anxiety: an examination of emotional leakage. *J. Abnorm. Psychol.* 86, 306–314. doi: 10.1037/0021-843x.86.3.306

Weeks, J. W., Heimberg, R. G., and Heuer, R. (2011). Exploring the role of behavioral submissiveness in social anxiety. *J. Soc. Clin. Psychol.* 30, 217–249. doi: 10.1521/jscp.2011.30.3.217

Witkower, Z., Tracy, J. L., Cheng, J. T., and Henrich, J. (2020). Two signals of social rank: prestige and dominance are associated with distinct nonverbal displays. *J. Pers. Soc. Psychol.* 118, 89–120. doi: 10.1037/pspi0000181

Wood, J. A. (2006). NLP revisited: nonverbal communications and signals of trustworthiness. *J. Pers. Sell. Sales Manag.* 26, 197–204. doi: 10.2753/pss0885-3134260206

Conflict of Interest: JB is affiliated with Discern Science International, a for-profit entity that develops systems for credibility assessment.

The remaining 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.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 17 February 2021
doi: 10.3389/fpsyg.2021.624108

Who's Leading This Dance?: Theorizing Automatic and Strategic Synchrony in Human-Exoskeleton Interactions

Gavin Lawrence Kirkwood, Christopher D. Otmar and Mohemmad Hansia*

Department of Communication, University of California, Santa Barbara, Santa Barbara, CA, United States

Wearable robots are an emerging form of technology that allow organizations to combine the strength, precision, and performance of machines with the flexibility, intelligence, and problem-solving abilities of human wearers. Active exoskeletons are a type of wearable robot that gives wearers the ability to effortlessly lift up to 200 lbs., as well as perform other types of physically demanding tasks that would be too strenuous for most humans. Synchronization between exoskeleton suits and wearers is one of the most challenging requirements to operate these technologies effectively. In this conceptual paper, we extend interpersonal adaption theory (IAT) to the exoskeleton context and explicate (a) the antecedents that are most likely to shape synchrony in human-exoskeleton interactions, (b) automatic and strategic synchrony as adaptive behaviors in human-exoskeleton interactions, and (c) outcome variables that are especially important in these processes. Lastly, we offer a discussion of key methodological challenges for measuring synchrony in human-exoskeleton interactions and offer a future research agenda for this important area.

Keywords: synchrony, human-machine interaction, exoskeletons, emerging technologies, methodological challenges, human-robot interaction

OPEN ACCESS

Edited by:
Marianne Schmid Mast,
University of Lausanne, Switzerland

Reviewed by:
Alexander Refsum Jensenius,
University of Oslo, Norway
Lucia Maria Sacheli,
University of Milano-Bicocca, Italy

****Correspondence:***
Gavin Lawrence Kirkwood
gavinkirkwood@ucsb.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *30 October 2020*
Accepted: *18 January 2021*
Published: *17 February 2021*

Citation:
Kirkwood GL, Otmar CD and Hansia M (2021) Who's Leading This Dance?: Theorizing Automatic and Strategic Synchrony in Human-Exoskeleton Interactions. Front. Psychol. 12:624108. doi: 10.3389/fpsyg.2021.624108

INTRODUCTION

Synchrony has intrigued non-verbal communication researchers for several decades (Bernieri et al., 1988; Kendon, 1990; Bernieri and Rosenthal, 1991; Burgoon et al., 1995). According to interpersonal adaption theory (IAT), non-verbal synchrony is a type of reciprocal adaption that involves rhythmic patterns during an interaction where dyads coordinate their behaviors interdependently through matching, motor mimicry, and mirroring (Burgoon et al., 1995, 2014). Burgoon et al. (1995) explained that behavioral matching and motor mimicry are "in response to a stimulus and is often directed toward another person, mirroring is the imitation of another's body movements" (p. 26). In other words, synchrony involves two parties engaging in an interaction similarly as a result of the coordination in their behavioral patterns (Burgoon et al., 1995; Fujiwara et al., 2020). The main types of non-verbal synchrony include simultaneous behaviors between interactants, interaction rhythms that occur over the course of an interaction, and behavioral meshing that creates a meaningful whole [Bernieri et al., 1988; Bernieri and Rosenthal, 1991; see also the complementary scholarship on joint action in Knoblich et al. (2011)]. Past research on synchrony in human dyads has shown that non-verbal synchronous behaviors are used to signal interest, involvement, rapport, similarity, or approval (Kendon, 1970; Warner, 1992; Tickle-Degnen and Gavett, 2003), resulting in highly synchronous exchanges being mutually rewarding experiences for the interactants.

Empirical research investigating synchrony as a predictor of rapport, similarity, and approval in interpersonal interactions has also been leveraged in the field of human-robot interaction (HRI) (Kendon, 1970; Warner, 1992; Tickle-Degnen and Gavett, 2003; Hasnain et al., 2013; Bartneck et al., 2020b). Recent developments in robotics include making human-robot interactions reflect synchronous interactions between human dyads (Prepin and Revel, 2007; Hasnain et al., 2013). For example, the adaptable robotics for interaction analysis (ADRIANA) platform enables a robot to detect movements in human users and synchronize its movements automatically in real-time. The field of HRI has largely adopted the assumption that when robots automatically synchronize their movement to users, users will feel that interactions with these technologies are more natural and similar to human interactions (Hasnain et al., 2013).

Wearable robots are an emergent technology that has the potential to reshape relationships between humans and robots through the process of synchrony (de Looze et al., 2016). Although some organizations might prefer to develop autonomous robots that replace humans, exoskeletons offer the opportunity to combine the strength, precision, and performance of machines with the intelligence, agility, and creativity of a human workforce. Exoskeletons are defined as, "a wearable, external mechanical structure that enhances the power of a person" (de Looze et al., 2016, p. 671). According to Zaroug et al. (2019) exoskeletons are an emerging form of wearable robots in which synchrony challenges are especially crucial and salient for functionality.

There are many different types of designs for exoskeletons (e.g., tailored for lower limbs, full body suits) that can be passive or active. Passive exoskeletons do not have a power source, instead these devices rely on counterweights to collect energy from the wearer's own movements. This design is primarily used to support healthy postures or prevent injuries in repetitive work tasks (de Looze et al., 2016). In contrast to passive exoskeletons, active exoskeletons have a power source that can be used to dramatically augment human abilities or performance in physical tasks (Zaroug et al., 2019). Active exoskeletons were originally developed for military use including Raytheon's XOS 2 powered armored suit which provided protection, enhanced lifting power, and improved moving capabilities for wearers (Kopp, 2011). However, recent trends in active exoskeleton development are geared toward developing suits for industries that place heavy physical demands on workers and have high risk of injury (e.g., automobile manufacturing plants, distribution warehouses). A full-body active exoskeleton currently being developed for industry is the Guardian XO suit by Sarcos Corp (2019). The Guardian XO allows humans to easily lift up to 200 lbs. and offers features that allow the wearer to perform highly precise tasks with heavy tools or industry-specific equipment.

Synchrony can be difficult to achieve in human-exoskeleton interactions because it requires that an active exoskeleton can accurately detect when a wearer initiates movement and understand what type of movement the wearer wants the suit to perform. The wearer also will experience challenges in synchronizing their movement with the suit including how long to wait for the exoskeleton to respond to the wearer's movements, knowing how to move when an exoskeleton is performing a task (such as lifting a heavy item), and knowing when it is appropriate to initiate new movements. For instance, if a wearer is trying to lift a 200-lb piece of equipment they will need to initiate movement with their arms and wait for the exoskeleton to respond and pick up the item. Although there is likely to be variation in lag times across different active exoskeleton devices, even short lag times still require a wearer to be mindful of how their body movements may disrupt suit functionality. In this paper we leverage active exoskeletons as an illustrative example for other forms of wearable technology because of the high-stakes nature of synchrony in this context. These stakes include safety concerns from giving exoskeleton wearers such dramatically high levels of physical strength as well as how UX may reshape traditional blue-collar industries in which these suits are adopted.

Aligning with the theme of this special issue, it is clear that non-verbal communication scholars can fill significant knowledge gaps in human-exoskeleton interactions; however, non-verbal theories will need to be extended in order to do this important work. This paper frames the adaptation patterns between exoskeleton and human as non-verbal synchrony—opposed to the broader term of coordination—because we are specifically interested in how a person's rhythms are set into motion by active exoskeletons (Condon and Ogston, 1966; Kendon, 1990; Burgoon et al., 1995). We leverage an IAT framework to theorize non-verbal human-exoskeleton synchrony because it grounds these interactions in a communicative lens and offers rich heuristic value for explicating (a) the antecedents that are most likely to shape synchrony in human-exoskeleton interactions, (b) automatic and strategic synchrony as adaptive processes in human-exoskeleton interactions, and (c) outcome variables that are especially important in these processes. Understanding these antecedents, processes, and outcomes are important for exploring both how to make experiences with active exoskeletons more satisfying to users as well as how to increase user efficacy in the workplace. Lastly, we offer a discussion of key methodological challenges for measuring synchrony in human-exoskeleton interactions and offer directions for future research.

INTERPERSONAL ADAPTION THEORY FRAMEWORK

IAT builds upon previous adaption and coordination theories—such as expectancy violations theory (EVT; Burgoon and Hale, 1988) and communication accommodation theory (CAT; Giles, 1973)—while also leveraging biological principles, cognitive arousal, and social norms to explain adaptive dyadic behavior during interactions (Burgoon et al., 1995, 2017). Within IAT, synchrony is conceptualized as an adaptation behavior in which two people coordinate their behaviors interdependently through mirroring, matching, or reciprocity (Burgoon et al., 2014). Early research in synchrony found that synchronous behavior can be used to signal interest, involvement, rapport, similarity, or approval (Kendon, 1970). According to Burgoon et al. (1995) synchrony typically involves automatic biological responses, but can also be used strategically as it can, "function to regulate interaction and facilitate speech processing as well as express relational and emotional states" (p. 25). This suggests that IAT may be useful to further understand the regulated interactions between humans and exoskeletons. IAT provides a useful lens to understand the coordination in human-exoskeleton interactions through non-verbal synchrony—especially given the important role that strategic synchrony plays in this adaptive dyadic behavior (Burgoon et al., 1995).

IAT and Emerging Technologies

Coordination is a fundamental part of satisfying interactions. The degree of coordination often predicts a variety of positive social and biological outcomes for the beneficiary. As an umbrella term that represents a broad range of concepts, coordination can be conceptualized as either communicative (e.g., mutual influencing interactions) or non-communicative (e.g., crew rowing) based upon the behaviors of the interactants. We use the term *adaption* to refer to the coordination of behaviors that are non-random and patterned in timing and form (Bernieri and Rosenthal, 1991) using a communicative lens (Burgoon et al., 1995). Although interpersonal adaption is often studied in face-to-face (Burgoon et al., 1995) and computer-mediated encounters (Dunbar et al., 2014), there is a growing body of scholarship that urges social scientists to further engage with HRI (Van Erp and Toet, 2013; Bartneck et al., 2020a). Given that all interpersonal interactions have a degree of coordination (Gatewood and Rosenwein, 1981; Bernieri and Rosenthal, 1991; Chartrand and Bargh, 1999), it is essential to construct a framework to better explain why, how, and to what effect interacting with emerging technologies have for humans.

Emerging technologies are increasingly designed to adapt to user preferences, understand and predict human behavior, and create optimal conditions for human-machine collaboration through a variety of approaches including machine learning via neural networks (Burrell, 2016), therefore making a theoretical framework for adaptive behaviors in HRI urgent. Unlike technologies where synchrony is automatic—such as robots designed with the ADRIANA platform—humans have more agency to synchronize with technologies they wear and operate. We recognize that exoskeletons are not the only emerging technology where synchrony is relevant, but it is clear that exoskeletons are a context in which synchrony is especially important for user safety and suit functionality. The stakes are higher for synchrony in human-exoskeleton interactions compared to other emerging technologies due to the closer proximal distance between users and suits. Additionally, unlike other technologies, how exoskeletons are designed will impact whether synchrony occurs automatically or if wearers are able to choose how and why they synchronize with the suit. This makes exoskeletons a helpful context for distinctions between strategic and automatic synchrony processes. Therefore, exoskeletons provide a prime exemplar for scholars to expand and rethink what it means to synchronize during an interaction and why particular antecedents of the user will predict a variety of outcomes to increase UX, comfortability, and efficacy with the technology.

One central assumption of IAT is that actors perform reciprocal or compensatory behaviors in response to the behaviors of the other partner in the interaction. Individuals are typically compelled toward reciprocal adaptions—such as synchrony—due to biological pressures and social expectations (e.g., politeness norms). IAT offers a detailed review of the conditions in which interactants are likely to partake in these reciprocal or compensatory behaviors based upon three antecedents that the person enters an interaction with: requirements (R), expectations (E), and desires (D).

Requirements, Expectations, Desires

Requirements (R) are the individual perceptions of what needs to happen at any point during the interaction. These are often based on biological factors such as proximity to the speaker in order to hear what they are saying. Interactants also have particular *expectations* (E) of the interactions. Expectations can be thought of as socially based anticipations of the encounter and the communicator (Floyd and Burgoon, 1999). For example, many individuals expect their partner will have a particular level of interpersonal skills and abide by the social norms according to the culture, relationship, or profession. However, expectations can also be based upon past experiences with the person or previous knowledge they may have about the interactant. Finally, individuals enter interactions with particular *desires* (D) about what the interaction should accomplish by its conclusion. For example, an employee may enter an interaction with their supervisor to resolve issues with how work tasks are distributed.

Interaction Position

Requirements (R), expectations (E), and desires (D) coalesce in an overall evaluation called the *interaction position* (IP). This position is a "valenced behavioral predisposition" for an individual's own interaction behavior or what is anticipated from an interactional partner (Burgoon et al., 1995, p. 271). In other words, the IP is the perception an individual has about the interaction—and this perception is comprised of R, E, and D. Although IAT states that R of an interaction take precedence over D and E, requirements are often satisfied during an interaction, bringing D and E to play a more prominent role in determining the IP (Floyd and Burgoon, 1999). For instance, if an exoskeleton suit-wearer's R is met (e.g., the suit fits their body), the suit aligns with their E (e.g., the suit helps them accomplish work tasks), and satisfies their D (e.g., the suit is comfortable to wear), then we can expect the suit-wearer to have a positively valanced IP.

The IP is then compared to the *actual* (A) communication performed by the person during the interaction, which will determine whether the following interaction is patterned by reciprocal or compensatory adaptions. According to Burgoon et al. (2017), if A is more desirable than the IP, a partner is more likely to reciprocate the behavior, however, if the actual behavior is less desirable than the IP, the partner is predicted to compensate. The compensatory and reciprocal

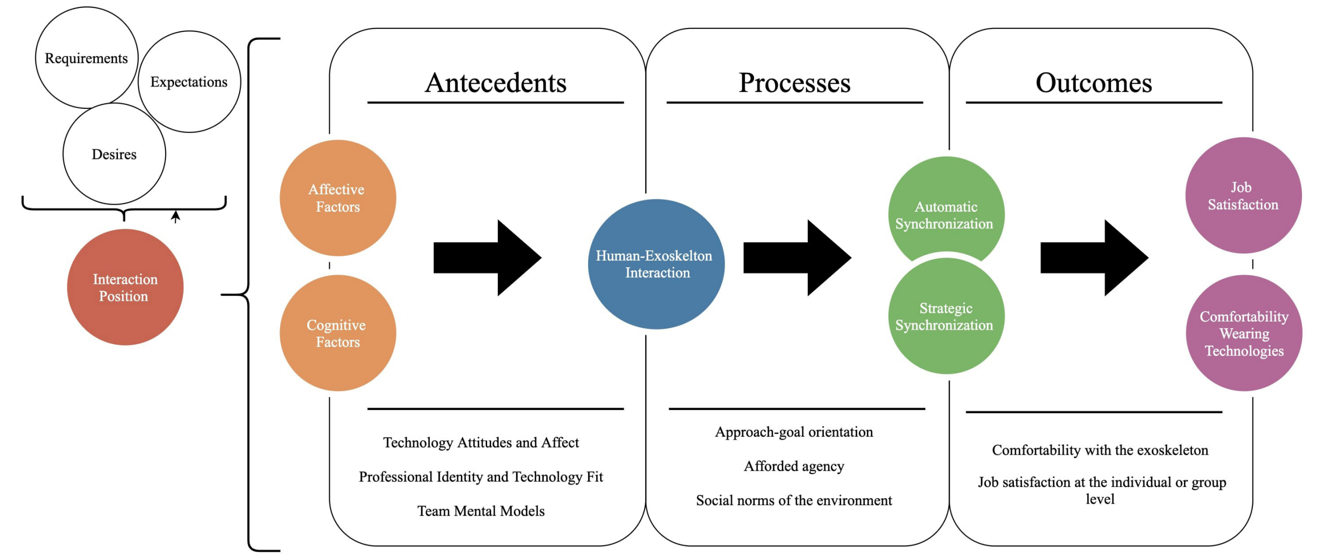


FIGURE 1 | Chart detailing antecedents, processes, and outcomes that are likely to impact automatic and strategic synchronization in human-exoskeleton interaction.

predictions are primarily based upon the magnitude of the discrepancy between the IP and the actual communication of the partner (A). Although small deviances are often tolerated during interactions, large deviances can lead to further assessment of the discrepancies' valence. IAT argues that large discrepancies should move toward whichever adaption pattern is more positively valanced for the interaction. Therefore, when the A > IP partners should display reciprocal behaviors (e.g., synchrony) and when A < IP, receivers should respond with compensatory behaviors (e.g., dissynchrony).

Non-verbal Synchrony as Adaptation

Although these behaviors are typically unconscious, there are cases when participants of an interaction will consciously try to synchronize their behaviors with their partner. According to Burgoon et al. (1995) synchrony involves automatic biological responses, but can also be used strategically as a conscious adaptive behavior. According to IAT, a positive valence (A > IP) of the interaction between the user and the exoskeleton will lead to a reciprocal adaption—including synchrony—and may play a critical role in increasing the wearer's trust, rapport, and comfortability with this technology. However, the relationship between an exoskeleton and the wearer is complex and synchronization with an exoskeleton places unique physiological and psychological demands upon on the human throughout the interaction (Knight and Baber, 2005). Seemingly antithetical to human synchrony, human-exoskeleton synchrony suffers from the lack of a *mutually* rewarding experience and is one-sided. However, without optimal synchrony between the wearer and the exoskeleton, the user may grow tired and frustrated of using the machine and ultimately resort to not using the technology—defeating the central purpose of the exoskeleton to aid with physically demanding tasks.

The encounter between the user and exoskeleton is especially sensitive because of the close proximity between the wearer and the suit. Due to the exoskeleton often alleviating the pressures of physical labor (Upasani et al., 2019), the exoskeleton is close to the body and has the potential to violate an individual's space expectations. The skin is an especially important channel for social communication and "robot-initiated contact implies that the robot will enter the person's intimate space" (Chen et al., 2014, p. 141). Indeed, haptic contact with a machine in the workplace may have physical and psychological consequences (Upasani et al., 2019), such as claustrophobic feelings which could result in a rise in cortisol, skin irritation from adjusting the machine, or feeling stressed that the exoskeleton cannot be removed. Therefore, a priority researchers and practitioners alike should be to determine the antecedents, processes, and outcomes of the reciprocal adaption between exoskeleton and wearer in order to avoid the negative consequences of wearing an exoskeleton (see **Figure 1**). Important antecedents in these contexts include affective factors (e.g., feelings about or fears of new technologies) and cognitive factors (e.g., perceptions about wearable robotics). Processes include both automatic and strategic non-verbal synchronization behaviors between the human wearers and exoskeletons. Lastly, outcomes for synchrony in the exoskeleton context include comfortability with exoskeletons and overall job satisfaction.

ANTECEDENTS IN HUMAN-MACHINE SYNCHRONY

Unlike traditional forms of equipment used for heavy lifting in industrial or manufacturing environments (e.g., forklifts, dollies, or various carts), exoskeletons are a wearable technology (de Looze et al., 2016). Although traditional equipment (e.g.,

TABLE 1 | Applying Knight and Baber's (2005) comfort dimensions to the exoskeleton context.

Comfort dimension	Wearer concerns	Exoskeleton application
Emotion	"How do I feel when wearing the suit? "How do I feel when other people see me wearing the suit?"	The enhanced lifting capabilities may make wearers feel empowered. Wearers may feel insecure when wearing the suit in front of coworkers, especially if colleagues view the suit as unnecessary or as a crutch.
Attachment	"How often do I feel the suit moving on its own?" "When I initiate a movement, is there a lag time before the machine responds?"	Active exoskeletons use sensors to interpret the wearer's movements and act accordingly. Lag times between the wearer initiating a movement and the machine responding can feel constraining because the wearer would not want any additional movement to disrupt the process.
Harm	"Do I experience any pain or discomfort when wearing the suit?" "Do I have any disabilities or injuries that would cause discomfort in the suit?" "Does the suit accommodate my body type?"	Exoskeleton suits will range in their ability to accommodate all body types. Even if the exoskeleton is designed as "one-size fits all," wearers who have sustained injuries or who have disabilities may be especially vulnerable to pain or discomfort.
Perceived Change	"Do I notice the suit moving while I'm wearing it?" "Is suit movement distracting or disorienting in any way?"	A feeling of disorientation may increase cognitive workload for wearers. If wearers are required to give increased focus while wearing the suit, it may compromise their ability to focus on work tasks. High levels of distraction or disorientation may mean that other coworkers or intelligent assistants will be required to help the wearer work effectively.
Movement	"Do I have free range of motion in the suit?" "Does the suit allow me to move in all the ways I need to work effectively?" "Does the suit restrict my ability to communicate with my coworkers?"	Exoskeleton suits are likely to vary in the range of motion they offer to wearers. To work effectively, wearers need to be able to move in ways that allow them to complete their work tasks. If range of motion is compromised, then a wearer's non-verbal communication might be compromised. Special consideration should be given in contexts where an exoskeleton reduces the social cues (e.g., gestures) required to communicate effectively.
Anxiety	"Does wearing the suit trigger any unique fears or anxieties?" "Am I afraid of or uncomfortable around new technologies?"	Wearing an exoskeleton may trigger unique anxieties or fears from individuals, such as claustrophobia. Other wearers may have anxieties and fears of the technology itself, which can be exacerbated when wearing the suit.

forklifts) often require an operator-tool dynamic with a clear physical distinction between operator and equipment, exoskeleton wearers have no physical distance between their bodies and the suit. The experience of wearing a technology involves multiple bodily sensations and can even make wearers feel as if the technology is an appendage of their own body (Smelik et al., 2016). For example, in a study exploring the affective impacts of wearable solar panels, Smelik et al. (2016) found that some participants reported the added spatial dimensions of the suit increased feelings of personal empowerment. According to Knight and Baber (2005), wearable technologies involve a myriad of cognitive and affective factors to consider for the safety and comfortability of wearer. Knight and Baber (2005) created a typology to operationalize the cognitive and affective factors for people who operate wearable computers including emotion, attachment, harm, perceived change, movement, and anxiety. Together, these six dimensions are likely to influence a wearer's ability to automatically or strategically synchronize with an exoskeleton (see **Table 1** for specific exoskeleton applications for each dimension) but this is not an exhaustive list of relevant antecedents in this context. In addition to updating and extending Knight and Baber's (2005) typology to an exoskeleton context we also explicate other important affective and cognitive antecedents that are important in human-exoskeleton interactions.

Affective Antecedents

Synchrony during human-to-human interactions typically indicates positive affect, but we must first discuss the relationship between humans and exoskeletons. Emerging technologies regularly incite feelings of uncertainty and fear in humans (although this may stem from a lack of knowledge or motivation) and wearable robots such as exoskeletons are likely to be no exception. Fear of technology and robots are important to examine to gain a better understanding for how incorporating exoskeletons into a workplace can influence the humans involved. For the better part of a century, robots have been framed in fiction as intrusive and dangerous (Szollosy, 2015). According to Szollosy (2015), negative depictions of robots are presented more often to reflect human anxieties or uncertainties rather than any true technological developments. Because of these underlying anxieties that many individuals carry, wearable robot designers must consider how the technology will be received by the public— including what dominant expectations of the technology are held. In particular, robots often arouse strong emotions from people including fears of deskilling in the workplace, the loss of employment, or even larger existential threats to humanity (Szollosy, 2015).

Previous studies have investigated attitudes humans have toward robots. For instance, Nomura et al. (2008) conducted experiments where a human and a robot participated in basic interactions (i.e., meeting, self-disclosure, and physical contact). Negative attitudes and avoidance behavior exhibited from participants were measured. Negative attitudes influenced behavior toward robots as participants with negative attitudes spent significantly less time talking and touching the robot. In addition, gender differences were found as men who had negative attitudes about robots had higher instances of avoidance behaviors. On the other hand, the results also suggested that repeated interaction with robots can decrease avoidance behaviors over time. According to Nomura et al. (2008), these findings demonstrate that attitudes, perceptions, and other factors such as gender are important to consider in HRI research.

In addition to attitudes toward robotics that are influenced by larger cultural discourses or differences in gender, there are also affective dimensions that are unique to wearable technologies (Knight and Baber, 2005). Each of the dimensions that Knight and Baber (2005) offered have an affective impact on a wearer's experience with an active exoskeleton. The emotional dimension addresses the ways that wearable technology can make people feel when wearing the device and how they feel when others observe them wearing the suit. The attachment dimension addresses whether wearers feel the suit moving; an example would be if the suit or device moves autonomously or if users have control over suit movements. The harm dimension references any discomfort or pain that could arise as a result of the wearable device. The harm dimension includes high levels of individual variation as wearers who have sustained workplace injuries or disabilities are likely to experience this dimension differently than other wearers. The perceived change dimension encompasses any way that the wearable technology makes individuals feel different than how they would normally feel without the device (e.g., feelings of distraction, disorientation). The movement dimension addresses the ways in which a wearable technology can restrict the wearer's ability to move. Finally, the anxiety dimension addresses the remaining affective factors (e.g., claustrophobia) that could cause feelings of insecurity for the wearer. Taken together, these five dimensions, as well as larger cultural attitudes and perceptions of robotic technology, are likely to influence a wearer's emotional state when wearing an active exoskeleton.

Cognitive Antecedents

Identity

One fundamental cognitive factor that is likely to shape synchrony between humans and exoskeletons in workplace settings is the wearer's sense of professional identity. According to Thoits and Virshup (1997) people develop meanings and expectations associated with work tasks based on how they understand their professional identities. These identities are shaped by social groups (in this case the groups at work) and the understanding of the self in relation to others (Tajfel and Turner, 1979). When it comes to interacting with technology, the technology's fit with a worker's identity tends to be just as important as a technology's fit to a task (Lin, 2012). According to Lin (2012) workers not only use technology that is necessary to complete a task but also use technologies that are consistent with how they view their work identities. It is important to recognize that social groups influence the task-technology and identity-technology fit: If a technology is perceived to not fit a certain identity, then using it to accomplish a task that goes against the expression of that identity would be deemed inappropriate (Fulk, 1993). For example, Upasani et al. (2019) found that agricultural workers were less likely to use exoskeleton technologies that, "do not seem to be work-related, and that are more 'medical' in their appearance" (p. 5). In other words, use of the technology violated the workers' understanding of their identities because they did not view a medical device as relevant to their work or role in the organization.

A wearer's sense of professional identity is likely to impact whether individuals synchronize with exoskeletons. When considering group membership and teamwork, identity often plays a large role in facilitating cooperation. CAT can be used to understand the relationship between intergroup dynamics and synchrony (Giles and Ogay, 2007). As a foundational theory of intergroup communication, CAT explains how people communicate and modify their communication toward different individuals. These modifications are made to converge or mimic the style of the other interactant, as well as the behaviors that meet the other interactants' perceived needs. Accommodation often occurs more when both participants in the interaction share a similar or compatible group identity.

According to Bernhold and Giles (2020), mimicry occurs with a goal for association as it overlaps with convergence and is synonymous with accommodation behaviors. Bernhold and Giles defined mimicry as the "unconscious imitation or mirroring of various nonverbal behaviors" (p. 62). If mimicry has the same results and goals as convergence, based on accommodation research, it can be suggested that identity would also influence synchrony in interactions. Particularly when discussing the success of a group, research on identity has illustrated that synchrony among individuals promotes prosociality. Prosociality can be defined as cooperation within individual dyads or between larger groups (Batson, 1998). Reddish et al. (2014) found that prosociality improved through synchrony among groups regardless of differences in group identity. The link between synchrony and accommodation has also been examined in a variety of contexts including parent communication with infants (baby talk) and communication between romantic partners (Locke, 1993; Lee et al., 2010). Other areas where these similarities have been researched include professional communication, and persuasion (Buller and Aune, 1988, 1992; Sparks, 1994). Accommodation and mimicry have been considered distinct but related concepts according to nonverbal research (Bernhold and Giles, 2020). Although research suggests that synchrony is likely more vital to group dynamics than perceived identity, it is still important to understand the identity affiliation or goal of the parties involved.

Team Mental Models

In addition to individual attitudes and perceptions toward a technology, group attitudes and perspectives will likely influence synchrony processes with exoskeletons in the workplace. According to Klimoski and Mohammed (1994) group members are related through shared cognition through team mental models (TMM). TMM is the idea that when working together, groups have conceptualizations and mental models that are either shared or compatible between group members. Research indicates that positive TMMs are positively associated with team coordination processes and overall performance (Mathieu et al., 2000; Fisher et al., 2012). As research suggests, if organizations can pinpoint general and contextual variables that can be linked to TMMs, and provide training and interventions to optimize TMMs, then one can anticipate highly coordinated and successful teams (Mathieu et al., 2000). During team compilation, team members interpret and obtain knowledge regarding their individual role, contextual social dynamics, what the task consists of, and what each team member brings to the group. This leads to an understanding of how they fit together and how they can accomplish tasks at hand. The higher degree of comprehension for these concepts, the more likely the TMM is positive. Positive TMM has been found to occur when team roles are understood early in team formation (Pearsall et al., 2010). If these same principles are addressed when an exoskeleton is designed and used, the results should be positive. The central distinction in the human-exoskeleton context would be that one member of the team is an active exoskeleton and would require the other members to be open to collaborating with it.

When the technology that is implemented is not introduced early and is seen as a threat to a team member's importance to the team, the results and perceptions may be negative. We suggest that mitigating potential tension between exoskeleton technology and team members includes early exposure, technical briefings, and plenty of hands-on experience. As with any technology, allowing time to first acclimate with the exoskeleton prior to implementation would give team members a better understanding of what the tech can and cannot accomplish. This then would give them an understanding on how to implement the tech into their TMM without feeling threatened that the suit will undermine the importance of human abilities or expertise in the group.

PROCESSES

Identifying the underlying processes of dyadic behavior has been a central aim for many nonverbal scholars that study coordination and adaption (e.g., Cappella, 1991; Arundale, 1996; Andersen, 1998). This makes sense considering that processual features are often thought of as the central component of human interaction [see Hewes (1979) for comments on process in social interaction research]. Identifying these features requires scrutiny of the simultaneous signaling as well as signal detection while the human and robot coordinate with one another. Patterson (2019) suggests that a key element to understanding the underlying process of HRI may be rooted in the goal compatibility between the human and robot. Complementary goals between humans and robots may increase the effectiveness and comfortability of the technology with the user. For the exoskeleton wearer, synchrony with the suit should be encouraged as the primary goal when wearing the device.

The seamless coordination of humans and technology could increase affect and trust in the machine, and therefore increase the likelihood that the wearer will be able to utilize and benefit from using an exoskeleton. Another factor that is unique to the exoskeleton-wearer interaction is the balance between the agency of the human and the abilities of the machine. If synchrony is to be the main goal during the coordination of the HRI, both the ability of the wearer to effectively control the robot and the capability of the exoskeleton to appropriately respond in kind to the wearer is key to the underlying process. It is not only important to understand the wearers' predispositions prior to using the exoskeleton and the consequences of the interaction, but critical to discuss how the synchronization unfolds throughout the course of the interaction. Burgoon et al. (1995) and Kellermann (1992) argue that adaption during interactions is mostly automatic but there is still some level of intent in every encounter. Therefore, the synchronization process can be both *automatic* and *strategic* representing both unconscious and conscious behaviors of exoskeleton wearers. Exoskeleton designers for private industry are still debating (a) what it means for humans to synchronize with exoskeletons, (b) how much control users should feel over an exoskeleton, and (c) how optimal synchrony can be achieved in these interactions. It remains to be seen which aspects of active exoskeletons are going to be designed to automatically synchronize with wearer movements or if these technologies will have some control over wearers in these interactions.

Automatic Synchrony

Automaticity during an interaction are the features that are involuntary and unconscious by the agents (Kellermann, 1992). Therefore, automatic synchrony occurs without reflection from the wearer of the exoskeleton and they allocate very few cognitive resources to the behavior being enacted. These underlying biological mechanisms are not based upon social or cultural variations because they are more fundamental and rudimentary to the human communication process (Cappella, 1991). From the robotic side of the interaction, these are the automatic processes that the machine engages in to achieve its pre-determined goal. The importance of automatic synchrony in the HRI is noteworthy because it provides a sense of normalcy and, possibly, positive affect (such as feelings of confidence) throughout the exoskeleton interaction. Indeed, speech convergence tends to be associated with positive affect and the disruption of this convergence is often seen as jarring and abnormal (Feldstein et al., 1982). The more automatic the synchrony, the more engaged the user is expected to become while interacting with the exoskeleton. Additionally, the less time that the user needs in order to synchronize with the exoskeleton, the more cognitive resources freed up in order to focus on work tasks. In other words, the more automatic the synchrony between the wearer and the exoskeleton, the more the user will be able to concentrate on achieving the particular goal.

Cappella (1991) describes two different types of automated patterns during interactions: (1) stimulation regulation and (2) emotional responsiveness. Although originally proposed between two humans, the biological origins of these automatic patterns may still bring insight in what the exoskeleton wearer requires (R) expects (E), and desires (D) of the robot, therefore influencing the interactional position (IP) of the wearer. Stimulation regulation is the dyadic process in which a person controls the others' expressed level of activation. An example provided by Cappella (1991) is the tempo of the conversation—often measured by the rate of speech and the quickness of response. Bartneck et al. (2020a) argue contingency anthropomorphization of social robots can help users feel like the machine is partaking in appropriate regulation. For example, if the robot detects motion, "it should briefly look toward the origin of the movement" (p. 54). Similarly, the robot could be designed to improve stimulation regulation by using information about the user's previous motion patterns in order to tailor the exoskeleton experience. The stimulation regulation of the exoskeleton—or the responsiveness to the wearer throughout the time wearing the device—may be highly predictive of the user experience since it is likely that the user will desire (D) the robot to reflect the stimulation regulation expected (E) in dyadic human interaction.

The second automatic pattern is emotional responsiveness throughout the course of the interaction. Satisfying communication is often directly tied to partners successfully communicating felt emotions during an interaction (Andersen and Guerrero, 1998). Cappella (1991) describes this emotional responsiveness as the tendency to approach and withdraw from the emotional state of another. Although the wearer of the exoskeleton is the only part of the dyad that has biological origins, it is possible that the wearer of the exoskeleton is still expecting emotional responsiveness from the robot throughout the interaction. For example, if a user's body begins to stiffen because they are in pain while using the exoskeleton, it is critical that the exoskeleton is able to respond in kind to this new development, opposed to passively operating as if the emotional state of the wearer has not changed. Emotional responsiveness may also come from skin-conductance sensors or increase in heartbeat that indicate stress from the wearer (Bartneck et al., 2020a). Both stimulation regulation and emotional responsiveness as automatic processes can significantly influence the IP of the user and can increase the degree of synchrony depending on the actual behavior (A) of the device.

Strategic Synchrony

Non-verbal synchrony can also be strategic and directed by the actor's goals. For example, a worker may be deliberate in their attempts to coordinate their movements with the exoskeleton in order to quickly finish a task and, in turn, increase productivity. Under these conditions, the wearer of the exoskeleton is intentional in their ability to adapt the machine and synchronize their movements in order to achieve a particular goal. Burgoon et al. (1995) explained that strategic synchrony regulates interactions and helps express relational and emotional states. If the wearer of the exoskeleton has positive affect toward the machine, the wearer may strategically attempt to synchronize movements with the robots in order to effectively accomplish the task that the exoskeleton and the wearer are jointly working on. The opposite may also be true. Increases in negative emotional states throughout an interaction have been associated with dissynchrony between adults and infants (Bernieri et al., 1988). It is possible that negative emotional states that unfold throughout an interaction between the exoskeleton and the wearer may cascade into an increasingly dyssynchronous encounter. However, there are underlying processes that take place as the interaction unfolds that motivate the wearer to strategically synchronize with the exoskeleton. We argue that three central motivations to strategically synchronize during the interaction with the exoskeleton are the (a) the levels of agency afforded to the exoskeleton wearer, (b) the goals of the exoskeleton wearer, and (c) social norms of the environment.

Agency

When we argue that agency is afforded to an exoskeleton user, we apply Gibson's (1986) theoretical concept of affordances to the exoskeleton context. Simply put, technological affordances refer to the intersection of what people believe they can accomplish with a technology and the technological features that either enable or constrain those goals. Treem and Leonardi (2013) explained that the technological affordance perspective is useful when exploring technology use because it "helps to explain why people using the same technology may engage in similar or disparate communication and work practices" (p. 146). In the exoskeleton context, the technical features in equipment design as well as the setting in which these technologies are adopted will likely impact the amount of agency afforded to human wearers.

According to Banks and de Graaf (2020), levels of agency afforded to humans and machines fundamentally impact human-machine collaboration. Agency typically refers to the ability of social actors that stem from resources, responsibilities, and capacity to reflect on situational context (Giddens, 1979). Within an exoskeleton context, it is important to recognize that strategic synchrony is the only form of synchrony that involves agency for human wearers. This means that in strategic synchrony contexts, workers have the ability to decide what ways they want to synchronize with the exoskeleton and how to enact those behaviors. In contrast, automatic synchrony means that regardless of what a wearer desires, synchrony will be achieved in the interaction. Although automatic synchrony gives lower levels of agency to the exoskeleton over the wearer, it could offer the seamless connection that wearers desire in the workplace or it may result in workers feeling disempowerment or a lack of control in their profession.

Goals

According to Lin et al. (2018), goals "shape people's behavior and direct their efforts toward different outcomes" (p. 314). Decades of research on interpersonal goals has shown that particular goal orientations often predict positive and negative affect toward interactions and relationships (Gable and Berkman, 2008). These orientations can also influence the interactants' non-verbal communication behaviors during an encounter, as well as an individual's understanding of the interaction once it has ended (Caughlin, 2010). Gable (2006) argued that two main goals drive most interpersonal interactions: approach goals and avoidance goals. Approach-goals tend to include positively valanced intentions and individuals seek to gain rewards from the interactions (e.g., affection). Avoidance-goals are characterized by evading threats during an interaction and are typically motivated by apprehension of conflict or failure. Individuals who are wearing exoskeletons that are more prone to avoidance-goals when interacting with robots could have a more negative experience. Further, they may be less likely to strategically synchronize their movements with the exoskeleton. However, wearers of exoskeletons who are more approach-goal oriented may be more likely to engage with the exoskeleton and strategically synchronize their movements to pursue the rewards of accomplishing the task. Apprehension to technology is likely a main predictor of an individuals' goal orientation throughout an interaction.

Norms

Social norms drive or constrain behavior and tend to be universally understood by particular members of a group (Horne, 2001). Emerging from interactions with other group members, social norm behaviors foster member expectations of themselves and others (Cialdini and Trost, 1998; Korte, 2010). Cialdini et al. (1990) even argued that these expectations, or "standards," typically develop from observing others (i.e., descriptive norms). For example, treating robots as humanlike could be understood as a social norm since workers are more likely to perceive the machines they are working with as human if they see others doing the same (Bartneck et al., 2020a). The degree of motivation to strategically synchronize with an exoskeleton might depend upon the social norms of the environment in which the individual is situated. Therefore, if the social norm in the workplace is to strategically synchronize with exoskeletons, it is possible that the worker will be more inclined to follow suit.

Strategic Synchrony as Cooperation

Both the social norms of the environment and goals of exoskeleton wearers are primary motivators of strategic synchrony and can be explained by marrying the conditional cooperation norm and the reinforcement of cooperation model (Reddish et al., 2014). The conditional cooperation norm proposes that modern society functions from the underlying norm of cooperation. Simply stated, individuals are more likely to contribute if others in their environment are also contributing. Further, the higher the contribution rates observed by members, the more likely they will also contribute (Frey and Meier, 2004). The increase in contribution creates cooperation within a system, environment, or workplace. Similarly, from a goals-perspective, the reinforcement of cooperation model (Reddish et al., 2014) explains why spatial alignment amplifies cooperative responses from participants. Originally developed to understand shared intentionality during music and dance performances (Reddish et al., 2014), this model suggests that when there is a common goal to synchronize, the perception produces immediate feedback to the actor that cooperation is taking place. Increases in the feelings of joint rhythmic coordination reinforces successful cooperation and leads to participants feeling perceived similarity, trust, and confidence in their partners (Launay et al., 2016).

Of course, cooperation is typically a two-sided and mutually rewarding experience. However, extending IAT and strategic synchrony to human-exoskeleton interactions provides an opportunity to conceptualize cooperation from this new perspective. In contrast to human-human interactions, in the human-exoskeleton context perceived reciprocity from the exoskeleton may not influence the desire for cooperation unless users heavily anthropomorphize these technologies (Bartneck et al., 2020a). In this section we have conceptualized cooperation norms as the social norms that employees have toward exoskeletons at the team, group, or organizational level. If the social norm is to cooperate by strategically synchronizing with the exoskeleton, it is likely that this will motivate an individual to produce behaviors that foster a goal of cooperation between the wearer and the exoskeleton which will, in turn, reinforce the synchronization. This may be especially true if the individual is approach-goal oriented when wearing the exoskeleton. In sum, underlying strategic processes that unfold during the interaction to create synchrony can be explained by the individual goals and social norms of the environment in which the encounter takes place.

OUTCOMES

When considering wearer synchrony and the use of active exoskeletons in the workplace, two main outcomes are especially salient. The first main outcome is wearer comfortability with the exoskeleton and the second main outcome is overall job satisfaction. Although there is currently no empirical research that explores the relationship between human-exoskeleton synchrony and these outcome variables, we offer some ideas on how automatic and strategic synchrony may influence these outcomes.

Comfortability With Wearable Technologies

Knight and Baber (2005) explained that it can be difficult for designers to create wearable technologies that multiple stakeholder groups feel comfortable wearing. These challenges are partly attributed to individual variations between users, especially as we expect to see large variations among users in the antecedent variables we propose. When we refer to comfortability in the exoskeleton context, we recognize two main distinctions. One aspect of comfortability involves how comfortable wearers feel when wearing the exoskeleton and the second aspect refers to how comfortable users feel when using the exoskeleton in a work environment. When it comes to comfortability when using the exoskeleton, affective and physiological factors are especially important as uncomfortability or pain can lead to musculoskeletal disorders for wearers.

For comfortability when using the suit in the workplace, automatic synchrony or strategic synchrony likely will influence whether employees want to use the exoskeleton and how they want to use it. For instance, automatic synchrony processes may make individuals feel less burdened on a cognitive level which could help them pay closer attention to their surroundings and feel more comfortable using the exoskeleton around coworkers. However, the opposite may also be true. Automatic synchrony could may make wearers feel burdened to understand how the machine is functioning, resulting in apprehension toward using the suit in the workplace. Strategic synchrony could also have nuanced impacts on comfortability for using the exoskeleton in work tasks. On one hand, the ability to strategically synchronize with an exoskeleton may help users feel more empowered and in control of the suit's movements leading to feelings of comfortability with the suit around coworkers. However, strategic synchrony may also place too many cognitive demands on the wearer and decrease their ability to perceive their surroundings which could lead to apprehension when wearing the suit.

Job Satisfaction

We expect that automatic and strategic synchrony could also have nuanced impacts on job satisfaction for wearers. Organizational research showcases that when employees feel a lack of control, agency or autonomy in the workplace, they are more likely to experience stress, burnout, and report decreased levels of overall of job satisfaction (Chen and Silverthorne, 2008; Mahon, 2014). If strategic synchrony can help increase feelings of autonomy and an employee's internal locus of control, then employees who can choose to engage in synchrony may be more satisfied in their work. But if strategic synchrony is too costly on a worker's physical and cognitive energy, they may not feel in control of the suit which could lead to stress and lower levels of overall job satisfaction. There are also nuanced possibilities on the impact of automatic synchrony and overall job satisfaction. For instance, if workers feel more in control of the tasks they complete with an exoskeleton in automatic synchronization conditions, then we expect overall job satisfaction to increase. However, if automatic synchrony compromises feelings of autonomy and control for wearers then overall job satisfaction is likely to decrease. Non-verbal scholars are well-positioned to examine the tradeoffs that employees make in automatic and strategic synchronization conditions and how these processes ultimately impact working conditions for wearers.

METHODOLOGICAL CHALLENGES

Exoskeletons are likely to dramatically shift how work is conducted in traditional blue-collar industries. Although there has yet to be research on synchrony between humans and wearable technologies—including exoskeletons—there are two key methodological challenges that non-verbal researchers will face in this area. The two key challenges are determining the relationship between perceptions of synchrony and actual instances of synchrony, and the technical challenge of separating human and machine datapoints for analysis.

Perceptions in Synchrony

One key challenge of synchrony research in HRI contexts is to understand whether the perceptions of synchrony that wearers report actually match levels of synchrony with exoskeleton. Although perceptions of synchrony are the least expensive and easiest data to obtain, it may not provide the complete picture of how synchrony functions in HRI. However, automated wavelet spectrum analysis in non-verbal communication research could be a useful tool for overcoming this specific challenge (Fujiwara and Daibo, 2016). Fujiwara and Daibo (2016) explained that early research on synchrony involved human coders observing interactions and coding for synchrony behaviors which was ultimately time-intensive and costly. In contrast to these traditional methods, researchers that use wavelet spectrum analysis leverage in-depth sensors and can automate the coding of synchrony behaviors. When conducting research on synchrony in wearable technologies, human coders would not be able to distinguish movements of the human wearer from the exoskeleton suit because of close proxemics, so an automated method such as wavelet spectrum analysis will likely be needed to conduct this research.

Separating Datapoints

The second main methodological challenge involves the ability to separate datapoints from the wearer's and suit's movements and collect both types of data for analysis. Active exoskeletons are already being designed with many sophisticated sensors that detect, anticipate, and react to wearer movements so the main challenge would be to log wearer movements independently of exoskeleton movements. One way this could be possible is to have sensors on the wearer that are independent from sensors (such as sensors on the clothing a wearer has underneath the exoskeleton suit) on the exoskeleton so that both types of data could be analyzed as separate units. The separation of these datapoints would be crucial for researchers to be able to use automated synchrony methods that do not require manual human coding [see Fujiwara and Daibo (2016) for an example of automated wavelet spectrum analysis in synchrony research]. Overcoming these two methodological challenges are crucial for conducting synchrony research in human-exoskeleton interactions.

FUTURE RESEARCH AGENDA

If research on synchronization in human-machine dyads is in its infancy, then research on synchronization in human-exoskeleton dyads is still in conception. We encourage non-verbal communication researchers to critically engage with wearable technologies and explore how traditional non-verbal communication theories can be extended to new contexts. We propose two key areas of future research that will help shape knowledge of synchrony between humans and exoskeletons including testing and adding complexity to the IAT framework we propose.

Testing the IAT Framework

We encourage non-verbal communication scholars who are interested in human-exoskeleton interactions to empirically test the antecedents, processes, and outcomes we offer in this piece. In designing this program of research, testing the interaction position including the requirements, expectations, and desires that users have before interacting with an exoskeleton would be a critical first step to testing this overall framework.

Antecedents

Next, we encourage researchers to explore the antecedent variables that are likely to have the biggest impact on synchrony in human-exoskeleton interactions. In the area of affective antecedents, prior HRI research has shown that humans may be predisposed with negative attitudes, anxieties, or negative affect toward robotic technologies (Nomura et al., 2008; Bartneck et al., 2020a). Scales such as the negative attitudes toward robots scale (NARS) or the robot anxiety scale (RAS) have been tested and used as ways to gauge attitudes and anxieties toward robots (Nomura et al., 2008). Although they can give insight for exoskeleton technology, these measures are still specific to fully automated robots. More research much be done to understand whether these attitudes, anxieties, and affect toward autonomous robots are transferable to wearable technologies more generally and specifically exoskeletons. A logical step in exoskeleton research would be to determine differences that people may have when discussing negative attitudes and anxieties toward robots and exoskeletons. If they shown to be similar, then many of the implications the scales have for HRI could be extended to exoskeleton research. Once the scales are also proven to be valid indicators of attitudes and anxieties toward exoskeletons, a potential program of research can be conducted to further determine how different attributes of humans influence how they score on these scales. The transferability of these concepts is especially important to explore considering the complex physiological, affective, and cognitive factors in wearable technologies (Knight and Baber, 2005).

Also, worth addressing is Asher's et al. (2020) research on how to lower levels of anxiety in HRI. Through an analysis of videos, Asher et al. explored how individuals with social anxiety did not improve non-verbal synchrony when having closeness-generating conversations but did improve when having small-talk conversations. This line of inquiry could potentially give insight on how to design interventions that can lower anxiety between humans and wearable robots as well as increase positive attitudes or affect toward these technologies. Isolating the relationship between these interventions and synchrony will be crucial for understanding how organizations should orient employees toward exoskeletons as well as how trainings can help users improve their ability to synchronize with the technology.

In regard to cognitive antecedents, we urge researchers to explore how organizations can help teams develop more positive TMMs during active exoskeleton adoption (Mathieu et al., 2000). This area of research must involve understanding how beliefs about individual roles, perceptions of tasks, and collaboration norms within traditional blue-collar work changes with active exoskeleton adoption at the team, department, or organizational level. Researchers interested in these issues should also consider when is the proper time to introduce interventions designed to increase the positivity of TMMs in this context. Although we know from Pearsall et al. (2010) that positive TMM is more likely to occur when the understanding of roles occurs early in the team's history, we do not know how early these interventions should be introduced to be most effective. For instance, in the active exoskeleton context, it is unclear when employees should be exposed to the technology before it is integrated in work practices as well as how much time teams should generally have to test the technology without worrying about hitting performance metrics typically required in their work.

Processes

Exploring the relationships between agency in automatic and strategic synchrony is also potentially a fruitful program of research. Although past organizational research showcases examples where employees need appropriate levels of agency and autonomy to enjoy their work, we simply do not have enough information to understand whether this is transferable to the use of active exoskeletons in the workplace (Chen and Silverthorne, 2008; Mahon, 2014). Research in this area should specifically explore whether users of wearable technology feel more agency in their roles if the technology automatically synchronizes with their movements or if workers feel more empowered when they can strategically synchronize with the technology. This line of inquiry is especially complex because it questions the relationship between active exoskeleton use and professional identity. Another complex dimension in this program of research is the tradeoff that employees make between receiving enhanced physical capabilities when synchronizing with active exoskeletons and the costs of cognitive energy that wearers experience when using exoskeletons to complete tasks.

Outcomes

Lastly, when designing a program of research to test the IAT framework it is important that researchers explore how different types (automatic or strategic) and levels of synchrony between humans and exoskeletons impact the outcome variables we suggest. For the outcome of comfortability for exoskeletons wearers, we have extended Knight and Baber's (2005) typology of comfortability dimensions for wearable technologies to the active exoskeleton context and have provided concrete examples of how these dimensions could be relevant in these interactions (See **Table 1**). We suggest that this line of inquiry first be conducted from an inductive or exploratory approach as there may be some important comfort dimensions relevant to the human-exoskeleton context that have not been mentioned in prior scholarship or research.

When designing a program of research addressing levels of job satisfaction, we know from prior organizational research that when employees feel a lack of control, agency or autonomy in the workplace, they are more likely to experience stress, burnout, and report decreased levels of overall of job satisfaction (Chen and Silverthorne, 2008; Mahon, 2014); but we do not the extent to which automatic or strategic types of synchrony impact levels of agency, autonomy, and control that workers perceive and experience. Special consideration should also be paid to variations in professional identity and industry affiliation play in predicting the relationship between type of synchrony (whether automatic or strategic) and agency, autonomy, and control in the workplace.

We expect that individuals who are used to high levels of agency, autonomy, and control (such as trainers, supervisors, or managers) may be more sensitive to changes in these variables and can be more susceptible to changes in job satisfaction when active exoskeletons are adopted. This may be partially attributed to the ways that active exoskeleton adoption can disrupt expertise in organizations. Past research on robotics have shown that when robots take larger roles in complex tasks, expertise in organizations can be disrupted in both positive and negative ways (Davenport, 2018; Beane, 2019). In an ethnographic case study on a cadre of initiate surgeons, Beane (2019) found that the new collaborative relationships with robots in surgery interrupted the normal training process for surgeons and required that they pre-maturely chose specific expertise. Due to the emerging nature of active exoskeleton technology in blue-collar industries researchers should explore (a) how disruptive these technologies will be and (b) how the disruptive nature will impact job satisfaction in these environments across different types and groups of employees.

Adding Complexity to the IAT Framework

The main challenge of theorizing about a cutting-edge technology such as active exoskeletons is that the complexity of frameworks that can be introduced in these contexts has limitations. Although we hope that the robust IAT framework we apply to the human-exoskeleton context helps inspire new and provocative types of research, we recognize that our framework is not an all-exhaustive list of the important concepts and variables for this context. We briefly mention two concepts that could be introduced to add more complexity to the IAT framework we propose.

Entrainment

Entrainment is defined as, "a process that leads to temporal coordination of two actors' behavior, in particular, synchronization, even in the absence of a direct mechanical coupling." (Knoblich et al., 2011, p. 63). It is important for researchers to consider how entrainment and temporal dimensions vary across different types of powered exoskeletons and the goals or motivations of the wearer. For instance, some types of powered exoskeletons are designed for medical rehabilitation for wearers who have sustained serious injuries. For instance, the Indego personal suit enables individuals with spinal cord injuries to stand and walk independently (Parker Hannifin Corp, 2019). Active exoskeletons designed to help injured individuals walk are very different than active exoskeletons designed to give workers super-human levels of strength in the workplace. It is clear that differences in the design of these medical active exoskeletons and motivations of wearers will impact the process of entrainment, the temporal nature of tasks that the exoskeleton is used for, and the type of synchrony (automatic or strategic) that is available in this context.

Process Interactions

Given that designers of active exoskeletons for private industry are still debating how responsive and how much control these technologies should have during interactions, we have conceptualized automatic and strategic synchrony as two distinct processes. However, it is possible that wearers may find that certain parts of an exoskeleton automatically synchronize with their movements more than others, or that performing some types of tasks give them more control over the exoskeleton. We recommend that researchers who are interested in synchrony in the active exoskeleton context should be open to considering how both types of synchrony may be present in the same piece of technology as well as how the interactions between these processes influences the antecedents and outcomes we mention.

CONCLUSION

Emerging technologies are becoming increasingly complex, not only in how the technology operates but also in how these technologies make people *feel.* Advancements in emerging technologies such as active exoskeletons illustrate that collaborative relationships between humans and machines are likely to become more important across a variety of professional settings. We have applied IAT to human-exoskeleton interactions in order to offer an in-depth and illustrative example for how traditional non-verbal communication theories can be reimagined in new technological contexts, but we certainly do not think these efforts should only be scoped to exoskeleton technologies. Certainly, there are a multitude of technological contexts, non-verbal communication variables, and methodological challenges that should be considered by non-verbal communication researchers and practitioners alike. We hope that the initial insights we have provided help inspire researchers to keep interrogating non-verbal communication theories in a rapidly changing world and continue to ensure our field has the relevance needed to meet the challenges of the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

GK leveraged expertise with emerging technologies to conceptualize the central argument for this piece, including how to extend IAT to the human-exoskeleton context. CO leveraged expertise in synchrony research to explicate key concepts and variables that were foundational in our theoretical argument and designed figures to help clarify theoretical arguments. MH leveraged expertise in nonverbal communication to identify important antecedents in human-exoskeleton interactions. All authors contributed to the article and approved the submitted version.

FUNDING

This project was funded by the National Science Foundation (Award Number 1839946) but the opinions and findings are the work of the authors and should not be considered U.S. Government policy or position. The funding agency had no role in the writing of the report; and in the decision to submit the article for publication.

REFERENCES

Andersen, P. A. (1998). "The cognitive valence theory of intimate communication," in *Progress in Communication Sciences: Vol. 14. Mutual Influence in Interpersonal Communication Theory and Research in Cognition, Affect, and Behavior*, eds M. Palmer and G. A. Barnett (Stamford, CT: Ablex), 39–72.

Andersen, P. A., and Guerrero, L. K. (1998). "Principles of communication and emotion in social interaction," in *Handbook of Communication and Emotion*, eds P. Andersen and L. Guerrero (San Diego, CA: Academic Press), 49–96.

Arundale, R. B. (1996). "Indexing pattern over time: criteria for studying communication as a dynamic process," In *Dynamic Patterns in Communication Processes*, eds J. Watt and C. A. van Lear (Thousand Oaks, CA: Sage), 95–118.

Asher, M., Kauffmann, A., and Aderka, I. M. (2020). Out of Sync: nonverbal synchrony in social anxiety disorder. *Clin. Psychol. Sci.* 8, 280–294. doi: 10.1177/2167702619894566

Banks, J., and de Graaf, M. (2020). Toward an agent-agnostic transmission model: synthesizing anthropocentric and technocentric paradigms in communication. *Hum. Mach. Commun.* 1, 19–36. doi: 10.30658/hmc.1.2

Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., and Šabanović, S. (2020a). *Human Robot Interaction: An Introduction.* Cambridge: Cambridge University Press.

Bartneck, C., Kulic, D., and Croft, E. (2020b). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Sock. Robots* 1, 71–81. doi: 10.1007/s12369-008-0001-3

Batson, C. D. (1998). "Altruism and prosocial behavior," in *The Handbook of Social Psychology, Vol. 2*, eds D. Gilbert, S. T. Fiske, and G. Lindzey (Boston: McGraw-Hill), 282–316.

Beane, M. (2019). Shadow learning: building robotic surgical skill when approved means fail. *Adm. Sci. Q.* 64, 87–123. doi: 10.1177/0001839217751692

Bernhold, Q. S., and Giles, H. (2020). Vocal accommodation and mimicry. *J. Nonverbal Behav.* 44, 41–62. doi: 10.1007/s10919-019-00317-y

Bernieri, F., Reznick, J. S., and Rosenthal, R. (1988). Synchrony, pseudo-synchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions. *J. Pers. Soc. Psychol.* 54, 243–253. doi: 10.1037/0022-3514.54.2.243

Bernieri, F., and Rosenthal, R. (1991). "Interpersonal coordination: behavioral matching and interactional synchrony," in *Foundations of Nonverbal Behavior*, eds R. S. Feldman and B. Rime (Cambridge: Cambridge University Press), 401–432.

Buller, D. B., and Aune, R. K. (1988). The effects of vocalics and nonverbal sensitivity on compliance: a speech accommodation theory explanation. *Hum. Commun. Res.* 14, 301–332. doi: 10.1111/j.1468-2958.1988.tb00159.x

Buller, D. B., and Aune, R. K. (1992). The effects of speech rate similarity on compliance: application of communication accomodation theory. *Western J. Commun.* 56, 37–53. doi: 10.1080/10570319209374400

Burgoon, J. K., Dunbar, N. E., and Giles, H. (2017). "Interaction coordination and adaption," in *Social Signalprocessing,* eds J. K. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli (Cambridge: Cambridge University Press), 78–96.

Burgoon, J. K., Dunbar, N. E., and White, C. (2014). "Interpersonal adaptation," in *Interpersonal Communication: Handbooks of Communication Science,* ed C. R. Berger (Boston: De Gruyter Mouton), 225–248.

Burgoon, J. K., and Hale, J. L. (1988). Nonverbal expectancy violations theory: model elaboration and application to immediacy behaviors. *Commun. Monogr.* 55, 58–79. doi: 10.1080/03637758809376158

Burgoon, J. K., Stern, L. A., and Dillman, L. (1995). *Interpersonal Adaption Dyadic Interaction Patterns.* Cambridge: Cambridge University Press.

Burrell, J. (2016). How the machine 'thinks:' understanding opacity in machine learning algorithms. *Big Data Soc.* 3, 1–12. doi: 10.1177/2053951715622512

Cappella, J. N. (1991). The biological origins of automated patterns of human interaction. *Commun. Theory* 1, 4–35. doi: 10.1111/j.1468-2885.1991.tb00002.x

Caughlin, J. P. (2010). A multiple goals theory of personal relationships: conceptual integration and program overview. *J. Soc. Pers. Relat.* 27, 824–848. doi: 10.1177/0265407510373262

Chartrand, T. I., 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, J., and Silverthorne, C. (2008). The impact of locus of control on job stress, job performance and job satisfaction in Taiwan. *Leadersh. Organ. Dev. J.* 29, 572–582. doi: 10.1108/01437730810906326

Chen, T. L., King, C. H. A., Thomaz, A. L., and Kemp, C. C. (2014). An investigation of responses to robot-initiated touch in a nursing context. *Int. J. Sock. Robot.* 6:1411161. doi: 10.1007/s12369-013-0215-x

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *J. Pers. Soc. Psychol.* 58, 1015–1026. doi: 10.1037/0022-3514.58.6.1015

Cialdini, R. B., and Trost, M. R. (1998). "Social influence: social norms, conformity, and compliance," in *The Handbook of Social Psychology, 4th ed*, *Vol. 2*, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York, NY: McGraw-Hill), 151–192.

Condon, W. S., and Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *J. Nervous Mental Dis.* 143, 338–457. doi: 10.1097/00005053-196610000-00005

Davenport, T. H. (2018). *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work.* Cambridge: The MIT Press.

de Looze, M. P., Bosch, T., Krause, F., Stadler, K. S., and O'Sullivan, L. W. (2016). Exoskeletons for industrial application and their potential effects on physical work load. *Ergonomics* 59, 671–681. doi: 10.1080/00140139.2015.1081988

Dunbar, N. E., Jensen, M. L., Tower, D. C., and Burgoon, J. K. (2014). Synchronization of nonverbal behaviors in detecting mediated and non-mediated deception. *J. Nonverbal Behav.* 38, 355–376. doi: 10.1007/s10919-014-0179-z

Feldstein, S., Konstantareas, M., Oxman, J., and Webster, C. D. (1982). The chronography of interactions with autistic speakers: an initial report. *J. Commun. Disord.* 15, 451–460. doi: 10.1016/0021-9924(82)90018-1

Fisher, D. M., Bell, S. T., Dierdorff, E. C., and Belohlav, J. A. (2012). Facet personality and surface-level diversity as team mental model antecedents: implications for implicit coordination. *J. Appl. Psychol.* 97, 825. doi: 10.1037/a0027851

Floyd, K., and Burgoon, J. K. (1999). Reacting to nonverbal expressions of liking: a test of interaction adaptation theory. *Commun. Monogr.* 66, 219–239. doi: 10.1080/03637759909376475

Frey, B., and Meier, S. (2004). Social comparisons and pro-social behavior: testing 'conditional cooperation' in a field experiment. *Am. Econ. Rev.* 94, 1717–1722. doi: 10.1257/0002828043052187

Fujiwara, K., Bernhold, Q. S., Dunbar, N. E., Otmar, C. D., and Hansia, M. (2020). Comparing manual and automated coding methods of nonverbal synchrony. *Commun. Methods Meas.* doi: 10.1080/19312458.2020.1846695. [Epub ahead of print].

Fujiwara, K., and Daibo, I. (2016). Evaluating interpersonal synchrony: wavelet transform toward an unstructured conversation. *Front. Psychol.* 12:516. doi: 10.3389/fpsyg.2016.00516

Fulk, J. (1993). Social construction of communication technology. *Acad. Manage. J.* 36, 921–950. doi: 10.2307/256641

Gable, S. L. (2006). Approach and avoidance social motives and goals. *J. Pers.* 74, 175–222. doi: 10.1111/j.1467-6494.2005.00373.x

Gable, S. L., and Berkman, E. T. (2008). "Making connections and avoiding loneliness: approach and avoidance social motives and goals," in *Handbook of Approach and Avoidance Motivation*, ed A. Elliot (New York, NY: Psychology Press), 203–216.

Gatewood, J. B., and Rosenwein, R. (1981). Interactional synchrony: genuine or spurious? A critique of recent research. *J. Nonverbal Commun.* 6, 12–29. doi: 10.1007/BF00987933

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Mahwah, NJ: Erlbaum.

Giddens, A. (1979). *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis.* Berekley: Macmillian.

Giles, H. (1973). Accent mobility: a model and some data. *Anthropol. Linguist.* 15, 87–105.

Giles, H., and Ogay, T. (2007). "Communication accommodation theory," in *Explaining Communication: Contemporary Theories and Exemplars*, eds B. B. Whaley and W. Samter (Lawrence Erlbaum Associates Publishers), 293–310.

Hasnain, S. K., Mostafaoui, G., Salesse, R., Marin, L., and Gaussier, P. (2013). "Intuitive HR interaction based on unintentional synchrony: a psycho-experimental study," in *Joint International Conference on Developmental Learning and Epigenetic Robotic*, 1–7.

Hewes, D. E. (1979). The sequential analysis of social interaction. *Q. J. Speech* 65, 56–73. doi: 10.1080/00335637909383458

Horne, C. (2001). "Sociological perspectives on the emergence of norms," in *Social Norms*, eds M. Hechter and K. D. Opp (New York, NY: Russell Sage Foundation), 3–34.

Kellermann, K. L. (1992). Communication: inherently strategic and primarily automatic. *Commun. Monogr.* 59, 288–300. doi: 10.1080/03637759209376270

Kendon, A. (1970). Movement coordination in social interaction: some examples described. *Acta Psychol.* 32, 1–25. doi: 10.1016/0001-6918(70)90094-6

Kendon, A. (1990). *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.

Klimoski, R., and Mohammed, S. (1994). Team mental model: construct or metaphor?. *J. Manage*. 20, 403–437. doi: 10.1177/014920639402000206

Knight, J. F., and Baber, C. (2005). A tool to assess the comfort of wearable computers. *Hum. Factors* 47, 77–91. doi: 10.1518/0018720053653875

Knoblich, G., Butterfill, S., and Sebanz, N. (2011). "Psychological research on joint action: theory and data," in *Psychology of Learning and Motivation, Vol. 54*, eds K. D. Federmeierand E. R. Schotter (Academic Press), 59–101.

Kopp, C. (2011). *Exoskeletons for Warriors of the Future*. Defence Focus. Retrieved from http://www.ausairpower.net/PDF-A/DT-Exoskeletons-Sep-2011.pdf

Korte, R. (2010). First, get to know them: a relational view of organizational socialization. *Hum. Resource Dev. Int.* 13, 27–43. doi: 10.1080/13678861003588984

Launay, J., Tarr, B., and Dunbar, R. I. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology* 122, 779–789. doi: 10.1111/eth.12528

Lee, C.-C., Black, M., Katsamanis, A., Lammert, A., Baucom, B., Christensen, A., et al. (2010). "Quantification of prosodic entrainment in affective spontaneous spoken interactions in married couples," in *Proceedings of InterSpeech* (Makuhari: International Speech Communication Association), 793–796.

Lin, W.-F., Gosnell, C. L., and Gable, S. L. (2018). Goals, emotions, and the effort to be responsive during couple interactions. *Motiv. Emot.* 43, 313–324. doi: 10.1007/s11031-018-9731-1

Lin, W. S. (2012). Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. *Int. J. Hum. Comput. Stud*. 70, 498–507. doi: 10.1016/j.ijhcs.2012.01.006

Locke, J. L. (1993). "The role of the face in vocal learning and the development of spoken language," in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, eds B. De Boysson-Bardies and S. de Schonen (Norwell, MA: Kluwer Academic Publishers), 317–328.

Mahon, P. R. (2014). A critical ethnographic look at paediatric intensive care nurses and the determinants of nurses' job satisfaction. *Int. Critic. Care Nurs.* 30, 45–53. doi: 10.1016/j.iccn.2013.08.002

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., and Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *J. Appl. Psychol.* 85:273. doi: 10.1037/0021-9010.85.2.273

Nomura, T., Kanda, T., Suzuki, T., and Kato, K. (2008). Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes towards robots. *IEEE Trans. Robot.* 24, 442–451. doi: 10.1109/TRO.2007.914004

Parker Hannifin Corp (2019). *Indego Therapy Resources*. Retrieved from http://www.indego.com/indego/en/Indego-Therapy

Patterson, M. L. (2019). A systems model of dyadic nonverbal interaction. *J. Nonverbal Behav.* 43, 111–132. doi: 10.1007/s10919-018-00292-w

Pearsall, M. J., Ellis, A. P. J., and Bell, B. S. (2010). Building the infrastructure. The effects of role identification behaviors on team cognition development and performance. *J. Appl. Psychol.* 95, 192–200. doi: 10.1037/a0017781

Prepin, K., and Revel, A. (2007). Human-machine interaction as a model of machine-machine interaction: how to make machines interact as humans do. *Adv. Robot.* 21, 1709–1723. doi: 10.1163/156855307782506192

Reddish, P., Bulbulia, J., and Fischer, R. (2014). Does synchrony promote generalized prosociality? *Relig. Brain Behav*. 4, 3–19. doi: 10.1080/2153599X.2013.764545

Sarcos Corp (2019). *Guardian XO Full-Body Powered Exoskeleton*. Retrieved from https://www.sarcos.com/products/guardian-xo-powered-exoskeleton

Smelik, A., Toussaint, L., and Van Dongen, P. (2016). Solar fashion: an embodied approach to wearable technology. *Int. J. Fash. Stud.* 3, 287–303. doi: 10.1386/infs.3.2.287_1

Sparks, B. (1994). Communicative aspects of the service encounter. *Hospital. Res. J.* 17, 39–50. doi: 10.1177/109634809401700205

Szollosy, M. (2015). "Why are we afraid of robots? the role of projection in the popular conception of robots," in *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide,* eds J. Romporti, E. Zackova, J. Kelemen (New York, NY: Springer), 121–131.

Tajfel, H., and Turner, J. C. (1979). "An integrative theory of intergroup conflict." In *The Social Psychology of Intergroup Relations*, eds W. G. Austin and S. Worchel (New York, NY: Brooks/Cole), 33–47.

Thoits, P. A., and Virshup, L. K. (1997). "Me's and we's: Forms and functions of social identities," in *Rutgers Series on Self and Social Identity, Vol. 1,* eds R. D. Ashmore and L. J. Jussim (Oxford University Press), 106–133.

Tickle-Degnen, L., and Gavett, E. (2003). "Changes in nonverbal behavior during the development of therapeutic relationship," in *Nonverbal Behavior in Clinical Settings*, eds P. Philippot, R. S. Feldman, and E. J. Coats (New York, NY: Oxford University Press), 75–110.

Treem, J. W., and Leonardi, P. M. (2013). Social media use in organizations: exploring the affordances of visibility, editability, persistence, and association. *Ann. Int. Commun. Associ.* 36, 143–189. doi: 10.1080/23808985.2013.11679130

Upasani, S., Franco, R., Niewolny, K., and Srinivasan, D. (2019). The potential for exoskeletons to improve health and safety in agriculture—perspectives from service providers. *IISE Trans. Occup. Ergonom. Hum. Factors* 7, 222–229. doi: 10.1080/24725838.2019.1575930

Van Erp, J. B. F., and Toet, A. (2013). "How to touch humans. guidelines for social agents and robots that can touch," in *Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (Geneva: IEEE Computer Society), 780–785.

Warner, R. M. (1992). Sequential analysis of social interaction: assessing internal vs. social determinants of behavior. *J. Pers. Soc. Psychol.* 63, 51–60. doi: 10.1037/0022-3514.63.1.51

Zaroug, A., Proud, J. K., Lai, D. T. H., Mudie, K., Billing, D., and Begg, R. (2019). "Overview of computational intelligence (CI) techniques for powered exoskeletons," in *Computational Intelligence in Sensor Networks*, eds B. Mishra, S. Dehuri, B. Panigrahi., A. Nayak., B. Mishra, H. Das (Berlin: Springer), 353–383.

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 February 2021
doi: 10.3389/fpsyg.2021.605928

Check for updates

Implications for Emotion: Using Anatomically Based Facial Coding to Compare Emoji Faces Across Platforms

Jennifer M. B. Fugate[1]*† and Courtny L. Franco[2]†

[1] Department of Psychology, University of Massachusetts Dartmouth, Dartmouth, MA, United States, [2] Department of Communication and Information Science, University of Alabama, Tuscaloosa, AL, United States

Emoji faces, which are ubiquitous in our everyday communication, are thought to resemble human faces and aid emotional communication. Yet, few studies examine whether emojis are perceived as a particular emotion and whether that perception changes based on rendering differences across electronic platforms. The current paper draws upon emotion theory to evaluate whether emoji faces depict anatomical differences that are proposed to differentiate human depictions of emotion (hereafter, "facial expressions"). We modified the existing Facial Action Coding System (FACS) (Ekman and Rosenberg, 1997) to apply to emoji faces. An equivalent "emoji FACS" rubric allowed us to evaluate two important questions: First, *Anatomically, does the same emoji face "look" the same across platforms and versions?* Second, *Do emoji faces perceived as a particular emotion category resemble the proposed human facial expression for that emotion?* To answer these questions, we compared the anatomically based codes for 31 emoji faces across three platforms and two version updates. We then compared those codes to the proposed human facial expression prototype for the emotion perceived within the emoji face. Overall, emoji faces across platforms and versions were not anatomically equivalent. Moreover, the majority of emoji faces did not conform to human facial expressions for an emotion, although the basic anatomical codes were shared among human and emoji faces. Some emotion categories were better predicted by the assortment of anatomical codes than others, with some individual differences among platforms. We discuss theories of emotion that help explain how emoji faces are perceived as an emotion, even when anatomical differences are not always consistent or specific to an emotion.

Keywords: emoji faces, emotion perception, facial action coding system, electronic platforms, facial expressions

Edited by:
José-Miguel Fernández-Dols, Autonomous University of Madrid, Spain

Reviewed by:
Anna Tcherkassof, Université Grenoble Alpes, France
Patrícia Arriaga, University Institute of Lisbon (ISCTE), Portugal

***Correspondence:**
Jennifer M. B. Fugate
jfugate@umassd.edu

†ORCID:
Jennifer M. B. Fugate
orcid.org/0000-0003-0831-4234
Courtny L. Franco
orcid.org/0000-0002-8878-1846

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *13 September 2020*
Accepted: *18 January 2021*
Published: *25 February 2021*

Citation:
Fugate JMB and Franco CL (2021) Implications for Emotion: Using Anatomically Based Facial Coding to Compare Emoji Faces Across Platforms. Front. Psychol. 12:605928. doi: 10.3389/fpsyg.2021.605928

IMPLICATIONS FOR EMOTION: USING ANATOMICALLY BASED FACIAL CODING TO COMPARE EMOJI FACES ACROSS PLATFORMS

Emojis, which are now incorporated into people's everyday channels of nonverbal communication, are assumed to represent, or at least resemble, human facial depictions of emotion (hereafter, referred to as "facial expressions"). Despite many studies that show that including emojis alters emotional content, only one study (Franco and Fugate, 2020) has examined whether emoji faces are perceived as a discrete emotion. A logical next step is to explore whether emoji faces are structurally "equivalent" among platforms and version updates, and whether emoji faces actually resemble prototypical facial expressions (in physical appearance).

In this paper, we adapted the Facial Action Coding System (FACS) (Ekman and Rosenberg, 1997) to systematically compare emoji faces with respect to facial movements, called action units (AUs). Although AUs are built on changes in the underlying facial musculature, movements can be inferred in still faces based on deviations from a baseline pose. We used this adaptive new system to code 31 emojis on their physical appearance across two different versions of three electronic platform carriers (Apple iOS 9.1, Apple iOS 13.3, Google Android 6.0, Google Android 10, Samsung TouchWiz 5.1 and Samsung One UI 1.5). We then systematically compared the AUs within and between emojis across platforms and versions. We also used previously collected data of participants' emotion category assignment for each emoji (Franco and Fugate, 2020) to see whether the emoji AUs conformed to those proposed for human facial expressions, according to the literature (Cordaro et al., 2018).

A Brief Primer on Emotion Theory and "Basic" Facial Expressions

According to some theories of emotion, facial expressions dissociate themselves reliably among emotions (Tomkins, 1962; Izard, 1991, 1992, 2013; Ekman, 1992, 2016; Matsumoto et al., 2008; Brosch et al., 2010; Sauter et al., 2011; for an extensive review see Barrett et al., 2019). Furthermore, under this view, a "basic" set of emotions are viewed as innate and universal among individuals (Ekman and Cordaro, 2011). According to these views, emotions are also biologically based and evolutionary-preserved, such that facial expressions have evolutionary significance and are shared with taxonomically related species (van Hooff, 1962; Ekman, 1972, 1992; Matsumoto, 1989; deWaal, 2003; Parr et al., 2007). This view comes mainly from similar mimetic facial musculature that is highly conserved across primates (Huber, 1931; Parr et al., 2007). The human and non-human mimetic facial musculature have been anatomically mapped by a system of action units that are shared between species (e.g., chimpFACS: Parr et al., 2007; and MaqFACS: Parr et al., 2010). Although the early work on emotion perception and facial musculature focused on six "basic" emotions (e.g., anger, disgust, sadness, happiness, fear, and surprise) (Ekman, 1972, 1992; Ekman et al., 1983; for reviews, see Elfenbein et al., 2002; Keltner et al., 2016, 2019), more recent research has proposed more than twenty "basic" emotions might exist based on self-report (Cowen and Keltner, 2017). Fourteen of these emotions show at least some consistency in the AUs identified for the emotion prototype across multiple studies (see **Table 1**). In our initial research (Franco and Fugate, 2020), we explored the nine most common emotions (plus envy).

Cordaro et al. (2018) collected free-response facial and bodily responses to emotional statements from Chinese, Indian, Japanese, Korean, and American individuals. Of the emotions investigated in the current paper, surprise, contentment, fear, and anger all had over 50% overlap with the proposed emotional prototype based solely on AUs (89%, 80%, 71%, 67%, respectively), whereas contempt, sadness, and disgust showed less than 50% overlap with the proposed emotional prototype (33%, 33%, 18%, respectively). From these similarities and differences, they concluded that approximately 50% of an individual's overall expressed facial movements represent the universal prototype, whereas another 25% are due to the culture's "emotional dialect." Some of these structural changes in the facial musculature are known to be less diagnostic (e.g., wide eyes) and are shared among emotion categories (e.g., fear and surprise) (see Keltner et al., 2019; for an excellent review of human facial "expressions" and emotion, see Barrett et al., 2019). Despite such cultural variations, many of these researchers still continue to accept the universality of human facial expressions (hence the term "expressions" rather than facial "movements") and have introduced the International Core Pattern (ICP) of AUs for each of the emotions (see **Table 1**).

Other researchers highlight the cultural variation of facial expressions while still prescribing to a correspondence between facial expressions and emotion. For example, Elfenbein et al. (2007) and Elfenbein (2013) coined the "dialect theory of emotional expression," which posits that emotional expressions have regional or linguistic dialects (Elfenbein et al., 2007).

Other theories of emotion treat emotions as *products* of a person's brain to categorize more generalized affective information, which alone is not diagnostic of a particular emotion category. For instance, the Theory of Constructed Emotion (formerly known as Psychological Constructionism) posits that emotions are constructed through a person's conceptual knowledge within a given context (Barrett, 2006a,b, 2017). According to this theory, there are likely to be no distinctive or prescriptive emotional indicators for a specific emotion in the face (e.g., AUs). Even though frowns and smiles provide differences in structural information, perceivers must learn to associate them with sadness and happiness. Such associations are learned when another person labels the face with an emotion word (e.g., sad or happy), or a person uses situational knowledge to contextualize the information (Betz et al., 2019; for reviews, see Lindquist et al., 2016; Barrett, 2017; Lindquist, 2017). Therefore, a person's conceptual knowledge and context play a large role in the formation of facial depiction-emotion associations, and by extension, would likely also contribute to the perception of emotion from emoji faces. These ideas are consistent with how people develop electronic communication skills, more generally. That is, people develop an understanding of what another means through experience with others and feedback on that information (Ling, 2010; Liu and Yang, 2016).

Considering that emoji faces were designed to convey emotional content and to (presumably) resemble human facial expressions, it is worth comparing whether software companies' depictions actually capture the physical resemblance to certain facial expressions. Little scholarship provides insight to why multiple variations of the same emoji exist in the first place (Bailey, 2018; Lee, 2018). That is, there is little information on how individual emoji were "translated" across platforms, only that there is one translation through the Unicode system (Toratani and Hirayama, 2011; Unicode, 2020). Much like how

TABLE 1 | AU Prototypes across Literature.

	Ekman et al., 1983	Keltner and Buswell, 1997	Shiota et al., 2003	Matsumoto et al., 2008	Du et al., 2014	Keltner and Cordaro, 2015	Cordaro et al., 2016	Cordaro et al., 2018 **(ICP reported)**
Amusement			6,12,26 or 27, 55 or 56*	–	–	6,7,12, 25,26,53		6,7,12, 16,25,53*
Anger ∧	4,5,7,23			4,5 or 7, 22, 23,24	4,7,(10),(17),(23), 24	4,5,17, 23,24		4,7
Awe			1,5,26 or 27,57*	–	1,2,(4),5,(20),25, (26)	–		1,2,5,12,25,53*
Contempt∧	12,14			12,14				4,14,25
Contentment∧						12,43	12,43	12,43*
Sex/Desire (Love) ∧						19,25,26, 43		6,7,12, 25
Disgust∧	9,15,16			9 or 10, (25 or 26)	(4),9,10,17, (24)	7,9,19, 25,26		4,6,7,9, 10,25,26*
Embarrassment		12,24,51, 54,64		–	–	7,12,15,52,54,64		6,7,12, 25,54*
Fear∧	1,2,4,5,7, 20,26			1,2,4,5, 20, (25 or 26)	1, (2),4,(5),20,25, (26)	1,2,4,5,7,20,25		1,2,5,7, 25*
Happiness∧	6,12			6,12	(6),12, 25	6,7,12, 25,26		6,7,12, 16,25,26*
Pride			6,12,24,53*	–	–	53,64		7,12,53*
Sadness∧	1,4,15[1]			1, (4),15, (17)	(1),4, (6),(11),15,(17)	1,4,6,15,17		4,43,54
Shame		54,64		–	–	54,64		4,17,54
Surprise∧	1,2,5,26			1,2,5,25 or 26	1,2,(5), 25,26	1,2,5,25,26		1,2,5, 25

*We compared our codes to the last column. * = additional head or posture movement indicated, but no AUs identified for such. ∧ = used in the current paper. Envy is also used, but there are no proposed AUs for this emotion.*

[1]The original paper listed the "15" as a "5". We believe this to be a mistake and therefore corrected it.

facial expressions vary across cultures, emoji sets vary across platforms. Thus, the specific renderings of an emoji belong to web creators or web developers (e.g., Apple, Samsung, Google, Facebook, Twitter, etc.). Too much overlap between the "same" emoji on different platforms and/or version updates might be considered copyright infringement and could result in litigation (Bailey, 2018). While the Unicode has one "translation" for each emoji across platforms and version updates, it is up to developers and web creators to decide exactly *what* each translated emoji will look like.

Some researchers have alluded to the fact that emojis are artistic creations or creative expressions (Lee, 2018). To this end, emojis are considered art and not meant to be realistic depictions (of facial movements, in this case). As with any work of creative art, it is therefore up to the artist to communicate the intention even when the representation is not apparent[1].

Previous Literature of Emoji Emotional Perception

People perceive emoji faces similarly to human emotion faces. For example, Gantiva et al. (2019) found that emoji faces produced similar neural responses to real faces observed during face-to-face communication. In another study, Yuasa et al. (2011) found that emojis and human facial expressions elicited similar brain activity in the right and left inferior frontal gyri. Other areas within the brain, known to be important in processing emotional faces (e.g., right fusiform gyrus), were not significantly activated by emojis, however. A recent fMRI study investigated memory retrieval for emotional emoji faces and found significant activation within the inferior frontal gyrus, amygdala, and right temporal pole (Chatzichristos et al., 2020).

A growing body of research aims to understand how people use emojis to relay emotional sentiment on social media platforms, such as Twitter and Facebook. In 2015, researchers categorized 96,269,892 tweets by emotional content to find overarching patterns of emoji sentiment (reported in Wolny, 2016). For example, researchers categorized tweets containing grinning emoji faces as having positive sentiment. The study reduced approximately 90% of all emojis into just four emotion categories: happy, sad/unhappy, undecided/skeptical, and surprise/shock (Wolny, 2016). The results suggested that many different emojis can be used interchangeably to communicate an emotion. In a more recent and even larger study, Felbo et al. (2017) conducted a sentiment analysis on 1,246 million tweets containing one of 64 common emojis. They examined emoji occurrences to learn sentiment, emotion, and sarcasm. They found that emoji use was structured by a combination of linguistic and social contexts, as well as cultural convention.

Only a handful of empirical research has investigated the relationship between perceived emotion category and emoji faces, however (Oleszkiewicz et al., 2017; Betz et al., 2019; Franco and Fugate, 2020). Oleszkiewicz et al. (2017) asked children to view real human and emoji faces and identify the emotion. Children assigned human faces and emoji faces with high agreement to the categories "happy" and "sad," yet there was only low agreement for the other basic emotions (e.g., fear, anger, surprise, and disgust) for both emoji and human faces. Another study (Betz et al., 2019) found that emotion words served as a context for perceiving emotions from the Finch faces, emoji-like faces created by Pixar illustrator, Matt Jones (Jones, 2017). This particular set of emoji faces was created based on Darwin's depictions of basic emotional "expressions" in man and animals (Darwin, 1872/2005). Despite these faces being created to specifically exemplify emotional "expressions," participants had overall low agreement about which emotion was displayed unless they were forced to choose from a provided emotion word. In our previous work, we found only about half of the emojis explored were assigned at statistically higher rates to one emotion category compared to another, and less than one sixth of the faces were specific to an emotion category (meaning that they were not also affiliated with another emotion at similar levels) (Franco and Fugate, 2020).

A handful of studies have shown that emoji rendering differences among electronic platforms may lead to miscommunication and misinterpretation (Miller et al., 2016; Tigwell and Flatla, 2016; Miller Hillberg et al., 2018; Rodrigues et al., 2018). Miller Hillberg et al. (2018) found that 25% of Twitter users were unaware that emojis' appearances change depending on a user's electronic platform. Additionally, 20% of users reported that they would edit their emoji selection or tweet after being shown rendering differences. And, in our previous research, we found that there were significant differences in what emotions people associated most intensely with an emoji face, depending on the electronic platform they viewed them on (Franco and Fugate, 2020).

In another study assessing electronic platform differences, users evaluated a randomized subset of 20 emoji faces on two platforms for their esthetic appeal, familiarity, visual complexity, concreteness, valence, arousal, and meaningfulness (Rodrigues et al., 2018). Users also provided a free response as to what they thought the emoji meant or what emotion they thought it represented. Although the individual free responses were not provided for each emoji in the article, overall agreement (after coding for similarities) of responses was slightly greater for iOS emojis (66.78%) than for the same emojis on Android (64.95%). Moreover, iOS ratings on "meaningfulness amount" were statistically higher for iOS (when bootstrapped) compared to those for Google Android.

The Current Study

In this paper, we first adapted the Facial Action Coding System (FACS) to systematically compare emoji faces with respect to anatomical-based changes (AUs). We compared 31 emojis (spanning ten emotions) on their appearance across three electronic platform carriers, each with two different version updates (Apple iOS 9.1, Apple iOS 13.3, Google Android 6.0, Google Android 10.0, Samsung TouchWiz 5.1, and Samsung One UI 1.5) (see **Table 2**). The creation of a coding rubric for schematic faces is essential in order to compare anatomically and reliably the renderings of an emoji across platforms and versions. Therefore, our goal in creating such a rubric was to be

[1] We thank a reviewer for pointing this out.

TABLE 2 | Emojis for both Apple Platform Versions with FACS Code and Perceived Emotion (Additional Platforms Below).

Updated unicode name (8/2020)	Initial unicode name (9/2017)	Apple iOS 13.3	FACS Code Apple iOS 13.3	Apple iOS 9.1	FACS Code Apple iOS 13.3	Emotion most frequently perceived (Franco and Fugate)	ICP prototype code (Cordaro et al., 2018)
FACE WITH TEARS OF JOY	FACE WITH TEARS OF JOY	😂	1 + 12 + 25 + 26 + 63 + crying	😂	1 + 12 + 25 + 26 + 63 + crying	*Happy*	*6 + 7 + 12 + 16 + 25 + 26*
SMILING FACE	WHITE SMILING FACE	☺️	6 + 12 + 25 + 63	☺️	6 + 12 + 25 +63	Happy	*6 + 7 + 12 + 16 + 25 + 26*
LOUDLY CRYING FACE	LOUDLY CRYING FACE	😭	1 + 25 + 26 + 63 + crying	😭	1 + 25 + 26 + 63 + crying	*Sad*	*4 + 43 + 54*
SMILING FACE WITH SMILING EYES	SMILING FACE WITH SMILING EYES	😊	6 + 12 + 25 + 63	😊	6 + 12 + 25 + 63	*Happy*	*6 + 7 + 12 + 16 + 25 + 26*
GRINNING FACE WITH SMILING EYES	SMILING FACE WITH OPEN MOUTH AND SMILING EYES	😄	12 + 25 + 26 + 63	😄	12 + 25 + 26 + 63	*Happy*	*6 + 7 + 12 + 16 + 25 + 26*
GRIMACING FACE	GRIMACING FACE	😬	20 + 25 + 26	😬	20 + 25 + 26	***Fear***	*1 + 2 + 5 + 7 + 25*
WEARY FACE	WEARY FACE	😩	1 + 4 + 15 + 25 + 26 + 64	😩	1 + 4 + 15 + 25 + 26 + 64	***Envy***	–
SMIRKING FACE	SMIRKING FACE	😏	12 + 61	😏	12 + 61	***Love***	*6 + 7 + 12 + 25*
WINKING FACE	WINKING FACE	😉	1 + 12 + 25 + 46	😉	1 + 12 + 25 + 46	*Love*	*6 + 7 + 12 + 25*
BEAMING FACE WITH SMILING EYES	GRINNING FACE WITH SMILING EYES	😁	12 + 25 + 26 + 63	😁	20 + 25 + 26 + 63	***Contempt***	*4 + 14 + 25*
UNAMUSED FACE	UNAMUSED FACE	😒	15 + 25 + 61	😒	15 + 25 + 61	*Envy*	–
GRINNING FACE WITH BIG EYES	SMILING FACE WITH OPEN MOUTH	😃	5 + 12 + 25 + 26	😃	5 + 12 + 25 + 26	***Happy***	*6 + 7 + 12 + 16 + 25 + 26*
PENSIVE FACE	PENSIVE FACE	😔	1 + 4 + 64	😔	1 + 4 + 64	*Sad*	*4 + 43 + 54*
FLUSHED FACE	FLUSHED FACE	😳	1 + 2 + 5 + 6	😳	1 + 2 + 5 + 6	***Surprise***	*1 + 2 + 5 + 25*
CRYING FACE	CRYING FACE	😢	1 + 15 + 25 + crying	😢	1 + 15 + 25 + crying	*Sad*	*4 + 43 + 54*
RELIEVED FACE	RELIEVED FACE	😌	1 + 12 + 25 + 64	😌	1 + 12 + 25 + 64	***Calm***	*12 + 43*
DISAPPOINTED FACE	DISAPPOINTED FACE	😞	15 + 25 + 64	😞	15 + 25 + 64	***Sad***	*4 + 43 + 54*
BEAMING SQUINTING FACE	SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES	😆	12 + 25 + 26 + 43	😆	12 + 25 + 26 + 43	***Happy***	*6 + 7 + 12 + 16 + 25 + 26*
GRINNING FACE	GRINNING FACE	😀	12 + 25 + 26	😀	12 + 25 + 26	***Happy***	*6 + 7 + 12 + 16 + 25 + 26*
CONFUSED FACE	CONFUSED FACE	😕	15	😕	15	***Sad***	*4 + 43 + 54*
EXPRESSIONLESS FACE	EXPRESSIONLESS FACE	😑	7 + 20	😑	7 + 20	***Anger***	*4 + 7*

(Continued)

TABLE 2 | Continued

Updated unicode name (8/2020)	Initial unicode name (9/2017)	Apple iOS 13.3	FACS Code Apple iOS 13.3	Apple iOS 9.1	FACS Code Apple iOS 13.3	Emotion most frequently perceived (Franco and Fugate)	ICP prototype code (Cordaro et al., 2018)
ANGRY FACE	ANGRY FACE		4 + 15 + 25		4 + 15 + 25	*Anger*	*4 + 7*
PERSEVERING FACE	PERSEVERING FACE		1 + 4 + 15 + 25 + 43		1 + 4 + 15 + 25 + 43	***Disgust***	*4 + 6 + 7 + 9+10 +25 + 26*
NEUTRAL FACE	NEUTRAL FACE		20		20	***Calm***	*12 + 43*
CONFOUNDED FACE	CONFOUNDED FACE		23 + 25 + 43		1 + 4 + 23 + 25 + 43	***Disgust***	*4 + 6 + 7 + 9+10 +25 + 26*
FACE WITHOUT MOUTH	FACE WITHOUT MOUTH		–		–	*Surprise*	*1 + 2 + 5 + 25*
FACE WITH OPEN MOUTH	FACE WITH OPEN MOUTH		25 + 26		25 + 26	*Surprise*	*1 + 2 + 5 + 25*
WORRIED FACE	WORRIED FACE		1 + 2 + 15 + 17 + 25		1 + 2 + 15 + 17 + 25	***Fear***	*1 + 2 + 5 + 7 + 25*
HUSHED FACE	HUSHED FACE		1 + 2 + 25 + 26		1 + 2 + 25 + 26	***Surprise***	*1 + 2 + 5 + 25*
FROWNING FACE WITH OPEN MOUTH	FROWNING FACE WITH OPEN MOUTH		15 + 25 + 26		15 + 25 + 26	***Fear***	*1 + 2 + 5 + 7 + 25*
KISSING FACE	KISSING FACE		18		6 + 18 + 63	*Love*	*6 + 7 + 12 + 25*

Updated unicode name (8/2020)	Initial unicode name (9/2017)	Google Android 10.0	FACS Code Google Android 10.0	Google Android 6.0	FACS Code Google Android 6.0	Emotion most frequently perceived (Franco and Fugate)	ICP prototype code (Cordaro et al., 2018)
FACE WITH TEARS OF JOY	FACE WITH TEARS OF JOY		1 + 12 + 25 + 26 + 63 + crying		12 + 25 + 63 + crying	*Happy*	6 + 7 + 12 + 16 + 25 + 26
SMILING FACE	WHITE SMILING FACE		6 + 12 + 25 + 63		12 + 25	***Calm***	12 + 43
LOUDLY CRYING FACE	LOUDLY CRYING FACE		1 + 15 + 25 + 26 + 63 + crying		10 + 15 + 25 + 26 + 43 + crying	*Sad*	4 + 43 + 54
SMILING FACE WITH SMILING EYES	SMILING FACE WITH SMILING EYES		6 + 12 + 25 + 63		6 + 12 + 63	*Happy*	6 + 7 + 12 + 16 + 25 + 26
GRINNING FACE WITH SMILING EYES	SMILING FACE WITH OPEN MOUTH AND SMILING EYES		12 + 25 + 26 + 63		12 + 25 + 63	*Happy*	6 + 7 + 12 + 16 + 25 + 26
GRIMACING FACE	GRIMACING FACE		20 + 25 + 26		20 + 25 + 26	***Fear***	1 + 2 + 5 + 7 + 25
WEARY FACE	WEARY FACE		1 + 4 + 15 + 25 + 26 + 64		10 + 15 + 25 + 26 + 64	***Sad***	4 + 43 + 54
SMIRKING FACE	SMIRKING FACE		1 + 12 + 61		12 + 61	***Contempt***	4 + 14 + 25
WINKING FACE	WINKING FACE		12 + 25 + 46		12 + 46	*Love*	6 + 7 + 12 + 25

(Continued)

TABLE 2 | Continued

Updated unicode name (8/2020)	Initial unicode name (9/2017)	Google Android 10.0	FACS Code Google Android 10.0	Google Android 6.0	FACS Code Google Android 6.0	Emotion most frequently perceived (Franco and Fugate)	ICP prototype code (Cordaro et al., 2018)
BEAMING FACE WITH SMILING EYES	GRINNING FACE WITH SMILING EYES		12 + 25 + 26 + 63		6 + 12 + 25 + 26 + 63	***Happy***	6 + 7 + 12 + 16 + 25 + 26
UNAMUSED FACE	UNAMUSED FACE		1 + 15 + 25 + 61		15 + 25 + 26 + 61	*Envy*	–
GRINNING FACE WITH BIG EYES	SMILING FACE WITH OPEN MOUTH		5 + 12 + 25 + 26		12 + 25 + 26	***Calm***	12 + 43
PENSIVE FACE	PENSIVE FACE		1 + 4 + 64		15 + 64	*Sad*	4 + 43 + 54
FLUSHED FACE	FLUSHED FACE		1 + 2 + 5 + 6		6	***Calm***	12 + 43
CRYING FACE	CRYING FACE		1 + 15 + 25 + crying		15 + 64 + crying	*Sad*	4 + 43 + 54
RELIEVED FACE	RELIEVED FACE		1 + 12 + 25 + 64		12 + 63 + sweating	***Fear***	1 + 2 + 5 + 7 + 25
DISAPPOINTED FACE	DISAPPOINTED FACE		15 + 25 + 64		15 + 25 + 64	***Sad***	4 + 43 + 54
BEAMING SQUINTING FACE	SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES		12 + 25 + 26 + 43		6 + 12 + 25 + 26 + 43	***Love***	6 + 7 + 12 + 25
GRINNING FACE	GRINNING FACE		12 + 25 + 26		12 + 25	***Happy***	6 + 7 + 12 + 16 + 25 + 26
CONFUSED FACE	CONFUSED FACE		1 + 2 + 4 + 15 + 25		4 + 20 + 15 + 64	***Disgust***	4 + 6 + 7 + 9 + 10 + 25 + 26
EXPRESSIONLESS FACE	EXPRESSIONLESS FACE		7 + 20		7	***Calm***	12+ 43
ANGRY FACE	ANGRY FACE		4 + 15		4 + 15 + 64	*Anger*	4 + 7
PERSEVERING FACE	PERSEVERING FACE		1 + 4 + 15 + 25 + 43		6 + 15 + 43	***Anger***	4 + 7
NEUTRAL FACE	NEUTRAL FACE		20		12	***Calm***	12 + 43
CONFOUNDED FACE	CONFOUNDED FACE		23 + 43		10 + 64	***Sad***	4 + 43 + 54
FACE WITHOUT MOUTH	FACE WITHOUT MOUTH		–		–	*Surprise*	1 + 2 + 5 + 25
FACE WITH OPEN MOUTH	FACE WITH OPEN MOUTH		5 + 25 + 26		25 + 26	*Surprise*	1 + 2 + 5 + 25
WORRIED FACE	WORRIED FACE		1 + 4 + 15 + 25		4 + 15 + 64	***Envy***	–
HUSHED FACE	HUSHED FACE		1 + 4 + 15 + 25 + 26		1 + 4 + 25 + 26	***Surprise***	1 + 2 + 5 + 25
FROWNING FACE WITH OPEN MOUTH	FROWNING FACE WITH OPEN MOUTH		15 + 25 + 26		4 + 15 + 25 + 26 + 64	***Disgust***	4 + 6 + 7 + 9 + 10 + 25 + 26
KISSING FACE	KISSING FACE		18		18	*Love*	6 + 7 + 12 + 25

(Continued)

TABLE 2 | Continued

Updated unicode name (8/2020)	Initial unicode name (9/2017)	Samsung One UI 1.5	FACS Code Samsung One UI 1.5	Samsung TouchWiz 5.1	FACS Code Samsung TouchWiz 5.1	Emotion most frequently perceived (Franco and Fugate, 2020)	ICP Prototype code (Cordaro et al., 2018)
FACE WITH TEARS OF JOY	FACE WITH TEARS OF JOY		1 + 12 + 25 + 26 + 63 + crying		1 + 4 + 12 + 25 + 26 + 63 + crying	*Happy*	6 + 7 + 12 + 16 + 25 + 26
SMILING FACE	WHITE SMILING FACE		1 + 2 + 6 + 12 + 25 + 26 + 63		12	***Happy***	6 + 7 + 12 + 16 + 25 + 26
LOUDLY CRYING FACE	LOUDLY CRYING FACE		1 + 63 + 25 + 26 + crying		1 + 4 + 7 + 15 + 25 + 26 + crying	*Sad*	4 + 43 + 54
SMILING FACE WITH SMILING EYES	SMILING FACE WITH SMILING EYES		6 + 12 + 63		1 + 2 + 6 + 12 + 63	*Happy*	6 + 7 + 12 + 16 + 25 + 26
GRINNING FACE WITH SMILING EYES	SMILING FACE WITH OPEN MOUTH AND SMILING EYES		12 + 25 + 26 + 63		1 + 2 + 6 + 12 + 25 + 26 + 63	*Happy*	6 + 7 + 12 + 16 + 25 + 26
GRIMACING FACE	GRIMACING FACE		20 + 25 + 26		4 + 10 + 17 + 25 + 26 + 41	***Anger***	4 + 7
WEARY FACE	WEARY FACE		1 + 4 + 15 + 17 + 25 + 26 + 64		1 + 4 + 15 + 25 + 41	***Disgust***	4 + 6 + 7 + 9+10 +25 + 26
SMIRKING FACE	SMIRKING FACE		1 + 2 + 12 + 25 + 61		12 + 25 + 64	***Calm***	12 + 43
WINKING FACE	WINKING FACE		12 + 25 + 26 + 46		1 + 2 + 12 + 25 + 26 + 46	*Love*	6 + 7 + 12 + 25
BEAMING FACE WITH SMILING EYES	GRINNING FACE WITH SMILING EYES		12 + 25 + 26 + 63		1 + 2 + 6 + 12 + 25 + 26 + 63	***Happy***	6 + 7 + 12 + 16 + 25 + 26
UNAMUSED FACE	UNAMUSED FACE		15 + 25 + 61		20 + 61	*Envy*	–
GRINNING FACE WITH BIG EYES	SMILING FACE WITH OPEN MOUTH		12 + 25 + 26		1 + 2 + 12 + 25 + 26	***Happy***	6 + 7 + 12 + 16 + 25 + 26
PENSIVE FACE	PENSIVE FACE		1 + 4 + 64		1 + 4 + 15 + 62	*Sad*	4 + 43 + 54
FLUSHED FACE	FLUSHED FACE		1 + 2 + 5 + 6		1 + 4 + 15 + 25 + 26	***Sad***	4 + 43 + 54
CRYING FACE	CRYING FACE		1 + 15 + 17 + 25 + 26 + crying		1 + 4 + 20 + 25 + crying	*Sad*	crying
RELIEVED FACE	RELIEVED FACE		1 + 2 + 12 + 64		1 + 2 + 12 + 64	***Calm***	12 + 43

(Continued)

TABLE 2 | Continued

Updated unicode name (8/2020)	Initial unicode name (9/2017)	Samsung One UI 1.5	FACS Code Samsung One UI 1.5	Samsung TouchWiz 5.1	FACS Code Samsung TouchWiz 5.1	Emotion most frequently perceived (Franco and Fugate, 2020)	ICP Prototype code (Cordaro et al., 2018)
DISAPPOINTED FACE	DISAPPOINTED FACE		15 + 25 + 64		1 + 2 + 15 + 61	***Envy***	–
BEAMING SQUINTING FACE	SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES		12 + 25 + 26 + 43		1 + 2 + 12 + 25 + 63	*Happy*	6 + 7 + 12 + 16 + 25 + 26
GRINNING FACE	GRINNING FACE		12 + 25 + 26		6 + 12 + 25 + 26	***Surprise***	1 + 2 + 5 + 25
CONFUSED FACE	CONFUSED FACE		15		1 + 15	***Fear***	1 + 2 + 5 + 7 + 25
EXPRESSIONLESS FACE	EXPRESSIONLESS FACE		7 + 20		7 + 20	***Contempt***	4 + 14 + 25
ANGRY FACE	ANGRY FACE		4 + 15		4 + 15 + 17 + 25	*Anger*	4 + 7
PERSEVERING FACE	PERSEVERING FACE		1 + 4 + 15 + 25 + 43		1 + 4 + 15 + 17 + 25 + 26 + 43	***Fear***	1 + 2 + 5 + 7 + 25
NEUTRAL FACE	NEUTRAL FACE		20		20	***Contempt***	4 + 14 + 25
CONFOUNDED FACE	CONFOUNDED FACE		10 + 43		23 + 43	***Disgust***	4 + 43 + 54
FACE WITHOUT MOUTH	FACE WITHOUT MOUTH		nothing		1 + 2 + 6	*Surprise*	1 + 2 + 5 + 25
FACE WITH OPEN MOUTH	FACE WITH OPEN MOUTH		25 + 26		25 + 26	*Surprise*	1 + 2 + 5 + 25
WORRIED FACE	WORRIED FACE		1 + 4 + 15 + 25 + 26 + 17		1 + 4 + 15 + 17 + 25 + crying	***Sad***	4 + 43 + 54
HUSHED FACE	HUSHED FACE		1 + 2 + 25 + 26		1 + 4 + 25 + 26	***Fear***	1 + 2 + 5 + 7 + 25
FROWNING FACE WITH OPEN MOUTH	FROWNING FACE WITH OPEN MOUTH		15 + 17 + 25 + 26		1 + 4 + 20 + 25 + 26	*Fear*	1 + 2 + 5 + 7 + 25
KISSING FACE	KISSING FACE		18		1 + 4 + 6 + 22 + 25 + 64	*Love*	6 + 7 + 12 + 25

Bolded emotions are those which were perceived as different across platforms.

able to use it to address two fundamental questions about the relationship between the physical renderings of emoji faces on different among platforms (and versions), and their relationship to human facial expressions.

The first goal was to assess objectively *physical appearance*. Specifically, are the anatomical-based changes (AUs) of an emoji face the same across electronic platforms and version updates? Said another way: *Anatomically, does the same emoji face "look" the same across platforms and versions?* After using our adaptive emoji-FACS rubric to code each emoji face, we then systematically compared the distribution and frequencies of AUs across emoji faces by platforms and versions.

> ***Hypothesis 1a:*** Action units (as an objective measure of facial coding) should reflect the known perceptual differences users encounter when an emoji is sent from another platform. For the same set of emojis, AU counts and distributions should differ among platforms/versions.

> ***Hypothesis 1b:*** If emoji faces represent facial "expressions," those faces perceived as the same emotion across platforms should be more similar in AU counts and distributions compared to those which are perceived as different emotions.

The second goal was to assess *emotional meaning*. Specifically, do the anatomical-based changes of an emoji face reflect those proposed for human facial depictions of emotion? Said another way: *Do emoji faces perceived as an emotion category resemble human facial depictions of the same emotion category?* To assess this goal, we compared the AUs we coded for emoji faces to the the prototypical AUs (ICPs) described for facial expressions (according to the literature, see Cordaro et al., 2018).

> ***Hypothesis 2a:*** If emoji faces resemble human facial "expressions," then there should be a high correspondence among AUs for an emoji face (perceived as an emotion) and the human facial depiction for that emotion.

> ***Hypothesis 2b:*** If emoji faces resemble human facial "expressions," then the AUs should significantly predict (classify) the perceived emotion category.

METHODS

Stimuli Sets of Emojis

We began by using the 31 emojis from Apple iOS 9.1 (hereafter called Apple 9.1), Google Android 6.0, and Samsung TouchWiz 5.1 (hereafter called Samsung Wiz) that were identified as belonging to ten different emotion categories in Franco and Fugate (2020). The emotions investigated in that paper were ten of those listed as being basic emotions, and included anger, calm (called contentment according to Cordaro et al., 2018), contempt, fear, envy[2], disgust, happiness, love (called sex/desire according to Cordaro et al., 2018), sadness, and surprise. All emojis were represented in the Unicode Standard System (see **Table 2** for Unicode name). We then added the equivalent, most up-to-date (at the time this project began) emojis from each of these platforms (also listed in **Table 2**). Therefore, we used 31 emojis, which were represented on each of two versions for the three platforms (e.g., Apple 9.1, Apple iOS 13.3 (hereafter Apple 13.3); Google Android 6.0, Google Android 10.0; and Samsung Wiz, Samsung One UI 1.5 (hereafter Samsung One) (Emojipedia, 2020). All emoji face names are referred to by the newer, updated Unicode name.

Coding of AUs

Both coders were certified FACS-coders, with over 25 years of combined experience, who completed their training with Erika Rosenberg and used the FACS Investigator Guide to code[3].

The first author set some initial guidelines as to what was considered "baseline" for schematic faces. The initial guidelines included the following marks as "baseline": (1) circle eyes, as long as not oval or extra-large; (2) straight line mouths, as long as not elongated; (3) straight line eyebrows (when present; not all emojis have eyebrows and marks were only considered eyebrows if there was also an eye). The second coder agreed to these assumptions. Both coders agreed to not code intensities of AUs or to code head movements or miscellaneous codes[4]. Both coders initially coded unilateral movements, but later dropped right and left designations in the final codes for simplicity[5].

Both coders independently came up with a list of AUs that they could conceivably code. This included 25 AUs (in chronological order): 1 (inner brow raise), 2 (outer brow raise), 4 (brow lowerer), 5 (upper lid raise), 6 (cheek raiser)[6], 7 (lid tightener), 10 (upper lip raiser), 12 (lip corner puller), 14 (dimpler), 15 (lip corner depressor), 16 (lower lip depressor), 17 (chin raiser), 18 (lip pucker), 20 (lip stretch), 22 (lip funneler), 23 (lip tightener), 25 (lip part), 26 (jaw drop), 41 (lid droop), 43 (eyes closed), 46 (wink), 61 (eyes left), 62 (eyes right), 63 (eyes up), and 64 (eyes down).

Both coders then produced a depiction(s) of each AU and sent it to one another. Together, they combined different variations for each AU. There was some initial debate over AU 10, 23, and 22. Renderings for all three of these AUs were agreed upon after discussion (see **Table 3** for final depictions). After discussion, the two coders came to agreement through conversation, and eventually both used **Table 3** as the final coding rubric.

AUs in which neither coder could conceive of what it might look like were not included in the rubric. These included: AU 9 (nose wrinkle; no noses in emojis), 13 (sharp lip puller; unable to distinguish from AU 12 or 14), 24 (lip press; unable to distinguish from AU 23), 27 (mouth stretch; unable to distinguish between open mouth, AU 25 and AU 26), 28 (lip

[2] Envy does not have an ICP.

[3] face-and-emotion.com/dataface/facs/guide/FACSIV1.html

[4] AU 54 is one of the ICP prototype AUs.

[5] Only a small percentage (less than 5%) of total codes contained R/L.

[6] Indicated with cheek blushing, no cheeks otherwise.

TABLE 3 | Coding Rubric.

Eyes right = 62	
Eyes left = 61	
Eyes looking up (soft curve) = 63	
Eyes wide = 5	
Eyes looking down (soft curve) = 64	
Eye lid drop = 41	
Cheek raiser = 6	*If rosy coloration to cheeks added*
Eyes slit (any angle or flat) = 7 eyes any slit any distance apart (code eyebrows separately)	
Eyebrows up only interior (inner brow raise) = 1 (any thickness of brow)	
Eyebrows up arched = 1 + 2 (any thickness of brow)	
Eyebrows up with brow lower = 1 +4 (any thickness of brow and any distance apart)	
Brow lower = 4	(code eyebrows separately)
Dimpler = 14	
Arched down mouth open = 15 + 25	
Arched down mouth = 15	
Arched up mouth open = 12 + 25	
Arched up mouth = 12	
Arched down open mouth with depressor = 15 + 16 + 25	
Mouth pulled straight across (elongated) = 20	
Eyes closed = 43	> < **x x**
Wink = 46	> or < one eye open and one slanted
Chin Raiser = 17	or if under mouth (code mouth separately)
Lip Tightener = 23	
Upper Lip Raiser = 10	
Lip pucker = 18	**} or {**
Lip funneler = 22 + 25	
Circle mouth or open mouth space (with or without teeth) = 25 + 26	

The coders looked through the basic set of AU codes (excluding miscellaneous and head positions) to see which AUs they both could conceivably imagine what that AU might look like in a schematic form. This included 25 AUs (in chronological order): 1, 2, 4, 5, 6, 7, 10, 12, 14 (none noted), 15, 16 (none noted), 17, 18, 20, 22, 23, 25, 26, 41, 43, 46, 61, 62, 63, 64. Of the AUs indicated for the ICP emotion prototypes for the nine (note envy does not have a code) emotions investigated, only AU 9 did not have a code, and AU 54 was not included because no head positions were included. AUs in which neither coded could conceive of what it might look like, were not included: AU 9 (nose wrinkle; no noses in emojis), 13 (sharp lip puller), 24 (lip press, unable to distinguish might be how different from 23), 27 (mouth stretch), 28 (lip suck), 45 (blink), 65 (walleye), and 66 (Crosseye). "Absence" codes were not used [AU 70,71 (brows and eyes not visible, respectively), and 72 (lower face not visible)]. Circle eyes are considered "normal," whereas straight line (un-elongated) mouths are considered "normal." Eyebrows only assumed if eye also present. Straight line eyebrows are considered normal. Straight line mouth (not elongated) is considered "normal." Head positions and miscellaneous codes not included.

suck; unable to imagine), 45 (blink; no movement), 65 (walleye; unable to imagine), and 66 (crosseye; unable to imagine). "Absence" codes were not used (AU 70, 71) (brows and eyes not visible, respectively), and 72 (lower face not visible). Of the AUs indicated for the ICP emotion prototypes for the nine emotions investigated, only AU 9 and AU 54 were not included in the rubric.

Finally, both coders noted that there were two additional "embellishments" that were seen regularly on emoji faces and might be important to code: this include a tear (which was called the *crying code)* and a "tear" but alongside the upper face (not eye) (which was called the *sweating code)*. Although there is no AU for crying or sweating in FACS, tears and sweat have been proposed as possible emotional outputs.

Reliability

Each coder first coded ten random emoji faces (from different platforms and versions). The first coder compared the AUs between her and the other coder. Reliability was greater than 89%, and the coders resolved any disagreements, which resulted in 100% agreement on the final code for the first ten emojis in the file.

Each coder then used the rubric to code the rest of the emoji faces, which were presented randomly by platform, one emoji per page in a file. There were two files total, which divided the earlier version from the later version. The first author then calculated the reliability for each emoji face, for each platform, for each version. Reliability was calculated by scoring a "1" for any AUs indicated by only one coder and a "2" for any AUs agreed upon by both coders. The total number of AUs counted was then added. Finally, the summed count of AUs from the coders was divided by the AUs counted multiplied by two. The overall reliability between coders on Apple 9.1 was 75% (ranging from 67 to 100% across faces, $n = 9$ faces had perfect agreement); on Apple 13.3, 94% (ranging from 70% to 100%, $n = 19$ had perfect agreement); on Google Android 6.0, 88% (ranging from 50% to 100%, $n = 12$ had perfect agreement); on Google Android 10.0, 96% (ranging from 83% to 100%, $n = 22$ had perfect agreement); on Samsung Wiz, 88% (ranging from 67% to 100%, $n = 10$ had perfect agreement); and finally on Samsung One, 93% (ranging from 63 to 100%, $n = 20$ had perfect agreement). The overall reliability between the coders across platforms and versions was 91%. In cases in which the codes did not match, the first author made the final decision and included it as the "final code" in **Table 2**. The second coder approved the final codes.

RESULTS

Overall Use of AUs

Twenty seven coded AUs (including *crying* and *sweating*) were identified on the coding rubric. **Table 4** shows the percentage of time each AU was coded across all platforms/versions. AU 14 and AU 16 were never coded in any emoji face. Statistical significance was conducted with an alpha of .05 two-tailed, unless indicated otherwise.

Analysis 1a: Counts and Distribution of AUs

Across All Platforms and Versions

Both Apple 9.1 and 13.3 used 20 of the 25 AUs across emoji faces. Both platforms did not use AUs 10, 22, 41, 62, or the sweating code (**Table 4**).

Google Android 6.0. used 18 AUs, and Google Android 10.0 version used 19 AUs. Neither version used AU 17, AU 22, AU 41, or AU 62. Google Android 6.0 also did not use AU2, AU 5, or AU 23, whereas Google Android 10.0 also did not use AU10 or *sweating*.

Samsung Wiz used 22 AUs and Samsung One used 20 AUs. Neither used *sweating*. Samsung Wiz did not use AU5 and AU 18, whereas Samsung One did not use AU 22, AU 23, AU 41, and AU 62.

To assess Hypothesis 1a overall, we compared the overall AU count from the 25 AUs for which we had data across the three platforms and versions.

Because our data was not normally distributed, we used a Kruskal-Wallis test for both the overall AU count and AU distribution. For the overall AU count, there was a significant difference among the platforms/versions, $H(5) = 11.844$, $p < 0.05$ (Mean rank Apple 9.1 = 93.76; Apple 13.3 = 89.27; Google Android 6.0 = 74.68; Google Android 10.0 = 94.42; Samsung Wiz = 119.50; Samsung One = 89.37). When controlling for multiple comparisons (Bonferroni), only Google Android 6.0 and Samsung Wiz differed statistically from each other, $U(2) = -44.823$, $p < 0.05$.

To test differences in the distribution of individual AUs, we again performed a Kruskal-Wallis test on each AU. Three AUs differed statistically across platforms/versions. The first was AU 1, $H(5) = 27.980$, $p < 0.05$, in which Google Android 10.0 and Samsung Wiz differed statistically (controlling for multiple comparisons) ($p < 0.05$). AU 2 also differed, $H(5) = 15.157$, $p < 0.05$, in which Google Android 6.0 and Samsung Wiz differed (controlling for multiple comparisons) ($p < 0.05$). AU 10 also differed across platforms/versions, $H(5) = 12.333$, $p < 0.05$, but no follow-up comparisons remained significant after controlling for multiple comparisons. The distribution of AU 17 was marginally significant across platforms/versions, $H(5) = 10.932$, $p = 0.053$.

Individual Platforms

We next investigated whether the AUs differed between the older and newer versions of emojis for each platform.

Apple Versions

Between the two versions of Apple (9.1 and 13.3), there were very few obvious physical differences between the corresponding emojis. The one exception was the original "beaming face with smiling eyes," which was replaced with the "grinning face with smiling eyes," and was physically quite different (see **Table 2**).

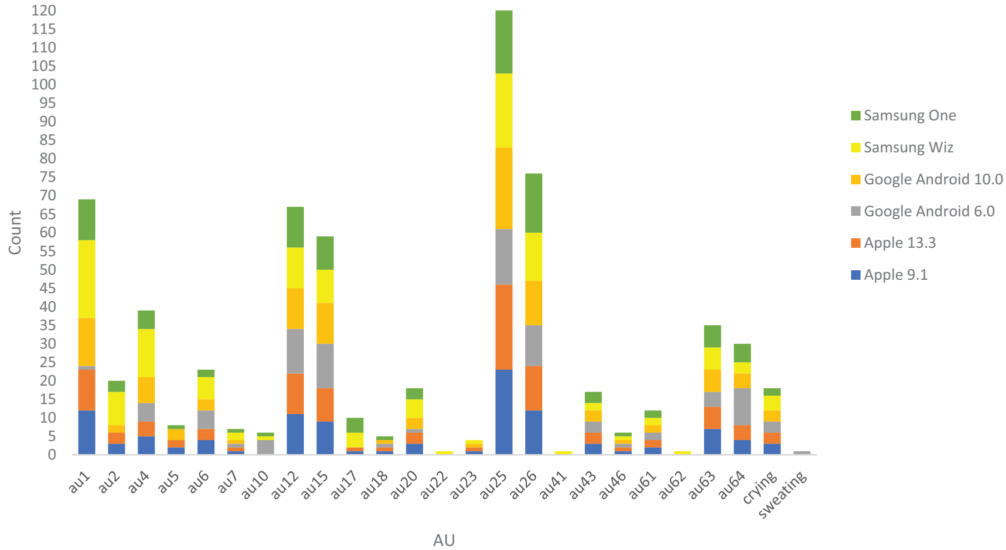
There was no difference between the Apple versions on overall count of AUs, $U(2) = 455.5$, $p > 0.05$. None of the individual AUs between Apple versions were significant either.

Android Versions

The majority of emojis between the two versions of Google Android (6.0. and 10.0) were noticeably different from just looking at them. Most apparent was that the gum-drop shaped head of the original version was replaced with the more standard circle head. Thus, version 10.0 appeared more similar to the other platforms. In addition, the newer version used the yellow-orange color variation of faces seen in the other platforms and versions. The large red "blob mouths" were replaced with lines, again converging with the other platforms and versions.

The difference between the overall AU count, however, was not statistically significant, $U(2) = 591.00$, $p > 0.05$. AU 1 and AU 10

TABLE 4 | AU counts by Platform/Versions.



differed significantly between the two versions, $U(2) = 666.50$ and $U(2) = 418.5$, $ps < 0.05$.

Samsung Versions

About half of the emojis looked noticeably different between the two versions of Samsung (Wiz and One) (see **Table 2**). Specifically, the newer version had emojis looking straight on, whereas the older version had several emojis with head turns and tilts.

The difference between the overall AU count was statistically significant, $U(5) = 336.00$, $p < 0.05$. AU 1 and AU 4 differed between the two versions, $U(2) = 325.50$ and $U(2) = 356.5$, $ps < 0.05$. AU 2 was marginally significant, $U(2) = 387.5$, $p = 0.056$.

Summary

With respect to Hypothesis 1a, some platforms and versions had different total AU counts and different distributions of AUs. Overall, Google Android 6.0 had the fewest countable AUs ($n = 19$), even though it only had only slightly fewer AUs than most of the other platforms.

Between the two versions of Apple, there was no difference in the overall AU count or distribution of AUs overall. Between the two versions of Google Android, although there were no significant differences across overall AU count, there were differences in the counts for two individual AUs. Finally, there was a significant difference for both overall AU count between versions of Samsung, and three individual AUs differed statistically (or marginally so).

For a more detailed look of AU correlations between versions/platforms by individual emoji face, we refer the reader to **Table 5** which lists the correlation coefficients (based on Spearman's rho).

Analysis 1b: Correlation Across AUs for Faces Perceived as the Same Emotion vs. Different Emotion(s)

Twelve emoji faces were perceived as the same emotion across all three platforms. Twelve faces were perceived as a different emotion on one of the three platforms (i.e., two platforms shared a perceived emotion). Seven additional faces were perceived as a different emotion on *each of the three* platforms (see **Table 2**).

To test Hypothesis 1b, we compared the overall AU count and distribution of AUs across those emoji faces that were perceived as the same emotion ($n = 35$) vs. those perceived differently ($n = 58$) (on at least one other platform).

For the overall AU count, there was not a significant difference between the AUs for same- and differently-perceived emotions using a Mann Whitney U-test, $U(1) = 0.011$, $p > 0.05$. It is also worth noting that of the emoji faces perceived as the same emotion across platforms, 50.0% had a correlation among AUs exceeding 75%. The rate was barely less (45%) for emoji faces perceived as different emotions.

The distribution of AUs between same- and differently-perceived emotions was significantly different for three AUs, however: AU 46, $U(1) = 928.0$; AU 63, $U(1) = 847.5$; *crying*, $U(1) = 789.0$, $ps < 0.05$.

Summary

Overall, Hypothesis 1b was not supported: the distribution of AUs was not statistically different among faces perceived as the same emotion compared to faces perceived as a different emotion(s). Three AUs were significantly different between same- and differently-perceived emotions, however, suggesting that there are some AUs that might be helpful in distinguishing certain emotions from others (thus increasing the agreement of emotion category perception).

Note that of the 12 emoji faces which were perceived differently on one platform (but the same on the two others),

TABLE 5 | Correlation Coefficients for each Emoji Face among all AUs by Platforms and Versions.

Unicode Name (8/2020)	Apple 9.1 and Apple 13.3	Google Android 6.0 and Google Android 10.0	Samsung Wiz and Samsung One	Apple 9.1 and Google Android 6.0	Apple 9.1 and Samsung Wiz	Google Android 6.0 and Samsung Wiz	Faces Perceived as Same Emotion Across Platforms	Faces Perceived as Different Emotion Across Platforms
	Between versions of the same platform			Between older versions of each platform (for which perception data exists) (see right columns)				
Angry Face	1.0	0.700	0.677	0.623	**0.874**	0.513	Angry	–
Beaming Face with Smiling Eyes	1.0	0.874	0.703	**0.874**	0.703	*0.804*	–	Contempt: A – Happy: G & S
Beaming Squinting Face	1.0	0.874	0.333	**0.874**	*0.333*	**0.257**	–	Happy: A & S – Love: G
Confounded Face	0.740	−0.803	0.458	−0.141	**0.592**	−0.083	–	Disgust: A – Sad: G & S
Confused Face	1.0	0.333	0.693	0.469	**0.693**	0.277	–	Sad: A – Disgust: G – Fear: S
Crying Face	1.0	0.513	0.196	0.513	**0.603**	0.129	Sad	–
Disappointed Face	1.0	1.0	0.180	***1.0***	0.180	0.180	–	Sad: A & G – Envy: S
Expressionless Face	1.0	0.693	1.0	0.693	**1.0**	0.693	–	Anger: A – Calm: G – Contempt: S
Face with Open Mouth	1.0	0.799	1.0	**1.0**	**1.0**	**1.0**	Surprise	–
Face with Tears of Joy	1.0	0.778	0.902	0.778	**0.902**	0.703	Happy	–
Face without Mouth (no AUs coded)	NA	NA	NA	NA	NA	NA	Surprise	–
Flushed Face	1.0	0.469	0.129	**0.469**	0.062	−0.098	–	Surprise: A – Calm: G – Sad: S
Frowning Face with Open Mouth	1.0	0.740	0.333	**0.740**	*0.435*	0.505	–	Fear: A & S – Disgust: G
Grimacing Face	1.0	1.0	0.374	***0.435***	0.374	0.428	–	Fear: A & G – Angry: S
Grinning Face	1.0	0.799	0.847	*0.799*	**0.847**	0.677	–	Happy: A & G – Surprise: S
Grinning Face with Big Eyes	1.0	0.847	0.740	**0.847**	*0.603*	0.740	–	Happy: A & S – Calm: G
Grinning Face with Smiling Eyes	1.0	0.833	0.677	**0.833**	0.677	0.778	Happy	–
Hushed Face	1.0	0.874	0.677	*0.705*	**1.0**	**1.0**	–	Surprise: A & G – Fear: S
Kissing Face	0.554	1.0	−0.110	**0.554**	0.088	−0.110	Love	–
Loudly Crying Face	1.0	0.513	0.196	0.428	0.584	0.491	Sad	–
Neutral Face	1.0	−0.040	1.0	*−0.040*	**1.0**	−0.040	–	Calm: A & G – Contempt: S
Pensive Face	1.0	0.847	0.740	0.348	**0.513**	0.277	Sad	–
Persevering Face	1.0	0.435	0.804	0.435	**0.804**	0.324	–	Disgust: A – Anger: G – Fear: S
Relieved Face	1.0	0.180	1.0	0.180	***0.409***	0.180	–	Calm: A & S – Fear: G
Smiling Face	1.0	0.677	0.365	0.677	*0.469*	**0.693**	–	Happy: A & S – Calm: G
Smiling Face with Smiling Eyes	1.0	0.847	0.703	**0.847**	0.703	0.595	Happy	–
Smirking Face	1.0	0.799	0.435	**1.0**	0.348	0.348	–	Love: A – Contempt: G – Calm: S
Unamused Face	1.0	0.705	0.348	**0.847**	0.348	0.277	Envy	–
Weary Face	1.0	0.659	0.584	**0.659**	**0.659**	0.257	–	Envy: A – Sad: G – Disgust: S
Winking Face	1.0	0.799	0.778	0.677	**0.778**	0.527	Love	–
Worried Face	1.0	0.513	0.738	0.129	**0.659**	0.374	–	Fear: A – Envy: G – Sad: S

For columns labeled "Between older versions of each platform," bold values represent the highest correlation coefficient between two platforms. For faces that were perceived as different emotions(s) among these platforms, the predicted highest correlation coefficient (based on shared perception) is in italics. For only the "relieved face" and the "disappointed face" was the prediction supported. A = Apple 9.1; G = Google Android 6.0; S = Samsung Wiz.

only two faces showed relatively lower correlations among AUs compared to those perceived as the same emotion (see **Table 5**).

Analysis Set 2a: Correspondence Between AUs for ICP Prototypes and Perceived Emotion

In our previous study, 228 English-speaking participants chose to which emotion category(ies) each of the 31 emojis belonged (Franco and Fugate, 2020). Participants randomly received all 31 emoji faces from either the Apple 9.1, Google Android 6.0, or Samsung Wiz platform. Emojis were shown individually for ten emotions (presented as words). Participants could indicate up to three emotion categories for each emoji face. Once an emotion category was selected, participants indicated the strength of that relationship on a 10-point Likert scale. Participants did not need to choose more than one emotion, but they needed to select at least one for each emoji face. For the purposes of this paper, we used the most frequent emotion category that participants indicated for each emoji face (for each of the three platforms). These results are also part of Table 2 in the Supplementary Files of that article Franco and Fugate (2020).

Table 6 presents the percentage of time each AU was used for each perceived emotion across platforms.

Across All Platforms and Versions

Of the 15 AUs identified for the ICP prototypes for the emotions we explored, we did not code for two: AU 9 and AU 54. Although we came up with a code for AU 14 and AU16, we never coded any instances of either. Therefore, we were able to compare codes on the 11 AUs common to the ICP prototypes and our emoji faces. We removed faces perceived as envy from these analyses, as there is no ICP prototype for envy.

To assess Hypothesis 2a, we used a Wilcoxon signed-ranks test to compare the distribution of AUs between the ICP prototype and our emoji faces. There was a significant difference using the Z transformation statistic, $Z(87) = -5.15$, $p < 0.05$ (mean rank ICP prototype AUs = 24.59, mean rank coded AUs = 39.32). Therefore, the distribution of AUs between the ICP prototypes overall and our coded AUs was different. Hypothesis 2a was not supported.

Individual Platforms

We next analyzed the distribution of these AUs by platform.

Apple

There was a significant difference between the distribution of AUs between the ICP prototype and our emoji faces, using the Z transformation statistic, $Z(29) = -3.92$, $p < 0.05$ (mean rank ICP prototype AUs = 5.0, mean rank coded AUs = 13.57). Thus, Hypothesis 2a was not supported on the Apple 9.1 platform.

Google Android

Between the ICP prototypes and our emoji faces on the Google Android platform, there was also a significant difference, $Z(29) = -3.67$, $p < 0.05$ (mean rank ICP prototype AUs = 8.13 and mean rank coded AUs = 14.48). Thus, Hypothesis 2a was not supported on the Google Android 6.0 platform.

Samsung

Lastly, between the ICP prototypes and our emoji faces on the Samsung platform, there was not a significant difference, $Z(29) = -0.859$, $p > 0.05$ (mean rank ICP prototype AUs = 10.17, mean rank coded AUs = 11.63). Therefore, only Samsung Wiz used AUs similarly to the ICP prototypes (across all emotions). Thus, overall, Hypothesis 2a was only supported for one platform.

Prototype AUs by Emotion

To further explore the correlation and importance of AUs for each emotion prototype as it related to the perceived emotion, we next separated the results by emotion. The following numbers represent how many emoji faces were perceived as each emotion (across the three platforms): anger ($n = 6$), calm ($n = 9$), contempt ($n = 4$), disgust ($n = 5$), envy ($n = 6$), fear ($n = 9$), happy ($n = 19$), love ($n = 8$), sad ($n = 17$), and surprise ($n = 10$) (see **Table 2**).

Anger

Only one face was perceived across the three platforms as anger: "angry face." Each platform had an additional face perceived as angry. The ICP prototype for anger is AU 4 and AU7 (see **Table 1**). A multinomial regression using AU 4 and AU 7 as predictors to obtain perceived emotion was significant, $X^2(18, n = 93) = 44.1$, $p = 0.001$, Nagelkerke = .382 (McFadden = 0.108). AU 4 was a significant predictor of perceived emotion overall, $X^2(9, n = 93) = 34.8$, $p < 0.001$, but not for anger. Only the absence of AU 4 predicted the emotion happy, $B = 3.59$ (SE = 1.37), Wald = 6.87, $p < 0.01$), and surprise, $B = 2.90$ (SE = 1.39), Wald = 4.35, $p < 0.05$. In fact, there were no classifications to anger using these two AUs (but see classification rate using all AUs, Hypothesis 2b below).

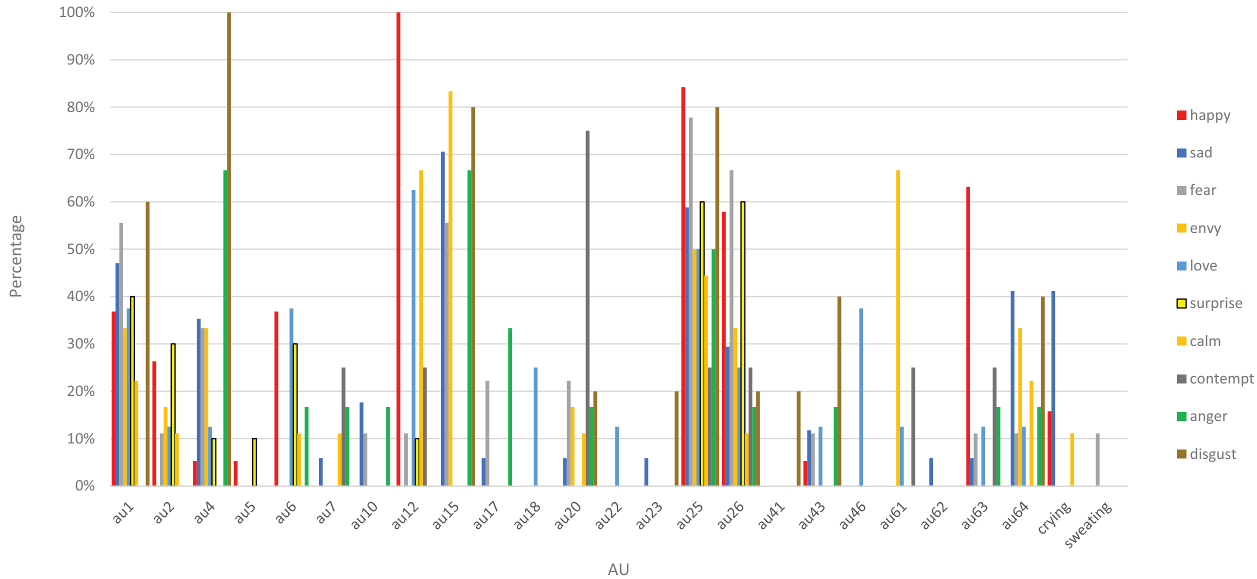
Calm

No faces were perceived as calm across the three platforms. Apple had two faces perceived as calm, Google Android had five faces, and Samsung had two faces. The ICP prototype for calm is AU 12 and AU 43 (see **Table 1**). A multinomial regression using AU 12 and AU 43 as predictors was significant, $X^2(18, n = 93) = 91.4$, $p < 0.001$, Nagelkerke = 0.634 (McFadden = 0.225). AU 12 was a significant predictor of perceived emotion overall, $X^2(9, n = 93) = 81.45$, $p < 0.001$, but not for calm. The absence of AU 12 significantly predicted surprise, $B = 2.89$ (SE = 1.27), Wald = 5.19, $p < 0.05$, and fear, $B = 2.90$ (SE = 1.33), Wald = 4.79, $p < 0.05$. There were no classifications to calm with these two AUs, however (but see classification rate using all AUs, Hypothesis 2b below).

Contempt

No faces were perceived as contempt across the three platforms. Apple and Google Android had one face perceived as contempt, and Samsung had two faces. The ICP prototype for contempt is AU 4, AU 14, and AU 25 (see **Table 1**). AU 14 was never coded. A multinomial regression using AU 4 and AU 25 as predictors was significant, $X^2(18, n = 93) = 46.73$, $p < 0.001$, Nagelkerke = 0.400 (McFadden = 0.115). As mentioned before, AU 4 was a significant predictor of emotion. AU 25 was not a

TABLE 6 | Percentage of AUs (as a total of number of AUs) by Emotion.



significant predictor or perceived emotion overall, although the presence of AU 25 significantly predicted happy, $B = -2.726$ (SE = 1.32), Wald = 4.29, $p < 0.05$. There were no classifications to contempt with these two AUs (but see classification rate using all AUs, Hypothesis 2b below).

Disgust

No faces were perceived as disgust across the three platforms. Apple, however, had two faces perceived as disgust; Google Android had two, and Samsung had one face perceived as disgust. The ICP prototype for disgust is AU 4, AU 6, AU 7, AU 9, AU 10, AU 25, and AU 26 (see **Table 1**). We did not have a code for AU 9. A multinomial regression using these six AUs as predictors was significant, $X^2(54, n = 93) = 108.57$, $p < 0.001$, Nagelkerke = 0.698 (McFadden = 0.267). In addition to AU 4 and AU 12, which were previously identified as significant predictors, AU 6, AU 10, and AU 26 were also now identified as significant predictors overall, $X^2(9, n = 93) = 19.9$, $p < 0.05$; $X^2(9, n = 93) = 17.45$, $p < 0.001$; $X^2(9, n = 93) = 26.23$, $p < 0.01$, respectively. None of the AUs significantly predicted any individual emotion, however, but the classification of disgust was 60% using these AUs (but see Hypothesis 2b).

Fear

No faces were perceived across the three platforms as fear, yet three faces on Apple, two on Google Android, and four on Samsung were perceived as fear. The ICP prototype for fear is AU 1, AU 2, AU 5, AU 7, and AU 25 (see **Table 1**). A multinomial regression using these five AUs as predictors was not significant, $X^2(45, n = 93) = 52.00$, $p > 0.05$, Nagelkerke = 0.434 (McFadden = 0.128). AU 1 and AU 2 were marginally significant predictors of perceived emotion overall, however, $X^2(9, n = 93) = 16.58$, $p = 0.056$, and $X^2(9, n = 93) = 16.59$, $p = 0.056$, respectively. None of the AUs significantly predicted any individual emotion, and there were no correct classifications to fear (but see Hypothesis 2b).

Happy

Only three emoji faces were perceived as happy across all the three platforms: "face with tears of joy," "smiling face with smiling eyes," and "grinning face with smiling eyes." An additional four emojis were perceived as happy on Apple, an additional two emoji faces on Google Android, and an additional four emoji faces on Samsung. The ICP prototype for happy is AU 6, AU 7, AU 12, AU 16, AU 25, and AU 26 (see **Table 1**). A multinomial regression using these six AUs as predictors was significant, $X^2(45, n = 93) = 134.65$, $p < 0.001$, Nagelkerke = 0.775 (McFadden = 0.331). AU 6, AU 12, and AU 26 were identified as significant predictors (as previously mentioned), although none of them predicted any individual emotion. Despite this, these AUs classified happiness 94.7% (the same as when all 11 coded AUs were added to the model, see Hypothesis 2b, below).

Love

Two faces were perceived across the three platforms as love: "winking face" and "kissing face." Both Apple and Google Android had one additional face perceived as love. The ICP prototype for love is AU 6, AU 7, AU 12, and AU 25 (see **Table 1**). A multinomial regression using these four AUs as predictors was significant, $X^2(36, n = 93) = 116.87$, $p < 0.001$, Nagelkerke = 0.725 (McFadden = 0.288). AU 6 and AU 12 had been previously identified as significant predictors, and maintained here. The absence of AU 12 predicted fear, $B = 3.11$ (SE = 1.42), Wald = 4.82, $p < 0.05$), and surprise, $B = 3.03$, (SE = 1.37), Wald = 4.93, $p < 0.05$. Zero percent of faces were classified to love (but see Hypothesis 2b).

Sad

Three faces were perceived as sad across the three platforms: "loudly crying face," "pensive face," and "crying face." Apple, however, had two additional faces perceived as sad, whereas Google Android and Samsung had an additional three faces each perceived as sad. The ICP prototype for sad is AU 4, AU 43, and AU 54 (see **Table 1**). We did not code for AU 54. A multinomial regression using these two AUs as predictors was significant, $X^2(18, n = 93) = 45.03$, $p < 0.001$, Nagelkerke = 0.389 (McFadden = 0.111). As previously indicated, AU 4 was a significant predictor of perceived emotion overall. The absence AU 4 significantly predicted happy, $B = 2.35$, (SE = 1.15), Wald = 4.19, $p < 0.05$. These AUs predicted sadness 47.1% (which was substantially lower than when all 11 AUs were included, see Hypothesis 2b).

Surprise

Two faces were perceived as surprise across the three platforms: "face without mouth" and "face with open mouth." Apple had two additional faces perceived as surprise, whereas Google Android and Samsung had one additional face each. The ICP prototype for surprise is AU 1, AU 2, AU 5, and AU 25 (see **Table 1**). A multinomial regression using these four AUs as predictors was not significant, $X^2(36, n = 93) = 44.60$, $p > 0.05$, Nagelkerke = 0.386 (McFadden = 0.110). AU 2 was a significant predictor overall (as previously indicated). These AUs classified surprise 20% (substantially lower than with all 11 AUs, see Hypothesis 2b).

Summary

To summarize, AUs that were significant predictors overall of an emotion category (although not specifically which one) included AU 1, AU 2, AU 4, AU 6, AU 10, AU 12, and AU 26. None of the 11 AUs represented in the ICP prototypes for the emotions we studied predicted any one emotion category specifically, except AU 25 which predicted happy. Interestingly, AU 25 is part of the ICP prototype for *all but* three of the emotions we studied, yet we only found that its presence predicted happy. Of the other AUs, only AU 10 is thought to be specific (disgust)[7].

Analysis 2b: Perceived Emotion Classification

To test Hypothesis 2b, we used a multinomial logistic regression to test whether the 11 AUs from the ICP prototypes could better predict the perceived emotion category across emoji faces. We also compared individual platforms/versions.

Across Platforms

We first computed the MLR on the 11 shared AUs across platforms for all ten emotions. The dependent variable was the perceived emotion category. The model produced was significant, $X^2(99, n = 93) = 226.261$, $p < 0.001$, Nagelkerke = 0.924 (McFadden = 0.557). Likelihood ratio tests were significant for seven AUs: AU 1 ($X^2(9) = 42.63$, $p < 0.001$); AU 4 ($X^2(9) = 45.55$, $p < 0.001$); AU 6 ($X^2(9) = 21.10$, $p < 0.05$); AU 7 ($X^2(9) = 21.90$, $p < 0.05$); AU 12 ($X^2(9) = 57.84$, $p < 0.001$); AU 25 ($X^2(9) = 17.67$, $p < 0.05$); AU 26 ($X^2(9) = 29.38$, $p = 0.001$). None of the individual emotions were significantly predicted, however, with these AUs.

The overall classification rate of emotions to their predicted category was 58.1%. **Table 7** shows the classification matrix. Overall, *happy* was the best classified at 94.7% ($n = 19$). One incorrect classification was assigned to calm. *Sad* had the next best classification rate at 88.2% ($n = 17$). One incorrect classification went to fear. *Anger* had a classification rate of 83.3% ($n = 6$), with incorrect classification assigned to disgust. *Disgust* had a classification rate of 80% ($n = 5$), with one incorrect classification assigned to sad. *Surprise* had a 60.0% classification rate ($n = 9$). Incorrect classifications were mainly assigned to sad, followed by one each to fear and to happy. *Fear* had a 33.3% classification rate ($n = 9$): Fear was misclassified mainly as sad and surprise, followed by a tie between calm and envy. C*alm* had a classification rate of 22.2% ($n = 9$). Incorrect classifications were mainly assigned to happy, followed by a tie between surprise, anger, and sad. *Envy* had a poor classification rate at 16.7% ($n = 6$). Incorrect classifications were mainly assigned to sad, followed by fear, love, and disgust. *Love* also had a poor classification rate at 12.5% ($n = 8$): Love was misclassified as calm, followed by happy, and then sad and surprise. *Contempt* had the worst classification rate (0.0%, $n = 4$), with incorrect classifications split among sad, surprise, calm, and anger. These results are generally in line with classification rates of AUs to human emotion categories. Specifically, individual instances of faces perceived as happy, anger, and fear contain more of the prototypical AUs, compared to contempt, sadness, and disgust, which generally show less overlap with the proposed codes (Cordaro et al., 2018).

Individual Platforms

We next compared the three platforms. Apple and Samsung both produced marginally significant models: Apple, $X^2(90) = 112.83$, $p = 0.052$, Nagelkerke = 0.987 (McFadden = 0.841); Samsung, $X^2(90) = 112.58$, $p = 0.054$, Nagelkerke = 0.987 (McFadden = 0.849). The model for Google Android was not significant, $X^2(81) = 196.76$, $p > 0.05$, Nagelkerke = 0.968 (McFadden = 0.717). Interesting, however, when comparing the AIC values, the best fit was Google Android. This is likely because there were fewer AUs coded for Google Android, but of those, there was slightly better classification: Google Android AIC = 196.92, followed by Apple AIC = 208.05, and AIC Samsung = 210.87. The lower the value, the "better" fit of the model.

Yet, Samsung had the highest overall classification rates, with 83.9% (Samsung: range = 0% disgust to 100% for happy, fear, envy, surprise, and anger). Apple had an overall classification rate of 80.6% (range: 0% for calm and contempt to 100% for fear, envy, love, anger, and disgust). By comparison, Google Android had a correct classification rate of 64.5% (range 0% for fear and contempt to 100% for sad).

[7] AU 15 is not included in the ICP for sadness, despite all previous studies noted in **Table 1**. AU 20 is also not included for fear, despite all previous studies noted in the **Table 1**.

TABLE 7 | Classification Matrix using all 11 AUs common to ICP prototypes for Studied Emotions.

	Happy	Sad	Fear	Envy	Love	Surprise	Calm	Contempt	Anger	Disgust
Happy	94.7	0.0	0.0	0.0	0.0	0.0	5.3	0.0	0.0	0.0
Sad	0.0	88.2	11.8	0	0.0	0	0	0	0	0
Fear	0.0	22.2	33.3	11.1	0.0	22.2	11.1	0.0	0.0	0
Envy	0.0	33.3	16.7	16.7	16.7	0	0.0	0.0	0	16.7
Love	25.0	12.5	0.0	0	12.5	12.5	37.5	0.0	0.0	0.0
Surprise	10.0	20.0	10.0	0	0.0	60.0	0.0	0.0	0.0	0.0
Calm	44.4	11.1	0.0	0	0.0	11.1	22.2	0	11.1	0.0
Contempt	0.0	25.0	0.0	0	0.0	25.0	25.0	0	25.0	0.0
Anger	0.0	0.0	0.0	0.0	0	0.0	0.0	0	83.3	16.7
Disgust	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0	80.0

Summary

To summarize, 11 AUs were better at predicting perceived emotions than only the ones in the ICP prototype for each emotion. There were differences in how well each platform classified individual emotions from AUs. For example, Google Android only had two faces perceived as fear but did not classify either correctly, whereas Apple and Samsung had three and four faces perceived as fear and classified them all correctly. Apple did a poor job classifying calm (0%) (n = 2), but classification was 50% (n = 2) on Samsung and 60% (n = 5) on Google Android. Finally, Samsung did a poor job classifying disgust (0%) (n = 1), but Google Android had a 50% classification rate (n = 2) and Apple had a 100% classification rate (n = 2).

DISCUSSION

In this manuscript, we created an adapted emoji-FACS system to explore whether emoji faces (from an anatomical perspective) look similar across platforms/versions, and whether the anatomical configurations are shared with human expressions. Although FACS was not designed for nonhuman faces, it has been adapted and validated for a number of species over the years (e.g., chimpFACS, Parr et al., 2007; and MaqFACS, Parr et al., 2010). Clearly emoji faces are not human (or nonhuman faces), but they are perceived as faces with emotional content.

Once we established the emoji-FACS rubric, the first goal was to systematically compare AUs for emoji faces across platforms and versions. Although emoji faces were designed for the purpose of communicating emotional information, there is little agreement about what specific emotion an individual face is perceived as. We found that different platforms and versions not only often relied on different AUs, but also often that the frequency of AUs was different across platforms and versions (Hypothesis 1a). In addition, faces perceived as the same emotion and those perceived as different emotion(s) were equally diverse in their use and distribution of AUs (Hypothesis 1b). In a few instances, certain AU counts did differ between faces perceived as the same vs. different emotion(s), but this could be attributed to the fact that these AUs were only present in one emotion category and had good predictive validity (e.g., tears for sadness and winks for love).

The second goal was to assess whether emoji-coded AUs were similar to the AUs in the ICP prototypes for the same perceived emotion. Across platforms and versions, we found that AUs common to emotion prototypes were used in emoji faces, but AUs did not predict *specific* emotion categories (Hypothesis 2a). Similar results were found when we included all the AUs in our model to predict emotion category, although overall classification rates increased when we did so. Our model was moderately good at predicting emotion: The average across categories was 58.1%. Specifically, happy, sad, anger and disgust were best predicted overall, but there were substantial differences among platforms in the individual emotion classification rates (Hypothesis 2b). Google Android showed the least predictive ability, yet it produced the best fitting model of the three platforms. This was likely because it used fewer AUs, but used them in more consistent ways. None of the AUs predicted a specific emotion category, however, except AU 25. Rather than outright predicting a specific emotion category, individual AUs seemed to narrow down to what emotion category an emoji face might belong by knowing *what category it is not*. Thus, the majority of AUs only give some predictive validity.

Although we did not test a model which included all 26 of our codable AUs as predictors of emotion category (rather than the 11 AUs shared with ICP prototypes), there is little doubt that some of these additional AUs would have been significant predictors (e.g., crying was only used in faces perceived as sad, and AU 18, AU 22, and AU 46 were only used in faces perceived as love). Thus, it is reasonable to assume that some AUs we coded (even though not part of the ICP prototypes for human facial expressions) are specific to an emotion category.

This finding is consistent with the results of a recent study using emoji-like faces (Betz et al., 2019). In that study, participants were asked to which emotion category each face belonged. Faces were either presented in the context of emotion words or not. Overall, adding emotion words increased emotion agreement for these faces, as adding emotion words increases the agreement among raters for human facial depictions of emotion (for reviews, see Lindquist and Gendron, 2013; Lindquist et al., 2016; Barrett et al., 2019). Yet, some emojis in that study were less affected by the context of words. For instance, people largely agreed (without any context) that the face with wide eyes and a gaping mouth was surprise, even without the added context of emotion words.

The Theory of Constructed Emotion (Barrett, 2017; Barrett et al., 2019) suggests that the human brain is constantly predicting what a stimulus is (e.g., a face) and to what emotion category it might belong (e.g., anger or fear). It recognizes that emotion perception (and the perception of categories of the mind, more generally) is the product of such predictions. According to this view, people perform a type of "affective calculus" in which their brain is constantly predicting (based on provided labels, situational context, and previous knowledge) what a stimulus is and to what category it belongs (see also Betz et al., 2019). Of course, predictions are built (at least initially) on information from the world-in the case of emotion perception, from the information our body senses either within ourselves or other people. Some of these changes can become associated with emotional meaning when occurring in a specific context. Perhaps then we can best think of emoji faces (much like human faces) as providing a starting point for more refined predictions. Faces, like voices, bodily postures, and the like aren't diagnostic of emotions, but they can help to narrow the outcome of our brain's predictions. We might then think of this core set of AUs (plus perhaps a few other which might be specific to emoji faces) as helping to narrow which emotion category a face belongs. This idea seems particularly in line with our findings that the core AUs did not predict a specific emotion well (much in the same way AUs do not predict specific emotions from a human face very well), but they contributed to the process. Although we did not test this theory specifically, in future studies adding a context (whether verbal or pictorial) should facilitate perception and therefore increase agreement among raters as to which emotion category an emoji face belongs.

These findings are also consistent with Channel Expansion Theory in Communication (Carlson and Zmud, 1999), in which exposure to electronic communication enhances a receiver's knowledge about those platforms and thus refines possible interpretations. Indeed, receivers develop their computer-mediated communication skills through experience with others using the same medium and the feedback they receive from others. Therefore, experience with online communication (in which emoji faces are used) allows receivers to develop and ultimately better convey information, such as that about emotion (Gudykunst, 1997).

Implications

So, what does this mean for computer programmers in charge of the physical renderings of emoji faces? Two things jump to mind. The first is that programmers should be aware of just how different "equivalent" emojis really are in terms of their appearance. They must also be aware that, more often than not, "equivalent" emoji faces not only look different, but are also perceived as different emotion categories.

Second, emoji faces do not appear in isolation. Although a single emoji can be sent or texted to an individual, it is in reference to something either explicitly communicated or implicitly understood between the two parties. Therefore, regardless of the individual theory of emotion to which a person ascribes, there is likely interplay between a face and the context (e.g., Trope, 1986; Aviezer et al., 2008, 2011). Future work should therefore also consider the usage of emojis in context and elaborate on how the context can affect emotional interpretation (see Walther and D'Addario, 2001; Kelly, 2015).

Limitations and Future Directions

This research has several limitations. The first is the selection of possible choices (and number) of emotions and emojis. This study used 31 emojis (depicted on three electronic platforms). The Unicode system now has 3,136 emoji characters, 92 of which are emoji faces (Unicode, 2020). We also only included three major platforms, and there are many others, including Facebook, Twitter, and WhatsApp. In addition, we only investigated ten emotions, and as noted, other researchers have proposed more (Cordaro et al., 2018).

Another limitation is that we only adapted 26 codes from the FACS system, which includes more than 65 (with head and eye positions). While ICP prototypes, however, only use a subset (15 AUs, not including body postures which are sometimes included), our codes included only 11 of these 15. One way that we tried to address this was by including additional AUs. This included things like eye gaze, lip puckers and funnelers, *crying*, and including other potential "candidates" for specific emotion AUs (e.g., AU 20 and AU 15).

Perhaps the largest limitation is that we used only the ICP prototypes from Westerns, and such configurations do not likely apply to displays from Eastern countries (Cordaro et al., 2018). For example, East Asian models show less distinction between emotions (see also Jack et al., 2012). Related to this limitation, our perceived emotions from emoji faces came from English-speakers who all resided in the United States and were mainly between 18 and 24 years old (see Franco and Fugate, 2020).

We recommend that future empirical research on emojis both broadens the repertoire of emojis (also opens up to additional platforms) and also considers the perceived emotion given from non-Western individuals. Ultimately, however, it is in the hands of the programmers to decide how to translate an emoji in newer versions across platforms. That said, we strongly advocate that programmers also consider the role that emoji labels play (e.g., "confused" face, "disappointed" face) as they might be in opposition to the perceived emotion. Moreover, we strongly advocate that the field of emotion, and general nonverbal communication as a whole, explore the role that the perceiver's conceptual knowledge and that situational cues play in interpreting the rudimentary structural information that exists in the face.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Emoji Faces and Coding at OSFHome: https://osf.io/9dzfw/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of MA - Dartmouth IRB # 16.052. A. Kareberg, Office of Institutional Compliance. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

JF performed all coding and analyses. Both authors conceived of the idea and wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank Brandy Burkett for her help with FACS coding and the adapted emoji FACS coding rubric. Emotion perception categories were previously published in the Journal of Nonverbal Behavior in 2020.

REFERENCES

Aviezer, H., Dudarev, V., Bentin, S., and Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion* 11, 1406–1414. doi: 10.1037/a0023578

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. *Psychol. Sci.* 19, 724–732. doi: 10.1111/j.1467-9280.2008.02148.x

Bailey, J. (2018). *Copyright and Emoji. Plagiarism Today*. Available online at: https://www.plagiarismtoday.com/2018/04/10/copyright-and-emojis/ (accessed April 10, 2018).

Barrett, L. F. (2006a). Solving the emotion paradox: categorization and the experience of emotion. *Pers. Soc. Psychol. Rev.* 10, 20–46. doi: 10.1207/s15327957pspr1001_2

Barrett, L. F. (2006b). Are emotions natural kinds? *Perspect. Psychol. Sci.* 1, 28–58. doi: 10.1111/j.1745-6916.2006.00003.x

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* 12, 1–23. doi: 10.1093/scan/nsx060

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol. Sci. Public Interest* 20, 1–68. doi: 10.1177/1529100619889954

Betz, N., Hoemann, K., and Barrett, L. F. (2019). Words are a context for mental inference. *Emotion* 19, 1463–1477. doi: 10.1037/emo0000510

Brosch, T., Pourtois, G., and Sander, D. (2010). The perception and categorisation of emotional stimuli: a review. *Cogn. Emot.* 24, 377–400. doi: 10.1080/02699930902975754

Carlson, J. R., and Zmud, R. W. (1999). Channel expansion theory and the experiential nature of media richness perceptions. *Acad. Manag. J.* 42, 153–170. doi: 10.5465/257090

Chatzichristos, C., Morante, M., Andreadis, N., Kofidis, E., Kopsinis, Y., and Theodoridis, S. (2020). Emojis influence autobiographical memory retrieval from reading words: An fMRI-based study. *PLoS One* 15:e0234104.

Cordaro, D. T., Brackett, M., Glass, L., and Anderson, C. L. (2016). Contentment: Perceived completeness across cultures and traditions. *Rev. Gen. Psychol.* 20, 221–235. doi: 10.1037/gpr0000082

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., and McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion* 18, 75–93. doi: 10.1037/emo0000302

Cowen, A. S., and Keltner, D. (2017). Self-report capture 27 distinct categories of emotion bridged by continuous gradients. *Proc. Natl. Acad. Sci.U.S.A.* 114, E7900–E7909.

Darwin, C. (1872/2005). *The Expression of the Emotions in Man and Animals*. Philadelphia, PA: Running Press.

deWaal, F. B. M. (2003). Darwin's legacy and the study of primate visual communication. *Ann. N.Y. Acad. Sci.* 1000, 7–31. doi: 10.1196/annals.1280.003

Du, S., Tao, Y., and Martinez, A. M. (2014). Compound facial expressions of emotion. *Proc. Natl. Acad. Sci.* 111, E1454–E1462. doi: 10.1073/pnas.1322355111

Ekman, P. (1972). "Universal and cultural differences in facial expression of emotions," in *Proceedings of the Nebraska Symposium on Motivation*, ed. J. Cole (Lincoln, NE: University of Nebraska Press), 207–283.

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200.

Ekman, P. (2016). What scientists who study emotion agree about. *Perspect. Psychol. Sci.* 11, 31–34. doi: 10.1177/1745691615596992

Ekman, P., and Cordaro, D. (2011). What is meant by calling emotions basic. *Emot. Rev.* 3, 364–370. doi: 10.1177/1754073911410740

Ekman, P., Levenson, R., and Friesen, W. (1983). Autonomic nervous system activity distinguishes among emotions. *Science* 221, 1208–1210. doi: 10.1126/science.6612338

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

Elfenbein, H. A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emot. Rev.* 5, 90–96. doi: 10.1177/1754073912451332

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

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

Emojipedia (2020). Smileys and People: Emojis for Smileys, People, Families, Hand Gestures, Clothing and Accessories. Available online at: https://emojipedia.org/people (accessed May, 2020)

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., and Lehmann, S. (2017). "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 1615–1625. doi: 10.18653/v1/D17-1169

Franco, C. L., and Fugate, J. M. B. (2020). Emoji face renderings: Exploring the role emoji platform differences have on emotional interpretation. *J. Nonverb. Behav.* 44, 301–328. doi: 10.1007/s10919-019-00330-1

Gantiva, C., Sotaquira, M., Araujo, A., and Cuervo, P. (2019). Cortical processing of human and emoji faces: an ERP analysis. *J. Behav. Inf. Technol.* 39, 935–943. doi: 10.1080/0144929X.2019.1632933

Gudykunst, W. B. (1997). Cultural variability in communication: an introduction. *Commun. Res.* 24, 327–348. doi: 10.1177/009365097024004001

Huber, E. (1931). *Evolution of Facial Musculature and Facial Expression*. Baltimore, MD: Johns Hopkins Press.

Izard, C. E. (1991). *The Psychology of Emotions*. Berlin: Springer Science & Business Media.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol. Rev.* 99, 561–565.

Izard, C. E. (2013). *Human Emotions*. Berlin: Springer Science & Business Media.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., and Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proc. Natl. Acad. Sci.* 109, 7241–7244. doi: 10.1073/pnas.1200155109

Jones, M. (2017). *Matt Jones—IMDb*. Available online at: https://www.imdb.com/name/nm2886030/

Kelly, C. (2015). *Do You Know What I Mean >:(A linguistic Study of the Understanding of Emoticons and Emojis in Text Messages.* urn:nbn:se:hh:diva-27611. Halmstad: halmstad university.

Keltner, D., and Buswell, B. N. (1997). Embarrassment: its distinct form and appeasement functions. *Psychol. Bull.* 122, 250–270. doi: 10.1037/0033-2909.122.3.250

Keltner, D., and Cordaro, D. T. (2015). Understanding multimodal emotional expressions: Recent advances in basic emotion theory. Emotion Researcher. Available online at: http://emotionresearcher.com/understanding-multimodal-emotional-expressions-recent-advances-in-basic-emotion-theory/

Keltner, D., Sauter, D., Tracy, J., and Cowen, A. (2019). Emotional expression: advances in basic emotion theory. *J. Nonverb. Behav.* 43, 133–160. doi: 10.1007/s10919-019-00293-3

Keltner, D., Tracy, J., Sauter, D., Cordaro, D., and McNeil, G. (2016). "Expression of emotion,"in *Handbook of Emotions*, eds L. F. Barrett, M. Lewis, J. M. Haviland-Jones (New York, NY: Guilford Press).

Lee, S. J. (2018). Emoji at MoMA: considering the "original emoji" as art. *First Monday* 23, 1–12. doi: 10.5210/fm.v23i9.9401

Lindquist, K. A. (2017). The role of language in emotion: existing evidence and future directions. *Curr. Opin. Psychol.* 17, 135–139. doi: 10.1016/j.copsyc.2017.07.006

Lindquist, K. A., and Gendron, M. (2013). What's in a word? Language constructs emotion perception. *Emot. Rev.* 5, 66–71. doi: 10.1177/1754073912451351

Lindquist, K. A., Gendron, M., Satpute, A. B., Barrett, L., Lewis, M., and Haviland-Jones, J. (2016). "Language and emotion: Putting words into feelings and feelings into words," in *Handbook of Emotions*, eds J. M. Haviland–Jones, M. Lewis and L. F. Barrett (New York, NY: Guilford Press), 579–594.

Ling, R. (2010). Texting as a life phase medium. *J. Comput. Med. Commun.* 15, 277–292. doi: 10.1111/j.1083-6101.2010.01520

Liu, D., and Yang, C. (2016). Media niche of electronic communication channels in friendship: a meta-analysis. *J. Comput. Med. Commun.* 21, 451–466. doi: 10.1111/jcc4.12175

Matsumoto, D. (1989). Cultural influences on the perception of emotion. *J. Cross Cult. Psychol.* 20, 92–105. doi: 10.1177/0022022189201006

Matsumoto, D., Keltner, D., Shiota, M., O'Sullivan, M., and Frank, M. (2008). "What's in a face? Facial expressions as signals of discrete emotions," in *Handbook of Emotions*, eds Lewis, M., Haviland-Jones, J. M., and Barrett, L. F, (New York, NY: Guilford Press).

Miller Hillberg, H., Levonian, Z., Kluver, D., Terveen, L., and Hecht, B. (2018). "What I see is what you don't get: the effects of (not) seeing emoji rendering differences across platforms," *Proceedings of the ACM on Human-Computer Interaction (CSCW)*. 2, 124. doi: 10.1145/3274393

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., and Hecht, B. (2016). "Blissfully happy" or "ready to fight": varying interpretations of emoji," in *Proceedings of ICWSM*, Cologne, 11.

Oleszkiewicz, A., Frackowiak, T., Sorokowska, A., and Sorokowski, P. (2017). Children can accurately recognize facial emotions from emoticons. *Comput. Hum. Behav.* 76, 372–377. doi: 10.1016/j.chb.2017.07.040

Parr, L. A., Waller, B. M., Burrows, A. M., Gothard, K. M., and Vick, S. J. (2010). MaqFACS: a muscle-based facial movement coding system for the Rhesus Macaque. *Am. J. Phys. Anthropol.* 143, 625–630. doi: 10.1002/ajpa.21401

Parr, L. A., Waller, B. M., and Vick, S. J. (2007). New developments in understanding emotional facial signals in chimpanzees. *Curr. Direct. Psychol. Sci.* 16, 117–122. doi: 10.1111/j.1467-8721.2007.00487.x

Rodrigues, D., Prada, M., Gaspar, R., Garrido, M. V., and Lopes, D. (2018). Lisbon emoji and emoticon database (LEED): norms for emoji and emoticons in seven evaluative dimensions. *Behav. Res. Methods* 50, 392–405. doi: 10.3758/s13428-017-0878-6

Sauter, D. A., LeGuen, O., and Haun, D. B. M. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion* 11, 1479–1483. doi: 10.1037/a0025336

Shiota, M. N., Campos, B., and Keltner, D. (2003). The faces of positive emotion: Prototype displays of awe, amusement, and pride. *Ann. N. Y. Acad. Sci.* 1000, 296–299. doi: 10.1196/annals.1280.029

Tigwell, G. W., and Flatla, D. R. (2016). "Oh that's what you meant!: Reducing emoji misunderstanding," in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct – MobileHCI '16,* New York, NY, 859–866. doi: 10.1145/2957265.2961844

Tomkins, S. (1962). *Affect Imagery Consciousness: Volume I: The Positive Affects*. New York, NY: Springer Publishing Company.

Toratani, Y., and Hirayama, M. J. (2011). "Psychological analysis of emoticons used for e-mails on cellular phones," in *Proceedings of the International Conference on Mobile IT Convergence*, (Piscataway, NJ: *IEEE*), 49–53.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychol. Rev.* 93:239.

Unicode (2020). *Full Emoji list, Unicode.com*. Available online at: https://unicode.org/emoji/charts/full-emoji-list.html (accessed May, 2020).

van Hooff, J. A. R. A. M. (1962). "Facial expressions in higher primates," in *Proceedings of the Symposium of the Zoological Society of London* 8, 97–125.

Walther, J. B., and D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Soc. Sci. Comput. Rev.* 19, 324–347. doi: 10.1177/089443930101900307

Wolny, W. (2016). "Emotion analysis of Twitter data that use emoticons and emoji ideograms," in *Proceedings of the International Conference on Information Systems Development: complexity in Information Systems Development*, eds J. Gołuchowski, M. Pańkowska, C. Barry, M. Lang, H. Linger, and C. Schneider (Katowice: University of Economics in Katowice), 476–483.

Yuasa, M., Saito, K., and Mukawa, N. (2011). Brain activity when reading sentences and emoticons: An fMRI study of verbal and nonverbal communication. *Electron. Commun. Jpn.* 94, 17–24. doi: 10.1002/ecj.10311

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 February 2021
doi: 10.3389/fpsyg.2021.612923

The Expressive Triad: Structure, Color, and Texture Similarity of Emotion Expressions Predict Impressions of Neutral Faces

Daniel N. Albohn and Reginald B. Adams Jr.**

Department of Psychology, The Pennsylvania State University, University Park, PA, United States

Previous research has demonstrated how emotion resembling cues in the face help shape impression formation (i. e., emotion overgeneralization). Perhaps most notable in the literature to date, has been work suggesting that gender-related appearance cues are visually confounded with certain stereotypic expressive cues (see Adams et al., 2015 for review). Only a couple studies to date have used computer vision to directly map out and test facial structural resemblance to emotion expressions using facial landmark coordinates to estimate face shape. In one study using a Bayesian network classifier trained to detect emotional expressions structural resemblance to a specific expression on a non-expressive (i.e., neutral) face was found to influence trait impressions of others (Said et al., 2009). In another study, a connectionist model trained to detect emotional expressions found different emotion-resembling cues in male vs. female faces (Zebrowitz et al., 2010). Despite this seminal work, direct evidence confirming the theoretical assertion that humans likewise utilize these emotion-resembling cues when forming impressions has been lacking. Across four studies, we replicate and extend these prior findings using new advances in computer vision to examine gender-related, emotion-resembling structure, color, and texture (as well as their weighted combination) and their impact on gender-stereotypic impression formation. We show that all three (plus their combination) are meaningfully related to human impressions of emotionally neutral faces. Further when applying the computer vision algorithms to experimentally manipulate faces, we show that humans derive similar impressions from them as did the computer.

Keywords: face perception, emotion expression, machine learning, impression formation, facial expresions

OPEN ACCESS

Edited by:
José-Miguel Fernández-Dols, Autonomous University of Madrid, Spain

Reviewed by:
Chaona Chen, University of Glasgow, United Kingdom
Ferenc Kocsor, University of Pécs, Hungary

****Correspondence:***
Daniel N. Albohn
daniel.albohn@chicagobooth.edu
Reginald B. Adams Jr.
regadams@psu.edu

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *01 October 2020*
Accepted: *18 January 2021*
Published: *25 February 2021*

Citation:
Albohn DN and Adams RB Jr (2021) The Expressive Triad: Structure, Color, and Texture Similarity of Emotion Expressions Predict Impressions of Neutral Faces. Front. Psychol. 12:612923. doi: 10.3389/fpsyg.2021.612923

INTRODUCTION

Decades of research have revealed facial expressions to be a powerful vehicle for social communication. Humans are so tuned to reading dynamic displays from the face that overt expressions tend to influence stable trait impressions (Knutson, 1996). Indeed, years of research suggests that even emotion resembling cues in otherwise neutral faces have a powerful impact on impressions formed (e.g., emotion overgeneralization effect; Zebrowitz et al., 2010). Recent attention has been aimed at training computers to automatically recognize human emotion from the face. These advancements have equipped researchers with a powerful set of tools for exposing new theoretical insights and creating novel societal applications based on this work. Commercially available face reading programs (e.g., FaceReader, Affectiva, and AFDEX) are now widely used across a variety of diverse settings such as classroom observation, user-end experience, human-robot interaction, virtual reality, and marketing.

We know that even faces devoid of overt expression contain a surprising amount of socially relevant information despite being objectively categorized and subjectively posed as neutral (at least in terms of affect). Static physical features such as gender-related appearance and age-related changes in the face alter perceptions and impressions (Zebrowitz et al., 2010; Adams et al., 2016; Albohn and Adams, 2020b). Further, first impressions based on so-called "neutral" faces tend to be consistent across different observers. This suggests that, on some level, individuals are attuned to similar socially relevant cues from which they draw similar judgments. Such judgments are at least in part attributable to the reading of emotion resembling cues that are confounded with gender and age (Adams et al., 2016). The mere resemblance of a face to an expression powerfully influences a wide array of trait impressions of others (Adams et al., 2012). For instance, simply moving the eyebrows to be lower on a non-expressive face leads to greater dominance and anger attributions, whereas moving the eyebrows higher yields greater submissiveness and fear attributions (Keating et al., 1977; Laser and Mathie, 1982). Also, shortening or lengthening the distance between the eyes and mouth results in perceptions of anger and sadness, respectively (Neth and Martinez, 2009). These findings contribute to a growing body of evidence that shows that incidental emotion-resembling cues, and in some cases subtle expressivity lingering on a subjectively non-expressive face, powerfully influence impressions (Zebrowitz et al., 2010; Adams et al., 2016; Albohn and Adams, 2020a).

Various computer vision techniques have complimented the host of findings that suggests information can be derived from non-expressive faces. For instance, Zebrowitz et al. (2007) trained a neural network to detect actual baby vs. adult faces, and then applied this model to detecting such cues in surprise, anger, happy, and neutral expressions. They found that the model detected babyfacedness in surprise expressions, and maturity in anger expressions due to similarities in brow position. Likewise, these researchers later found that both gender and race (Zebrowitz et al., 2010) as well as age (Palumbo et al., 2017) cues in otherwise affectively neutral faces were recognized by the neural network as containing emotion cues. Along these same lines, Said et al. (2009) trained a Bayesian classifier to detect expressions in faces and then applied it to images of neutral faces that had been rated on a number of personality traits (e.g., trustworthy and dominance). Said et al. (2009) found that the trait ratings of the faces were meaningfully correlated with the perceptual resemblance the faces had with certain expressions. These results speak to a mechanism of perceptual overlap whereby expression and identity cues trigger similar processes due simply to their physical resemblance.

STRUCTURE, COLOR, AND TEXTURE FACE METRICS

While the past few decades have seen an increase in the development and use of machine learning methods within person perception, most of this work has focused exclusively on evaluating (separately) each metric/feature's influence on model *performance*, rather than how the metric/feature relates back to human visual perception and impression formation. As such, model evaluation is based on absolute performance (i.e., minimizing error from ground truth) rather than attempting to understand how computer vision relates to human vision. Prior research suggests that structure, color, and texture of the face are all important metrics for face identification (Sinha et al., 2006). It stands to reason that each of these metrics has its own influence on human facial emotion recognition as well as downstream impression formation.

Myriad research has examined how facial structural resemblance to emotion expressions relates to impression formation (i.e., emotion overgeneralization). Seminal work has shown that faces with cues that incidentally resemble emotional expressions are subsequently evaluated in terms of that emotional expression (see, e.g., Zebrowitz and McDonald, 1991; Marsh et al., 2005; Zebrowitz et al., 2007). Structural resemblance to a specific expression on a non-expressive (i.e., neutral) face powerfully influences trait impressions of others. In one study, facial structural resemblance to anger expressions was correlated with threatening personality traits (e.g., dominant), and resemblance to happy expressions was correlated with positive traits (e.g., caring; Said et al., 2009). In another study, manipulating neutral faces to structurally resemble anger and fear influenced a whole host of physical, emotional, and person perception impressions, including ones with non-obvious links to emotion such as anger-resembling cues yielding relatively greater impressions of shrewdness and fear-resembling cues yielding relatively greater impressions of intuitiveness (Adams et al., 2012).

Critically, related work has shown that facial structural resemblances to different expressions are related to stereotypes and biases associated with gender. Zebrowitz et al. (2010) first trained connectionist neural network models to discriminate between an expressive (anger, happy, and surprise) face and a neutral face. Next, they applied the trained classifier to neutral faces varying in race and gender. Zebrowitz et al. (2010). In terms of gender, they found that female faces structurally resembled surprise expressions more than male faces. Similarly, male faces structurally resembled anger expressions more than female faces. Finally, male faces structurally resembled happy expressions more than female faces (see, e.g., Becker et al., 2007; Hess et al., 2009; note that this latter finding contrasts with several other prior studies that female faces structurally resemble happy expressions more than male faces). This last point highlights the complex interaction between gender and emotion and suggests that some gendered impressions are influenced from bottom up cues (e.g., metrics of the face), whereas others are overridden by top down stereotypes. In the case of happiness, male faces have been found to structurally resemble happy expressions more than female faces, yet rating studies have shown that more masculine faces are often rated as less trustworthy than feminine faces (see, e.g., Todorov et al., 2008; Adams et al., 2015).

Although less studied than face structure, there is evidence that other features such as color and texture are also important for face identification, impression formation, and emotion judgements (Russell et al., 2006; Sinha et al., 2006). For example, researchers have found that increasing the luminance difference between the eyes and mouth results in more attributions of that face appearing female, while decreasing contrast in the eye and mouth regions result in greater perceptions of the face appearing male (Russell, 2003, 2009). These results are underscored by the observation that women often use cosmetics to increase the contrast of the eyes and lips with the rest of their face and that observers rate women with high facial contrast dimorphism as appearing more attractive (Rhodes, 2006). Face color also appears to have a defining influence on perceptions of religiosity (Rule et al., 2010) as well as judgments of health and attractiveness (Pazda et al., 2016; Thorstenson et al., 2017; Perrett et al., 2020). In both cases, perceptions of "healthier" skin drove impressions and classifications.

Lastly, there has been relatively little work examining the influence of face texture on emotion and impression formation. Most previous research examining how face texture shapes perception has either done so directly by showing that face texture relates to perceived health and attractiveness, or indirectly by showing face cues that are related to skin texture also influence perception (e.g., aging cues). Most research that has examined facial skin texture directly has distinguished it from face skin color. For example, many researchers include in their definition of skin texture components such as skin elasticity (e.g., sagging, wrinkles, smoothness), dermatosis issues (e.g., acne, sun damage, freckles, pores), and facial hair features (e.g., eyebrow thickness). Given that humans are particularly adept at surface property perception (see, e.g., Klatzky et al., 1987), it makes sense that texture would also influence face judgements.

Direct investigations into the influence of surface properties of the face on perception have shown that it is related to perceived health, trustworthiness, and other related traits. Tan et al. (2018) used Gabor wavelet analysis to decompose the texture of a patch of skin on the cheek into three components. They then had participants rate full face pictures on its perceived health. Results showed a significant relationship between perceived health and each Gabor features. Examination of each Gabor feature showed that the three features related to perceived health appeared to be the number of red spots on the skin (less was viewed as healthier), scarring and holes (less perceived as healthier), and roughness (smoother viewed as healthier).

Taken together, facial structure, color, and texture have been shown to be important for both face identification and classification. Individuals appear to use these face cues not only to differentiate faces from non-faces, but also when making judgements, such as the gender or health of a face. All three cues have been shown to both independently and in combination influence emotion expression perception or related personality traits. Critical to the current research, however, is that relatively little research has examined each face metrics' independent and combined contribution to impression formation. Of the work that has examined each metric's contribution, it has done so using either only a subset of features or without the aid of machine learning technology. Thus, there are critical gaps in the literature that the current research attempts to address: What is the relative importance of each face feature for expression classification? How are these features (independently and in combination) related to human-derived impressions? And can they be used to capture subtle or incidental emotion-resemblance on non-expressive faces?

OVERVIEW OF CURRENT WORK

Previous work has shown that emotion expressions can be predicted via the structure of the face alone, and that a neutral face's structural resemblance to emotion is related to human impressions. However, to the best of our knowledge, no research has extended this work beyond structure, shown meaningful relationships with human impressions beyond just model output, or used the model output to manipulate faces to confirm whether what machine learning algorithms are using to make accurate predictions also drive human impressions. The current work proposes an end-to-end experimental pipeline in which we replicate and extend previous work, while also providing novel experiments to address the current gaps in the literature.

We chose to focus initially on gender-related emotion stereotypes and related trait impressions because these are highly validated and robust effects reported in the literature over the last couple decades (see Adams et al., 2015 for review). Further, recent findings suggest there are visual confounds in emotionally expressive and gender-related appearances cues (Becker et al., 2007; Hess et al., 2009). Thus, the closest to a ground truth to use as an initial test of our algorithm are gender-related emotion and trait impressions. These include the predictions that male neutral faces would be more resembling of anger expressions, which would give rise to more dominant trait attributions, whereas female faces would be more resembling of fear expressions, which would give rise to more affiliative trait impressions. We also predict that, like in previous research, male faces will counter-stereotypically resemble happy expressions more than female faces (Zebrowitz et al., 2010; Adams et al., 2015). Once establishing that, we sought to directly examine the broader theoretical assertion that humans meaningfully utilize emotion-resembling cues in the face when forming impressions by not only showing that computer derived emotion cues mediate human impressions, but by experimentally manipulating faces using algorithmically-derived facial cues and showing that humans utilize those cues to arrive at same impressions predicted by the computer.

We have broken our central thesis into three levels (see **Figure 1**). The first level (Study 1) attempts to replicate and extend previous work by training models on the structure, color, and texture of the face to see if each was a meaningful predictor of emotion expression. Level 1 also validates our trained models by showing that they predict that male and female neutral faces

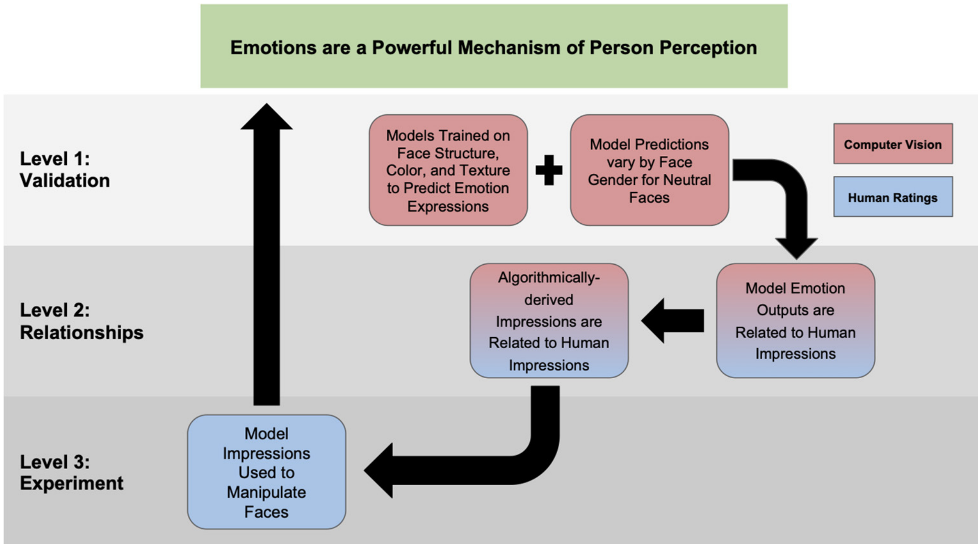


FIGURE 1 | Theoretical roadmap for the current set of studies. Our overall central thesis is presented at the top of the figure. Study 1 corresponds to Level 1; Studies 2 & 3 correspond to Level 2; Study 4 corresponds to Level 3.

vary in emotion expressivity in an expected pattern (see Adams et al., 2015). Next, in Level 2, we first replicate previous work and show that our model outputs are meaningfully related to human impressions. More importantly in this study, we also provide novel experimental evidence that the emotion output from our models can be used to algorithmically compute higher-order impressions and that these predict similar human impressions. This finding suggests that humans are, at least in part, using emotions to form their non-emotion, higher-order impressions. Lastly, Level 3 aims to directly test this assertion by taking the trained model outputs and using them to reverse-engineer the actual structure, color, and texture cues that the machine used to derive higher-order impressions (based on its previously trained emotion associations) and show that when faces are altered to resemble these features, humans make similar judgements. By manipulating the faces, we isolate the features available to human participants. Thus, in doing so, we can conclude that any systematic influences on impressions must be driven by the same computer derived emotion-resembling that humans are also using to make their judgements.

In summary, prior research has supported theoretical assertions that humans utilize emotion-resembling cues in neutral faces when deriving higher order impressions. The current work attempts to directly test this supposition, by first showing that computers can use emotion-resembling cues in the face to predict not just human impressions of emotionality, but of higher order person perception. Finally, by systematically manipulating the emotion cues that computers use to arrive at these impressions back into human faces, we aim to isolate the cues available to human impression formation, in order to confirm that the cues computers are using to predict human perception corresponds, at least in part, with the cues humans are using. In doing so, we are able to directly test theoretical assumptions of emotion overgeneralization driving human impressions by using new computer-driven technological advances.

STUDY 1

Study 1 applied a machine learning model trained on emotion expressions (see **Supplementary Material 1**) to neutral faces varying in gender. An additional purpose of Study 1 was to first assess the relative utility of using face metrics beyond structure to predict facial emotion, and then to apply the trained model to neutral faces to assess relative differences in the structure, color, and texture emotion outputs by face gender. Study 1 attempted to better understand previous gender-emotion overgeneralization results by utilizing a broader set of face metrics (in addition to structure) important for emotion classification, namely color, texture, and a combination model. Based on the prior literature, we expected that female faces would be found to resemble fear more than male faces, and that male faces would be found to resemble anger more. Behavioral findings have supported there also being a confound for female faces to resemble happiness more than males (Becker et al., 2007; Hess et al., 2009), but one prior computer vision study found the opposite, so we aimed to further examine this effect here.

METHOD

Model Training

The full procedure for training the structure, color, texture, and their combination models are reported in **Supplementary Material 1**. Briefly, structure, color, and texture metrics were extracted from the interior portion of the face for several thousand faces varying in emotion expressivity

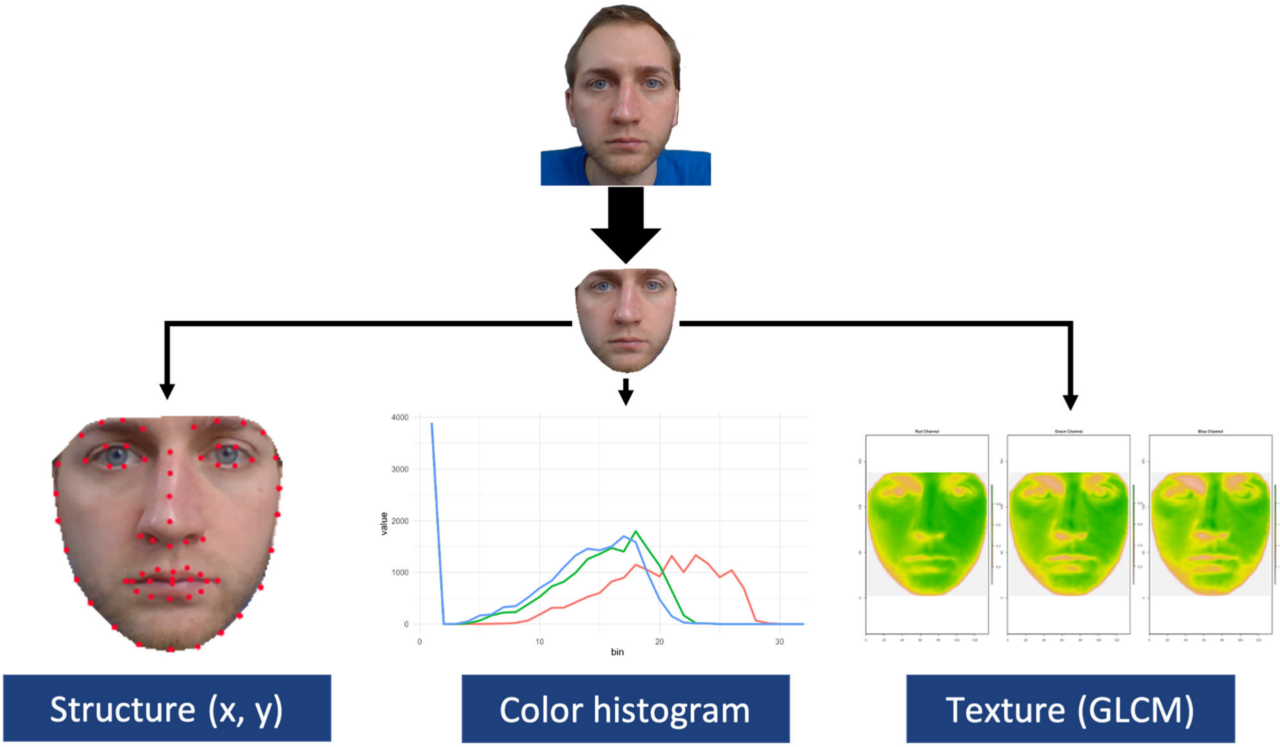


FIGURE 2 | Visual representation of feature extraction of structure, color, and texture from the internal portion of the face.

TABLE 1 | Standardized regression estimates for mediation models presented in Studies 2 & 3.

Mediation	Path estimates (standardized)				Indirect effect	
	a (se)	b (se)	c (se)	c' (se)	ab	95% CIs
STUDY 2						
Face gender → Anger output → Dominance ratings	0.19** (0.07)	0.32*** (0.07)	0.36*** (0.07)	0.30** (0.07)	0.06*	[0.01, 0.12]
Anger output → Masculine-Feminine → Dominance Ratings	0.26*** (0.06)	0.42*** (0.06)	0.38*** (0.07)	0.27*** (0.06)	0.11*	[0.05, 0.17]
Face gender → Happy output → Trustworthy ratings	0.19* (0.07)	−0.01 (0.07)	−0.36*** (0.07)	−0.36*** (0.07)	0	[−0.03, 0.03]
Happy output → Masculine-Feminine → Trustworthy ratings	0.19*** (0.07)	−0.45*** (0.07)	−0.08 (0.07)	0.01*** (0.07)	−0.09*	[−0.16, −0.02]
STUDY 3						
Face gender → Dominance output → Dominance Ratings	0.24** (0.07)	0.32*** (0.07)	0.36*** (0.07)	0.29*** (0.07)	0.08*	[0.03, 0.14]
Dominance output → Masculine-Feminine → Dominance ratings	0.30*** (0.07)	0.41*** (0.06)	0.39*** (0.07)	0.26*** (0.07)	0.12*	[0.06, 0.19]
Face gender → Affiliation output → Trustworthy ratings	−0.12 (0.07)	0.09 (0.07)	−0.36*** (0.07)	−0.35*** (0.07)	−0.01	[−0.04, 0.01]
Affiliation output → Masculine-Feminine → Trustworthy ratings	−0.16* (0.07)	−0.44*** (0.07)	0.13ᵗ (0.07)	0.06*** (0.07)	0.07*	[0, 0.14][b]

*[t] p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001; [b] This CI includes 0 due to rounding.*

(see **Figure 2**). Each model was trained on numerous machine learning algorithms, including a stacked ensemble of multiple models. Each model was assessed via its test accuracy, which was computed on a separate set of faces that each model had not been trained on. The best performing model was retained for each face metric. We also computed a weighted, combined model by summing each model's emotion output weighted by their test accuracy. Each metric model reached a test accuracy that was statistically above chance. However, the weighted, combined model reached an accuracy of nearly 90%, which was statistically higher than all of the other models, $\chi^2(0) = 101.49$, $p < 0.001$ (see **Table 1**). This suggests that each face metric uniquely contributed to overall performance. Each model, plus an interactive GUI (see **Figure 6** in Discussion), are available on the first author's website (https://www.daniel-albohn.com/) for research purposes.

Face Stimuli

White neutral faces varying in gender were taken from the Chicago Face Database (Ma et al., 2015). Neutral faces selected for this study were not used during the training or testing phase of the machine learning models (see **Supplementary Material 1**). Each individual face was subjected to the feature extraction procedure detailed above, resulting in structure, color, texture, and combined face metrics for each face. The automated feature extraction procedure led to a total sample of 93 male and 90 female neutral faces.

Computer Vision

For the sake of brevity, we only consider the weighted, combined model results in the main text but consider important results across all three metrics in the Discussion. However, the linear mixed effects regressions and significant pairwise comparisons for the structure, color, and texture individual models, as well as a graphical summary, are presented in **Supplementary Material 2**.

The weighted, combined model revealed stereotypical gender-emotion associations: Female faces resembled fear [$t_{(1,086)} = 8.5$, $p < 0.001$, CIs [0.09, 0.15]] more than male faces, and male faces resembled anger [$t_{(1,086)} = -3.8$, $p < 0.001$, CIs [−0.08, −0.03]] and happy [$t_{(1,086)} = -2.63$, $p = 0.009$, CIs [−0.06, −0.01]] expressions more than female faces. The full linear mixed effects model results for each metric are presented in **Supplementary Material 3**.

DISCUSSION

Study 1 used the trained machine learning models (see **Supplementary Material 1**) to examine how each face metric was related to a set of neutral faces. The results from Study 1 replicated previous computational and behavioral work examining emotion stereotypes related to gender.

Across structure and texture female faces resembled fear expressions to a greater degree than male faces. Similarly, the structure and texture of male faces resembled anger expressions to a greater degree compared with female faces.

Interestingly, the structure of female faces had slightly greater resemblance to happy expressions, but the color and texture and combination of all three metrics for male faces were more similar to happy expressions compared to female faces. This finding is particularly interesting given stereotypes related to women and smiling. Indeed, a slight smile on a woman is seen as neutral, whereas a similarly intense smile on a man is rated as happy (see, e.g., Bugental et al., 1971). It makes sense that of the three metrics, structure would show the expected gender stereotype of female neutral faces appearing more happy-like since facial landmarks pick up on gross shape information, whereas color and texture capture more nuanced differences (skin tone, aging cues, etc.). For example, if a female's "neutral" face is slightly smiling (i.e., minor upturned corners of the lips), the structure face metric is the most likely candidate to capture this change and use it in the model to make predictions. On the other hand, if someone has stereotypic happy-appearing cues such as "rosy cheeks" or natural "crows' feet" at the corners of their eyes, these are features that color and texture are likely to capture and utilize to make predictions. It is entirely possible that in our test set males had more of these subtle "smile-resembling cues" compared to females.

Overall, the models created replicated and extended previous research when applied to neutral faces varying in gender. Specifically, previous research has shown gender differences with regard to emotion expression resemblance, with male neutral faces more similar to anger expressions and female neutral faces more similar to fear expressions. The current work largely confirms this observation, but across more varied and specific face metrics.

STUDY 2

Study 1 examined each model's ability to predict human impressions of neutral faces varying in gender. The goal of Study 2 was to determine the utility of using individual and combined face metrics for predicting subtle emotional content in neutral faces. It is important to examine each model's predictive power on human ratings to assess each model's ecological validity. That is, an accurate machine learning model might classify expressions with a high precision, but it may not be able to classify/predict human responses to the same degree. This is particularly important for the current work as each model was specifically trained on low-level face metrics so that they might be able to detect subtle emotion cues in neutral faces. If each model is correlated with a corresponding and related human impression rating, it suggests that humans are–at least in part–using similar face features to make their judgements about the individual. In addition to examining correlational effects, Study 2 also examined whether structure, color, and texture resemblance to emotions are casual variables through mediation.

We established in the previous study that the trained models predicted the expected outcome of results for faces varying in gender, and thus only present the results from the weighted, combined model in the main text. Further, the weighted, combined model had the highest accuracy out of the three metrics examined, suggesting that it has the most utility in terms of predictive power. However, the results for each individual model (structure, color, and texture) are presented in **Supplementary Material 4** and considered in Study 2's Discussion.

METHOD

Face stimuli were the same as Study 1.

The Chicago Face Database supplies normed rating data from human participants on a number of different impressions and features. For example, each (neutral) face in the database was rated on how "feminine/masculine" the face appeared. Each neutral face was rated by a minimum of 20 raters ($M = 43.74$). This normed data has been successfully used in recent publications (see, e.g., Hester, 2018) with a high degree of success. The norming data supplied by Ma et al. (2015) in the Chicago Face Database serves as the human impression ratings for Study 2. Of all the suppled norming data, only a subset of emotions and impressions theoretically related to the present work and informed by gender-emotion stereotypes were examined

(Johnson et al., 2012; e.g., Adams et al., 2015). Specifically, anger-dominant and happy-trustworthy emotion-impression pairs were examined and interpreted in detail. Further, we examine the relationship between all algorithm emotion outputs and human impressions of dominance, trustworthiness, anger, and happy collapsed across gender via correlations.

Following procedures suggested by previous relevant work (e.g., Zebrowitz et al., 2010), emotion-trait pairs were examined via mediation to see if the machine learning model outputs for the weighted, combined model (and structure/color/texture in the **Supplementary Material**) mediated the relationship between actual face gender and human impressions. Additionally, we also examined whether perceived masculinity-femininity mediated the relationship between the weighted, combined model emotion output and human ratings. We focused on examining these two mediation models since each explains a different, theoretically important point related to the confounded nature of gender and emotion. Significant findings for the first mediation model (algorithm emotion output as the mediator) would suggest that faces varying in gender have different structure, color, and texture similarity to emotions, and that individuals use these cues that vary by gender to inform their impressions. On the other hand, significant findings for the second mediation model (perceived masculinity-femininity) would suggest that individuals are using structure, color, and texture resemblance to emotion expressions to guide their perceptions of gender, which in turn influence the perceivers' impressions of the face on related impressions.

For each mediation analysis gender was coded as 0 = "Female" and 1 = "Male." Perceived face gender was computed by taking supplied CFD masculinity and femininity ratings [$r_{(181)} = -0.96$], multiplying the femininity ratings by -1 and adding it to the masculinity ratings such that higher scores on the computed masculine-feminine were indicative of higher masculinity ratings, and lower scores were indicative of higher femininity ratings. All reported mediation indirect effects are estimated with 10,000 bootstrapped samples, and beta coefficients are standardized. Regression coefficients and standard errors for each mediation are reported in **Table 1**.

RESULTS

Dominance

The algorithm output for similarity to anger expressivity partially mediated the relationship between actual face gender and human ratings of dominance, beta = 0.06, CIs [0.01, 0.12]. Neutral male faces appeared more angry-like and faces that appeared more like angry expressions were subsequently rated as higher in dominance.

Similarly, masculine-feminine ratings partially mediated the relationship between algorithm anger output and dominance ratings, beta = 0.11, CIs [0.05, 0.17]. Neutral faces that the algorithm predicted to be more angry-like were rated higher in masculinity, and more masculine appearing faces were rated higher in dominance.

Trustworthy

The algorithm output for similarity to happy expressivity did not mediate the relationship between actual face gender and human ratings of trustworthiness, beta = 0, CIs [−0.03, 0.03].

However, masculine-feminine ratings mediated the relationship between algorithm happy output and trustworthiness ratings, beta = −0.09, CIs [−0.16, −0.02]. Neutral faces that the algorithm predicted to be more happy-like were rated higher in masculinity, but more masculine appearing faces were rated lower in trustworthiness ratings.

That the machine-derived happy output did not mediate the relationship between face gender and trustworthiness ratings, coupled with the fact that more happy-appearing faces were more masculine is a slightly unexpected finding. These results are discussed more thoroughly in the Discussion.

Correlations

Correlations between the weighted, combined model and human ratings largely revealed the predicted pattern of results. Machine-derived anger output positively correlated with dominance, masculine-feminine, and anger, and negatively correlated with trustworthy ratings. Interestingly, happy output only correlated with masculine-feminine ratings, with more masculine faces appearing happier. Further, happy output did not correlate with human ratings of happiness. While this may seem odd at first pass, it should be noted that these are correlations with the weighted, combined model.

While not directly related to the present set of studies, significant correlations for the other emotion outputs deserve mention. The more a face was predicted to be expressing disgust, the more masculine and less trustworthy it appeared, mirroring the correlations found for anger output. Similarly, predicted fear output negatively correlated with dominance, masculine-feminine, and anger, while positively correlating with trustworthiness ratings. Finally, sad and surprise output negatively correlated with dominance, while surprise output negatively correlated with masculine-feminine. Taken together, these additional emotion output correlations largely follow the pattern of results expected for gender-emotion stereotypes. Dominance emotions (anger, disgust) positively predict masculinity, dominance, and anger, whereas submissive emotions (fear, surprise) positively predict femininity and trustworthiness. **Figure 3** reports all of the correlations.

DISCUSSION

Study 2 used the trained models from Study 1 to examine the unique and combined ability to predict human-provided trait impressions of neutral faces varying in gender. Only the model output for resemblance to anger expressivity mediated the relationship between actual face gender and human ratings of dominance. Specifically, male neutral faces appeared more anger-like, and more anger-appearing neutral faces were rated higher in dominance. Conversely, happy expression resemblance did not mediate the relationship between face gender and ratings of trustworthiness.

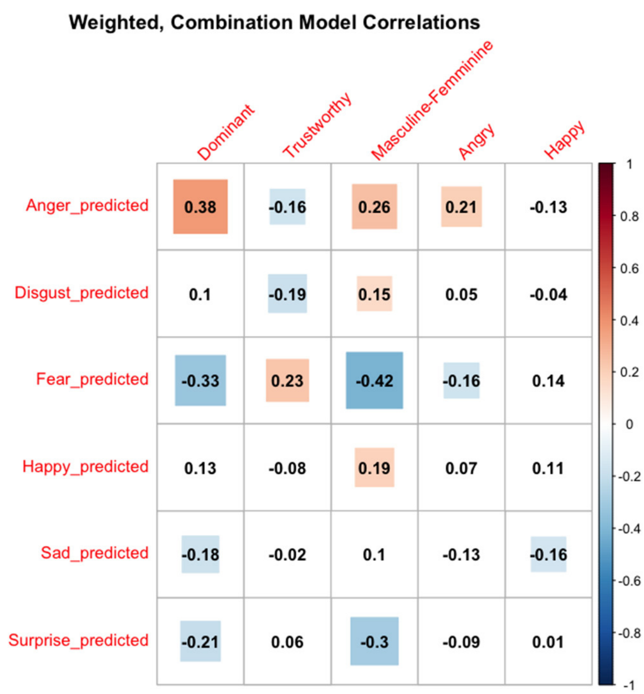


FIGURE 3 | Correlations between weighted combination model predictions and human impressions. Machine-derived output is on the Y-axis, whereas human ratings are on the X-axis. Colored cells are significant. Hot colors are significant positive correlations and cool colors are significant negative correlations.

On the other hand, the masculine-feminine ratings mediated the relationship between the model emotion output and human impressions. Specifically, regardless of gender, faces that appeared more angry-like across structure, color, and texture were rated higher in masculinity, and neutral faces rated higher in masculinity were perceived as more dominant. Further, across color, texture, and the weighted model faces that appeared happier were also rated as more masculine, yet the more masculine a face appeared the lower it was rated in trustworthiness. This is in line with previous work, and would suggest that the relationship between happiness and trustworthiness is moderated by actual gender or perceived masculinity-femininity (Adams et al., 2015).

Correlations between the model's anger output and human impressions revealed an expected pattern of results. Anger output positively correlated with dominance, masculine-feminine, anger, and happy ratings, and negatively correlated with trustworthy ratings. On the other hand, resemblance to happy expressions only predicted masculinity ratings.

An examination of the individual correlations for the structure, color, and texture model outputs revealed a similar pattern of results for anger and disgust; they were all correlated with dominant emotions and impressions. However, it was only happy structure that was significantly positively correlated with trustworthiness ratings. Happy color similarity was marginally correlated with happy human ratings, but not with trustworthiness ratings. Finally, happy texture appeared to be correlated with dominant emotions and impressions, much like anger and disgust (see **Supplementary Material 4** for the correlation charts for all three metrics).

The observation that faces that resembled happy expressions were rated as more masculine (and in the case of texture dominant-oriented impressions/emotions), yet these same faces were seen as less trustworthy is intriguing and deserves further speculation. These results may be due to an expectancy bias. That is, females are expected to display more happiness than males, and their neutral faces appear more similar to happy expressions. Thus, a female neutral face that appears happy-like across all the metrics examined would be seen as more "neutral" than a similar appearing male face (Fabes and Martin, 1991; Zebrowitz et al., 2010). Indeed, males that have neutral faces that resemble happy expressions might be granted particular attributes *because* they violate expectations. This observation makes logical sense: males who smile (or appear "smiley") will be rated as appearing more trustworthy than females or males who do not smile. This is consistent with a classic study conducted by Bugental et al. (1971; "Perfidious feminine faces") in which they found that children perceive their fathers' verbal messages to be friendlier and more approving when delivered with a smile, but no such effect was found for perceptions of their mothers. This finding shows the importance of considering both the phenotypic and stereotypic contributions to face derived trait impressions.

Despite the inconsistent results for happy/trustworthy with the combined model, it should be noted that the expected pattern emerged for the structure mediation model (see **Supplementary Material 4**). The color and texture output models showed a similar pattern of results as the weighted, combined model, again suggesting a similar confounded gender-emotion effect. Indeed, examining the correlations between happy expression similarity across the three face metrics and each gender revealed that there were few, and negative correlations for females (r's < 0.1, p's > 0.392), but there were meaningful relationships for males. Specifically, there were marginal-to-significant correlations between structure similarity ($r = 0.14$, $p = 0.182$) and color similarity ($r = 0.27$, $p = 0.009$) to happy expressions and human ratings of trustworthiness. These relationships only being present for males strengthens our argument that males may indeed by granted counter-stereotypic traits at a higher rate than females simply due to their gender.

Taken together, these results largely conform with previous work that has shown similar emotion-overgeneralization results between face gender and structural similarity to emotion expressions (e.g., Zebrowitz et al., 2010). However, these results extend previous work by showing that there is a causal relationship between the face metric (structure, color, and texture) similarity to expressions and gendered impressions.

Overall, the models created in Study 1 replicated and extended previous research that has shown actual and perceived gender differences with regard to emotion expression expectations and expression resemblance, with the most robust effect in our data being that male neutral faces were more similar to anger expressions, and thus rated as more dominant. The predicted relationships between emotion expression resemblance and impressions also occurred, with the largest effects seen for anger and disgust similarity increasing power-oriented impressions, and fear and surprise resemblance increasing attributions of submission impressions. In sum, results across neutral face similarity to specific emotions and its relationship to human impressions replicates and extends past work.

STUDY 3

Study 2 showed that the emotion output from the trained models were related to human impressions, and in the case of dominance was even causal in explaining the relationship between gender and stereotypic human impressions. Given the significant findings of Study 2, Study 3 aimed to show that higher-order impressions can be algorithmically computed from the emotion output of the machine learning models and that these algorithmically-computed impressions are related to human impressions.

Research across multiple decades suggests that two powerful dimensions of human impression formation are dominance/power and affiliation/trustworthiness. For example, Knutson (1996) extended Wiggins' interpersonal circumplex (Wiggins et al., 1988) to show that emotion expressions fall within a two-dimensional dominance/affiliation face space. Later, Todorov et al. (2008) showed that specific traits could be represented within a two-dimensional dominance and trustworthy face space. Further, Todorov et al. (2008) showed that computer-generated neutral faces at the extremes of the trustworthy dimension mimicked expressive features: +3 SD trustworthy neutral faces appeared happy, and −3 SD trustworthy neutral faces appeared angry. Given the importance of these dimensions to person perception research, Study 3 focuses on the impressions of dominance and trustworthiness/affiliation.

METHOD

Stimuli were the same male and female neutral faces used in Study 2 from the CFD face database (Ma et al., 2015).

Human impressions used in Study 3 were the same as those used in Study 2 (i.e., provided by the CFD face database). The machine-derived impressions were computed from the emotion output of the weighted, combined model detailed in Study 1 (see below).

Dominance and affiliation were algorithmically computed for each neutral face and derived from the emotion resemblance metrics. Estimation of dominance and affiliation was accomplished by following a similar procedure reported by Knutson (1996), but in reverse. Whereas, Knutson (1996) computed the spatial location of each emotion expression in dominance/affiliation face space, here the opposite approach was taken. Specifically, the face space emotion expressions scores found by Knutson (1996) (see **Table 2**) were multiplied by the emotion expression output from the weighted, combined machine learning model built in Study 1 to algorithmically project each neutral face onto a two-dimensional dominance by affiliation social face space.

TABLE 2 | The relative dominance and affiliation values for each emotion expression found by Knutson (1996).

Face space	Anger	Disgust	Happy	Fear	Sad	Surprise
Dominance	1	0.6	1	−0.5	−1	−0.5
Affiliation	−1.5	−1	2	0.5	0.1	0.1

Specifically, dominance, $\hat{D}$, was computed by multiplying the emotion expression value by the dominance value (y axis) found by Knutson (1996), Y_d, with the weighted, combined model emotion output, I_j, and summing across all emotions, i, such that

$$\hat{D} = \sum_{i=1} Y_{id} * I_{ij}$$

Similarly, affiliation, $\hat{A}$, was computed by multiplying the emotion expression value by the affiliation value (x axis) found by Knutson (1996), X_a, with the weighted, combined model emotion output, I_j, and summing across all emotions, i, such that

$$\hat{A} = \sum_{i=1} X_{ia} * I_{ij}$$

RESULTS

Like Study 2, algorithmically computed dominance and affiliation were assessed to see if they mediated the association between human ratings of dominance/trustworthiness and gender, and if perceived masculinity-femininity mediated the relationship between the algorithmically-derived impressions and human ratings. **Table 1** presents the standardized regression coefficients for each mediation.

Dominance

Algorithmically computed dominance scores partially mediated the relationship between face gender and dominance ratings, beta = 0.08, CIs [0.03, 0.14]. Neutral male faces appeared more dominant and as faces appeared more dominant-like within face space they were subsequently rated as higher in dominance.

Similarly, human masculine-feminine ratings partially mediated the relationship between algorithmically computed dominance scores and human ratings of dominance, beta = 0.12, CIs [0.06, 0.19]. Regardless of actual gender, faces that were higher on algorithmically-derived dominance were perceived by humans to be higher in masculinity, and faces higher in masculinity were perceived by humans to be higher in dominance.

Affiliation

Algorithmically computed affiliation scores did not mediate the relationship between face gender and trustworthy ratings, beta = −0.12, CIs [−0.04, 0.01]. However, human ratings of masculinity-femininity did mediate the relationship between algorithmically computed affiliation scores and human ratings of trustworthiness, beta = −0.16, CIs [0.003, 0.14]. Specifically, regardless of actual face gender, faces that were higher on computed affiliation were seen as less masculine, and more masculine-appearing faces were rated lower on trustworthiness. In other words, faces that were higher in computed affiliation were seen by humans as more feminine, and more feminine faces were rated as overall more trustworthy.

DISCUSSION

Study 3 showed that the emotion output from the machine learning models could be used to algorithmically derive higher-order impressions that were meaningfully related to similar human impressions. Indeed, algorithmically computed dominance scores mediated the relationship between face gender and human ratings of dominance. Further, human ratings of masculine-feminine mediated the relationship between the machine outputs of dominance and affiliation, and dominance and trustworthiness, respectively. Together with the results from Study 2, this suggests that humans are partially using facial metric features in the face to derive their impressions of dominance and trustworthiness ratings of neutral faces. This conforms with research that suggests emotion expressions are a powerful mechanism that individuals use to form impressions of others, particularly when other information is absent, as is the case with neutral faces (see, e.g., Zebrowitz et al., 2010; Adams et al., 2012; Albohn et al., 2019; Albohn and Adams, 2020a).

One reason why algorithmically computed affiliation did not mediate the relationship between face gender and trustworthiness may be due to how the scores were algorithmically derived. That is, the dominance model had matching algorithmically computed and human impressions, whereas the affiliation model had algorithmically computed affiliation scores but human trustworthiness scores. While affiliation and trustworthiness are highly correlated (and in some cases used interchangeably), this may have added additional noise to the model. It is also possible, given the findings of Study 1 and 2 whereby masculinity appears to be related to perceptions of masculinity, that the affiliation effects based on emotion are more tenuous. Future work should account for this shortcoming by algorithmically calculating trustworthy scores or having humans rate faces on affiliation. Similarly, stronger models may allow for more nuanced relationships to emerge.

STUDY 4

Study 2 showed the relative and combined importance of each face metric on various human-provided impressions, and Study 3 showed that the emotion output from the machine learning models could be used to algorithmically estimate higher order impressions. Study 4 attempted to extend these results by using an experimental design through which neutral faces were psychophysically manipulated to appear more or less like a specific impression in a systematic manner. The goal of the psychophysical manipulation was to physically manipulate the face stimuli with explicit intention of creating psychological changes in the subjective perception of the participants.

Study 4 aimed to show that higher-order impressions can be algorithmically computed from the emotion output of machine learning models and that these algorithmically-computed impressions are related to human impressions. Without direct comparisons across trait dimensions and within face identity, one cannot make any conclusions about whether it was the individual face features (as assessed via machine learning) that drive human impressions formation.

It was predicted that when a neutral face is transformed to appear more like a neutral face that has been classified by a machine learning algorithm into dominance/affiliation face space as based on resemblance to specific expressions (e.g., angry/happy) it will be rated by humans in a similar manner based on the same physical properties (structure, color, texture) of the face that the machine used to make its classification.

METHOD

Study 4 manipulated faces to appear more like faces that would appear in Wiggins' circumplex using dominance and affiliation as criterion for this procedure. Again, dominance and affiliation were selected due to the fundamental nature of these attributes on person perception (Knutson, 1996; Todorov et al., 2008).

Using dominance-affiliation face space has the added benefit of each quadrant of the face space roughly corresponding to a specific and dissociable impression that can then be derived from the face. For example, the upper-left quadrant corresponds to a high dominance, low affiliative face and these faces are often rated as appearing angry (for review and examples see, Knutson, 1996).

Participants

A total of 216 (91 male, 122 female, one gender nonconforming, 2 other; $M = 20.05$, $SD = 4.19$) undergraduate participants participated online in the study in exchange for course credit. **Table 3** reports participant race and ethnicity. Participants were allowed to select all that applied.

Stimuli and Transformation Procedure

White neutral faces were extracted from the CFD (Ma et al., 2015), FACES (Ebner et al., 2010), NIMSTIM (Tottenham et al., 2009), RAFD (Langner et al., 2010), MR2 (Strohminger et al., 2016), and FACES (Minear and Park, 2004) databases resulting in a total of 446 neutral faces. Each neutral face was subjected to the same pipeline as reported in Study 1 & 3 to calculate the predicted emotion expression classification values using the weighted, combined model and dominance and affiliation scores.

After dominance and affiliation scores were computed for each face, this information was used to calculate each faces' distance and angle from the origin in face space. Euclidean distance, $\hat{E}$, was calculated as

$$\hat{E} = \sqrt{(x_2 - x_1)^2 + \left(y_2 - y_1\right)^2}$$

where x_2 and y_2 were set to the origin, or $(0,\ 0)$, and x_1 and y_1 were set as affiliation and dominance values, respectively. The

TABLE 3 | Participant demographics for Study 4.

Race	Latinx/Hispanic	Not latinx/Hispanic	Not reported
Asian	0	20	0
Black	2	10	0
Black, White, Asian, Native American/American Indian	0	1	0
White	8	158	0
White, Asian	0	6	0
White, Asian, Native Hawaiian/Pacific Islander	0	1	0
White, Native American/American Indian	1	0	0
White, Other	1	0	0
Native American/American Indian	0	1	0
Not Reported	0	0	1
Other	2	4	0

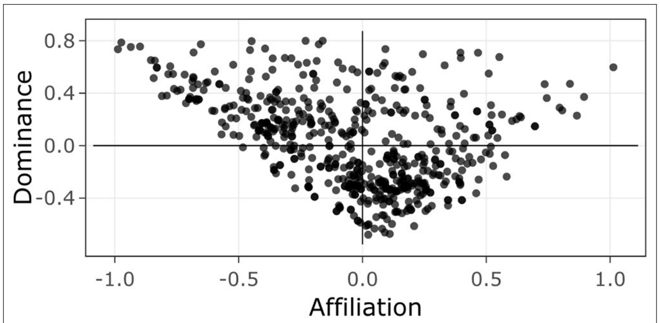


FIGURE 4 | Distribution of neutral faces in computed dominance (y) and affiliation (x) social face space.

angle from the origin, θ, was calculated by first computing the radians from the arctangent function via

$$\theta = atan2(y,\ x)$$

where x and y were each faces' affiliation and dominance values, respectively, and then converted to degrees. **Figure 4** shows the distribution of faces within each quadrant.

Faces were selected for the transformation procedure following a stepwise process. First, four male ($M = 0.14$) and four female ($M = 0.08$) neutral faces from the CFD database were selected by locating faces that had the shortest distance from the origin ([0, 0] in computed face space). Next, these eight faces were morphed with a random subset of neutral faces that fell within the middle range of each quadrant of the calculated face space.

Each of the eight neutral faces was morphed with a randomly selected neutral face from each quadrant. This process was repeated four times for each of the eight faces, resulting in 240 total neutral faces. Faces were selected from the larger set if 1) they had a distance from the origin >0.15, and 2) fell within the middle portion of each quadrant based on their angle from the origin. These parameters were adopted in order to guarantee that each face fell within a reasonable position within the computed face space (i.e., not too close to the origin, or near the edges of the face space quadrant). Specifically, faces in quadrant one (upper right) were between 10 and 80°. Faces in quadrant two (upper left) were between 100 and 170°. Faces in quadrant three (lower left) were between 190 and 260°. Faces in quadrant four (lower right) were between 190 and 350°.

After each neutral face was classified into a quadrant, the top 20 male and 20 female neutral faces from each quadrant were placed into a pool to be randomly selected for transformation with the close-to-origin neutral faces identified in the previous step. Similarly, the top 20 anger and 20 joy neutral faces were acquired from the full set of images by taking the highest rated neutral faces for each expression and gender.

The structure, color, and texture of the randomly selected images from each quadrant, as well as anger and happy, were transferred onto the close-to-origin neutral faces at a 50–50 split using PsychoMorph v.6 (see Tiddeman et al., 2005). Close-to-origin and randomly selected neutral faces were gender matched before each transform. After transformation, the faces were cropped to a standardized size and visually inspected for artifacts. Unrealistic appearing images were manually manipulated to appear more genuine or were discarded. Manual inspection of the images reduced the final set of manipulated and morphed images to 221. Example morphed images are shown in **Figure 5**.

Participant Procedure

Participants completed the rating portion of the study using an online participant recruitment platform run by the authors' host University. Participants were randomly presented with 70 faces from the full set of 221 faces. Each face was presented focally with a Likert-type scale underneath the image ranging from "1—Not at all" to "7—Very much." Participants were asked to rate each face on how much each person appeared "angry," "happy," "trustworthy," "dominant," "healthy," "attractive," "babyish," and "smart." At the beginning of the experiment participants read instructions that included a short definition of each trait. Each trial consisted of rating each face on all of the traits before moving onto the next transformed image. Individual ratings were separated by a 100 ms fixation cross. Ratings for each stimulus were randomized between stimuli, and participants were instructed when a new face rating block was about to occur. Participants then filled out basic demographic information before debriefing and returned to the online participation platform.

RESULTS

Only the results for the four quadrants are presented in the main text. Results for anger and happy psychophysical transforms are presented in **Supplementary Material 5** but considered in the Study Discussion.

Data Preprocessing and Analysis Strategy

Due to the nature of online studies (e.g., lack of accountability, inattention, etc.), some preprocessing of the data was necessary. First, participants were dropped from analyses if their responses had little variability across all of the trials, specifically a standard deviation <0.4 ($n = 2$). Second, individual participant trials were eliminated if the trial reaction time was <50 ms (< 6.6%) or >10,000 ms (< 1.8%). These two elimination criteria ensured that only reliable responses from participants who were paying attention were analyzed. Additionally, 1 was serially subtracted from each participants' individual trial response so that responses ranged from 0 to 6, allowing for greater interpretability.

Results are reported as separate linear mixed-effects regression models for each quadrant because analyzing a full model that contained every possible comparison (six quadrants by seven trait ratings by two face genders) would result in 84 comparisons and significantly reduce power. Each model compares ratings of the original, close-to-origin neutral face with ratings for a specific quadrant. The models include fixed effects for each block/valence (i.e., impression rating) and stimulus gender and random effects for each participant. Lastly, the models include each image's average attractiveness ratings as covariates (see, Zebrowitz et al., 2010). **Supplementary Material 6** reports the full linear mixed effects regressions for each quadrant.

FIGURE 5 | Examples of transformed stim rated for dominance (top) and trustworthiness (bottom). Left column depicts transforms rated low in the trait. Right column depicts transforms depicted high in the trait.

Quadrant I: High Dominance, High Affiliation

As expected, there was a main effect of rating, $F_{(6,22183.79)} = 102.11$, $p < 0.001$, $\eta^2 = 0.41$. No other main effects were significant. However, there were significant interactions. First, there was an interaction between face type and rating [$F_{(6,22183.79)} = 6.33$, $p < 0.001$, $\eta^2 = 0.03$]. Quadrant I neutral transforms were rated higher than close-to-origin neutral faces on dominance, estimate = −0.19, $SE = 0.08$, $t_{(162.76)} = -2.25$, $p = 0.026$, CIs [−0.35, −0.02], and happy, estimate = −0.19, $SE = 0.08$, $t_{(162.76)} = -2.3$, $p = 0.023$, CIs [−0.36, −0.03]. Quadrant I faces were also rated lower on babyishness, estimate = 0.28, $SE = 0.08$, $t_{(162.76)} = 3.37$, $p < 0.001$, CIs [0.12, 0.45].

There was also a significant interaction between impression rating and gender [$F_{(6,22183.79)} = 17.31$, $p < 0.001$, $\eta^2 = 0.07$]. Female faces were rated higher than male faces on babyishness, estimate = 0.34, $SE = 0.08$, $t_{(163.13)} = 4.08$, $p < 0.001$, CIs [0.18, 0.5], and trustworthiness, estimate = 0.17, $SE = 0.08$, $t_{(163.13)} = 2.02$, $p = 0.045$, CIs [0, 0.33]. Male faces were rated higher in dominance than female faces, estimate = −0.5, $SE = 0.08$, $t_{(163.13)} = -6.03$, $p < 0.001$, CIs [−0.67, −0.34].

Finally, there was a three-way interaction between face type, impression rating, and gender [$F_{(6,22183.79)} = 4.91$, $p < 0.001$, $\eta^2 = 0.02$]. Examining this three-way interaction revealed a number of gender-specific interactions. Specifically, quadrant I female faces were rated as less babyish than close-to-origin neutral faces, estimate = 0.35, $SE = 0.12$, $t_{(159.4)} = 3$, $p = 0.003$, CIs [0.12, 0.58]. Quadrant I female faces were also rated as happier [estimate = −0.32, $SE = 0.12$, $t_{(159.4)} = -2.75$, $p = 0.007$, CIs [−0.55, −0.09]], healthier [estimate = −0.23, $SE = 0.12$, $t_{(159.4)} = -1.93$, $p = 0.056$, CIs [−0.46, 0.01]], and (marginally) more trustworthy [estimate = −0.23, $SE = 0.12$, $t_{(159.4)} = -1.93$, $p = 0.055$, CIs [−0.46, 0]] than their close-to-origin counterparts. Male quadrant I faces were only rated as more dominant than close-to-origin neutrals, estimate = −0.28, $SE = 0.12$, $t_{(166.16)} = -2.34$, $p = 0.021$, CIs [−0.51, −0.04].

Quadrant II: High Dominance, Low Affiliation

There was a main effect of rating, $F_{(6,22652.59)} = 95.24$, $p < 0.001$, $\eta^2 = 0.44$. No other main effects were significant. However, there was a significant interaction between rating and gender [$F_{(6,22652.59)} = 18.73$, $p < 0.001$, $\eta^2 = 0.09$]. Female faces were rated higher than male faces on babyishness, estimate = 0.36, $SE = 0.09$, $t_{(144)} = 4.08$, $p < 0.001$, CIs [0.18, 0.53], and health, estimate = −0.29, $SE = 0.09$, $t_{(144)} = -3.32$, $p < 0.001$, CIs [−0.46, −0.12]. Male faces were rated higher in dominance, estimate = −0.53, $SE = 0.09$, $t_{(144)} = -6.11$, $p < 0.001$, CIs [−0.71, −0.36].

While there was not a three-way interaction between face type, rating, and gender [$F_{(6,22652.59)} = 0.52$, $p = 0.796$, $\eta^2 = 0$], *post hoc* exploratory analyses were still computed and analyzed. As expected, male quadrant II transforms were rated higher on dominance than close-to-origin male faces, estimate = −0.3, $SE = 0.12$, $t_{(146.35)} = -2.42$, $p = 0.017$, CIs [−0.55, −0.06]. No other pairwise comparisons reached significance.

Quadrant III: Low Dominance, Low Affiliation

There was a main effect of rating, $F_{(6,21041.39)} = 85.26$, $p < 0.001$, $\eta^2 = 0.42$. No other main effects were significant. There was a significant interaction between rating and gender [$F_{(6,21041.39)} = 25.28$, $p < 0.001$, $\eta^2 = 0.13$]. Female faces were rated higher than male faces on babyishness, estimate = 0.53, $SE = 0.08$, $t_{(201.08)} = 6.81$, $p < 0.001$, CIs [0.38, 0.69], and health, estimate = −0.22, $SE = 0.08$, $t_{(201.08)} = -2.76$, $p = 0.006$, CIs [−0.37, −0.06]. Male faces were rated higher in dominance, estimate = −0.52, $SE = 0.08$, $t_{(201.08)} = -6.6$, $p < 0.001$, CIs [−0.67, −0.36].

Finally, there was a three-way interaction between face type, rating, and gender ($F_{(6,21041.39)} = 2.53$, $p = 0.019$, $\eta^2 = 0.01$). Quadrant III female faces were marginally rated as more happy than close-to-origin neutral faces, estimate = −0.19, $SE = 0.11$, $t_{(197.7)} = -1.73$, $p = 0.085$, CIs [−0.41, 0.03]. Quadrant III male faces were only rated as angrier [estimate = −0.22, $SE = 0.11$, $t_{(203.62)} = -1.97$, $p = 0.05$, CIs [−0.44, 0]] and more dominant [estimate = −0.26, $SE = 0.11$, $t_{(203.62)} = -2.38$, $p = 0.018$, CIs [−0.48, −0.05]] than their close-to-origin counterparts.

Quadrant IV: Low Dominance, High Affiliation

For quadrant IV faces there was a main effect of rating, $F_{(6,21871.16)} = 91.05$, $p < 0.001$, $\eta^2 = 0.4$. No other main effects were significant. However, there was a significant interaction between rating and gender [$F_{(6,21871.16)} = 17.91$, $p < 0.001$, $\eta^2 = 0.08$]. Female faces were rated higher than male faces on babyishness, estimate = 0.39, $SE = 0.09$, $t_{(152.34)} = 4.52$, $p < 0.001$, CIs [0.22, 0.56], and marginally higher on trustworthiness, estimate = 0.14, $SE = 0.09$, $t_{(152.34)} = 1.66$, $p = 0.099$, CIs [−0.03, 0.31]. Male faces were rated higher in dominance, estimate = −0.49, $SE = 0.09$, $t_{(152.34)} = -5.71$, $p < 0.001$, CIs [−0.66, −0.32], and marginally higher on health, estimate = −0.17, $SE = 0.09$, $t_{(152.34)} = -1.96$, $p = 0.052$, CIs [−0.34, 0].

Finally, there was a three-way interaction between face type, rating, and gender [$F_{(6,21871.16)} = 3.14$, $p = 0.004$, $\eta^2 = 0.01$]. Examining this three-way interaction revealed one significant pairwise comparison: Quadrant IV female faces were rated as healthier than close-to-origin neutral faces, estimate = −0.24, $SE = 0.12$, $t_{(149.02)} = -1.98$, $p = 0.049$, CIs [−0.48, 0].

DISCUSSION

Study 4 experimentally and psychophysically manipulated the structure, color, and texture of neutral faces with other faces that were reliably and highly categorized by the weighted, combined model as resembling a specific expression or a given trait (within an estimated two-dimensional face space model based on dominance and affiliation). It was predicted that when a neutral face was experimentally manipulated to structurally, color-wise, and texturally resemble a different neutral face that highly resembled an expression/trait that it would be perceived in much the same manner as if it had an overt expression or was explicitly rated as high in that trait.

Results across four quadrants within the face-space dimension largely supported the predicted pattern of results. Specifically, quadrant I transform faces (high dominant, high affiliative) were rated as more dominant, happy, and less babyish than unaltered neutral faces. These results were largely driven by male transforms being rated as more dominant, and female faces being rated as happier and less babyish. These results largely follow from the observation that quadrant I faces typically appear more "smiley" and appear to have healthy skin color (carotenoids; see, e.g., Perrett et al., 2020). These effects are further corroborated by the results from the happy transforms (see **Supplementary Material 6**). Happy transforms–particularly female faces–were rated as healthier, more trustworthy, and more intelligent. These results follow previous research that shows that individuals who express positive emotions are endowed with more positive traits (e.g., the halo effect).

Quadrant II male transforms (high dominance, low affiliative) were rated as more dominant than unaltered faces. Similarly, male anger transforms were rated as more dominant and less babyish than unaltered neutral faces (see **Supplementary Material 6**). It appears that the structure, color, and texture of neutral faces that appear more anger-like influence ratings of dominance more than directly changing ratings of anger. This may be due to the fact that making an emotion expression judgment about a neutral face is harder than making a higher order impression judgment. That is, rating a neutral appearing face on dominance is easier for participants than rating a neutral face on how "angry it appears" because–by virtue of its definition–a neutral face is low in emotional expressivity.

Quadrant III male transforms (low dominance, low affiliation) were rated as angrier and more dominant than unaltered neutral faces. This pattern of results is not entirely unexpected, as male faces are more dominant to begin with. Thus, observers might only be using (negative) valence information to make judgements about the faces. Lastly, quadrant IV female transforms (low dominance, high affiliation) were rated as less healthy than unaltered neutral faces.

Despite significant results, there are still a number of limitations that deserve discussion. One unexpected finding was that anger transforms were not rated as less trustworthy than unaltered neutral faces. This may be due to the fact the faces chosen were less trustworthy to begin with before manipulation. Indeed, neutral faces are often rated as more negative to begin with (Lee et al., 2008). Again, this is underscored by the observation that only happy transforms were rated as significantly higher in trustworthiness. Further, re-running the current analyses with more stimuli or fewer, more direct comparisons could raise the power of each model. This would help to raise the significance of marginally significant comparisons, or non-significant comparisons that are in the correct and predicted direction. Indeed, male anger transforms were rated as numerically *less* trustworthy than close-to-origin neutral faces, yet this comparison failed to reach significance [estimate = −0.13, $SE = 0.12$, $t_{(138.34)} = -1.08$, $p = 0.282$, CIs [−0.37, 0.11]].

TABLE 4 | Summary of findings and significance across the four studies presented in the current research.

Study	Finding	Significance
1	Face **structure**, **color**, and **texture**, and their **weighted combination** are reliable predictors of facial affect; each metric varies by gender in an expected manner	**Validates** model; **extends previous work** showing gender differences in facial structure to texture and color as well
2	All three metrics **correlate** with, and in some cases, **mediate the relationship** between face gender and **human impressions**	Provides correlational evidence that the **metrics used by machine learning** to predict emotions **relates to human impressions** in an expected manner
3	**Algorithmically-derived** impressions of dominance and affiliation are related to **human impressions** of dominance and trustworthiness	Demonstrates that **higher-order impressions can be derived from machine learning** output trained on emotions
4	**Algorithmically-derived** impressions can be used to **reverse-engineer** important **structure, color**, and **texture** features in neutral faces	**Experimentally demonstrates** that metrics machine learning models use to **predict emotions** are also **used by humans** to **form impressions**

GENERAL DISCUSSION

The present research proposed an extension of previous computer vision work that has examined the structural resemblance of neutral faces to specific expressions and personality traits (Said et al., 2009; Zebrowitz et al., 2010). While previous work has examined the utility (**Figure 1**, Level 1) and correlational relationships (**Figure 1**, Level 2) between machine learning output and human impressions, no work to our knowledge has taken machine learning output and constructed higher-order impressions (**Figure 1**, Level 2) or used that output to manipulate faces to experimentally show that humans are using similar metrics as machines to form impressions (**Figure 1**, Level 3), let alone an end-to-end experimental pipeline. To this end, our work adds novel insight into a growing body of literature which shows that emotion expressions are a powerful mechanism of person perception (see **Table 4** for a summary of findings and significance).

Across four studies we replicated and extended prior work by showing that similarities in structure, color, and texture (as well as their weighted combination) to expressions vary across neutral facial appearance associated with actual and perceived face gender in a largely stereotypic manner. Further, this work provides evidence that all three face metrics examined (plus their combination) predict human impressions of emotionally neutral faces similar to what would be expected from overt expressions. Finally, in a test of this experimentally, we showed that when neutral faces are psychophysically manipulated to alter their structure, color, and texture they yield similar patterns of impression biases, underscoring that each feature the algorithms used and learned to make accurate predictions was–at least in part–what was used by humans to arrive at similar judgements.

Study 1 introduced four machine learning models that were able to accurately predict emotion expressions from the structure, color, and texture of faces. These face metrics were selected to ensure that the models would be able to use low-level features to predict the expressive content of faces that were minimally- or non-expressive. That is, it was predicted that training models to use fundamental face metrics such as structure, color, and texture would create models more sensitive to the emotional content of faces expressing little or no emotion. All of the metric models performed with above chance levels of accuracy on a separate test set of expressions. Combining all three metrics into a single weighted model yielded the highest accuracy. The combined emotion recognition accuracy of these models was nearly 90%, statistically significantly higher than any of the three models individually, suggesting that each metric and their features uniquely contributed to performance of emotion recognition.

Study 1 also showed that there were structure, color, and texture differences across neutral faces that varied by gender. Overall, the results from Study 1 suggest that a more holistic view of person perception can be gained by examining individual face metrics/features as well as their combination. Male faces showed greater resemblance to power-oriented expressions (e.g., anger, happy) across each metric and female faces showed greater resemblance to fear expressions across each metric.

Study 2 revealed that structure, color, and texture resemblance to emotion expressions were related to human impressions in a similar gender-stereotypic manner: resemblance to anger and disgust expressions predicted power-oriented impressions, while resemblance to fear and surprise expressions predicted affiliative-oriented impressions. Similarly, the model output for anger expressions mediated the relationship between face gender and ratings of dominance, while human ratings of masculinity-femininity mediated the relationship between model outputs of anger/happy and dominance/trustworthiness, respectively.

Study 3 revealed that the emotion output from the trained machine learning models could be used to calculate higher-order impressions of neutral faces, and that these impressions were causally related to similar human judgements of the face. Specifically, algorithmically-derived dominance acted as the mediator between face gender and human ratings of dominance, and perceived gender acted as the mediator for both algorithmically-computed dominance/affiliation and human ratings of dominance/trustworthiness, respectively. In sum, Study 3 showed that dominance and affiliation could be reliably computed from the anger, disgust, fear, happy, sad, and surprise emotion output of the machine learning models, and that these algorithmically computed scores were related to similar human impressions through perceived face gender.

The results of Study 2 and 3 are important for several reasons. First, it showed that the emotion output from the models is meaningful and interpretable (i.e., not a “black box”). Second, it showed that humans are partially using emotion resemblance across all three face metric channels to make impressions of neutral faces. Having a tool that can predict subtle emotionality and impressions of neutral faces is an important tool for researchers and practitioners of affective science.

Lastly, Study 4 showed that when neutral faces were manipulated to resemble the structure, color, and texture of high and low dominance by high and low affiliation, anger, and happy expressions they were subsequently rated in a manner similar to faces naturally high/low in such attributes. These results provide the first experimental evidence showing that when faces are systematically manipulated to possess structure, color, and texture features of faces that incidentally or naturally have such features, they are judged in a similar, stereotypic manner. These results provide further evidence that individuals use fundamental face metrics–either separately or in combination–to make impression judgments of minimally- or non-expressive faces.

The results from Study 4 appear to be largest for quadrant I, II, anger, and happy transformed faces. This pattern of results is most likely due to both quadrants and both expressions being high dominant and high arousal, resulting in structure, color, and texture features that may be easier to identify. In conclusion, multiple comparisons between psychophysically manipulated and unaltered neutral faces support the primary hypothesis that face features the models learned are the same features that humans use to make impressions. These results experimentally replicated the correlational results reported in Studies 1 & 2.

Together, four studies highlight that while social visual perception at times can be accurate, emotion resembling features of the face can bias impressions and contribute to stereotypic evaluations (See Adams et al., 2017 for discussion). While there are myriad cues in the face beyond structure, color, and texture that can influence impressions, we believe that this is an important first step at disentangling the fundamental face metrics that appear to be influencing a perceivers' visual perception when making judgements. Despite not being able to specifically say what metrics are related to which impressions, the current results can definitely state that individuals appear to be using, at least partially, the structure, color, and texture cues related overt emotion expressions when judging others' faces.

The present results confirm that in addition to facial structure, color and texture related to emotion expressions are also important cues individuals use to make decisions. Research that has examined these cues in isolation demonstrated that they impact perceptions related to attractiveness and health (Pazda et al., 2016; Thorstenson et al., 2017; Perrett et al., 2020) as well as gender (Russell, 2003, 2009). It is certainly possible face color and texture influence impressions related to gender or health in a largely associative manner, much like how face structure has been shown to influence impressions via emotion overgeneralization (see, Heerey and Velani, 2010; Kocsor et al., 2019). That is, health influences mood (see, Yeung, 1996 for review), and therefore healthy individuals are likely associated with positivity. In so much as individuals are able to extract and associate specific color and texture cues with health, it may provide reciprocal feedback associations that can give rise to specific stereotypes. Indeed, neutral faces that were transformed to appear happier in Study 4 were rated as healthier, suggesting a correspondence between expressive cues across structure, color, and texture and health.

This work also suggests that computer vision techniques can be used to successfully extract and predict the emotional content of faces. Further, it was shown that machine learning models can be used to predict emotionality from minimally-expressive and non-expressive faces. Indeed, it appears as though fundamental face metrics, including structure, color, and texture, can be used to make meaningful predictions about such faces. Examining face metrics separately allows for parsing the relative (and combined) contribution each has to face perception.

Reachable and Replicable Science

Machine learning is an important and rapidly expanding field of research within behavioral science. However, despite advances in the field it remains relatively inaccessible to non-programmers. Commercial applications have made strides in making machine learning programs straightforward for end-users by providing graphical user interfaces (GUIs) that allow for simple point-and-click operations (iMotions, 2019; Noldus Information Technology, 2019). However, most freely available and open source machine learning algorithms for person perception require some degree of coding or technological expertise. It is imperative that these tools and resources be available to all parts of the scientific community in order to advance research forward and answer both novel and old questions in new ways in a timely manner.

To this end, the current work includes an open source GUI (See **Figure 6**) written in R, JavaScript, and Python to utilize the structure, color, texture, and combined models in a point-and-click manner. The GUI is packaged as a shiny application residing in a Docker image, allowing for complete containerization (i.e., replicability across machines) so long as the end-user utilizes Docker on the host machine. The GUI allows for the user to easily upload pictures from their machine to be analyzed by each algorithm with moderate control over input parameters, such as the weight of each model in the aggregate predictive model. In addition to calculating expression estimates, the user is able to visualize each feature and obtain a computational estimate of where each face exists in the predicted two-dimensional social face space for every type of model. The user also has access to the computed data and can download it at any time straight from the GUI. All uploaded data to the app remains on the user's host machine, and no data is collected or stored by the app once it is shut down. Researchers can obtain access to this software by contacting the first author or by visiting https://www.daniel-albohn.com.

CONCLUSIONS

It is a testament to the human visual system that individuals are able to derive meaningful information from the face given its complexity. Faces are important for social interactions, oftentimes signaling internal states and potential behavior through both numerous facial configurations as well as incidental resemblance to such features. Indeed, faces are so fundamental to forecasting intended and potential behavior that individuals effortlessly derive information from faces that only *incidentally* resemble emotion expression or personality traits. Non-expressive faces carry a surprising amount of information that aid individuals in forming impressions of others. Yet, despite the incredible amount of social information contained

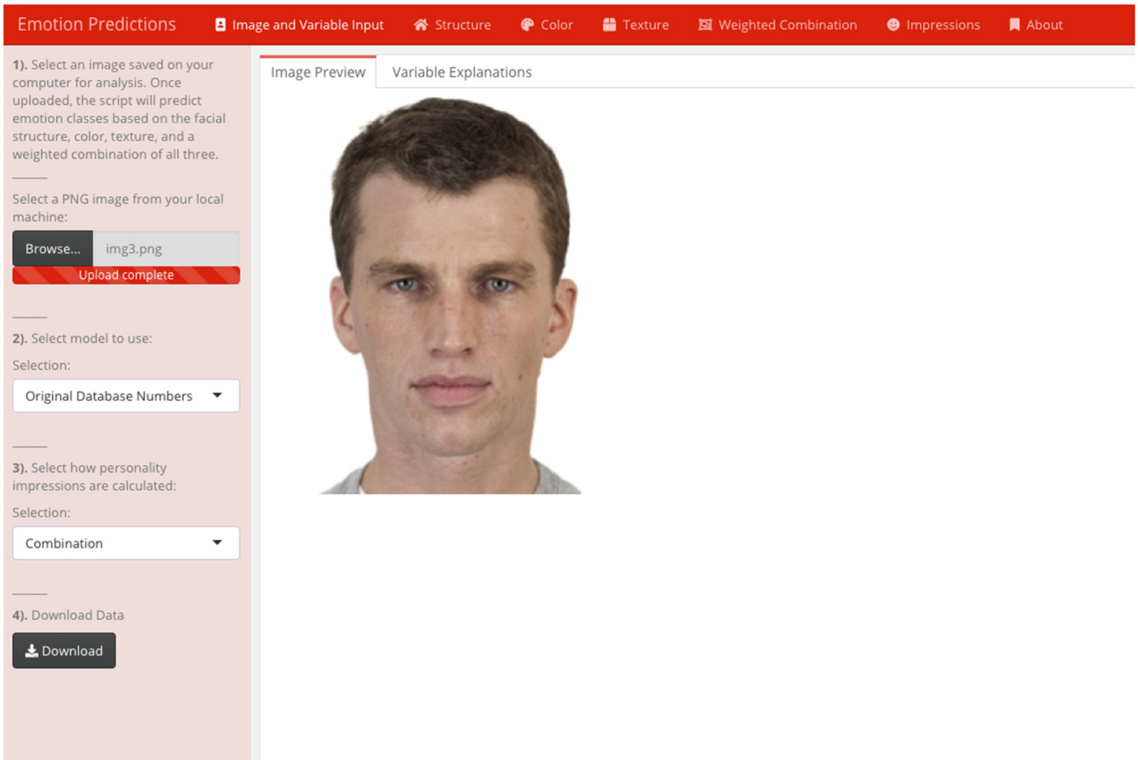


FIGURE 6 | Image of GUI designed to predict the emotion expression and impressions of an uploaded face image. End-users have access to each feature model (structure, color, and texture), the weighted model of all three features, and algorithmically computed impressions for the uploaded face image (see panel at top of the GUI).

in neutral displays, relatively little work has utilized state of the art computer vision programs to reliably and accurately extract emotion expression from the face to make predictions about human behavior.

One central thesis put forth in the current research was that computer vision algorithms could be used to derive emotional content from minimally and non-expressive faces, and that the emotional content of these faces was related to human impressions. Results across four studies support these assumptions, revealing that not only can machine learning be used to accurately predict subtle emotion expressivity from neutral faces, but that these learned emotion outputs were related to human impressions in meaningful ways. Thus, the current work can begin to answer the question of what *exactly* are the mechanisms that influence an individual's impressions?

Beyond the utility of using machine learning algorithms to further our understanding of human perception, the current work also demonstrates–at a fundamental and objective level–that emotions are a powerful mechanism of impression formation. So much so, in fact, that in the face of no overt expressivity (i.e., a neutral face) humans appear to be grasping for any sort of emotional meaning in the face to make an informed decision, whether that be resemblance to emotion through such visual channels as structure, color, texture, or some combination of all three. The notion that emotions have such an impact on human impressions underscores the importance of understanding them within the broader context of person perception and non-verbal behavior.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors upon request, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Office for Research Protections The Pennsylvania State University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR'S NOTE

DA is now at University of Chicago Booth School of Business.

AUTHOR CONTRIBUTIONS

DA conducted the research, performed the analyses, and wrote the initial draft of the manuscript under the supervision of RA. DA and RA both interpreted the results, revised and edited, commented on, and approved the final draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by a Heritage Dissertation Award and a Penn State Liberal Arts Research and Graduate Studies Office Dissertation Improvement Grant awarded to DA.

ACKNOWLEDGMENTS

We would like to thank Joseph Brandenburg, Lindsay Palmer, and Jose Soto for their helpful comments on early drafts of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.612923/full#supplementary-material

REFERENCES

Adams, R. B. Jr., Albohn, D. N., and Kveraga, K. (2017). Social vision: applying a social-functional approach to face and expression perception. *Curr. Dir. Psychol. Sci* 26, 243–248. doi: 10.1177/0963721417706392

Adams, R. B. Jr., Garrido, C. O., Albohn, D. N., Hess, U., and Kleck, R. E. (2016). What facial appearance reveals over time: when perceived expressions in neutral faces reveal stable emotion dispositions. *Front. Psychol.* 7:986. doi: 10.3389/fpsyg.2016.00986

Adams, R. B. Jr., Hess, U., and Kleck, R. E. (2015). The intersection of gender-related facial appearance and facial displays of emotion. *Emotion Rev.* 7, 5–13. doi: 10.1177/1754073914544407

Adams, R. B. Jr., Nelson, A. J., Soto, J. A., Hess, U., and Kleck, R. E. (2012). Emotion in the neutral face: a mechanism for impression formation? *Cogn. Emot* 26, 431–441. doi: 10.1080/02699931.2012.666502

Albohn, D. N., and Adams, R. B. Jr. (2020a). Emotion residue in neutral faces: implications for impression formation. *Soc. Psychol. Personal. Sci.* doi: 10.1177/1948550620923229

Albohn, D. N., and Adams, R. B. Jr. (2020b). Everyday beliefs about emotion perceptually derived from neutral facial appearance. *Front. Psychol.* 11:264. doi: 10.3389/fpsyg.2020.00264

Albohn, D. N., Brandenburg, J. C., and Adams, R. B. Jr. (2019). "Perceiving emotion in the 'neutral' face: a powerful mechanism of person perception," in *The Social Nature of Emotion Expression*, eds U. Hess and S. Hareli (Cham: Springer International Publishing), 25–47. doi: 10.1007/978-3-030-32968-6_3

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., and Smith, D. M. (2007). The confounded nature of angry men and happy women. *J. Personal. Soc. Psych.* 92, 179–190. doi: 10.1037/0022-3514.92.2.179

Bugental, D. E., Love, L. R., and Gianetto, R. M. (1971). Perfidious feminine faces. *J. Pers. Soc. Psychol.* 17, 314–318. doi: 10.1037/h0030586

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

Fabes, R. A., and Martin, C. L. (1991). Gender and age stereotypes of emotionality. *Personal. Soc. Psychol. Bull.* 17, 532–540.

Heerey, E. A., and Velani, H. (2010). Implicit learning of social predictions. *J. Exp. Soc. Psychol.* 46, 577–581. https://doi.org/10.1016/j.jesp.2010.01.003

Hess, U., Adams, R. B. Jr., Grammar, K., and Kleck, R. E. (2009). Face gender and emotion expression: are angry women more like men? *J. Vis,* 9, 1–8. doi: 10.1167/9.12.19

Hester, N. (2018). Perceived negative emotion in neutral faces: gender-dependent effects on attractiveness and threat. *Emotion* 19, 1490–1494. doi: 10.1037/emo0000525

iMotions (2019). AFFDEX. Available online at: https://imotions.com/biosensor/fea-facial-expression-analysis/

Johnson, K. L., Freeman, J. B., and Pauker, K. (2012). Race is gendered: how covarying phenotypes and stereotypes bias sex categorization. *J. Pers. Soc. Psychol.* 102, 116–131. doi: 10.1037/a0025335

Keating, C. F., Mazur, A., and Segall, M. H. (1977). Facial gestures which influence the perception of status. *Sociometry* 40, 374–378.

Klatzky, R. L., Lederman, S. J., and Reed, C. (1987). There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision. *J. Exp. Psychol. Gen.* 116:356.

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

Kocsor, F., Kozma, L., Neria, A. L., Jones, D. N., and Bereczkei, T. (2019). Arbitrary signals of trustworthiness – social judgments may rely on facial expressions even with experimentally manipulated valence. *Heliyon* 5:e01736. doi: 10.1016/j.heliyon.2019.e01736

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cogn. Emotion* 24, 1377–1388. doi: 10.1080/02699930903485076

Laser, P. S., and Mathie, V. A. (1982). Face facts: an unbidden role for features in communication. *J. Nonverbal Behav.* 7, 3–19. doi: 10.1007/BF01001774

Lee, E., Kang, J. I., Park, I. H., Kim, J.-J., and An, S. K. (2008). Is a neutral face really evaluated as being emotionally neutral? *Psychiatry Res* 157, 77–85. doi: 10.1016/j.psychres.2007.02.005

Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* 47, 1122–1135. doi: 10.3758/s13428-014-0532-5

Marsh, A. A., Adams, R. B. Jr., and Kleck, R. E. (2005). Why do fear and anger look the way they do? Form and social function in facial expressions. *Personal. Soc. Psychol. Bull.* 31, 73–86. doi: 10.1177/0146167204271306

Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* 36, 630–633. doi: 10.3758/bf03206543

Neth, D., and Martinez, A. M. (2009). Emotion perception in emotionless face images suggests a norm-based representation. *J. Vis.* 9, 5–5. doi: 10.1167/9.1.5

Noldus Information Technology (2019). *FaceReader 7.1*. Available online at: https://www.noldus.com/facereader?gclid=Cj0KCQjwj7v0BRDOARIsAGh37ipxkZpfWeDL9l_k4AEbMAWyYC596JVHVQ9KE91fhfgwbV-WAAlYGIAaAkNvEALw_wcB (Accessed April 9, 2020).

Palumbo, R., Adams Jr., R. B., Hess, U., Kleck, R. E., and Zebrowitz, L. (2017). Age and gender differences in facial attractiveness, but not emotion resemblance, contribute to age and gender stereotypes. *Front. Psychol.* 8:1704. doi: 10.3389/fpsyg.2017.01704

Pazda, A. D., Thorstenson, C. A., Elliot, A. J., and Perrett, D. I. (2016). Women's facial redness increases their perceived attractiveness: mediation through perceived healthiness. *Perception* 45, 739–754. doi: 10.1177/0301006616633386

Perrett, D. I., Talamas, S. N., Cairns, P., and Henderson, A. J. (2020). Skin color cues to human health: carotenoids, aerobic fitness, and body fat. *Front. Psychol.* 11:392. doi: 10.3389/fpsyg.2020.00392

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.* 57, 199–226. doi: 10.1146/annurev.psych.57.102904.190208

Rule, N. O., Garrett, J. V., and Ambady, N. (2010). On the perception of religious group membership from faces. *PLoS ONE* 5:e14241. doi: 10.1371/journal.pone.0014241

Russell, R. (2003). Sex, beauty, and the relative luminance of facial features. *Perception* 32, 1093–1107. doi: 10.1068/p5101

Russell, R. (2009). A sex difference in facial contrast and its exaggeration by cosmetics. *Perception* 38, 1211–1219. doi: 10.1068/p6331

Russell, R., Sinha, P., Biederman, I., and Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception* 35, 749–759. doi: 10.1068/p5490

Said, C. P., Sebe, N., and Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion* 9, 260–264. doi: 10.1037/a0014681

Sinha, P., Balas, B., Ostrovsky, Y., and Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE* 94, 1948–1962. doi: 10.1109/JPROC.2006.884093

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., and Heagins, T. B. (2016). The MR2: a multi-racial, mega-resolution database of facial stimuli. *Behav Res.* 48, 1197–1204. doi: 10.3758/s13428-015-0641-9

Tan, K. W., Tiddeman, B., and Stephen, I. D. (2018). Skin texture and colour predict perceived health in asian faces. *Evol. Huma. Behav.* 39, 320–335. doi: 10.1016/j.evolhumbehav.2018.02.003

Thorstenson, C. A., Pazda, A. D., Elliot, A. J., and Perrett, D. I. (2017). Facial redness increases men's perceived healthiness and attractiveness. *Perception* 46, 650–664. doi: 10.1177/0301006616680124

Tiddeman, B. P., Stirrat, M. R., and Perrett, D. I. (2005). Towards realism in facial image transformation: results of a wavelet MRF method: results of a wavelet MRF method. *Comp. Graph. Forum* 24, 449–456. doi: 10.1111/j.1467-8659.2005.00870.x

Todorov, A., Said, C. P., Engell, A. D., and Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends Cogn. Sci.* 12, 455–460. doi: 10.1016/j.tics.2008.10.001

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. doi: 10.1016/j.psychres.2008.05.006

Wiggins, J. S., Trapnell, P., and Phillips, N. (1988). Psychometric and geometric characteristics of the revised interpersonal adjective scales (IAS-r). *Multivariate Behav. Res.* 23, 517–530. doi: 10.1207/s15327906mbr2304_8

Yeung, R. R. (1996). The acute effects of exercise on mood state. *J. Psychosom. Res.* 40, 123–141.

Zebrowitz, L. A., Kikuchi, M., and Fellous, J.-M. (2007). Are effects of emotion expression on trait impressions mediated by babyfaceness? Evidence from connectionist modeling. *Personal. Soc. Psychol. Bull.* 33, 648–662. doi: 10.1177/0146167206297399

Zebrowitz, L. A., Kikuchi, M., and Fellous, J.-M. (2010). Facial resemblance to emotions: group differences, impression effects, and race stereotypes. *J. Pers. Soc. Psychol.* 98, 175–189. doi: 10.1037/a0017990

Zebrowitz, L. A., and McDonald, S. M. (1991). The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law Hum. Behav.* 15, 603–623.

Conflict of Interest: 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.

frontiers in Psychology

PERSPECTIVE
published: 20 April 2021
doi: 10.3389/fpsyg.2021.612664

Non-verbal Adaptation to the Interlocutors' Inner Characteristics: Relevance, Challenges, and Future Directions

Valerie Carrard[1,2*]

[1] Swiss Paraplegic Research (SPF), Nottwil, Switzerland, [2] Department of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

OPEN ACCESS

Edited by:
José-Miguel Fernández-Dols, Autonomous University of Madrid, Spain

Reviewed by:
Tessa Dekkers, University of Twente, Netherlands
Edward D. Sturman, State University of New York, United States

***Correspondence:**
Valerie Carrard
valerie.carrard@paraplegie.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *30 September 2020*
Accepted: *19 January 2021*
Published: *20 April 2021*

Citation:
Carrard V (2021) Non-verbal Adaptation to the Interlocutors' Inner Characteristics: Relevance, Challenges, and Future Directions. Front. Psychol. 12:612664. doi: 10.3389/fpsyg.2021.612664

Human diversity cannot be denied. In our everyday social interactions, we constantly experience the fact that each individual is a unique combination of characteristics with specific cultural norms, roles, personality, and mood. Efficient social interaction thus requires an adaptation of communication behaviors to each specific interlocutor that one encounters. This is especially true for non-verbal communication that is more unconscious and automatic than verbal communication. Consequently, non-verbal communication needs to be understood as a dynamic and adaptive process in the theoretical modeling and study of social interactions. This perspective paper presents relevance, challenges, and future directions for the study of non-verbal adaptation in social interactions. It proposes that non-verbal adaptability is more pertinently studied as adaptation to interlocutor's inner characteristics (i.e., expectations or preferences) than to interlocutor's behaviors *per se*, because behaviors are communication messages that individuals interpret in order to understand their interlocutors. The affiliation and control dimensions of the Interpersonal Circumplex Model are proposed as a framework to measure both the interlocutors' inner characteristics (self-reported) and the individuals' non-verbal responses (external coders). These measures can then be compared across different interactions to assess an actual change in behavior tailored to different interlocutors. These recommendations are proposed in the hope of generating more research on the topic of non-verbal adaptability. Indeed, after having gathered the evidence on average effects of non-verbal behaviors, the field can go further than a "one size fits all" approach, by investigating the predictors, moderators, and outcomes of non-verbal adaptation to the interlocutors' inner characteristics.

Keywords: behavioral adaptability, non-verbal behavior, expectations, preferences, social interaction

INTRODUCTION

> *"It is not the strongest of the species that survives, not the most intelligent that survives. It is the one that is the most adaptable to change."*—Charles Darwin

As stated in this famous quote attributed to Darwin, adaptation might be the most important quality for the survival of species, and this could still apply to modern human beings. As humans, we inexorably need to adapt to new situations, roles, and environments, and as social beings, we need to constantly adapt to each interactional partner we encounter. Every social encounter happens in a specific context, bears its specific goals, and involves specific interlocutors (i.e., interactional partner). Each of these elements requires an adaptation of communicative behaviors to achieve successful interactions. Street (1992) thus defines interpersonal communications as "processes of personal and mutual influence that unfold according to the characteristics of the individuals (e.g., attitudes, knowledge, communicative style) and the interactive processes related to how interactants adapt their communication to one another" (p. 1155). Consequently, communication behaviors need to be understood as a dynamic and adaptive process in the theoretical modeling of social interactions, but also in the way it is studied.

Still, many social psychology studies aim at identifying the behaviors that would relate to successful interactions overall, in spite of the interlocutor and situation at hand. Which communication style should physicians display? What is the best leadership behavior? How much should I smile during a job interview? The answer to these questions might be "it depends." Investigating the impact of communication behaviors from an average perspective is obviously important, because it provides overall guidelines that are more likely to trigger the intended output. However, a "one size fits all" perspective disregards the specificities of each interactional partner. For instance, giving a lot of information and establishing shared decision making might be the medical communication style that is linked to better patient outcomes on average, but some patients actually prefer less information or more passivity (Kiesler and Auerbach, 2006). Thus, a better way to achieve successful interaction outcomes may require promoting flexible adaptation instead of a set of behaviors to apply in every interaction. However, fewer studies focused on the beneficial effect of adapting one's behavior in social interaction. The literature on the subject comes mostly from the communication (Brennan and Hanna, 2009) and the medical interaction fields (Kiesler and Auerbach, 2006) and mainly focuses on the adaptation of verbal behaviors. In comparison, the adaptation of non-verbal behaviors (NVBs) has been scarcely investigated.

NVB is not merely an automatic outward display of inner states; rather, they are unambiguously social signals that are produced with a communicative purpose. Studies have indeed showed that NVBs are produced more intensely in social interactions and are directly linked to social consequences (Schmidt and Cohn, 2001). Furthermore, the establishment of critical interaction styles, such as power, relies largely on non-verbal signals, such as facial expressions, gestures, and spatial management (Hall et al., 2005). Behavioral adaptation of non-verbal signals seems especially critical, because they are more automatically processed in comparison to verbal communication (Choi et al., 2005).

This perspective paper strives to present relevance, challenges, and future directions for the study of non-verbal adaptation in social interactions in terms of mutual adjustment of NVB (and not in the Darwinian's sense of adaptation). First, it will be underlined that individuals do not merely adapt to the interlocutors' behaviors *per se*, but also to the individuals' interpretation of what the behaviors convey about the interlocutors' inner state (i.e., expectations or preferences). Then, the operationalization challenges of non-verbal adaptation to the interlocutor's inner characteristics will be discussed.

BEHAVIORAL CONTAGION, BEHAVIOR-TO-BEHAVIOR ADAPTATION, AND ADAPTATION TO THE INTERLOCUTOR'S INNER CHARACTERISTICS

Research on behavioral mutual influences in social interactions first focused on unintentional and automatic NVB contagion such as mimicry (i.e., imitation of speech inflections, facial expressions, and postures; Chartrand and Bargh, 1999) and interactional synchrony (i.e., timely coordination of verbal behvior and NVB; Condon and Ogston, 1967). Later theoretical models such as the Communication Accommodation Theory (Giles et al., 1987, 1991) conceptualize intentional behavior-to-behavior adaptation strategies such as convergence (displaying the same behaviors) and divergence (displaying the opposite behaviors) moves, used to engage or disengage from the interaction. Mimicry, synchrony, and behavior-to-behavior adaptation looking at similarities and dissimilarities between two interactants' behaviors have been extensively studied in different fields such as communication and clinical interactions (Hatfield et al., 2014; Leclère et al., 2014; Soliz and Giles, 2014). However, verbal behaviors or NVBs are not merely oral or visual features. As underlined by several theoretical models such as the Expectancy Violations Theory (Burgoon and Hale, 1988), the Sequential–Functional Model (Patterson, 1982), or the Interaction Adaptation Theory (Burgoon et al., 2007), behaviors are communication tools carrying a message, and the interactants will interpret the meaning of this message and adapt their response accordingly. Evidence supports that individuals adapt their behaviors to the interlocutor's inner characteristics, which are interpreted through displayed behaviors. For instance, research showed that children adapt their communication behaviors (e.g., initiation/response behaviors, gazing, voice frequency, number of words used, and length of vowels) according to their interactional partners' abilities (i.e., hearing or sight impairment) and preferences (wanting help or not; Ganea et al., 2018; Granlund et al., 2018; Gampe et al., 2019). Similarly, surgeons report adapting their guidance in decision-making according to their perception of patients' autonomy, communication competence, interpersonal style, and ability to manage illness (Dekkers et al., 2018). Thus, adaptation to interlocutors' inner characteristics as perceived through behavioral cues is a reality of social interaction.

The interlocutors' inner characteristics that one might make assumptions about and thus adapt to can be conceptualized as expectations (defined by social norms, roles, and the situation at hand) or preferences (defined by personality and emotional state) related to the interaction. For instance, one can interpret an interlocutor's increased interpersonal distance as conveying his or her expectations for a more formal interaction due to a hierarchically defined relationship or his or her preferences for a colder exchange, due to an introverted personality. The multiple behavioral input individuals receive from their interlocutors provide many cues that will add up and enable individuals to determine the behaviors expected or preferred by their interlocutors. In short, most of the adaptation, and especially the more conscious and intentional adjustments, will be based on how behaviors are interpreted, as cues of the interlocutor's inner characteristics. This process of adaptation to the interlocutor's inner characteristics has however been less investigated in comparison to behavior-to-behavior adaptation. This lack of research could be due to the scarcity of extant methodological guidelines available to study it.

RECOMMENDATIONS FOR THE INVESTIGATION OF NON-VERBAL ADAPTATION TO THE INTERLOCUTOR'S INNER CHARACTERISTICS

Investigating how individuals adapt their NVB to the inner characteristics of their interlocutors requires three steps of operationalization: (1) assessing the inner characteristics of the interlocutor, (2) assessing the non-verbal response of the individual, and (3) assessing how the NVB is adapted to the interlocutor's inner characteristics. Each of these three steps implies important methodological considerations in terms of collection method, timing of assessment, and operationalization.

Regarding the collection method, many studies rely on self-report. The inner characteristics (i.e., expectations or preferences) of the interlocutor, like any inner world variable, are indeed best measured with self-report. However, using self-reported measures of displayed behaviors is subject to many biases such as social desirability and recall bias (Paulhus and Vazire, 2007). Indeed, even when individuals are conscious of their behaviors, they are inaccurate in reporting them (Jones, 1991). Moreover, the rating of behaviors (one's own or the interactive partner's) is highly influenced by the overall impression of and satisfaction with the interaction (Kiesler and Auerbach, 2006). Thus, an observer coding is a more reliable operationalization of individuals' NVB.

The timing of the assessment is also critical to avoid biases. It is indeed important to note that the interlocutors' inner characteristics are more reliably measured before the interaction of interest, because a post-interaction assessment would be biased by the overall impression of the interaction and interaction outcomes would then be confounded with the measure of interlocutor's expectations or preferences.

Most importantly, the operationalization of interlocutor's inner characteristics and non-verbal answer must be chosen carefully. Assessing how non-verbal response is adapted to the expectations or preferences of an interlocutor implies the comparison of observed behaviors to some inner characteristics. To do so, both need to be assessed with similar operationalization. To this end, researchers can rely on a theoretical framework, which clusters interpersonal behaviors and attitudes according to their functions: the Interpersonal Circumplex Model (ICM; Wiggins and Trobst, 1997). The ICM proposes that two basic dimensions underlie all human interactions: control and affiliation. Control is the dimension pertaining to the verticality of human interaction going from dominance to submission, whereas affiliation represents the horizontality with a continuum from friendliness to hostility. The ICM has served as a theoretical model for many studies of personality (Smith, 1992; Smith et al., 1996) and interpersonal behaviors (Moskowitz et al., 2001; Kiesler and Auerbach, 2003; Newton et al., 2005), because the control and affiliation dimensions can be applied to describe both behavioral display and inner characteristics. Indeed, several validated questionnaires based on the ICM can be used to measure long-standing personality dispositions (e.g., the Interpersonal Checklist; LaForge and Suczek, 1955), interactional preferences (e.g., the Patient-Practitionner Orientation Scale; Krupat et al., 2000), and behaviors (e.g., the Impact Message Inventory or the Interpersonal Transactions-Revised; Kiesler, 1987; Kiesler and Schmidt, 1993). The most versatile instrument measuring the control and affiliation dimensions of the ICM is the Revised Interpersonal Adjective Scale (IAS-R; Wiggins et al., 1988). The IAS-R is composed of 64 adjectives describing interpersonal characteristics mapped on the ICM such as "unsympathetic" and "kind" for the affiliation continuum or "shy" and "assertive" for the control continuum. This scale is versatile, because with a small adaptation of the instructions, researchers can use it to assess expectations or preferences for an upcoming interaction as well as actual interactional behaviors displayed. With the following instruction: "Please indicate the extent to which the following adjectives correspond to your *expectations* regarding the behavior of your interlocutor in the upcoming interaction" or with the same instruction asking about *preferences*, researchers can measure the extent to which interlocutors expect or prefer an affiliative ("unsympathetic" or "kind") or controlling (e.g., "shy" or "assertive") interactional partner. These expectations or preferences can then be compared to the extent of affiliation and control the partner actually displayed during the interaction. Note that whether one wants to assess expectations or preferences will depend on the objective of the study and its context. Indeed, in some social interactions such as job interviews, expectations about the interlocutor's communication behavior seems more central than preference, whereas preferences might be more important in other context such as medical interactions, where patient preferences are critical, according to the currently recommended patient-centered approach. In any case, expectations are not synonymous to preferences, and the two might differ considerably. Thus, adaptation to preferences or adaptation to expectations should be measured separately and not aggregated.

The displayed NVB of the interactional partner can also be measured with the IAS-R adjectives as rated by external coders with the following instruction: "Based on the displayed behaviors, indicate the extent to which the following adjectives correspond to the interactional partner." Research has indeed used the IAS or a shortened version of it with external coders and reported satisfactory reliability and convergent validity (i.e., related to self-reported IAS or observed discrete behaviors; Gifford, 1994; Gifford and Hine, 1994; Muran et al., 1997). The use of the IAS-R by external coders does not assess the display of specific NVBs, but a global impression of overall interpersonal behaviors. Nevertheless, the frequency or duration of several discrete NVB can also be coded. The specific NVB can then be classified in clusters according to the overall control and affiliation dimensions of the ICM, as literature provide evidence for a dimensional conception of the ICM (Lorr and Strack, 1990). To guide which behavior relates to the control or affiliation dimensions of the ICM, one can rely on past research such as Gifford's (1991) mapping of NVB on the ICM, Kiesler and Auerbach's (2003) review of NVBs signaling affiliation and control, or Hall et al.'s (2014) meta-analysis of NVBs related to the vertical dimension of human interactions.

When both the individual's behavior and the interlocutor's inner characteristics have been assessed with measures relating to common overall dimensions, the extent to which one is adapted to the other can be estimated. In that regard, it is of utmost importance to consider that adaptability is not merely defined by similarities or dissimilarities between displayed NVB and inner characteristics of the interlocutor. Measuring similarity and dissimilarity does not tell whether the individuals were displaying their usual pattern of behaviors or actually changing it to fit a particular interlocutor. In order to measure individuals' ability to adapt his or her NVB, one needs to measure their NVB when interacting with at least two different interlocutors. The computation of adaptability scores needs to account for the extent to which the NVB of the individual fit the specific inner characteristics of different interlocutors. An evident measure of correspondence between a behavior variable and an inner characteristics variable across different interactions is a correlation. For example, a study of the effect of physicians' behavioral adaptability to patient preferences used such correlation method (Carrard et al., 2018). Several physicians were videotaped when interacting with four of their patients. Correlations were then computed for each physician between observer ratings of the physicians' behavior in each of their four videotaped interactions and the rating of each of the four corresponding patients' preferences (self-reported before the interaction). The higher the correlation estimate, the more the NVB displayed by a physician in each interaction corresponds to each patient's preferences. Then, the correlation estimates (e.g., Pearson's r transformed into Fisher's z to avoid added error to the analysis) were used as behavioral adaptability scores predicting interaction outcomes in order to test the beneficial effect of behavioral adaptation to interlocutor's inner characteristics (Carrard et al., 2018).

DISCUSSION

This perspective paper proposes five main postulations. First, research needs to take a step further than the "one size fits all" approach and thus study adaptation of behaviors, in order to predict better interactional outcomes. Second, the study of non-verbal adaptation is essential, because NVBs highly contribute to the communication between two interactants. Third, behavioral adaptability is more pertinently studied as adaptation to interlocutor's inner characteristics (i.e., expectations or preferences) than to interlocutor's behaviors *per se*. Fourth, the present paper proposes the ICM and its control and affiliation dimensions as the framework to measure both interlocutor's inner characteristics and adapted non-verbal response in a comparable way. Finally, measuring non-verbal adaptability to interlocutors' inner characteristics implies the assessment of different interactions, because it involves a change of NVB according to each interlocutor's specific inner characteristics.

Further research is needed to understand the process of non-verbal adaptation to the interlocutor's inner characteristics, its outcomes, predictors, and covariates. Some evidence suggest that a perceived match between individual's behaviors and interlocutor's expectations or preferences is related to better outcomes such as patient satisfaction (Street et al., 2012), better learning outcomes of students (Young et al., 2003), and more credibility of and attraction to interactional partner (Burgoon and Le Poire, 1993). However, further research on adaptability instead of match are needed, especially for the adaptation of NVB, to confirm its beneficial effect.

Eventually, the ability to display the behaviors that will match the interlocutors' expectations or preferences will depend on one's ability to infer these inner characteristics based on the interlocutor's behaviors. This skill called interpersonal accuracy has been shown to be related to more positive interaction outcomes in sales, clinical interactions, and the workplace (DiMatteo et al., 1986; Byron et al., 2007; Hall et al., 2009). It has been suggested that the relationship between interpersonal accuracy and positive interaction outcome is mediated by behavioral ability (Hall et al., 2016), and a study in patient–physician interactions provides some evidence of this mediation for female physicians (Carrard et al., 2018). Interestingly, a meta-analysis showed that interpersonal accuracy can be efficiently trained with short training sessions (Blanch-Hartigan et al., 2012). Future studies should confirm whether interpersonal accuracy is a predictor of behavioral adaptability, because such training would be an interesting avenue to improve communication competencies. Another predictor of behavioral adaptability is the possession of a large behavioral repertoire and the ability to flexibly change behavioral displays. However, the possibility of training behavioral repertoires or behavioral flexibility is still unknown.

Moreover, future studies should also consider investigating the potential covariates and moderators of non-verbal adaptability. For instance, it has been shown that women are more knowledgeable regarding NVB compared to men (Rosip and Hall, 2004). A meta-analysis further showed that women's interpersonal accuracy is more strongly linked to psychosocial functioning than men's (Hall et al., 2009). Additionally, a study in the physician–patient interaction context showed that the link between physician non-verbal adaptability to patient preferences is linked to more positive patient outcomes for females, but not for males (Carrard et al., 2018). Thus, gender, as well as other potential moderators such as age or culture, should be tested to better understand the non-verbal adaptability process and how it relates to better consultation outcomes.

This perspective paper did not strive to deliver an exhaustive review of the literature on the topic, but to provide some hints for the measure and study of non-verbal adaptation to interlocutor's inner characteristics. My hope is that the limited overview and recommendations presented promotes more research on the topic. Considering that the link between NVB and better interaction outcomes has been acknowledged, the field can move forward with the investigation of non-verbal adaptation to interlocutors' inner characteristics as the foundation of efficient social interactions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

As sole author, VC, completed all tasks related to the conception and writing of the submitted manuscript.

REFERENCES

Blanch-Hartigan, D., Andrzejewski, S. A., and Hill, K. M. (2012). The effectiveness of training to improve person perception accuracy: a meta-analysis. *Basic Appl. Soc. Psychol.* 34, 483–498. doi: 10.1080/01973533.2012.728122

Brennan, S. E., and Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Top. Cogn. Sci.* 1, 274–291. doi: 10.1111/j.1756-8765.2009.01019.x

Burgoon, J. K., and Hale, J. L. (1988). Nonverbal expectancy violations: model elaboration and application to immediacy behaviors. *Commun. Monogr.* 55, 58–79. doi: 10.1080/03637758809376158

Burgoon, J. K., and Le Poire, B. A. (1993). Effects of communication expectancies, actual communication, and expectancy disconfirmation on evaluations of communicators and their communication behavior. *Hum. Commun. Res.* 20, 67–96. doi: 10.1111/j.1468-2958.1993.tb00316.x

Burgoon, J. K., Stern, L. A., and Dillman, L. (2007). *Interpersonal Adaptation: Dyadic Interaction Patterns.* New York, NY: Cambridge University Press.

Byron, K., Terranova, S., and Nowicki Jr, S. (2007). Nonverbal emotion recognition and salespersons: linking ability to perceived and actual success. *J. Appl. Soc. Psychol.* 37, 2600–2619. doi: 10.1111/j.1559-1816.2007.00272.x

Carrard, V., Schmid Mast, M., Jaunin-Stalder, N., Junod Perron, N., and Sommer, J. (2018). Patient-centeredness as physician behavioral adaptability to patient preferences. *Health Commun.* 33, 593–600. doi: 10.1080/10410236.2017.1286282

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

Choi, Y. S., Gray, H. M., and Ambady, N. (2005). "The glimpsed world: unintended communication and unintended perception," in *The New Unconscious*, eds R. R. Hassin, J. S. Uleman, and J. A. Bargh (New York, NY: Oxford University Press), 309–333.

Condon, W. S., and Ogston, W. D. (1967). A segmentation of behavior. *J. Psychiatr. Res.* 5, 221–235. doi: 10.1016/0022-3956(67)90004-0

Dekkers, T., Melles, M., Mathijssen, N. M. C., Vehmeijer, S. B. W., and de Ridder, H. (2018). Tailoring the orthopaedic consultation: how perceived patient characteristics influence surgeons' communication. *Patient Educ. Couns.* 101, 428–438. doi: 10.1016/j.pec.2017.08.018

DiMatteo, M. R., Hays, R. D., and Prince, L. M. (1986). Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload. *Health Psychol.* 5:581. doi: 10.1037/0278-6133.5.6.581

Gampe, A., Wermelinger, S., and Daum, M. M. (2019). Bilingual children adapt to the needs of their communication partners, monolinguals do not. *Child Dev.* 90, 98–107. doi: 10.1111/cdev.13190

Ganea, N., Hudry, K., Vernetti, A., Tucker, L., Charman, T., Johnson, M. H., et al. (2018). Development of adaptive communication skills in infants of blind parents. *Dev. Psychol.* 54, 2265–2273. doi: 10.1037/dev0000564

Gifford, R. (1991). Mapping nonverbal behavior on the interpersonal circle. *J. Pers. Soc. Psychol.* 61, 279–288. doi: 10.1037/0022-3514.61.2.279

Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *J. Pers. Soc. Psychol.* 66, 398–412. doi: 10.1037/0022-3514.66.2.398

Gifford, R., and Hine, D. W. (1994). The role of verbal behavior in the encoding and decoding of interpersonal dispositions. *J. Res. Pers.* 28, 115–132. doi: 10.1006/jrpe.1994.1010

Giles, H., Coupland, N., and Coupland, J. (1991). "Accommodation theory: communication, context, and consequences," in *Contexts of Accommodation: Developments in Applied Sociolinguistics*, eds H. Giles, J. Coupland, and N. Coupland (New York, NY: Cambridge University Press), 1–68.

Giles, H., Mulac, A., Bradac, J. J., and Johnson, P. (1987). Speech accommodation theory: the first decade and beyond. *Ann. Int. Commun. Assoc.* 10, 13–48. doi: 10.1080/23808985.1987.11678638

Granlund, S., Hazan, V., and Mahon, M. (2018). Children's acoustic and linguistic adaptations to peers with hearing impairment. *J. Speech Lang. Hear. Res.* 61, 1055–1069. doi: 10.1044/2017_JSLHR-S-16-0456

Hall, J. A., Andrzejewski, S. A., and Yopchick, J. E. (2009). Psychosocial correlates of interpersonal sensitivity: a meta-analysis. *J. Nonverbal Behav.* 33, 149–180. doi: 10.1007/s10919-009-0070-5

Hall, J. A., Coats, E. J., and Smith LeBeau, L. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* 131, 898–924. doi: 10.1037/0033-2909.131.6.898

Hall, J. A., Latu, I. M., Carney, D. R., and Schmid Mast, M. (2014). "Nonverbal communication and the vertical dimension of social relations," in *The Psychology of Social Status*, eds J. T. Cheng, J. L. Tracy, and A. Cameron (New York, NY: Springer), 325–343.

Hall, J. A., Schmid Mast, M., and West, T. V. (2016). *The Social Psychology of Perceiving Others Accurately*. Cambridge: Cambridge University Press.

Hatfield, E., Bensmana, L., Thorntona, P. D., and Rapsona, R. L. (2014). New perspectives on emotional contagion: a review of classic and recent research on facial mimicry and contagion. *PsychOpen* 8, 159–179. doi: 10.5964/ijpr.v8i2.162

Jones, S. E. (1991). Problems of validity in questionnaire studies of nonverbal behavior: Jourard's tactile body-accessibility scale. *South J. Commun.* 56, 83–95. doi: 10.1080/10417949109372821

Kiesler, D. (1987). *Check List of Psychotherapy Transactions-Revised (CLOPT-R) and Check List of Interpersonal Transactions Revised (CLOIT-R)*. Richmond, VA: Virginia Commonwealth University.

Kiesler, D., and Schmidt, J. (1993). *The Impact Message Inventory: Form IIA Octant Scale Version*. Palo Alto, CA: Mind Garden.

Kiesler, D. J., and Auerbach, S. M. (2003). Integrating measurement of control and affiliation in studies of physician–patient interaction: the interpersonal circumplex. *Soc. Sci. Med.* 57, 1707–1722. doi: 10.1016/S0277-9536(02)00558-0

Kiesler, D. J., and Auerbach, S. M. (2006). Optimal matches of patient preferences for information, decision-making, and interpersonal behavior: evidence, models, and interventions. *Patient Educ. Couns.* 61, 319–341. doi: 10.1016/j.pec.2005.08.002

Krupat, E., Yeager, C. M., and Putnam, S. (2000). Patient role orientations, doctor-patient fit, and visit satisfaction. *Psychol. Health* 15, 707–719. doi: 10.1080/08870440008405481

LaForge, R., and Suczek, R. F. (1955). The interpersonal dimension of personality: an interpersonal check list. *J. Pers.* 24, 94–112. doi: 10.1111/j.1467-6494.1955.tb01177.x

Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., et al. (2014). Why synchrony matters during mother-child interactions: a systematic review. *PLoS ONE* 9:e113571. doi: 10.1371/journal.pone.0113571

Lorr, M., and Strack, S. (1990). Wiggins interpersonal adjective scales: a dimensional view. *Pers. Indiv. Differ.* 11, 423–425. doi: 10.1016/0191-8869(90)90227-I

Moskowitz, D., Pinard, G., Zuroff, D. C., Annable, L., and Young, S. N. (2001). The effect of tryptophan on social interaction in everyday life: a placebo-controlled study. *Neuropsychopharmacol* 25, 277–289. doi: 10.1016/S0893-133X(01)00219-6

Muran, J. C., Samstag, L. W., Jilton, R., Batchelder, S., and Winston, A. (1997). Development of a suboutcome strategy to measure interpersonal process in psychotherapy from an observer perspective. *J. Clin. Psychol.* 53, 405–420. doi: 10.1002/(sici)1097-4679(199708)53:5<405::aid-jclp1>3.0.co;2-m

Newton, T. L., Watters, C. A., Philhower, C. L., and Weigel, R. A. (2005). Cardiovascular reactivity during dyadic social interaction: the roles of gender and dominance. *Int. J. Psychophysiol.* 57, 219–228. doi: 10.1016/j.ijpsycho.2005.03.001

Patterson, M. L. (1982). A sequential functional model of nonverbal exchange. *Psychol. Rev.* 89, 231–249. doi: 10.1037/0033-295X.89.3.231

Paulhus, D. L., and Vazire, S. (2007). "The self-report method," in *Handbook of Research Methods in Personality Psychology*, eds R. Robins, R. Fraley, and R. Krueger (New York, NY: The Guilford Press), 224–239.

Rosip, J. C., and Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding accuracy. *J. Nonverbal Behav.* 28, 267–286. doi: 10.1007/s10919-004-4159-6

Schmidt, K. L., and Cohn, J. F. (2001). Human facial expressions as adaptations: evolutionary questions in facial expression research. *Am. J. Phys. Anthropol.* 44, 3–24. doi: 10.1002/ajpa.20001

Smith, T. W. (1992). Hostility and health: current status of a psychosomatic hypothesis. *Health Psychol.* 11, 139–150. doi: 10.1037/0278-6133.11.3.139

Smith, T. W., Limon, J. P., Gallo, L. C., and Ngu, L. Q. (1996). Interpersonal control and cardiovascular reactivity: goals, behavioral expression, and the moderating effects of sex. *J. Pers. Soc. Psychol.* 70, 1012–1024. doi: 10.1037/0022-3514.70.5.1012

Soliz, J., and Giles, H. (2014). Relational and identity processes in communication: a contextual and meta-analytical review of communication accommodation theory. *Ann. Int. Commun. Assoc.* 38, 107–144. doi: 10.1080/23808985.2014.11679160

Street, R. L., Elwyn, G., and Epstein, R. M. (2012). Patient preferences and healthcare outcomes: an ecological perspective. *Expert Rev. Pharmacoecon. Outcomes Res.* 12, 167–180. doi: 10.1586/erp.12.3

Street, R. L. Jr. (1992). Communicative styles and adaptations in physician-parent consultations. *Soc. Sci. Med.* 34, 1155–1163. doi: 10.1016/0277-9536(92)90289-3

Wiggins, J. S., Trapnell, P., and Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivar. Behav. Res.* 23, 517–530. doi: 10.1207/s15327906mbr2304_8

Wiggins, J. S., and Trobst, K. K. (1997). "When is a circumplex an "interpersonal circumplex?" The case of supportive actions," in *Circumplex Models of Personality and Emotions*, eds R. Plutchik and H. Conte (Washington, DC: American Psychological Association), 57–80.

Young, M. R., Klemz, B. R., and Murphy, J. W. (2003). Enhancing learning outcomes: the effects of instructional technology, learning styles, instructional methods, and student behavior. *J. Mark. Educ.* 25, 130–142. doi: 10.1177/0273475303254004

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

frontiers in Psychology

REVIEW
published: 29 April 2021
doi: 10.3389/fpsyg.2021.667326

Check for updates

Capturing Behavior in Small Doses: A Review of Comparative Research in Evaluating Thin Slices for Behavioral Measurement

Nora A. Murphy[1]* and *Judith A. Hall*[2]

[1] Department of Psychology, Loyola Marymount University, Los Angeles, CA, United States, [2] Department of Psychology, Northeastern University, Boston, MA, United States

Thin slices are used across a wide array of research domains to observe, measure, and predict human behavior. This article reviews the thin-slice method as a measurement technique and summarizes current comparative thin-slice research regarding the reliability and validity of thin slices to represent behavior or social constructs. We outline decision factors in using thin-slice behavioral coding and detail three avenues of thin-slice comparative research: (1) assessing whether thin slices can adequately approximate the total of the recorded behavior or be interchangeable with each other (representativeness); (2) assessing how well thin slices can predict variables that are different from the behavior measured in the slice (predictive validity), and (3) assessing how interpersonal judgment accuracy can depend on the length of the slice (accuracy-length validity). The aim of the review is to provide information researchers may use when designing and evaluating thin-slice behavioral measurement.

Keywords: thin slice, predictive validity, behavioral coding, nonverbal behavior analysis and synthesis, reliability

OPEN ACCESS

Edited by:
Miles L. Patterson,
University of Missouri–St. Louis,
United States

Reviewed by:
Philip Furley,
German Sport University
Cologne, Germany
Diana Whalen,
Washington University in St. Louis,
United States

***Correspondence:**
Nora A. Murphy
nora.murphy@lmu.edu

Specialty section:
This article was submitted to
Personality and Social Psychology,
a section of the journal
Frontiers in Psychology

Received: *12 February 2021*
Accepted: *24 March 2021*
Published: *29 April 2021*

Citation:
Murphy NA and Hall JA (2021)
Capturing Behavior in Small Doses: A Review of Comparative Research in Evaluating Thin Slices for Behavioral Measurement.
Front. Psychol. 12:667326.
doi: 10.3389/fpsyg.2021.667326

INTRODUCTION

Observing and measuring behavior is foundational to behavioral research (Greene, 1941; Vaughan, 1948). The measurement of behavior to understand features of communication and person perception is widespread across many domains such as psychology, sociology, medicine, and communication. In this article, we review the thin-slice method as a behavioral measurement technique and review comparative thin-slice research (Ambady et al., 2000; Slepian et al., 2014; Murphy et al., 2015).

Thin-slice methodology refers to utilizing a small excerpt from a longer behavioral stream. This means, for the researcher, either deciding at the outset to record or gather very limited amounts of behavior (for example, recording only the 1st min of an interaction even though the interaction is much longer), or making a later decision to analyze, or present to viewers, only short excerpts from all the recorded or transcribed material that one possesses. Typically, an interaction is video or audio recorded and then slices are extracted from those recordings or their respective transcripts. The interaction or "behavioral stream" can be of any length, and while there is no fixed definition of what constitutes a "thin slice," thin slices typically are under 5 min. The thin-slice excerpt then can be coded or rated for behaviors or characteristics of individuals (targets) in the interaction. Thin slices also may be shown to viewers who judge a target's state or trait, if the goal is to assess judgment accuracy. The idea is that the slice is representative of a target's behavior throughout the interaction and/or that the slice may reveal or predict a target's internal states, personality, or other social attributes. In this article, we review comparative thin-slice research involving dynamic stimuli[1], which typically involves comparisons about different slice lengths (Murphy, 2005; Murphy et al., 2015; Krzyzaniak et al., 2019), as well as examination of slice locations (Carney et al., 2007; Fowler et al., 2009; Wang et al., 2020).

Thin slices are used to code target behavior (i.e., how is the person behaving) or as stimuli in person perception research, wherein observers make inferences about targets based on their behavior. Behavioral researchers are usually drawn to thin-slice techniques out of sheer pragmatism—to ease coding burdens, reduce viewer time, and in general to make the best use of limited resources of time and patience among research personnel and research participants (Murphy, 2005, 2018). The practical benefits of the thin-slice method are clear. Given the inherent complexity of behavior, thin-slice methods ease the burden of behavioral measurement because measuring behavior is an arduous task. Various researchers' descriptions of dynamic behavioral coding include: "time-consuming," "labor-intensive," "tedious," "costly," "complex," "challenging," "painstaking," "mentally-straining," "inefficient," "serious commitment," and "daunting," among many other unfavorable terms (Gosling et al., 1998; Murphy, 2005; Black et al., 2013; Fujiwara and Daibo, 2014; Carcone et al., 2015). One way researchers deal with the time-consuming nature of behavioral coding is to ask coders or raters to watch or listen for several behaviors at the same time—for example, to simultaneously count smiles and head tilts, or to simultaneously rate anger, anxiety, and sadness (e.g., Wang et al., 2020). This may not be optimal because it divides the observer's attention and may encourage inflated correlations among the behaviors or attributes being coded or rated.

Choosing not to employ thin slices could exponentially increase coding or rating time, depending on what is being measured and the length of original recordings (Murphy, 2005). Some coding projects are impressively colossal in scale; for example, Bensing et al. (2008) employed two coders both of whom timed gaze by physicians toward their patients for the entirety of some 2,000 patient visits that averaged about 10 min each. Fairbairn et al. (2013) coded 7.9 million frames of video data from 92 participants engaged in 36-min interactions. For obvious reasons, therefore, researchers actively seek techniques to reduce the burdens of coding. Simply put, thin-slice behavioral measurement is easier than coding a longer behavioral stream.

Automated coding using software or equipment is another approach to reducing coding labor (Georgiou et al., 2011), because such systems can sometimes eliminate the human element and are unlikely to be limited by the duration of the stimuli. Existing technology can automatically extract nonverbal features such as prosody, turn-taking, pauses, gesturing, interactional synchrony, and nodding (Fujiwara and Daibo, 2014; Nguyen and Gatica-Perez, 2015; Lausberg and Sloetjes, 2016; Ramseyer, 2020). Machine-learning methods can train a computer model to recognize behavioral features (e.g., a smiling face) based on a small corpus of recorded behavior (Chakravarthula et al., 2021). Such automated approaches can considerably reduce manual coding time, though many still require human coders (Narayanan and Georgiou, 2013; Girard et al., 2015).

While automated methods are attractive, they are not a panacea for reducing coding burdens (Schmid Mast et al., 2015). These sophisticated methods rely on an interdisciplinary approach often involving computer scientists, statisticians, and behavioral researchers. Learning how to implement new software or equipment, which is often expensive, requires a steep learning curve. As these techniques advance, the time and training needed to learn such automatic approaches will likely decrease. For now, there is a trade-off between learning and paying for automated methods and the time to complete traditional manual coding. Another potential limitation of automated methods is the inability to code molar constructs that involve the extraction of meaning from an integration of behavioral cues (e.g., friendliness, anxiety, or competence). In fact, it can be an error to assume that exact measurement such as provided by automatic methods equates to *psychologically meaningful* measurement; behavioral researchers often want to know what movements mean, not just how often they happen or what they look like (Funder and Colvin, 1991; Blanch-Hartigan et al., 2018).

Thus, researchers may turn to thin slices as a desirable coding technique and the use of thin slices across many domains is a testament to the method's versatility. Yet, inherent to the thin-slice technique are questions about the reliability and validity of thin slices to represent behavior or social constructs such as: How well do thin slices capture the whole of a behavioral stream? Are they interchangeable with each other? How well do thin slice measurements correlate with different (external) variables, compared with the totality of the recorded behavior? How much does accuracy of judging targets' states and traits depend on the duration of the stimuli shown to perceivers?

While our present goal is to provide potentially useful information to academic researchers, there are real-world applications where more knowledge about thin slices could be important. As examples, Perrault (2020) compared slices of different lengths from introduction videos made by physicians for potential patients, in order to determine how long such videos should be in terms of viewers' responses and attention span. Hall et al. (2009, 2014), in studies of medical visits and corporate technical support calls respectively, found evidence for the importance of the very first min or two of the interaction in predicting important patient or client outcomes. In fact many studies using thin slices have been conducted in clinical psychology, medicine, and business, demonstrating that thin-slice research could have meaningful impact in domains far removed from the psychology lab.

Truth be told, however, there is little rhyme or reason to the many choices researchers make about thin-slice coding, such

[1] Although photos have been extensively used in person perception and impression formation research, there is a dearth of comparative thin-slice research involving photos—such as comparing photos to dynamic stimuli, or comparing longer or shorter exposures to photos. Thus, this paper focuses exclusively on dynamic stimuli.

as deciding an appropriate slice length and from where in the interaction should the slice be extracted. Using thin slices rests on an assumption that the methodology itself is a reliable and valid representation of behavior, but making these decisions is often an act of faith on the part of researchers. As with any measurement tool, empirical conclusions are only as strong as the reliability and validity of such methods (Flake et al., 2017). This article will review thin slices for behavioral measurement, describing thin-slice coding techniques and reviewing comparative research of several kinds. We cannot describe in detail the many studies that have employed thin slices, as there are far too many. Rather, we will focus on research that is aimed at understanding the *trade-offs* involved in longer vs. shorter slices, or slices from different temporal positions within the behavioral stream, so that researchers can exercise more rationality (i.e., go beyond pragmatism or guesswork) when designing and evaluating research. Efforts to establish reliability and validity of thin slices by comparing results obtained for slices of different lengths—what we call *comparative thin-slice research*—is a relatively new undertaking[2]. We will focus on three tacks in which comparative thin-slice research has been done: (1) assessing whether thin slices can adequately approximate the total of the recorded behavior or be interchangeable with each other (which we call *representativeness*), (2) assessing how well thin slices can predict variables that are different from the behavior measured in the slice (which we call *predictive validity*), and (3) assessing how interpersonal judgment accuracy can depend on the length of the slice (which we call *accuracy-length validity)*.

HISTORICAL ROOTS AND MODERN USES OF THIN-SLICE METHODS

The term "thin slice" was coined by Ambady and Rosenthal (1992) in a meta-analysis on correlations between thin-slice coding or rating and outcomes of interest, such as depression, ratings of teacher effectiveness, or medical patients' satisfaction[3]. That article (cited more than 2,200 times as of January, 2021)[4] demonstrated thin slices to be a method that is widely and justifiably used, and established the thin-slice methodology as a topic of research in its own right.

Of course, thin slices had been in use for many years before they received their name. For example, Waxer (1974) used 2-min silent video clips of individuals in psychiatric interviews to find out whether naïve viewers could recognize those with depression. Milmoe et al. (1967) used ratings of electronically filtered audio clips of under 2 min (and their respective transcripts) to predict therapists' success in referring alcoholics for treatment. Ekman et al. (1980) used both 1 and 2-min video slices for judgments of honest and dishonest interviewees. Hall and Braunwald (1981) obtained impression ratings of male and female speakers based on 10-s audio clips from television shows to find out if listeners could tell whether a male or female was being spoken to and what vocal qualities the speakers used.

To observe and measure behavior, researchers code and analyze interactions of all sorts, such as interactions between parent and child, relationship partners, or strangers in get-acquainted sessions. Coding schemes are often employed and typically capture many constructs but the schemes often involve lengthy training periods in addition to the actual coding time itself. For instance, the Living in Familial Environments coding system (LIFE) involves more than 40 separate coding units, such as caring and irritated, with one study describing a training protocol as lasting 6 months (Hops et al., 1987). As another example, the Motivational Interviewing Skills Code scheme (MISC; Miller and Rollnick, 2002) was used to assess therapist and client functioning and one study reported that training coders took more than 40 hr across 4 weeks (Moyers et al., 2003).

Motivated by easing coding burdens, some researchers only record relatively short episodes of behavior, instead of analyzing slices of longer recordings. At a technical level, such approaches do not involve thin slices as slices are not being extracted from a longer interaction. Yet, these measures do align with theoretical perspectives that short interactions (5 min or less) can reliably capture behavior and other social constructs. As one example, communication constructs such as cognitive sensitivity and responsivity were reliably captured from 5-min parent, child, and/or sibling interactions (Prime et al., 2014, 2015; Sokolovic et al., 2021). The research design was developed specifically to provide quick, cost-effective, and validated measures of communication styles and the authors concluded that using a thin-slice approach (i.e., using interactions <5 min) is a viable alternative to coding longer interactions. Yet, as valuable as such approaches may be, they do not answer the question of whether 5 min is better than, say, 1 or 2 min, or worse than 10 or 15 min[5].

THEORETICAL PERSPECTIVES ON BEHAVIORAL EXPRESSION WITHIN THIN SLICES

Beyond the practicality of thin slices, there is confidence to be gained in the legitimacy of thin slices to represent behavior as thin slices are related to a number of larger theories about behavior in social interactions. There is an evolutionary advantage to drawing inferences about a person from glimpses of their behavior. Just as primate expressions involving shrieks that signal anger and potential attack (Chevalier-Skolnikoff, 1973), a loud voice is perceived as and validly indicates social dominance in humans (Hall et al., 2005). These brief behavioral expressions (in essence, thin slices of behavior) provide information for a perceiver to act upon, potentially conferring an evolutionary advantage of making decisions about approach or avoidance, communication,

[2]One literature we do not address concerns the "acquaintance effect" (Connelly and Ones, 2010), where accuracy of personality judgment is compared between people who know the target for longer or shorter amounts of time.

[3]Ambady and Rosenthal (1992) referred to such correlations as reflecting the accuracy of thin-slice judgments, but in fact such correlations speak to predictive validity.

[4]Retrieved February 2, 2021 per Google Scholar.

[5]It is likely that many researchers compare a variety of different slice lengths or locations in the process of developing their methodology but do not report that entire process in their publications. Therefore, there may be informal comparative thin-slice research in existence that we are not aware of.

and further interaction (Zebrowitz and Collins, 1997). Being able to make judgments and decisions based on little information is required for functional daily life. For instance, racial bias is detectable via thin slices (Richeson and Shelton, 2005) and knowing such information is undoubtedly of value in making interaction decisions.

Theoretical support for thin slices is also evident in the bedrock of personality science, behavioral consistency, whereby an individual's behavior is consistent across situations and time. That is, traits will exist within an individual with some regularity across various situations, such as at work and at home (Epstein, 1979). Underlying this tenet is the idea that personality is evident in behavioral expressions, and a host of research supports this notion (Allport, 1937; Murphy, 2007; Leikas et al., 2012; Letzring et al., 2021). While acknowledging situational variance, research persistently supports behavioral consistency in the expression of personality (Funder and Colvin, 1991; Shoda, 1999; Fleeson and Law, 2015; Geukes et al., 2017). Thus, the use of thin slices, as small glimpses into a person's behavior, fits within the behavioral consistency premise in personality science.

While not quite a theory, Egon Brunswik applied his existing visual perception paradigm to social situations, and the Brunswik lens model is often applied in understanding social perception processes (Brunswik, 1956; Hall et al., 2019). The Brunswik lens model specifies that individual behavioral cues are related to impressions of a target as well as a target's actual personality or other characteristic of interest. Behavioral cues that relate to observer impressions of a target provide insight into observers' implicit theories about the trait or characteristic in question. Likewise, cues related to a target's measured state or trait provide insight into how states or traits are revealed through behavior. Measurements of behavior are an essential feature of the Brunswik lens model and thin slices provide that opportunity. In sum, beyond the practicality of using thin slices for behavioral coding, a number of larger theories of human interaction support the notion that thin slices are appropriate reflections of human behavior and, in turn, offer support to a researcher's decision to adopt thin slices as a behavioral measurement tool.

DECIDING TO USE THIN SLICES FOR BEHAVIORAL MEASUREMENT

As with any decision to use behavioral coding, a number of considerations should be taken into account deciding whether thin-slice coding and stimuli are appropriate to a research question or design (e.g., Baucom et al., 2017; Blanch-Hartigan et al., 2018). Here, we outline some topics a researcher might consider in making such decisions. At a basic level, any coded behavior or construct needs to be *observable* (Funder, 1995; Ambady et al., 2000). Is there existing evidence that the construct of interest is (potentially) observable at all? For instance, while extraversion may be easily judged from thin slices, agreeableness is harder to detect, and thus, potentially harder to code from thin slices (Ames and Bianchi, 2008). Likewise, the observability of a behavior or construct within the specific setting of the behavioral stream should be considered. Is it reasonable to expect that attraction or suggestibility would be evident in a medical setting or interactions with children? For instance, Whalen et al. (2020) found that obtaining thin-slice reliability for judging preschoolers' personality traits varied between structured tasks (e.g., unwrapping an empty box compared to telling a story). The authors noted that certain tasks may have restricted the range of expression, in turn making personality harder to observe and reliably measure from thin slices in those particular settings. Thus, a researcher should think about the potential observability of a behavior or construct of interest within the specific research setting.

When researchers design lens model studies, they must choose what behaviors to code and include in their models. There can be wide variation in the "success" of such a model, depending on whether the cues have a priori likelihood of being related to the criterion (trait or state) and/or to judges' impressions. When extensive prior research enables the researcher to pick highly diagnostic cues, they will likely find a great deal of evidence for how accurate judgments are mediated by specific cues (e.g., Laukka et al., 2013). In contrast, a study that is more exploratory might find there is accurate judgment but fail to identify any validly utilized cues because, presumably, judges were able to validly utilize cues that the researcher did not measure (e.g., Ruben and Hall, 2016). Though these examples do not speak to the intrinsic wisdom of using thin slices, whether the lens model has high explanatory strength is a matter of the coded cues' theoretical and empirical relevance to the trait or state being studied.

Researchers also should also consider the *consistency* and *frequency* of behavioral expressions across and within interactions and across settings. Some nonverbal behaviors seem to be expressed more consistently across interactions as compared to others. For instance, gaze and nods showed relatively consistent expression within interactions and across various interaction settings (e.g., zero-acquaintanceship dyadic interactions, job interviews, medical settings) (Patterson, 1973; Leikas et al., 2012; Murphy et al., 2015). However, more variability has been found for behaviors such as speaking time, indicating that (shorter) slices may not adequately capture or be representative of that behavior across an interaction or setting. Relatedly, the possible frequency of a behavior is pertinent. A behavior such as crossed arms may occur less frequently as a whole and thus capturing that behavior and its representativeness within slices may be less likely.

Another important decision is whether to code *molar constructs* and/or *micro behaviors*. Molar constructs refer to higher levels of abstraction and may be more holistic in nature. Examples of molar constructs include dominance, awkwardness, perceived intelligence, or pleasant style of speech. Often, such constructs would use Likert-style ratings for measurement. On the other hand, individual (micro) behaviors that are coded descriptively, not requiring much if any coder inference, may be more concrete or exact, as they usually represent specific behaviors or expressions, such as number of smiles, duration of eye gazing, or speaking time. While individual behaviors such as smiling or gazing also can be measured with Likert-style ratings (e.g., Briton and Hall, 1995; Wang et al., 2020), often such behaviors are measured as frequency counts or duration. The decision between coding molar constructs or micro behaviors depends entirely on what the researcher's interests are. Micro behaviors may suit some research goals whereas molar judgments (e.g., ratings of awkwardness, sincerity, or truthfulness) may better answer other research questions (Funder and Colvin, 1991; Leikas et al., 2012). Conceivably, optimal slice lengths might vary depending on how much inference coders or raters are required to make. One overarching aspect in aforementioned thin-slice coding decisions is whether topics of interest are affective states or traits. States are more temporary experiences while traits are more stable across time and situations (Augustine and Larsen, 2015). The questions of observability, frequency, duration, molar constructs, and micro behaviors may all tie into whether the construct of interest is a state or trait, though a wide array of both states and trait research has employed thin-slice methods, suggesting that thin slices may be appropriate to either.

COMPARATIVE THIN-SLICE RESEARCH

Three strands of comparative thin-slice research will be described in the following sections: *representativeness*, *thin-slice predictive validity*, and *accuracy-length validity*. It must be said at the outset that in all three of these domains there is much methodological variation across studies—in the behaviors that are measured and the constructs that are judged, the outcome variables used for prediction, specific measurement methods (e.g., molar vs. micro), slice lengths, temporal position of the slice, length of the total recorded behavior, and other variables, meaning that no one study can settle questions regarding optimal slice lengths or slice locations. All we can provide is an overview in the hope that some knowledge on these questions is better than no knowledge at all.

Also important is that, for most purposes, the relevant metric for interpretation is the magnitude of the correlations being compared, not whether they are statistically significant, as the latter is tied to the irrelevant (for these purposes) factor of sample size. The challenge for a researcher who is interpreting such correlations, or planning a study, is to decide whether the relevant correlations are big enough, or similar enough, according to their own criteria to justify using thin slices rather than the "total" behavior, whatever that may be. An illustration regarding predictive validity will help. Let us say that total interpersonal gaze in a 5-min interaction predicts observers' ratings of likeableness at $r = 0.30$, while the same correlation based on a 1-min slice of that 5-min interaction is $r = 0.26$. There is some loss in magnitude of prediction, to which the researcher would apply their own decision rule. One researcher might decide the loss is too much and opt to stick with the "total" gazing measurement, or perhaps decide to use a 2-min slice that yielded a predictive correlation of $r = 0.29$. Another researcher might decide that the loss of magnitude by using the 1-min slice is well worth the savings in personnel time and cost. Of course, the researcher can also conduct a power analysis to decide what the sample size should be if they want to conduct inferential statistical tests (Abraham and Russell, 2008). Ultimately, it is up to an individual researcher to decide which approach fits their needs.

Finally, for studies looking at representativeness and predictive validity, it goes without saying that a great deal depends on the psychometric quality of the behavioral coding (e.g., Moskowitz and Schwarz, 1982). Slices with strong intercoder reliability will show more promising evidence for the value of the slices than slices with weak intercoder reliability—so a general statement of "slices work well" or "slices don't work well" could easily be confounded by the psychometric quality of the slices, not the slice length or location *per se*. Poorly measured slices will not correlate well with each other or with other variables. Increasing the number of raters is an easy way to improve reliability, yet leads us back to the issue of labor and time. Other sources can provide more detail and guidance in assessing and improving psychometric quality (for example, by employing more and/or better trained coders) (Li et al., 1996; Rosenthal, 2005).

Representativeness

Representativeness refers to the ability of one slice to be interchangeable with another slice, or to adequately represent the "total" behavior. Both of these aspects of representativeness are addressed via correlations, often using the same statistics that researchers commonly used for assessing reliability (e.g., coefficient alpha, intraclass correlation, corrected part-whole correlations). For that reason, sometimes representativeness is referred to as reliability but sometimes it is referred to as validity when the question is the correlation between a slice and the total (because the total is operationally defined as the ground "truth" and the correlation of the slice to the total indicates the validity of the slice) (Murphy et al., 2015).

It is important to note that the reliability discussed here is different from the reliability of coders (as discussed in the previous section) or the internal consistency of items in a coding scheme. Also, readers should be clear that we are not concerned with comparisons of mean levels of any given coded or rated behavior (for example, whether the amount of smiling varies across slices). While that question is of great interest for some research purposes (e.g., Ruben et al., 2015), it does not speak to the representativeness of slices, which is assessed via inter-slice correlations or by slice-total correlations.

Inter-Slice Reliability (Interchangeability)

Here the question is whether the individuals whose behavior is measured maintain their rank in the distribution from slice to slice—for example, if Jacinda smiles a lot (relative to other people) in the first slice, does Jacinda also smile a lot (again, relative to others) in other slices? If this kind of reliability is high, it means that people's relative amount of the behavior is well captured in any slice.

Hall et al. (2009) extracted three 1-min slices from early, middle, and late in a 15-min medical interview and obtained raters' impressions of rapport in each slice. The three slices were strongly correlated with each other, ranging from 0.60–0.82, suggesting good interchangeability. It was not stated, however, whether the slices were rated consecutively (one after the other). If that was the case, there could be some inflation due to carryover of a rater's impression from one slice to the next.

Murphy et al. (2015) investigated inter-slice reliability based on data from four separate studies in which specific nonverbal behaviors were coded in 30-s or 1-min slices from video recorded interactions that originally ranged from 5 to 9 min in length (though not all studies coded all behaviors). Inter-slice reliability (as assessed with intraclass correlations) was strongest for gazing behavior but slices of gestures, nods, self-touch, and smiles also reached reasonable levels of inter-slice reliability, providing empirical evidence of slice interchangeably for those measured behaviors. However, speaking time was a notable exception in showing considerable variation across studies and there was no evidence that one slice of speaking time reasonably predicted any other speaking time slice within an interaction.

As mentioned, with higher inter-slice reliability, a researcher can be more confident that thin slices are appropriate. For planning stages, researchers could consider applying the Spearman–Brown formula to calculate how many slices are needed to achieve a given reliability level for the slices combined (based on estimates from pilot data or past studies) (Brown, 1910; Spearman, 1910). The formula could also be applied *post-hoc*, essentially examining intercorrelations to see if they correlate well enough to justify combining them into a "total" (This is conceptually analogous to calculating reliability with Cronbach's α.). For further information on using the Spearman-Brown formula to establish inter-slice reliability see Li et al. (1996) and Murphy et al. (2015).

In general, there is relatively little comparative research specifically investigating inter-slice reliability. While the above research provides some evidence of slice interchangeability, such research did not answer questions about slice-length comparisons (e.g., 20 vs. 40 s slices, or 1 vs. 3 min, slices, etc.) or other measured behaviors or macro constructs. What is quite clear is that future research is needed to investigate inter-slice reliability and the related questions.

Slice-Whole Validity

Slice-whole validity is another way of examining representativeness, not between slices but between any given slice and the totality of the measured behavior. In comparison to inter-slice reliability research, there is a larger set of studies investigating slice-whole validity. Murphy (2005) investigated the slice-whole validity of five specific nonverbal behaviors (gestures, nods, self-touches, smiles, time spent gazing at partner) by coding 50 participants who engaged in 15-min dyadic social interactions. Reliable judges coded three randomly selected 1-min slices. A separate set of judges coded the full 15-min interactions for each behavior. With the exception of self-touch, 1-min slices showed acceptable slice-whole validity (i.e., moderate to large effects) based on part-whole correlations (which removes possible inflation from including a given slice from the behavior total). And all behaviors reached acceptable slice-whole validity when adding two of the randomly selected slices together. (This also true for summing three slices except for nodding behavior, which showed a consistent *decrease* in representativeness as slices were added together). In a more extensive analysis of four studies, Murphy et al. (2015) (described in previous section) found modest to strong slice-whole validity for gaze, nods, and smiles; yet, there was little evidence of such validity in speaking time and gestures (see Murphy et al., 2015 Figure 2). The location of the extracted slice mattered; generally, slices extracted from the middle of interactions showed stronger validity than slices from the beginning or end of an interaction.

Evidence of slice-whole validity also was found in several studies involving clinical settings. In Perrault (2020), participants viewed a clinician video biography and rated the clinician on various constructs such as trust and liking. Results showed that 46-s slices were equally predictive of ratings across eight constructs when compared to 63 and 80-s slices (which constituted the "whole" interaction). A corpus of videotaped patient-counselor clinical sessions (20–30 min) had previously been coded for verbal constructs such as open-ended questions and affirmation (Carcone et al., 2015). Each session was divided into four equal segments and then 1- or 2-min slices were randomly selected from each segment. There was considerable variation depending on the construct but the authors concluded that using six 2-min slices from across the longer interaction sufficiently captured the whole interaction. And slice-whole validity dropped only slightly when using four 2-min slices (compared to six 2-min slices). In a similar study, acceptable levels of slice-whole validity was found for 5-min, and especially 10-min, slices extracted from motivational clinical interviews (a type of trained counseling style) in five of seven measured constructs (Klonek et al., 2015). And, importantly, the authors also investigated the location of extracted slices and found that 5-min slices extracted from after the first 5-min of the interaction had the strongest validity, supporting previous recommendations to avoid using slices from the very beginning of interactions (Murphy et al., 2015; Hirschmann et al., 2018).

In another study, across five measured constructs (liking, attention, coordination, trust, rapport), there was evidence of slice-whole validity for 1.5-min slices of patient-physician interactions, though the effects were modest in magnitude (Foster, 2015)[6]. Caperton et al. (2018) investigated the minimum slice length needed to capture the whole of therapist behavior during motivational interviewing sessions. Previously recorded sessions (average length = 28 min) had been coded for therapist utterances. The authors found using ∼8 min reliably captured the whole session. In another study, 10-min slices extracted from 40+ min mother-child interactions showed strong evidence of slice-whole validity in measuring maternal sensitivity and maternal feedback (Hirschmann et al., 2018). Such findings provide further support to the notion that smaller excerpts from longer behavioral streams can adequately represent some behaviors across an interaction.

As a whole, research generally suggests that excerpts shorter than their respective totals have the potential to reliably capture coded behaviors or constructs. But of course, any conclusions from the aforementioned slice-whole validity findings are qualified by the many other variables at play, including the

[6]Foster (2015) concluded that thin-slice ratings were *not* comparable to full interaction ratings. However, the magnitude of their effects (*r*s between 0.33 and 0.49) align with other comparative thin-slice research findings on slice-whole validity.

total length of the interaction, the type of interaction, and constructs being measured. And there is research refuting slice-whole validity. James et al. (2012) recorded interactions between mothers and their deaf children and found that 3-min slices of play behaviors were not representative of whole 18-min interactions. The authors concluded: "If we had used 3 min segments [slices] to code data then our conclusions would have differed considerably from the findings based on an entire play session" (p. 357), providing a cautionary message against thin-slice coding for certain behaviors in specific contexts and/or with targeted populations.

Thin-Slice Predictive Validity

The second comparative thin-slice tradition concerns predictive validity: How do slices fare, compared to total, in predicting a different outcome variable (that is, a variable that is different from the behavior that is measured in the slice)? As a hypothetical example, consider a study of nervousness and smiling during 15-min video recorded dyadic interactions. Targets complete a self-report measure of nervousness and judges count target smiles in 1-min slices across the entire 15-min interaction, enabling the researcher to calculate 1-min slices of smiling and the entire 15-min of smiling. If the correlation between target nervousness and the 1-min smile coding is close to or the same as the correlation between target nervousness and the 15-min smile coding, it would suggest that little to no predictive validity is lost by using the 1-min coding. We use the term "predictive" loosely, meaning the variable does not have to be measured literally after the behavior occurred. We also refer to the predicted variables as "outcome variables," without implying the "outcome" has a causal relationship to the behavior that is measured.

Ambady and Rosenthal (1992) conducted a meta-analysis of correlations between behaviors coded from excerpts and a wide array of outcome variables. The authors found no association between slice length (which varied from under 30 s to 5 min) and the strength of the predictive correlations. Although this study was groundbreaking, not only in introducing the term "thin slices" but in showing that thin slices can predict other variables, it was not optimal for testing the impact of slice length—because the analysis was necessarily a comparison between studies rather than within studies, meaning that both slice lengths and outcome variables were confounded with other study variables (sample characteristics, for example) and therefore made for an imprecise test of the slice-length question.

Ambady and Rosenthal (1993) began the tradition of comparing slice lengths *within* a study. In two studies of thin slices of teacher behavior predicting performance evaluations, they compared 2, 5, and 10-s excerpts and found that although the correlations for longer slices were stronger, the longer slices did not predict to the criterion variable of teacher effectiveness better than did the shorter slices at statistically-significant levels, indicating evidence of predictive validity.

In Roter et al. (2011), three 1-min slices of verbal behavior (selected from early, middle, and late, as well as combined) were as predictive of independent judgments of rapport between clinicians and patients as was coding of the full 15-min interaction; also the single 1-min slices were not much different than the 3-min combined slice. In a different analysis based on the same database, Hall et al. (2009; described above) obtained ratings of rapport from three 1-min video slices. The individual slices and their 3-min total were compared in terms of correlations with a wide range of other variables. In general there was some loss of predictive validity for the 1-min slice (which was the 1st min of the interaction) compared to all 3 min, with some variables showing a fairly strong loss of prediction. However, for a number of variables the loss was not great or even non-existent, such as coder ratings of interest, warmth, and respect, and analog patients' ratings of the clinician's competence, calmness, communication quality, and self-confidence, as well as patients' own satisfaction.

Tskhay et al. (2017) obtained ratings of charisma from 5, 15, and 30-s silent slices from a 1-min video. There was not much difference between the slices and the total for predicting independent ratings of leadership potential and several other variables including gender, eye contact, wearing glasses, and physical attractiveness.

In a study of job applicants, slices (<130 s in length) applicant audio cues (e.g., prosody, pauses, etc.) predicted observer-rated hireability impressions based on the whole interview (Nguyen and Gatica-Perez, 2015), indicating that shorter excerpts of a job interview could be predictive of interview outcomes. Additionally, the results indicated that no one slice was markedly more predictive than another, though the thin slices were always less predictive than full interview outcomes. Similar results were found in analyzing participants in audio-recorded game interactions, whereby temporal position of 1-min slices from the beginning, middle, or end showed that any slice was equally predictive of game performance (Lepri et al., 2009).

Murphy et al. (2019) and Wang et al. (2020) offered multivariable examinations of predictive validity by examining multiple behaviors and multiple outcome variables. Murphy et al. (2019) examined predictive validity in five studies for six nonverbal behaviors (nodding, smiling, gesturing, gazing, self-touch, and speaking time). While 1-min slices were somewhat worse in predicting a highly varied list of 33 outcome variables than the whole 5-min videos were, 2-min slices were nearly as predictive as the 5-min totals. Wang et al. (2020) collected self-rated, perceiver-rated, and objectively measured data within one study based on a 5-min interaction. One-min slices were rated for verbal and nonverbal behaviors via global impressions, using the same rater for all five slices and also using a different rater for each slice. For single slices, results indicated no clear pattern for optimal slice locations. In general, single slices had weaker predictive validity than the total (5 slices combined). However, slices of 2 or 3 min were, in general, equal to 5-min total in predictive validity. The magnitude of correlations was similar when same- vs. different-coder methodologies were compared.

As a whole, research on thin-slice predictive validity suggests that thin-slice measurements may adequately predict an outcome variable, in comparison to variable measurement of an entire interaction. But once again, the preceding findings are qualified by many existing contingencies such as different measured constructs, outcome variables, slice lengths, and total interaction lengths, among other considerations. Researchers using the same specific behaviors and/or variables as the aforementioned studies could examine specific findings for more details on how to conduct their own behavioral measurement.

Accuracy-Length Validity

There is a longstanding tradition of using thin slices as stimuli in studies of person perception accuracy. The study of accurate person perception can be considered a special case of predictive validity because accuracy is defined as the match or correlation between a perceiver's judgment and the criterion (i.e., the "correct answer" on a test item). Depending on the cue modality and the construct being judged, tests of interpersonal judgment accuracy vary considerably in the length of their stimuli, but not a great deal is known about the impact of variations in slice length. Many studies of emotion recognition use photographs, exposed for varying amounts of time, while others use dynamic stimuli of <20 s; example tests are the Diagnostic Analysis of Nonverbal Accuracy–Adult Prosody (DANVA2-AP; Baum and Nowicki, 1998), the Geneva Emotion Recognition Test (GERT; Schlegel et al., 2014), the Multimodal Emotion Recognition Test, MERT, Bänziger et al., 2009), and the Profile of Nonverbal Sensitivity (PONS; Rosenthal et al., 1979).

In the accuracy field, researchers' interest in slice length is twofold. First, for psychometric reasons a test developer might compare accuracy resulting from different slice lengths in order to create a test that has an optimal difficulty level. If the slices are too short, perhaps judgment accuracy is impossible, while if the slices are too long, the test might be too easy. For example, the PONS test was originally piloted with 5-s audiovisual clips, but this was reduced to 2 s to reach a psychometrically optimal difficulty level. Similarly, Gesn and Ickes (1999) reported choosing 15-s clips from dyadic interactions as stimuli in their judgment study, based on a process of trial and error in the piloting phase.

Slice length also has theoretical, not just methodological, interest to accuracy researchers because they want to know how accuracy of judging some attribute, state, or trait of target others may be related to the length of exposure to targets. Researchers interested in the validity of first impressions are especially likely to ask this question (Ambady and Skowronski, 2008). We know that people draw conclusions about others automatically and very quickly (Todorov, 2017). In other words, people rely intuitively on thin slices. In one study, almost a third of hiring managers reported making a decision about an applicant's suitability within the first 5 min of an interview (Frieder et al., 2016). And for many kinds of judgments, impressions are formed based on stimuli far shorter than that. Thus arises the question of how much time is needed to form an accurate judgment. Ambady et al. (1999) found that accuracy of judging sexual orientation was significantly greater for 10-s video clips than for 1-s clips. Rule and Ambady (2008) found that above-chance accuracy at judging sexual orientation could be obtained even when photographs were displayed for 50 ms but not for 33 ms. Some types of nonverbal affective stimuli can be judged very accurately from extremely minimal exposure length. The Japanese and Caucasian Brief Affect Recognition Test (JACBART; Matsumoto et al., 2000), which is based on prototypical, posed, "basic" emotions, achieved high judgment accuracy with exposures as short as 1/5 s. In contrast, spontaneously produced nonverbal affective expressions can be quite hard to judge, even at considerably longer exposures (Gesn and Ickes, 1999; Hall and Schmid Mast, 2007).

In the domain of personality judgment, researchers often use longer stimuli than those used for emotion recognition. In examining length of exposure, Blackman and Funder (1998) found that video clips of 25–30 min produced significantly greater accuracy in judging personality than clips of 5–10 min. However, there was almost equal accuracy for 5–10 min clips compared to 15–20 min. Letzring et al. (2006) found that interacting with someone for 3 hr did not produce more accurate personality judgment than interacting with someone for 50 min. And Fowler et al. (2009) looked at slices of 5, 10, and 20 s in a study of accuracy of judging psychopathy in criminal offenders. For one criterion measure of psychopathy, the *shortest* slice length produced the highest accuracy.

Carney et al. (2007) examined Big Five judgment accuracy for slices of 5, 20, 45, 60, and 300 s duration and found that extraversion and conscientiousness showed significant linear trends indicating increased accuracy for longer slices, and agreeableness showed a marginally significant linear trend. However, accuracy for neuroticism and openness to experience did not show a linear trend for slice length. Importantly, accuracy was above chance even at 5 s of exposure for all of the traits except agreeableness. There was a gain in accuracy in going from 60 to 300 s, but it was not substantive. Similar results were found in a study comparing 30-s, 1-, 3-, and 5-min slices in accurately perceiving personality traits (Krzyzaniak et al., 2019). When analyzing all traits combined, accuracy (referred to as distinctive accuracy within the study) did not improve with longer slice length, but there was notable exceptions within specific traits, suggesting that appropriate slice lengths depend on the construct being measured.

Hall et al. (2008), in a meta-analysis on interpersonal accuracy studies, performed an analysis of slice length across studies. Exposure lengths ranging from <1 s to 45 min (including studies using photographs) revealed no evident trend associating slice length to judgment accuracy. However, the studies varied widely in terms of what construct was being judged (emotion, traits, etc.), and between-studies comparisons of slice lengths are confounded by all of the other methodological differences between the studies, potentially obscuring the duration effect.

Only one study that we know of has looked at accuracy and slice length for textual material. Hall et al. (2021) divided college students' two-page personal narratives into fifths and looked at individual fifths (slices) as well as cumulative slices in terms of readers' accuracy of judging the Big Five traits. For extraversion, agreeableness, and openness to experience, longer cumulative slices produced more accurate judgment, but this was not consistently the case: for neuroticism all of the cumulative slices showed higher accuracy than the total narrative did, due to the fifth and final slice producing no accuracy at all.

An unexplored question is how the criterion used for judging what is the "correct answer" on such a test might affect accuracy overall, and, relevant to our current interests, accuracy for different slice lengths or locations. The criterion for studies of judging personality is typically the target's self-report of personality, sometimes supplemented with reports by friends or family. For judging attributes of people there are often objective criteria that can be used; for example, if hierarchical status is being judged, the researcher might have access to the organizational chart of the company in question. For judging affective states, a wide variety of criteria are used including giving the target an assigned emotion to portray either by posing or by re-enactment of a lived experience, manipulation of situational stimuli such as what kind of photo or movie the target is watching, and the consensus of viewers (including the researchers) as to what affective state is being shown.

(MANY) UNANSWERED QUESTIONS

Like automated methods, the thin-slice method is not a panacea to solving coding burdens. It is impossible to state that thin slices would work for all behaviors. In fact, the research is quite clear that there is considerable variability depending on a number of factors. The measured construct may not be reliably measured via thin slices, which may be due to the construct's consistency or frequency of expression (Leikas et al., 2012; Murphy et al., 2015). Unique settings and/or specialized populations may not allow for adequate capture of behaviors via thin slices, as shown by the lack of slice-whole validity in play behaviors of deaf children with their mothers (James et al., 2012). Alternative measurements of the same construct (e.g., counting vs. rating, or intervals vs. whole) might not always work equally well or show parity within thin-slice measurement (Blanch-Hartigan et al., 2018).

Conclusions about appropriate slice length and location of the slice also cannot be universally applied. One might predict that longer slices could yield higher levels of representativeness, predictive validity, and accuracy. Unless a behavior is manifested with extreme reliability over time, there is inevitably an information loss when using shorter excerpts to represent a longer interaction. However, research also shows that the magnitude of that loss could be negligible, depending on the measured construct and slice length (Murphy, 2005; Carcone et al., 2015; Murphy et al., 2015). Research on behavioral consistency also reiterates that aggregation of data (e.g., behavioral measurements) from across situations, targets, and judges or coders increases the reliability of findings (Epstein, 1979, Moskowitz and Schwarz, 1982). While some predictive validity research suggests that the temporal position of the slice may not matter (Lepri et al., 2009; Nguyen and Gatica-Perez, 2015), selecting slices from after the beginning of an interaction (e.g., after the 1st min) has some empirical support, as there is some literature indicating lower slice-whole validity from very beginning slices (Klonek et al., 2015; Murphy et al., 2015).

The source of behavioral streams is an area worth further investigation. Much of the research cited here involved thin slices extracted from video stimuli recorded in a laboratory. Yet, thin-slice work is relevant beyond laboratory interactions. Thin-slice research exists across a wide array of domains such as judgments of online social networks, televised soccer and sport matches, TED talks, teachers' classroom behavior, and prison interviews, among many content areas (Fowler et al., 2009; Pretsch et al., 2013; Furley and Schweizer, 2014; Stopfer et al., 2014; Gheorghiu et al., 2020). As more evidence accumulates for the reliability and validity of thin-slice methods, it will be important that future comparative thin-slice research investigate stimuli from beyond the laboratory.

The vast majority of thin-slice research, whether comparative in nature or otherwise, involves White individuals from European and/or American backgrounds. And almost all of that research is limited to young adults or children. There are cross-cultural comparisons of person perception processes using thin slices. Using 10-s slices extracted from 3-min interactions, Place et al. (2012) found consistent levels of accuracy in detecting speed daters' romantic interest in samples from the U.S., Germany, and China. Thin-slice research showed equivalent levels of accuracy in judging rapport in U.S. and Greek participants (Bernieri and Gillis, 1995) and consensus in personality impressions were found in both U.S. and Chinese participant samples (Albright et al., 1997). Such studies at least extend thin-slice work into broader population samples, but the numbers are few and far between. And such research does little to acknowledge racial and/or ethnic identities even within the measured samples (Roberts et al., 2020). We are not aware of any comparative thin-slice research involving participants who are not predominantly White, European, and/or American. It is quite clear that there are likely cultural factors of targets or even coders that could limit any generalizability into populations not previously studied (Masuda et al., 2020).

CONCLUSIONS

The thin-slice measurement technique itself is applicable to any behavioral domain, potentially even for behavioral measurement of non-human populations (Jamieson et al., 2017). Perhaps an expanded view of what constitutes a "thin slice" (beyond 5 min) is warranted given research on longer slices (e.g., 10 min) from lengthier interactions (e.g., >40 min) that shows similarities to findings examining shorter slices from briefer interactions (e.g., 30 s from 5 min) (Caperton et al., 2018; Hirschmann et al., 2018). At a conceptual level, there is evidence that thin slices reliably and validly measure behavior across various domains, including zero-acquaintanceship interactions and clinical settings.

At this initial stage, comparative thin-slice research provides some cautious optimism for researchers concerned with slice-whole validity and predictive validity, and those who use thin slices in interpersonal accuracy research. **Appendix A** is a representative list of cited studies on comparative thin-slice research. The **Appendix** is provided as potential resource for other thin-slice researchers who seek further information about reliability and validity of thin-slice measurement decisions. (It is important to note that the list is not intended to be exhaustive and the listed studies do not necessarily indicate support for the comparative construct). We also suggest reviewing the previously-mentioned factors listed in the "Deciding to

Use Thin Slices for Behavioral Measurement" section and using the Spearman-Brown formula as discussed in the "Representativeness" section (see also Li et al., 1996; Murphy et al., 2015).

Given the current replication crisis in psychology (and beyond), the use of sound research practices is now more important than ever (Schimmack, 2020). Without reliable and valid measurement, any conclusions based on such measurements are acutely curtailed, if not nullified (Flake et al., 2017; Eronen and Bringmann, 2021). Of course, it is inaccurate to state that thin slices can be used any time a researcher wishes to reduce coding burdens by coding shorter excerpts of behavior. And, every researcher needs to make their own decision about whether a given degree of representativeness, predictive validity, or accuracy-length validity is "good enough" for their research purposes. Measurement is never perfect; each researcher decides at what point their measurements satisfy their standards and their resources. We hope this article may be a potential resource for researchers considering using thin-slice behavioral measurement; by reviewing current comparative thin-slice literature, researchers could identify potential sources which may support the many decisions going into using thin slices to measure behavior.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the writing the manuscript and approved the submitted manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.667326/full#supplementary-material

REFERENCES

Abraham, W. T., and Russell, D. W. (2008). Statistical power analysis in psychological research. *Soc. Personal. Psychol. Compass* 2, 283–301. doi: 10.1111/j.1751-9004.2007.00052.x

Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., Winquist, L., et al. (1997). Cross-cultural consensus in personality judgments. *J. Pers. Soc. Psychol.* 72, 558–569. doi: 10.1037/0022-3514.72.3.558

Allport, G. W. (1937). *Personality: A Psychological Interpretation*. New York, NY: Holt.

Ambady, N., Bernieri, F. J., and Richeson, J. A. (2000). "Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream," in *Advances in Experimental Social Psychology, Vol. 32*, ed M. P. Zanna (Cambridge: MA: Academic Press), 201–271. doi: 10.1016/S0065-2601(00)80006-4

Ambady, N., Hallahan, M., and Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *J. Pers. Soc. Psychol.* 77, 538–547. doi: 10.1037/0022-3514.77.3.538

Ambady, N., and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol. Bull.* 111, 256–274. doi: 10.1037/0033-2909.111.2.256

Ambady, N., and Rosenthal, R. (1993). Half a minute: predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J. Pers. Soc. Psychol.* 64, 431–441. doi: 10.1037/0022-3514.64.3.431

Ambady, N., and Skowronski, J. J. (eds.). (2008). *First Impressions*. New York, NY: Guilford Publications.

Ames, D. R., and Bianchi, E. C. (2008). The agreeableness asymmetry in first impressions: perceivers' impulse to (Mis)judge agreeableness and how it is moderated by power. *Person. Soc. Psychol. Bull.* 34, 1719–1736. doi: 10.1177/0146167208323932

Augustine, A. A., and Larsen, R. J. (2015). "Personality, affect, and affect regulation," in *APA Handbook of Personality and Social Psychology, Volume 4: Personality Processes and Individual Differences*, eds M. Mikulincer, P. R. Shaver, M. L. Cooper, and R. J. Larsen (Washington, DC: American Psychological Association), 147–165. doi: 10.1037/14343-007

Bänziger, T., Grandjean, D., and Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion* 9, 691–704. doi: 10.1037/a0017088

Baucom, B. R. W., Leo, K., Adamo, C., Georgiou, P., and Baucom, K. J. W. (2017). Conceptual and statistical issues in couples observational research: rationale and methods for design decisions. *J. Fam. Psychol.* 31, 972–982. doi: 10.1037/fam0000370

Baum, K. M., and Nowicki, S. Jr. (1998). Perception of emotion: measuring decoding accuracy of adult prosodic cues varying in intensity. *J. Nonverbal Behav.* 22, 89–107. doi: 10.1023/A:1022954014365

Bensing, J. M., Verheul, W., and van Dulmen, A. M. (2008). Patient anxiety in the medical encounter: a study of verbal and nonverbal communication in general practice. *Health Educ.* 108, 373–383. doi: 10.1108/09654280810899993

Bernieri, F. J., and Gillis, J. S. (1995). The judgment of rapport: a cross-cultural comparison between Americans and Greeks. *J. Nonverbal Behav.* 19, 115–130. doi: 10.1007/BF02173170

Black, M. P., Katsamanis, A., Baucom, B. R., Lee, C.-C., Lammert, A. C., Christensen, A., et al. (2013). Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Commun.* 55, 1–21. doi: 10.1016/j.specom.2011.12.003

Blackman, M. C., and Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *J. Exp. Soc. Psychol.* 34, 164–181. doi: 10.1006/jesp.1997.1347

Blanch-Hartigan, D., Ruben, M. A., Hall, J. A., and Mast, M. S. (2018). Measuring nonverbal behavior in clinical interactions: a pragmatic guide. *Patient Educ. Couns.* 101, 2209–2218. doi: 10.1016/j.pec.2018.08.013

Briton, N. J., and Hall, J. A. (1995). Gender-based expectancies and observer judgments of smiling. *J. Nonverbal Behav.* 19, 49–65. doi: 10.1007/BF02173412

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Br. J. Psychol.* 3, 296–322. doi: 10.1111/j.2044-8295.1910.tb00207.x

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments, 2nd Edn*. Oakland, CA: University of California Press. doi: 10.1525/9780520350519

Caperton, D. D., Atkins, D. C., and Imel, Z. E. (2018). Rating motivational interviewing fidelity from thin slices. *Psychol. Addict. Behav.* 32, 434–441. doi: 10.1037/adb0000359

Carcone, A. I., Naar, S., Eggly, S., Foster, T., Albrecht, T. L., and Brogan, K. E. (2015). Comparing thin slices of verbal communication behavior of varying number and duration. *Patient Educ. Couns.* 98, 150–155. doi: 10.1016/j.pec.2014.09.008

Carney, D. R., Colvin, C. R., and Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *J. Res. Pers.* 41, 1054–1072. doi: 10.1016/j.jrp.2007.01.004

Chakravarthula, S. N., Baucom, B., Narayanan, S., and Georgiou, P. (2021). An analysis of observation length requirements for machine understanding of human behaviors from spoken language. *Comput. Speech Lang.* 66:101162. doi: 10.1016/j.csl.2020.101162

Chevalier-Skolnikoff, S. (1973). "Facial expression of emotion in nonhuman primates," in *Darwin and Facial Expression: A Century of Research in Review*, ed P. Ekman (Cambridge: MA: Academic Press), 11–89.

Connelly, B. S., and Ones, D. S. (2010). An other perspective on personality: meta-analytic integration of observers' accuracy and predictive validity. *Psychol. Bull.* 136, 1092–1122. doi: 10.1037/a0021212

Ekman, P., Friesen, W. V., O'Sullivan, M., and Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *J. Pers. Soc. Psychol.* 38, 270–277. doi: 10.1037/0022-3514.38.2.270

Epstein, S. (1979). The stability of behavior I: on predicting most of the people much of the time. *J. Pers. Soc. Psychol.* 37, 1097–1126. doi: 10.1037/0022-3514.37.7.1097

Eronen, M. I., and Bringmann, L. F. (2021). The theory crisis in psychology: how to move forward. *Perspect. Psychol. Sci*. doi: 10.1177/1745691620970586. [Epub ahead of print].

Fairbairn, C. E., Sayette, M. A., Levine, J. M., Cohn, J. F., and Creswell, K. G. (2013). The effects of alcohol on the emotional displays of Whites in interracial groups. *Emotion* 13, 468–477. doi: 10.1037/a0030934

Flake, J. K., Pek, J., and Hehman, E. (2017). Construct validation in social and personality research: current practice and recommendations. *Soc. Psychol. Personal. Sci*. 8, 370–378. doi: 10.1177/1948550617693063

Fleeson, W., and Law, M. K. (2015). Trait enactments as density distributions: the role of actors, situations, and observers in explaining stability and variability. *J. Pers. Soc. Psychol.* 109, 1090–1104. doi: 10.1037/a0039517

Foster, T. S. (2015). *The reliability and validity of the thin slice technique: Observational research on video recorded medical interactions [ProQuest Information and Learning]*. Dissertation Abstracts International: Section B: The Sciences and Engineering (Vol. 75, Issue 8–B(E)). Wayne State University, Detroit, MI, United States.

Fowler, K. A., Lilienfeld, S. O., and Patrick, C. J. (2009). Detecting psychopathy from thin slices of behavior. *Psychol. Assess* 21, 68–78. doi: 10.1037/a0014938

Frieder, R. E., Van Iddekinge, C. H., and Raymark, P. H. (2016). How quickly do interviewers reach decisions? An examination of interviewers' decision-making time across applicants. *J. Occupat. Organ. Psychol.* 89, 223–248. doi: 10.1111/joop.12118

Fujiwara, K., and Daibo, I. (2014). The extraction of nonverbal behaviors: using video images and speech-signal analysis in dyadic conversation. *J. Nonverbal Behav.* 38, 377–388. doi: 10.1007/s10919-014-0183-3

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychol. Rev.* 102, 652–670. doi: 10.1037/0033-295X.102.4.652

Funder, D. C., and Colvin, C. R. (1991). Explorations in behavioral consistency: properties of persons, situations, and behaviors. *J. Pers. Soc. Psychol.* 60, 773–794. doi: 10.1037/0022-3514.60.5.773

Furley, P., and Schweizer, G. (2014). The expression of victory and loss: estimating who's leading or trailing from nonverbal cues in sports. *J. Nonverbal Behav.* 38, 13–29. doi: 10.1007/s10919-013-0168-7

Georgiou, P. G., Black, M. P., Lammert, A. C., Baucom, B. R., and Narayanan, S. S. (2011). ""That's Aggravating, Very Aggravating": is it possible to classify behaviors in couple interactions using automatically derived lexical features?," in *Affective Computing and Intelligent Interaction: ACII 2011*, *Vol. 6974*, eds D'Mello, A. Graesser, B. Schuller, and J. C. Martin (Berlin Heidelberg: Springer), 87–96. doi: 10.1007/978-3-642-24600-5_12

Gesn, P. R., and Ickes, W. (1999). The development of meaning contexts for empathic accuracy: channel and sequence effects. *J. Pers. Soc. Psychol.* 77, 746–761. doi: 10.1037/0022-3514.77.4.746

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C. P., and Back, M. D. (2017). Trait personality and state variability: predicting individual differences in within- and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *J. Res. Pers.* 69, 124–138. doi: 10.1016/j.jrp.2016.06.003

Gheorghiu, A. I., Callan, M. J., and Skylark, W. J. (2020). A thin slice of science communication: are people's evaluations of TED talks predicted by superficial impressions of the speakers? *Soc. Psychol. Personal. Sci.* 11, 117–125. doi: 10.1177/1948550618810896

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

Gosling, S. D., John, O. P., Craik, K. H., and Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *J. Person. Soc. Psychol.* 74, 1337–1349. doi: 10.1037/0022-3514.74.5.1337

Greene, E. G. (1941). *Measurements of Human Behavior*. Danbury, CT: Odyssey Press. doi: 10.1037/12221-000

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., and Feinstein, B. A. (2008). Accuracy of judging others' traits and states: comparing mean levels across tests. *J. Res. Pers.* 42, 1476–1489. doi: 10.1016/j.jrp.2008.06.013

Hall, J. A., and Braunwald, K. G. (1981). Gender cues in conversations. *J. Pers. Soc. Psychol.* 40, 99–110. doi: 10.1037/0022-3514.40.1.99

Hall, J. A., Coats, E. J., and LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* 131, 898–924. doi: 10.1037/0033-2909.131.6.898

Hall, J. A., Harvey, S. E., Johnson, K. E., and Colvin, C. R. (2021). Thin-slice accuracy for judging Big Five traits from personal narratives. *Pers. Individ. Dif.* 171:110392. doi: 10.1016/j.paid.2020.110392

Hall, J. A., Horgan, T. G., and Murphy, N. A. (2019). Nonverbal communication. *Annu. Rev. Psychol.* 70, 271–294. doi: 10.1146/annurev-psych-010418-103145

Hall, J. A., Roter, D. L., Blanch, D. C., and Frankel, R. M. (2009). Observer-rated rapport in interactions between medical students and standardized patients. *Patient Educ. Couns.* 76, 323–327. doi: 10.1016/j.pec.2009.05.009

Hall, J. A., and Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion* 7, 438–446. doi: 10.1037/1528-3542.7.2.438

Hall, J. A., Verghis, P., Stockton, W., and Goh, J. X. (2014). It takes just 120 seconds: predicting satisfaction in technical support calls. *Psychol. Market*. 31, 500–508. doi: 10.1002/mar.20711

Hirschmann, N., Kastner-Koller, U., Deimann, P., Schmelzer, M., and Pietschnig, J. (2018). Reliable and valid coding of thin slices of video footage: applicability to the assessment of mother-child interactions. *J. Psychopathol. Behav. Assess.* 40, 249–258. doi: 10.1007/s10862-017-9630-x

Hops, H., Biglan, A., Sherman, L., Arthur, J., Friedman, L., and Osteen, V. (1987). Home observations of family interactions of depressed women. *J. Consult. Clin. Psychol.* 55, 341–346. doi: 10.1037/0022-006X.55.3.341

James, D. M., Wadnerkar, M. B., Lam-Cassettari, C., Kang, S., and Telling, A. L. (2012). Thin slice sampling of video footage for mother/child interaction: application to single cases. *J. Psychopathol. Behav. Assess.* 34, 351–360. doi: 10.1007/s10862-012-9282-9

Jamieson, L. T. J., Baxter, G. S., and Murray, P. J. (2017). Identifying suitable detection dogs. *Appl. Anim. Behav. Sci.* 195, 1–7. doi: 10.1016/j.applanim.2017.06.010

Klonek, F. E., Quera, V., and Kauffeld, S. (2015). Coding interactions in Motivational Interviewing with computer-software: what are the advantages for process researchers? *Comput. Human Behav*. 44, 284–292. doi: 10.1016/j.chb.2014.10.034

Krzyzaniak, S. L., Colman, D. E., Letzring, T. D., McDonald, J. S., and Biesanz, J. C. (2019). The effect of information quantity on distinctive accuracy and normativity of personality trait judgments. *Eur. J. Pers.* 33, 197–213. doi: 10.1002/per.2196

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., and Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion* 13, 434–449. doi: 10.1037/a0031388

Lausberg, H., and Sloetjes, H. (2016). The revised NEUROGES–ELAN system: an objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behav. Res. Methods* 48, 973–993. doi: 10.3758/s13428-015-0622-z

Leikas, S., Lönnqvist, J.-E., and Verkasalo, M. (2012). Persons, situations, and behaviors: consistency and variability of different behaviors in four interpersonal situations. *J. Pers. Soc. Psychol.* 103, 1007–1022. doi: 10.1037/a0030385

Lepri, B., Mana, N., Cappelletti, A., and Pianesi, F. (2009). Automatic prediction of individual performance from "thin slices" of social behavior. In *Proceedings of the 17th ACM International Conference on Multimedia* (Beijing), 733–736. doi: 10.1145/1631272.1631400

Letzring, T., Murphy, N. A., Allik, J., Beer, Zimmerman, J., and Leising, D. (2021). *The judgment of personality: an overview of current empirical research findings*. PsyArXiv. doi: 10.31234/osf.io/8t63v

Letzring, T. D., Wells, S. M., and Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *J. Pers. Soc. Psychol.* 91, 111–123. doi: 10.1037/0022-3514.91.1.111

Li, H., Rosenthal, R., and Rubin, D. B. (1996). Reliability of measurement in psychology: from Spearman-Brown to maximal reliability. *Psychol. Methods* 1, 98–107. doi: 10.1037/1082-989X.1.1.98

Masuda, T., Batdorj, B., and Senzaki, S. (2020). Culture and attention: future directions to expand research beyond the geographical regions of WEIRD cultures. *Front. Psychol.* 11:1394. doi: 10.3389/fpsyg.2020.01394

Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., et al. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *J. Nonverbal Behav.* 24, 179–209. doi: 10.1023/A:1006668120583

Miller, W. R., and Rollnick, S. (2002). *Motivational Interviewing: Preparing People for Change, 2nd Edn*. New York, NY: The Guilford Press. doi: 10.1097/01445442-200305000-00013

Milmoe, S., Rosenthal, R., Blane, H. T., Chafetz, M. E., and Wolf, I. (1967). The doctor's voice: postdictor of successful referral of alcoholic patients. *J. Abnorm. Psychol.* 72, 78–84. doi: 10.1037/h0024219

Moskowitz, D. S., and Schwarz, J. C. (1982). Validity comparison of behavior counts and ratings by knowledgeable informants. *J. Pers. Soc. Psychol.* 42, 518–528. doi: 10.1037/0022-3514.42.3.518

Moyers, T., Martin, T., Catley, D., Harris, K. J., and Ahluwalia, J. S. (2003). Assessing the integrity of motivational interviewing interventions: reliability of the motivational interviewing skills code. *Behav. Cogn. Psychother.* 31, 177–184. doi: 10.1017/S1352465803002054

Murphy, N. A. (2005). Using thin slices for behavioral coding. *J. Nonverbal Behav.* 29, 235–246. doi: 10.1007/s10919-005-7722-x

Murphy, N. A. (2007). Appearing smart: the impression management of intelligence, person perception accuracy, and behavior in social interaction. *Person. Soc. Psychol. Bull.* 33, 325–339. doi: 10.1177/0146167206294871

Murphy, N. A. (2018). "Using thin slices to investigate impression formation and to measure interpersonal behavior from recorded social interactions," in SAGE Research Methods Cases (Thousand Oaks, CA). doi: 10.4135/9781526437464

Murphy, N. A., Hall, J. A., Mast, M. S., Ruben, M. A., Frauendorfer, D., Blanch-Hartigan, D., et al. (2015). Reliability and validity of nonverbal thin slices in social interactions. *Person. Soc. Psychol. Bull.* 41, 199–213. doi: 10.1177/0146167214559902

Murphy, N. A., Hall, J. A., Ruben, M. A., Frauendorfer, D., Schmid Mast, M., Johnson, K. E., et al. (2019). Predictive validity of thin-slice nonverbal behavior from social interactions. *Person. Soc. Psychol. Bull.* 45, 983–993. doi: 10.1177/0146167218802834

Narayanan, S., and Georgiou, P. G. (2013). Behavioral signal processing: deriving human behavioral informatics from speech and language: computational techniques are presented to analyze and model expressed and perceived human behavior-variedly characterized as typical, atypical, distressed, and disordered-from speech and language cues and their applications in health, commerce, education, and beyond. *Proc. IEEE* 101, 1203–1233. doi: 10.1109/JPROC.2012.2236291

Nguyen, L. S., and Gatica-Perez, D. (2015). "I would hire you in a minute: thin slices of nonverbal behavior in job interviews," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15), Association for Computing Machinery* (Seattle, WA), 51–58. doi: 10.1145/2818346.2820760

Patterson, M. L. (1973). Stability of nonverbal immediacy behaviors. *J. Exp. Soc. Psychol.* 9, 97–109. doi: 10.1016/0022-1031(73)90002-4

Perrault, E. K. (2020). The diminishing returns for longer healthcare provider video biographies: a thin slice examination of patient decision-making. *Health Commun.* 36, 650–658. doi: 10.1080/10410236.2020.1733230

Place, S. S., Todd, P. M., Zhuang, J., Penke, L., and Asendorpf, J. B. (2012). Judging romantic interest of others from thin slices is a cross-cultural ability. *Evol. Human Behav.* 33, 547–550. doi: 10.1016/j.evolhumbehav.2012.02.001

Pretsch, J., Flunger, B., Heckmann, N., and Schmitt, M. (2013). Done in 60s? Inferring teachers' subjective well-being from thin slices of nonverbal behavior. *Soc. Psychol. Educ.* 16, 421–434. doi: 10.1007/s11218-013-9223-9

Prime, H., Browne, D., Akbari, E., Wade, M., Madigan, S., and Jenkins, J. M. (2015). The development of a measure of maternal cognitive sensitivity appropriate for use in primary care health settings. *J. Child Psychol. Psychiatry* 56, 488–495. doi: 10.1111/jcpp.12322

Prime, H., Perlman, M., Tackett, J. L., and Jenkins, J. M. (2014). Cognitive sensitivity in sibling interactions: development of the construct and comparison of two coding methodologies. *Early Educ. Dev.* 25, 240–258. doi: 10.1080/10409289.2013.821313

Ramseyer, F. T. (2020). Motion energy analysis (MEA): a primer on the assessment of motion from video. *J. Couns. Psychol.* 67, 536–549. doi: 10.1037/cou0000407

Richeson, J. A., and Shelton, J. N. (2005). Brief report: thin slices of racial bias. *J. Nonverbal Behav.* 29, 75–86. doi: 10.1007/s10919-004-0890-2

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., and Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspect. Psychol. Sci.* 15, 1295–1309. doi: 10.1177/1745691620927709

Rosenthal, R. (2005). "Conducting judgment studies: some methodological issues," in *The New Handbook of Methods in Nonverbal Behavior Research*, eds J. A. Harrigan, R. Rosenthal, and K. R. Scherer (Oxford: Oxford University Press), 199–234. doi: 10.1093/acprof:oso/9780198529620.003.0005

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., and Archer, D. (1979). *Sensitivity to Nonverbal Communication: The PONS Test*. Baltimore, MD: The Johns Hopkins University Press. doi: 10.1016/B978-0-12-761350-5.50012-4

Roter, D. L., Hall, J. A., Blanch-Hartigan, D., Larson, S., and Frankel, R. M. (2011). Slicing it thin: new methods for brief sampling analysis using RIAS-coded medical dialogue. *Patient Educ. Couns.* 82, 410–419. doi: 10.1016/j.pec.2010.11.019

Ruben, M. A., and Hall, J. A. (2016). A lens model approach to the communication of pain. *Health Commun.* 31, 934–945. doi: 10.1080/10410236.2015.1020261

Ruben, M. A., Hall, J. A., and Mast, M. S. (2015). Smiling in a job interview: when less is more. *J. Soc. Psychol.* 155, 107–126. doi: 10.1080/00224545.2014.972312

Rule, N. O., and Ambady, N. (2008). Brief exposures: male sexual orientation is accurately perceived at 50 ms. *J. Exp. Soc. Psychol.* 44, 1100–1105. doi: 10.1016/j.jesp.2007.12.001

Schimmack, U. (2020). A meta-psychological perspective on the decade of replication failures in social psychology. *Can. Psychol.* 61, 364–376. doi: 10.1037/cap0000246

Schlegel, K., Grandjean, D., and Scherer, K. R. (2014). Introducing the Geneva emotion recognition test: an example of Rasch-based test development. *Psychol. Assess.* 26, 666–672. doi: 10.1037/a0035246

Schmid Mast, M., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., and Choudhury, T. (2015). Social sensing for psychology: automated interpersonal behavior assessment. *Curr. Dir. Psychol. Sci.* 24, 154–160. doi: 10.1177/0963721414560811

Shoda, Y. (1999). A unified framework for the study of behavioral consistency: bridging person × situation interaction and the consistency paradox. *Eur. J. Person.* 13, 361–387. doi: 10.1002/(SICI)1099-0984(199909/10)13:5<361::AID-PER362>3.0.CO;2-X

Slepian, M. L., Bogart, K. R., and Ambady, N. (2014). Thin-slice judgments in the clinical context. *Annu. Rev. Clin. Psychol.* 10, 131–153. doi: 10.1146/annurev-clinpsy-090413-123522

Sokolovic, N., Borairi, S., Rodrigues, M., Perlman, M., and Jenkins, J. M. (2021). Validating an efficient measure of responsivity in father–child interactions. *Can. J. Behav. Sci.* 53, 84–89. doi: 10.1037/cbs0000185

Spearman, C. C. (1910). Correlation calculated from faulty data. *Br. J. Psychol.* 3, 271–295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Stopfer, J. M., Egloff, B., Nestler, S., and Back, M. D. (2014). Personality expression and impression formation in online social networks: an integrative approach to understanding the processes of accuracy, impression management and meta-accuracy. *Eur. J. Pers.* 28, 73–94. doi: 10.1002/per.1935

Todorov, A. (2017). *Face Value: The Irresistible Influence of First Impressions*. Princeton, NJ: Princeton University Press. doi: 10.2307/j.ctvc7736t

Tskhay, K. O., Zhu, R., and Rule, N. O. (2017). Perceptions of charisma from thin slices of behavior predict leadership prototypicality judgments. *Leadersh. Q.* 28, 555–562. doi: 10.1016/j.leaqua.2017.03.003

Vaughan, W. F. (1948). "Investigating social behavior," in *Social Psychology: The Science and Art of Living Together*, ed W. F. Vaughan (Danbury, CT: Odyssey Press), 65–96. doi: 10.1037/13255-003

Wang, M. Z., Chen, K., and Hall, J. A. (2020). Predictive validity of thin slices of verbal and nonverbal behaviors: comparison of slice lengths, locations, and rating methodologies. *J. Nonverbal Behav.* 45, 53–66. doi: 10.1007/s10919-020-00343-1

Waxer, P. H. (1974). Therapist training in nonverbal communication: I nonverbal cues for depression. *J. Clin. Psychol.* 30, 215–218. doi: 10.1002/1097-4679(197404)30:2<215::AID-JCLP2270300229>3.0.CO;2-Q

Whalen, D. J., Gilbert, K. E., Jackson, J. J., Barch, D. M., and Luby, J. L. (2020). Using a thin slice coding approach to assess preschool personality dimensions. *J. Pers. Assess.* 103, 214–223. doi: 10.1080/00223891.2020.1722140

Zebrowitz, L. A., and Collins, M. A. (1997). Accurate social perception at zero acquaintance: the affordances of a Gibsonian approach. *Personal. Soc. Psychol. Rev.* 1, 204–223. doi: 10.1207/s15327957pspr0103_2

Conflict of Interest: 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.

frontiers in Psychology

REVIEW
published: 27 July 2021
doi: 10.3389/fpsyg.2021.606548

Nonverbal Social Sensing: What Social Sensing Can and Cannot Do for the Study of Nonverbal Behavior From Video

Laetitia Aurelie Renier, Marianne Schmid Mast, Nele Dael and Emmanuelle Patricia Kleinlogel*

Department of Organizational Behavior – Faculty of Business and Economics (HEC), University of Lausanne, Lausanne, Switzerland

The study of nonverbal behavior (NVB), and in particular kinesics (i.e., face and body motions), is typically seen as cost-intensive. However, the development of new technologies (e.g., ubiquitous sensing, computer vision, and algorithms) and approaches to study social behavior [i.e., social signal processing (SSP)] makes it possible to train algorithms to automatically code NVB, from action/motion units to inferences. Nonverbal social sensing refers to the use of these technologies and approaches for the study of kinesics based on video recordings. Nonverbal social sensing appears as an inspiring and encouraging approach to study NVB at reduced costs, making it a more attractive research field. However, does this promise hold? After presenting what nonverbal social sensing is and can do, we discussed the key challenges that researchers face when using nonverbal social sensing on video data. Although nonverbal social sensing is a promising tool, researchers need to be aware of the fact that algorithms might be as biased as humans when extracting NVB or that the automated NVB coding might remain context-dependent. We provided study examples to discuss these challenges and point to potential solutions.

Keywords: nonverbal behavior, social sensing, coding, extraction, communication, technology, annotations, algorithm

Edited by:
Judee K. Burgoon, University of Arizona, United States

Reviewed by:
Ross W. Buck, University of Connecticut, United States
Sophie Van Der Zee, Erasmus University Rotterdam, Netherlands

****Correspondence:***
Laetitia Aurelie Renier
laetitia.renier@unil.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *15 September 2020*
Accepted: *21 June 2021*
Published: *27 July 2021*

Citation:
Renier LA, Schmid Mast M, Dael N and Kleinlogel EP (2021) Nonverbal Social Sensing: What Social Sensing Can and Cannot Do for the Study of Nonverbal Behavior From Video. Front. Psychol. 12:606548. doi: 10.3389/fpsyg.2021.606548

INTRODUCTION

Investigating nonverbal behavior (NVB), and in particular kinesics, namely face and body motions used in communication (Birdwhistell, 1955; Burgoon and Dunbar, 2018), involves observing social interactions and coding movements of participants in the face and the body. Manually coding NVB takes a considerable amount of time and resources because it means having coders sit in front of a video screen and, for instance, count the frequency of smiles, calculate the duration of gazing, code interruptions, or rate the target on a more global judgment (e.g., how dominant or deceiving) for many hours over many days. Moreover, this does not include the additional work of training the coders and establishing reliability among them.

Due to advanced growth in computer vision, new technologies and approaches (e.g., SSP, Vinciarelli et al., 2009a,b, 2012) have been developed to use and train algorithms to code NVB as action/motion units or as more global judgments (inferences) from videotaped individuals in social interactions (e.g., trustfulness). This has given rise to nonverbal social sensing, an approach that allows to automatize most of the NVB coding.

Once such algorithms are developed, they have the advantage of being scalable. Therefore, to the extent that researchers code the same NVB or judge the same inferences in different studies, such algorithms are valuable to researchers. Moreover, there is no standardized codebook detailing exactly how to code NVB (e.g., should smiling be assessed as a frequency, a duration, or a general impression about how much a person smiles on a scale of 1–5), which makes the comparison of results pertaining to NVB difficult across different studies. If more researchers used nonverbal social sensing, this field might gain in standardization and we might discover new insights that were not previously possible since the different coding methods would introduce too much noise to detect the signal. Furthermore, using nonverbal social sensing, when studying NVB, has the potential to reveal meaningful nonverbal patterns more easily (e.g., looking at the interaction partner while speaking, see Burgoon et al., 2014 for an example in detection of deception using computer-assisted coding and an algorithm to identify temporal patterns) instead of extracting only isolated NVB cues (e.g., duration of looking at the interaction partner and the number of speech turns of the target). These advantages might attract new researchers to study NVB, thus enriching and broadening the field.

The aim of this paper is to provide information and guidance to researchers who consider using nonverbal social sensing for their studies. We explained how nonverbal social sensing works, where we see the challenges of using it for the study, and how we recommend addressing such challenges. We illustrated these aspects with selective study examples.

In this paper, we focused on kinesics and the use of nonverbal social sensing based on video recordings (see Poppe, 2017 for an application of nonverbal social sensing beyond video recordings). Kinesics refers to two categories of NVB: (1) gesture and posture and (2) face and eye behavior (Vinciarelli et al., 2009a; the latter is also referred to as gaze, Harrigan, 2005, p. 137). Moreover, we focused on the extraction of NVB or inferences based on videotaped targets. We did not consider the sensor-based technologies, which require participants to wear sensors that register their NVB during the interaction task (see, e.g., Poppe et al., 2014; Rahman et al., 2019).

THE LEVEL OF NONVERBAL CODING: UNIT VS. INFERENCE

We studied the NVB coding on two different levels: action/motion units and kineme/inferences. An action/motion unit refers to specific body motions, such as muscle movements in the face and frequency or duration of a specific NVB (e.g., head motion and movement of the lips) or in the body (e.g., arm movement and leaning). As for "micro-kinesics," these units do not carry social meaning (see Birdwhistell, 1952). However, researchers are interested not only in specific nonverbal cues but also in inferences and the coding of global judgments based on NVB. Coders make inferences about trustworthiness, hireability, charisma, personality, or motivation of a target by observing the behaviors of the participant.

The lower the level of abstraction in coding, the more the interpretation of what the behavior means is already included in the coding, whereas higher levels of abstraction need interpretation and information about the context (see Birdwhistell, 1970). To illustrate, the number of smiles does not have much meaning attached to it. The meaning of smiling depends largely on the context. For instance, the simulation-of-smiles model (Niedenthal et al., 2010; Rychlowska et al., 2017) proposes to distinguish smiles according to their roles as follows: the smile that communicates positive emotions (enjoyment smile), the smile that suggests positive social intentions (affiliative smile), and the smile that reflects status or control (dominance smile). However, coding friendliness for instance (which might be based on smiling, but not exclusively) involves coding the meaning of the underlying NVB (e.g., smile, eye contact, and voice tone) to decide to what extent an individual appears friendly.

In summary, action/motion units can be coded relatively objectively, whereas inferences are more subjective because they need interpretation and are more context-dependent. This distinction between units and inferences, between objective and subjective measurements (Burgoon and Dunbar, 2018), is key in understanding the workings and challenges of nonverbal social sensing.

HOW NONVERBAL SOCIAL SENSING WORKS

Nonverbal social sensing originates in the field of SSP. SSP aims at automatically analyzing and synthesizing social signals (Vinciarelli et al., 2009b). SSP allows transforming raw input data (e.g., video recordings of people in social interactions) into social signals (i.e., units or inferences). Developing algorithms for nonverbal social sensing requires input data (i.e., videos of participants and ground truth). The videos refer to the material on which the algorithm is trained to extract and classify the NVB. The ground truth refers to the labels (e.g., manual coding or self-report) used as the standard of extraction or classification. The ground truth is either collected for the entire dataset or only on a subset (i.e., training set) of videos.

Ground truth data can be obtained in many different ways. For instance, satisfaction ratings of clients of a call center have been used as ground truth to train an algorithm to predict client satisfaction based on vocal cues of the call center employees (Zweig et al., 2006; Segura et al., 2016). When wanting to develop an algorithm that extracts personality, self-reports or other reports of personality can be used as the ground truth or expert assessments. When interested in developing algorithms that mimic human perception and judgment (e.g., perceived trustworthiness and hireability), we required human coders who are instructed and trained to perform the coding manually (i.e., manual annotations serve as the ground truth) or naïve raters who report their perception of the targets (e.g., source credibility ratings, Pentland, 2018).

We present below the general functioning of nonverbal social sensing in the following sections. We first present the application to NVB studies at the unit level. Second, we present two approaches to address NVB at the inference level.

Nonverbal Social Sensing at the Unit Level

At the action/motion unit level, nonverbal social sensing allows capturing a wide variety of nonverbal cues, such as micro-expressions, gestures, and movements. To illustrate, in the case of micro-expressions, the coding consists of extracting the frequency and the duration of muscle movements in the face, such as in the study of facial expressions. One of the most well-known and used classification methods to manually code facial expressions is the facial action coding system (FACS; Ekman and Friesen, 1978). When using the FACS, human coders note whether a facial action (i.e., activation of facial muscles such as lip corners going up or brow-raising) is present when coding a video. From this coding system, researchers develop algorithms to automatically recognize facial action units (AUs) from still records (Pantic and Rothkrantz, 2004) and moving records (Kapoor et al., 2003; Bartlett et al., 2006; Tong et al., 2006). As an application example, researchers used nonverbal social sensing to study the existence of cross-cultural differences in smiling (AU12) and brow furrowing (AU4) (McDuff et al., 2017). These researchers used automated extraction of these two units to study the effect of culture (i.e., individualist vs. collectivist), setting (i.e., home vs. lab), and gender on facial expressions. Their use of nonverbal social sensing enabled them to observe cultural (e.g., higher rate of brow furrowing in individualist culture than in collectivist culture) and gender differences (e.g., more smiling and less brow furrowing for women than men in both cultures, but more pronounced differences in individualist culture) at a lesser cost and on a larger scale (e.g., using a sample of 740,984 participants across 12 countries). Some of these researchers particularly worked on the development of algorithms for the detection of AU12 and AU4 and on a corpus of data for the study of spontaneous facial expressions (McDuff et al., 2013).

We might also need human coders at the unit level. In order to train an algorithm to extract the number of times a person nods in a video, we need to define which head movements qualify as a nod. This information is typically provided by human coders. We need several independent human coders to watch the same videos and to judge whether a given head movement is a nod, and then, we need to test for reliability (i.e., the extent to which the independent coders are consistent). The machine is then fed with this information together with the corresponding video, and from these two inputs, the machine can learn to detect head nods (e.g., Nguyen et al., 2012). Once trained, the algorithm will have learned to extract the features and classify them as action/motion units and can be used on new datasets. However, instead of measuring the ground truth, researchers might also rely on open-source tools such as OpenPose (i.e., body behavior; Cao et al., 2019) or OpenFace (i.e., face behavior; Baltrusaitis et al., 2018). OpenPose is an open-source library for multi-person detection providing real-time pose estimation (e.g., head, hand, foot, and face). OpenFace is also an open-source library designed to detect facial landmarks (e.g., facial AU, head pose, and eye-gaze). Both libraries are well-recognized tools for coding NVB as action/motion units enabling researchers to skip the training stage of nonverbal social sensing (for an application of OpenFace, see Burgoon et al., 2021).

Nonverbal Social Sensing at the Inference Level

At the inference level, NVB is coded according to its meaning, starting from the kineme to a higher-order inference. As examples of kinemes, we cite visual dominance—the ratio of the percentage of looking while speaking divided by the percentage of looking while listening (Dovidio et al., 1988)—or visual back-channeling—head nods while listening (Nguyen et al., 2012).

Nonverbal social sensing allows extracting data related to higher-order inferences or global judgments. For example, algorithms can capture how dominant or how trustworthy individuals are perceived through the measure of a combination of NVB (Burgoon and Buller, 1994; Hall et al., 2005; Mast et al., 2011). For instance, researchers used nonverbal social sensing to automatically predict the level of dominance of individuals during group interactions (Jayagopi et al., 2009) or their hireability (Naim et al., 2015). Other instances include the detection of personality traits (e.g., Pianesi et al., 2008; Batrinca et al., 2011), using personality recognition to improve automated detection of deception (An et al., 2018), or the detection of emotions based on body movements (Glowinski et al., 2008).

For higher-order inferences, the following two main approaches are currently pursued. In the first approach, the NVB is extracted automatically from the video input (as described for the motion unit extraction), and this extracted NVB is then linked with the ground truth. The machine is trained to first extract the nonverbal features (e.g., a nod and a smile) and only then learns to link those to the higher-order inferences (e.g., the classification of a target as friendly). For instance, to predict who gets hired for a job, the machine can first extract a set of specific NVB and then link it to the ground truth of hiring decisions. Another example is training a machine to predict social skills or personality (Biel et al., 2013; Muralidhar et al., 2018; Rasipuram and Jayagopi, 2018) or emotions (Ahn et al., 2010) based on previously extracted nonverbal cues. Again, the ground truth has to be measured (e.g., human coders assessing the personality of the people in the video or a self-report of their personality). The machine that extracted the NVB will link the extracted NVB to the ground truth. This approach allows identifying the NVB that is conductive of being hired (Frauendorfer et al., 2014; Nguyen et al., 2014; Muralidhar et al., 2016), which is important for training and the transparency of the decision-making. When predicting that a person is conscientious, this approach allows knowing which NVB pattern is responsible for this prediction.

In the second approach, the machine is fed with the video input and the ground truth (e.g., hireability) and learns to classify the videos into (not) hireable without involving the explicit extraction of NVB. This second approach relies on deep learning (see Mehta et al., 2019 for a review of the use of deep learning in the detection of personality traits). The machine is given the videos and the ground truth, which this time is an inference such as, for instance, how dominant a person behaves in a social interaction rated by external observers or the personality assessed *via* self-report. The machine learns the link between the training videos and the ground truth (i.e., annotated dataset). However, the researcher or user will not know which array of nonverbal cues the algorithm uses for the prediction. Does the machine judge people as dominant because they speak a lot, because of a loud tone of voice, because they move more, or because of their gender or skin color or any combination thereof? There is no way to be certain.

Using nonverbal social sensing for higher-order inferences by either first extracting the NVB or directly linking the videos to the ground truth (i.e., annotated dataset at the inference level) is a choice a researcher needs to make based on how important it is to know which behaviors are responsible for the inference. This approach might be considered less costly because researchers only need to feed the data to the machines without relying on human coders. However, the size of the dataset to be fed into the machine is large (i.e., hundreds of videos) and thus also potentially costly. Thus, the benefits and shortfalls of deep learning depend on the goals of the researchers. If they are interested in determining the behaviors responsible for the inferences, we cautioned researchers when using deep and unsupervised learning approaches given their black-box nature. However, if researchers are primarily interested in higher-order inferences, deep learning appears to be a suitable approach (e.g., Mehta et al., 2019). In between, supervised deep learning might also reduce the black-box aspect associated with unsupervised learning and might lead researchers to discover new patterns of behaviors and inferences (see LeCun et al., 2015). Finally, concerning lower-order inferences, advances in deep learning enable researchers to automatically extract human pose at a lesser cost (Mathis et al., 2018; Arac et al., 2019).

There are some corpora of annotated data concerning higher-order inferences available. For example, corpora of annotated data are available in the domain of group interaction studies (see Gatica-Perez, 2015 for a list of corpus), leadership emergence (corpus cited in Sanchez-Cortes et al., 2011, 2013), psychological distress (Gratch et al., 2014), or personality detection (Mana et al., 2007). These corpora might help reduce the cost of collecting the input data.

CHALLENGES WHEN USING NONVERBAL SOCIAL SENSING

Under this section, we highlight key challenges associated with the use of nonverbal social sensing for researchers. We additionally make suggestions to address them.

The Risk of Bias

Algorithms are often used because people think they are less biased. It is true that once the algorithm runs, it does not make a difference between, for example, women or men showing a certain behavior. It simply codes the behavior, whereas human coders might be affected by the gender of the person showing the behavior they are about to code. However, algorithms are only as good as the ground truth on which they are trained. In other words, if the ground truth is biased, the algorithm will be biased. The risk for biased ground truth is higher for predictions at the inference level than at the unit level because the former is a more subjective coding than the latter. Therefore, collecting ground truth on nodding is probably less biased than collecting ground truth on, for example, the hireability of a person for a job.

Bias might also plague algorithms that learn to detect patterns by themselves (i.e., unsupervised learning). For instance, algorithms might learn by themselves to discriminate women during the recruitment process (e.g., Dastin, 2018; Lambrecht and Tucker, 2019) without the developers or users being aware of this bias. To illustrate, an algorithm trained to select the best candidates for a job taught itself (i.e., based on the data fed to the algorithm) to discriminate against women during the recruitment process (Dastin, 2018). The algorithm extracted a rule based on the data it was fed (e.g., it detected a connection made between best candidates and males) and used the rule to make future judgments. This led Amazon to stop using its automated recruitment system. In the same vein, algorithms developed to attract new talents for STEM job opportunities targeted more men than women (Lambrecht and Tucker, 2019). As pointed out by Kleinberg et al. (2018), the training data might be "rooted in past discrimination" (Kleinberg et al., 2018, p. 116). Since the input data were biased, the output data were also biased.

Therefore, before using any established algorithms, researchers need to know what data the algorithm has been trained on to tentatively estimate the risk of bias. For example, if an algorithm has been trained to predict friendliness on videos showing mainly males from an individualistic culture, it is possible that the developed system will not offer accurate predictions for women or individuals from a collectivistic culture. In the same vein, researchers showed that algorithms trained on videos featuring only adults were biased in performing emotion recognition on a younger population (Howard et al., 2017). Researchers interested in developing their own algorithms also need to be critical about the input and output data used and created by their nonverbal social sensing system.

Biased decisions have important ethical ramifications. First, in the examples related to biased recruitment, the decision was made by a machine and not a human (see recommendations for trustworthy algorithms, High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019). Second, the algorithm ended up taking into account a feature protected by law (e.g., gender and ethnicity) to produce a decision that disadvantages the said group. Given that this subject is not the main focus of this study, we referred the reader to Kleinberg et al. (2018) for a discussion on the legal and ethical aspects of discrimination associated with the use of algorithms in the recruitment process and to Raghavan et al. (2020) for potential solutions and challenges.

Data Privacy

Another ethical issue is linked to data privacy. Social and computer scientists might not share the same ethical guidelines when studying NVB. This difference might be aggravated by open-science policies. For instance, social scientists, studying NVB based on video recordings of participants, need to ensure the anonymity of the participants and to disclose the specific use of the collected data. Meanwhile, computer scientists might not be required to do the same and to obtain the consent of participants to reuse their data. In this context, sharing data or developing corpora useful for future studies might be more difficult to achieve for social scientists than for computer scientists. Still, following the Facebook-Cambridge Analytica scandal, an ethical crisis related to data protection has also shaken computer scientists. In this context, researchers need to be attentive to ethical compliance across fields of research. In this vein, fostering collaborations between social and computer scientists might help in determining ethical guidelines that are common to both fields.

Concerning ethical algorithms, we suggested that social scientists, interested in the use of nonverbal social sensing systems, should be well-informed about policies related to artificial intelligence (AI). For instance, in Europe, a group of experts was commissioned to work on ethical guidelines for AI (Biel et al., 2013). The requirements for the so-called trustworthy AI are (1) human agency and oversight, (2) technical robustness and safety, (3) privacy data and governance, (4) transparency, (5) diversity, non-discrimination, and fairness, (6) environmental and societal wellbeing, and (7) accountability. As suggested by the High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019), these seven requirements should be addressed, and reflected upon, if adherence is not feasible.

Context-Dependency of Nonverbal Social Sensing

The quality of the output generated using nonverbal social sensing depends on the extent to which the data coded by the algorithm resemble the data on which the algorithm had been trained. To illustrate, if researchers use an algorithm that extracts head nods and this algorithm has been trained on videos featuring people sitting in front of a camera, but the video material for which the researchers want to use the algorithm shows people from the side, instead of a frontal view, involved in social interaction, it is likely that the algorithm will not perform that well.

For inferences, context-dependency is even more of an issue and the extent to which inferences are domain-specific or transversal is unclear. Will an algorithm trained to extract personality from videos of targets self-presenting during a job interview extract personality from videos of people self-presenting for a dating site with equal accuracy? Will an algorithm trained to extract trustworthiness from videos of targets giving a public speech perform equally well on videos of people answering job interview questions?

We suggested to scholars, who want to use nonverbal social sensing, to gather information about the W5 + (i.e., where, what, when, who, why, and how of the video input data the algorithm has been trained on, Vinciarelli et al., 2009b) and on potential moderators (i.e., culture, relationship, and gender, Burgoon and Dunbar, 2018). This information will enable the researcher to gauge whether the algorithm can be used for this study, as well as highlight boundary conditions or limitations of the developed algorithms for future applications.

Off-the-Shelf vs. Tailored Approaches

Some nonverbal social sensing systems are readily available (i.e., OpenPose and OpenFace to code NVB as a unit or systems such as FaceReader to code NVB in the face as more more global judgment). These systems are easy to use for people outside the field of computer science. We thus encouraged researchers interested in coding NVB as action/motion units to try well-known off-the-shelf open-source solutions (e.g., OpenPose and OpenFace). However, researchers need to keep in mind that off-the-shelf systems might not be suited for their specific study purposes. For example, a researcher might need data on the duration of an NVB while off-the-shelf systems provide data on its frequency.

Nonverbal social sensing systems to code NVB at the inference level are also available on the market (e.g., FaceReader or Affectiva for facial expression, and HireVue and Pymetrics for hireability). These commercial off-the-shelf systems come with a caveat. They typically do not provide information about the input data (i.e., videos and ground truth) on which the algorithms have been trained, making it impossible to gauge the reliability and the accuracy of the inferences for the dataset of the researcher. To illustrate, the HireVue algorithm automatically generates a score of hireability and a rank to help companies make their hiring decisions. With this type of off-the-shelf solution, several questions arise: Does the algorithm take into account the protected features? Is human agency respected? Is the process transparent enough? How is accuracy assessed? To assess the quality of the inferences obtained by the off-the-shelf solutions, the researchers have to manually code a portion of their data and compare it with the output of the algorithm to ensure that the algorithm performs at the expected level.

Hence, when using an off-the-shelf system to code NVB at the inference level, researchers need to have access to its input and output data. This is necessary to assess its reliability and algorithm performance. Researchers are also advised to verify that the system is compliant with the existing guidelines on the use of AI (see the recommendation of OECD of the Council on Artificial Intelligence—OECD AI Principles; High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019).

An alternative to the off-the-shelf solution is to become savvy in machine learning or to collaborate with computer scientists to develop an algorithm for automatic coding of NVB. These multidisciplinary collaborations can benefit both social and computer scientists by fostering the development of SSP and nonverbal social signals. Benefits have already been highlighted in the domain of neurosciences (Sedda et al., 2012). Social scientists can benefit from the technical expertise of computer scientists. Computer scientists can benefit from the expertise of social scientists in NVB studies (e.g., knowledge about taxonomies and key variables to take into account). Developing an algorithm to code for NVB is only a viable solution if the developed algorithm can be used for other research projects. This is because the generation of the input data (i.e., videos and ground truth) and the machine learning process are time and resource-intensive.

To help identify the best nonverbal social sensing approach, researchers need a clear research question. This will help them determine the type of data and method that is needed. We suggest two complementary reflections. First, the general approach to study NVB must be clarified and operationalized. In this domain, we suggest following the pragmatic guide developed by Blanch-Hartigan et al. (2018) to identify the input data and the data collection method. This step is crucial to identify whether nonverbal a social sensing system is appropriate for the research project. The questions to be answered are: Is computer vision sufficiently developed to extract the NVB? Does a model to predict global judgment already exist? and Is it necessary to create a new nonverbal social sensing system? Second, to refine choices about coding NVB decisions, we suggest that researchers clarify their coding approach (Burgoon and Dunbar, 2018). Determining NVB coding strategies directly affects nonverbal social sensing. For instance, researchers interested in kinesics at the dyadic level need at least two cameras to record each member of the dyad for data capture. Another example of a decision that needs to be taken (i.e., when, where, and by whom) concerns the granularity of the temporal dimension. To illustrate, OpenPose enables researchers to automatically code the NVB for each second of the interaction. Other issues that need to be addressed include whether an off-the-shelf solution is available to code the macro-behaviors and whether researchers are interested in objective or subjective measurements in coding NVB as a unit or an inference.

However, these new technologies are still in development. Moreover, they are not free of biases and their input and output data are highly context-dependent. At this stage, ubiquitous sensing and automated extraction only complement human coding and particular caution, and scrutiny about the quality of the algorithm, needs to be taken before one can use these sensing and extraction technologies.

Researchers assessing the usefulness of nonverbal social sensing for their study should ask themselves the following questions: Can I use an algorithm that is already developed or do I have to develop my own? If I have to develop my own, do I have the necessary competencies or the necessary collaboration partners with those competencies? When using an existing algorithm: (a) Is the video input data similar to the training dataset? (b) How is the ground truth obtained? and (c) Do I know on which NVB the inferences are based? To ensure the quality and accuracy of the coding done by the algorithm on the data gathered by the researchers, said researchers might want to consider manually coding a subset of the data and then compare the performance of the algorithm with the manual coding.

The more established and robust algorithms for NVB extraction become, the more attractive they are for researchers to use and the more they might advance the field of NVB studies. This is because using established and robust algorithm for the automatic coding of NVB will improve the comparability of NVB across studies and has the potential to attract more researchers into the field.

CONCLUSION

Nonverbal social sensing can extract NVB from videotaped social interactions or it can make inferences based on NVB in videotaped social interactions. Both of these outputs are highly relevant for researchers, and because such algorithms allow scalability, they might attract new researchers in the domain of NVB, contributing to the advancement of the field.

AUTHOR CONTRIBUTIONS

LR, MSM, ND, and EK conceived of the presented idea. LR and MSM refined the main ideas and proof outline. All authors contributed different parts of writing to the manuscript with LR being in charge of coordinating and integrating and writing the most extensive part of the manuscript. All authors discussed the initial content and LR, MSM, and EK contributed to the final manuscript.

FUNDING

SNF Sinergia, CRSII5-183564.

REFERENCES

Ahn, S. J., Bailenson, J., Fox, J., and Jabon, M. (2010). "Using automated facial expression analysis for emotion and behavior prediction," in *The Routledge Handbook of Emotions and Mass meDia*, eds K. Doveling, C. von Scheve, and E. Konijn (New York, NY: Routledge), 349–369.

An, G., Levitan, S. I., Hirschberg, J., and Levitan, R. (2018). "Deep personality recognition for deception detection," in *Proceedings Interspeech*, Hyderabad, 421–425. doi: 10.21437/Interspeech.2018-2269

Arac, A., Zhao, P., Dobkin, B. H., Carmichael, S. T., and Golshani, P. (2019). DeepBehavior: a deep learning toolbox for automated analysis of animal and human behavior imaging data. *Front. Syst. Neurosci.* 13:20. doi: 10.3389/fnsys.2019.00020

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. P. (2018). "Openface 2.0: facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, 59–66. doi: 10.1109/FG.2018.00019

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., and Movellan, J. R. (2006). Automatic recognition of facial actions in spontaneous expressions. *J. Mult.* 1, 22–35. doi: 10.4304/jmm.1.6.22-35

Batrinca, L. M., Mana, N., Lepri, B., Pianesi, F., and Sebe, N. (2011). "Please, tell me about yourself: automatic personality assessment using short self-presentations," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, New York, NY: ACM, 255–262. doi: 10.1145/2070481.2070528

Biel, J. I., Tsiminaki, V., Dines, J., and Gatica-Perez, D. (2013). "Hi YouTube! personality impressions and verbal content in social video," in *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*, New York, NY: ACM, 119–126. doi: 10.1145/2522848.2522877

Birdwhistell, R. L. (1952). *Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture*. Lexington, KY: University of Kentucky Press.

Birdwhistell, R. L. (1955). Background to kinesics. *Review Gen. Semant.* 13, 10–17.

Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body Motion*. Philadelphia: University of Pennsylvania Press.

Blanch-Hartigan, D., Ruben, M. A., Hall, J. A., and Schmid Mast, M. (2018). Measuring nonverbal behavior in clinical interactions: a pragmatic guide. *Patient Educ. Couns.* 101, 2209–2218. doi: 10.1016/j.pec.2018.08.013

Burgoon, J., and Dunbar, N. (2018). "Coding nonverbal behavior," in *The Cambridge Handbook of Group Interaction Analysis (Cambridge Handbooks in psychology)*, eds E. Brauner, M. Boos, and M. Kolbe (Cambridge: Cambridge University Press), 104–120. doi: 10.1017/9781316286302.007

Burgoon, J. K., and Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. *J. Nonverbal Behav.* 18, 155–184. doi: 10.1007/BF02170076

Burgoon, J. K., Proudfoot, J. G., Schuetzler, R., and Wilson, D. (2014). Patterns of nonverbal behavior associated with truth and deception: illustrations from three experiments. *J. Nonverbal Behav.* 38, 325–354. doi: 10.1007/s10919-014-0181-5

Burgoon, J. K., Wang, X., Chen, X., Pentland, S. J., and Dunbar, N. E. (2021). Nonverbal behaviors "speak" relational messages of dominance, trust, and composure. *Front. Psychol.* 12:624177. doi: 10.3389/fpsyg.2021.624177

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., and Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 172–186. doi: 10.1109/TPAMI.2019.2929257

Dastin, J. (2018). *Amazon Scraps Secret ai Recruiting Tool That Showed Bias Against Women*. London: Reuters.

Dovidio, J. F., Brown, C. E., Heltman, K., Ellyson, S. L., and Keating, C. F. (1988). Power displays between women and men in discussions of gender-linked tasks: a multichannel study. *J. Personal. Soc. Psychol.* 55, 580–587. doi: 10.1037/0022-3514.55.4.580

Ekman, P., and Friesen, W. (1978). *Facial Action Coding System*. Bel Air, CA: Consulting Psychologists Press.

Frauendorfer, D., Schmid Mast, M., Nguyen, L. S., and Gatica-Perez, D. (2014). Nonverbal social sensing in action: unobtrusive recording and extracting of nonverbal behavior in social interactions illustrated with a research example. *J. Nonverbal Behav.* 38, 231–245. doi: 10.1007/s10919-014-0173-5

Gatica-Perez, D. (2015). Signal processing in the workplace. *IEEE Signal Proc. Mag.* 32, 121–125. doi: 10.1109/MSP.2014.2359247

Glowinski, D., Camurri, A., Volpe, G., Dael, N., and Scherer, K. (2008). "Technique for automatic emotion recognition by body gesture analysis," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, 1–6. doi: 10.1109/CVPRW.2008.4563173

Gratch, J., Artstein, R., Lucas, G. M., Stratou, G., Scherer, S., Nazarian, A., et al. (2014). "The distress analysis interview corpus of human and computer interviews," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, 3123–3128.

Hall, J. A., Coats, E. J., and LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* 131, 898–924. doi: 10.1037/0033-2909.131.6.898

Harrigan, J. A. (2005). "Proxemics, kinesics, and gaze," in *The New Handbook of Methods in Nonverbal Behavior Research*, eds J. A. Harrigan, R. Rosenthal, and K. Scherer (Oxford: Oxford University Press), 137–198.

High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). *Ethics Guidelines for Trustworthy AI*. European Commission Report. Avaliable online at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed March 15, 2021).

Howard, A., Zhang, C., and Horvitz, E. (2017). "Addressing bias in machine learning algorithms: a pilot study on emotion recognition for intelligent systems," in *Proceedings of IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO)*, Austin, TX, 1–7. doi: 10.1109/ARSO.2017.8025197

Jayagopi, D. B., Hung, H., Yeo, C., and Gatica-Perez, D. (2009). Modeling dominance in group conversations using nonverbal activity cues. *IEEE Trans. Audio Speech Lang. Proc.* 17, 501–513. doi: 10.1109/TASL.2008.2008238

Kapoor, A., Qi, Y., and Picard, R. W. (2003). "Fully automatic upper facial action recognition," in *2003 IEEE International SOI Conference*, Newport Beach, CA, 195–202.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Sunstein, C. R. (2018). Discrimination in the age of algorithms. *J. Legal Anal.* 10, 113–174. doi: 10.1093/jla/laz001

Lambrecht, A., and Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Manag. Sci.* 65, 2966–2981. doi: 10.1287/mnsc.2018.3093

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., et al. (2007). "Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection," in *Proceedings of the 2007 Workshop on Tagging, Mining and Retrieval of Human Related Activity information*, (New York, NY: ACM), .9–14. doi: 10.1145/1330588.1330590

Mast, M. S., Hall, J. A., Cronauer, C. K., and Cousin, G. (2011). Perceived dominance in physicians: are female physicians under scrutiny? *Patient Educ. Couns.* 83, 174–179. doi: 10.1016/j.pec.2010.06.030

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., et al. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289. doi: 10.1038/s41593-018-0209-y

McDuff, D., Girard, J. M., and el Kaliouby, R. (2017). Large-scale observational evidence of cross-cultural differences in facial behavior. *J. Nonverbal Behav.* 41, 1–19. doi: 10.1007/s10919-016-0244-x

McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J., and Picard, R. (2013). "Affectiva-mit facial expression dataset (am-fed): naturalistic and spontaneous facial expressions collected," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Piscataway, NJ: IEEE, 881–888.

Mehta, Y., Majumder, N., Gelbukh, A., and Cambria, E. (2019). Recent trends in deep learning based personality detection. *Artif. Intel. Rev.* 53, 2313–2339.

Muralidhar, S., Nguyen, L. S., Frauendorfer, D., Odobez, J. M., Schmid Mast, M., and Gatica-Perez, D. (2016). "Training on the job: behavioral analysis of job interviews in hospitality," in *ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction*, (New York, NY: ACM), 84–91. doi: 10.1145/2993148.2993191

Muralidhar, S., Schmid Mast, M., and Gatica-Perez, D. (2018). A tale of two interactions: inferring performance in hospitality encounters from cross-situation social sensing. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 2, 1–24. doi: 10.1145/3264939

Naim, I., Tanveer, M. I., Gildea, D., and Hoque, M. E. (2015). "Automated prediction and analysis of job interview performance: the role of what you say and how you say it," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 1, Ljubljana, 1–6. doi: 10.1109/FG.2015.7163127

Nguyen, L. S., Frauendorfer, D., Schmid Mast, M., and Gatica-Perez, D. (2014). Hire me: computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. Multimed.* 16, 1018–1031. doi: 10.1109/TMM.2014.2307169

Nguyen, L. S., Odobez, J. M., and Gatica-Perez, D. (2012). "Using self-context for multimodal detection of head nods in face-to-face interactions," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, New York, NY: ACM, 289–292. doi: 10.1145/2388676.2388734

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

Pantic, M., and Rothkrantz, L. J. M. (2004). Facial action recognition for facial expression analysis from static face images. *IEEE Trans. Syst. Man Cybernet. Part B Cybernet.* 34, 1449–1461. doi: 10.1109/TSMCB.2004.825931

Pentland, S. J. (2018). *Human-Analytics in Information Systems Research and Applications in Personnel Selection*. Doctoral dissertation. Tucson, AZ: The University of Arizona. UA Campus Repository.

Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., and Zancanaro, M. (2008). "Multimodal recognition of personality traits in social interactions," in *Proceedings of the 10th International Conference on Multimodal Interfaces*, Chania, 53–60. doi: 10.1145/1452392.1452404

Poppe, R. (2017). "Automatic analysis of bodily social signals," in *Social Signal Processing*, eds J. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli (Cambridge, MA: Cambridge University Press), 155–167. doi: 10.1017/9781316676202.012

Poppe, R., Van Der Zee, S., Heylen, D. K., and Taylor, P. J. (2014). AMAB: automated measurement and analysis of body motion. *Behav. Res. Methods* 46, 625–633. doi: 10.3758/s13428-013-0398-y

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). "Mitigating bias in algorithmic hiring: evaluating claims and practices," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, NY, 469–481. doi: 10.1145/3351095.3372828

Rahman, A., Clift, L. G., and Clark, A. F. (2019). "Comparing gestural interfaces using kinect and OpenPose [Poster presentation]," *in Proceedings of the 37th Computer Graphics & Visual Computing, Bangor, United Kingdom*, Bangor, 103–104. doi: 10.2312/cgvc.20191264

Rasipuram, S., and Jayagopi, D. B. (2018). Automatic assessment of communication skill in interview-based interactions. *Mult. Tools Appl.* 77, 18709–18739. doi: 10.1007/s11042-018-5654-9

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

Sanchez-Cortes, D., Aran, O., and Gatica-Perez, D. (2011). "An audio visual corpus for emergent leader analysis," in *Workshop on Multimodal Corpora for Machine Learning: Taking Stock and Road Mapping the Future, ICMI-MLMI*, Alicante.

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., and Gatica-Perez, D. (2013). Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *J. Mult. User Interf.* 7, 39–53. doi: 10.1007/s12193-012-0101-0

Sedda, A., Manfredi, V., Bottini, G., Cristani, M., and Murino, V. (2012). Automatic human interaction understanding: lessons from a multidisciplinary approach. *Front. Hum. Neurosci.* 6:57. doi: 10.3389/fnhum.2012.00057

Segura, C., Balcells, D., Umbert, M., Arias, J., and Luque, J. (2016). *"Automatic speech feature learning for continuous prediction of customer satisfaction in contact center phone calls," in Proceedings of the International Conference on Advances in Speech and Language Technologies for Iberian Languages, Lisbon, Portugal*, (New York, NY: Springer International Publishing), 255–265.

Tong, Y., Liao, W., and Ji, Q. (2006). "Inferring facial action units with causal relations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, New York, NY, 1623–1629. doi: 10.1109/CVPR.2006.154

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2012). Bridging the gap between social animal and unsocial machine: a survey of social signal processing. *IEEE Trans. Affect. Comput.* 3, 69–87. doi: 10.1109/T-AFFC.2011.27

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009a). Social signal processing: survey of an emerging domain. *Image Vis. Comput.* 27, 1743–1759. doi: 10.1016/j.imavis.2008.11.007

Vinciarelli, A., Salamin, H., and Pantic, M. (2009b). "Social signal processing: understanding social interactions through nonverbal behavior analysis," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami, FL, 42–49. doi: 10.1109/CVPRW.2009.5204290

Zweig, G., Siohan, O., Saon, G., Ramabhadran, B., Povey, D., Mangu, L., et al. (2006). "Automated quality monitoring for call centers using speech and NLP technologies," *in Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume: Demonstrations, New York City, USA,* New York, NY, 292-295.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

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: frontiersin.org/about/contact

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