Saving Face in Front of the Computer? Culture and Attributions of Human Likeness Influence Users' Experience of Automatic Facial Emotion Recognition

Jan-Philipp Stein* and Peter Ohler

Chair of Media Psychology, Institute for Media Research, Chemnitz University of Technology, Chemnitz, Germany

In human-to-human contexts, display rules provide an empirically sound construct to explain intercultural differences in emotional expressivity. A very prominent finding in this regard is that cultures rooted in collectivism—such as China, South Korea, or Japan—uphold norms of emotional suppression, contrasting with ideals of unfiltered self-expression found in several Western societies. However, other studies have shown that collectivistic cultures do not actually disregard the whole spectrum of emotional expression, but simply prefer displays of socially engaging emotions (e.g., trust, shame) over the more disengaging expressions favored by the West (e.g., pride, anger). Inspired by the constant advancement of affective technology, this study investigates if such cultural factors also influence how people experience being read by emotion-sensitive computers. In a laboratory experiment, we introduce 47 Chinese and 42 German participants to emotion recognition software, claiming that it would analyze their facial micro-expressions during a brief cognitive task. As we actually present standardized results (reporting either socially engaging or disengaging emotions), we manipulate participants' impression of having matched or violated culturally established display rules in a between-subject design. First, we observe a main effect of culture on the cardiovascular response to the digital recognition procedure: Whereas Chinese participants quickly return to their initial heart rate, German participants remain longer in an agitated state. A potential explanation for this—East Asians might be less stressed by sophisticated technology than people with a Western socialization—concurs with recent literature, highlighting different human uniqueness concepts across cultural borders. Indeed, while we find no cultural difference in subjective evaluations of the emotion-sensitive computer, a mediation analysis reveals a significant indirect effect from culture over perceived human likeness of the technology to its attractiveness. At the same time, violations of cultural display rules remain mostly irrelevant for participants' reaction; thus, we argue that inter-human norms for appropriate facial expressions might be loosened if faces are read by computers, at least in settings that are not associated with any social consequence.

Keywords: display rules, collectivism, emotional suppression, affective computing, AFER, facial emotion recognition, cardiovascular activity, human uniqueness

INTRODUCTION

Throughout past millennia, the exchange of emotional expressions has mostly been reserved for the interaction between biological entities. As a result, emotional behavior is typically considered as a core concept of *human*—or, at the very least, animalistic—communication (Wegner and Gray, 2016). Recent breakthroughs in the field of affective computing, however, have started to contest this domain: Contemporary artificial intelligence not only incorporates emotional recognition algorithms, but may also possess the ability to emulate its own "affective state," reacting to a user's input in a supposedly emotional way. Due to these innovations, whole new research fields and industries have emerged all over the globe, including social robotics (e.g., Michaud et al., 2000), agent-based psychotherapy (e.g., Oker et al., 2015), and emotionally aware smartphone applications (e.g., Chen et al., 2015). A crucial feature for many of these technologies is automatic facial emotion recognition (AFER)—a camera-based form of facial analysis that gets more distinguished by the day (Doerrfeld, 2015; Gunes and Hung, 2016). AFER measures a wide range of movements in the user's facial muscles, including micro-expressions that are nearly undetectable to the human eye, before applying classification systems such as the Facial Action Coding System (Ekman and Friesen, 1978) to provide accurate interpretations of users' mood or short-term affective response. Depending on the variant of recognition software used, the final result may even offer a simultaneous quantification of several emotions—although the reliability of such systems remains the subject of critical debate.

Affective Technology vs. Human Uniqueness

Apart from the relentless advancement of emotion-sensitive technology itself, the public adoption of affective forms of human-computer interaction (HCI) has not proceeded without obstacles. In fact, recent studies have uncovered strong feelings of discomfort among participants who were presented with emotional computers (Gray and Wegner, 2012) and empathic digital agents (Stein and Ohler, 2017). Cross-national research implies that these effects might depend on cultural factors, as different religions and philosophical mindsets promote the importance of *human uniqueness* to a varying degree (Kaplan, 2004; Vess et al., 2012). Whereas most Western civilizations remain embedded in Christian principles of anthropocentrism—a philosophy that puts humans above all other creation—East Asian societies tend to have a less restricted idea of human distinctiveness, which allows the attribution of emotional experience to many different entities (Kitano, 2007; Kazuhiko, 2017). In consequence, Chinese or Japanese users may find it hardly problematic if a machine acquires typically "human" features such as the ability to recognize or express its own feelings; people in the West, on the other hand, are often socialized with a "Frankenstein syndrome" (Kaplan, 2004), therefore considering such technology as a threat to human nature itself (Złotowski et al., 2017). In practice, these arguments—although not entirely unchallenged (Haring

et al., 2014)—also offer an explanation for the notably higher acceptance of social robots in countries such as Japan or China (MacDorman et al., 2009; Li et al., 2010; Nomura et al., 2015).

The inclusion of philosophical concepts in empirical HCI studies certainly shows that users' acceptance of emotional technology reaches far beyond questions of programming or basic interface design. Still, due to the novelty and constant advancement of affective systems, many theoretical implications in terms of user experience have not yet been addressed in a sufficient way. A psychological construct that remains particularly under-researched in this regard are *emotional display rules*, which can be defined as behavioral criteria for the expression of emotions that stem from cultural socialization (Ekman and Friesen, 1969). While numerous studies have highlighted the importance of these sociocultural norms for human-to-human contexts, conclusive findings on the transferability of such rules to human-computer interactions are virtually absent from the academic discourse. The current study strives to contribute to this research gap, exploring whether the customary way of expressing emotions among humans may also affect the perception of emotion-sensitive software.

Emotional Display Rules

So far, psychological literature has not yet provided a decisive answer concerning the universality—or incongruity—of emotional experience across different cultures (Matsumoto and Ekman, 1989; Elfenbein and Ambady, 2002; Derntl et al., 2012; Hwang and Matsumoto, 2015). Nevertheless, scholars have unanimously acknowledged the importance of emotional display rules as a mediator of people's observable emotional behavior. This means that, while cultural socialization might not necessarily impact the conception of emotional states within the individual, it clearly determines which part of the subjective experience is presented to the environment (Matsumoto et al., 2008a). As a result, emotional display rules yield the power to make people feel accepted (or disregarded) by their cultural in-group, contributing profoundly to an individual's psychological well-being (Ford and Mauss, 2015).

Display rules have been shown to correlate with age (Camras, 1985; Underwood et al., 1992) and gender (Brody, 1997; Kring and Gordon, 1998), as well as several personality characteristics (Matsumoto, 2006). Most of all, however, they are influenced by factors of culture, building upon vastly different conventions, taboos, and expectations across ethnic groups, countries, and whole continents (Matsumoto et al., 2008a; Safdar et al., 2009). To structure these effects on a global level, researchers have frequently utilized the differentiation between *individualistic* and *collectivistic* cultures, which remains one of the most prominent frameworks in the field of cross-cultural psychology. In its original interpretation as a bipolar dimension (Hofstede, 1980), collectivism describes a culture's tendency to value interdependence, self-restraint, and in-group cohesion, in contrast to the individualistic emphasis on personal goals and self-expression. Based on these criteria, many East Asian cultures have typically been attributed with a strong collectivistic orientation, whereas the United States and several West European countries have been labeled as highly

individualistic societies (Hofstede, 2001). At the same time, an increasing number of authors have been contesting Hofstede's dichotomy as a general model of the East and the West (e.g., Oyserman et al., 2002; Takahashi et al., 2002; Parker et al., 2009), instead suggesting to split the singular construct into two independent traits (Triandis and Gelfand, 1998). In this evolved form, the IND-COL taxonomy is still used extensively for cultural comparisons, not least in the exploration of societal norms such as emotional display rules.

So far, a large number of studies have provided evidence that *emotional suppression* constitutes the overarching display rule in many countries with a collectivistic orientation (Matsumoto et al., 2008b), including China (Davis et al., 2012), South Korea (Kim and Sherman, 2007), Singapore (Moran et al., 2013), and Japan (Safdar et al., 2009). Unlike more individualism-centered societies in the West, these East Asian cultures have been shown to disregard unfiltered emotional displays, instead promoting the concealment of individual feelings for the sake of the collective social order (Markus and Kitayama, 1991; Matsumoto et al., 2008a). Even more so, cultural scientists have suggested that the according principles trace back as far as ancient Taoist and Confucian tradition (King and Bond, 1985; Ho, 1986) and are thus deeply embedded in the "cultural DNA" of the respective societies. As a result, the corresponding norms are usually internalized at an early age—which also explains why, unlike the adverse effects reported for Western samples, emotional suppression actually increases the subjective well-being (Matsumoto et al., 2004), academic performance (Chen et al., 2009), and psychological functioning (Soto et al., 2011) in members of Chinese or Japanese culture. Fascinatingly, this positive view on emotional concealment is also reflected by a variety of linguistic nuances, as East Asian languages provide an unmatched number of idioms to describe the act of hiding one's emotions, including the notion of "saving face" (Ho, 1976). Indeed, for the daily life of many East Asians, the metaphorical sense of "face" (as a form of social standing) remains inseparably intertwined with the literal face, as well-functioning regulatory mechanisms are deemed essential to avert public humiliation (Ho et al., 2004; Dong et al., 2013).

Display Rules and Types of Emotion

Apart from the well-replicated main effect of culture on emotional suppression norms, many cross-cultural researchers have directed their attention toward potential interaction effects between display rules, different audiences, and specific types of emotion. For instance, recent findings have shown that members of collectivistic cultures are particularly focused on differentiating between private and workplace contexts as they consider appropriate facial displays (Wang et al., 2012; Moran et al., 2013). According to a large-scale comparison of display rules across 32 countries by Matsumoto et al. (2008a), this might actually extend to general in- and out-group effects: Although participants from collectivistic societies reported less expressivity in general, they actually endorsed negative emotional displays toward strangers much more than participants with an individualistic background. This observation in turn connects to a growing body of literature about the "appropriateness" of

selected emotional states (e.g., Kitayama et al., 2000; Eid and Diener, 2001; Seo, 2011), which has suggested that individualistic cultures strongly favor displays of personal success (e.g., pride, joy), whereas collectivists prefer emotions that highlight interrelatedness—even if their expression emphasizes personal failure (e.g., guilt, shame). Considering the core principles of both cultural orientations, this actually makes perfect sense: Just as the visible acknowledgment of personal shortcomings highlights the investment in the collective well-being, turning guilt and embarrassment into *other-focused emotions* (Markus and Kitayama, 1991), displays of pride or happiness mostly serve to express private gain, therefore meshing with a more individualistic philosophy (i.e., *ego-focused emotions*). Lending further support to this conceptualization, recent research has revealed that East Asian participants actually appreciated guilt and shame as themes of "social engagement," while American individuals preferred "socially disengaging" displays of pride and anger (Kitayama et al., 2006; Boiger et al., 2013).

The Physiological Side of Emotional Expression

Anyone who has ever noticed their heart accelerating in fear or anticipation knows that emotional experience is not just an abstract product of the human psyche, but also heavily intertwined with physiological processes. In this regard, the most important interface of the human body can be found in the autonomic nervous system (ANS), which controls a multitude of unconscious bodily functions and mirrors the affective state of the individual through parameters such as blood pressure or skin conductance level (Kreibig, 2010). At the same time, previous research remains indecisive on the question whether physiological reflections actually offer insight in the quality—other than just the intensity—of emotional experience. While some authors argue that autonomic response patterns may in fact be emotion-specific (e.g., Levenson, 2003) or at least convey emotional valence (Bensafi et al., 2002; Brosschot and Thayer, 2003), recent research suggests that observable differences in bodily reactions merely deliver insight into underlying motivational systems (Mendes, 2016). The search for emotion-specific, autonomic reactions is further complicated by the fact that changes within the ANS are not only evoked by emotional states, but also reflect many other mental processes such as acute stress (e.g., Dickerson and Kemeny, 2004), higher levels of concentration (e.g., Wass et al., 2016), and ruminative thoughts (e.g., Ottaviani et al., 2009). On this account, it has become common practice to interpret increases of autonomic activity primarily as indicators of *arousal*, which can only be linked to specific emotions in controlled laboratory settings.

Returning to this study's main topic of emotional display rules, one may also wonder how emotional suppression efforts register in terms of physiological activity. However, findings on this subject have pointed into two directions: Just as some studies indicate a reduced physiological activation after suppression strategies (Zuckerman et al., 1981), others have provided potent arguments for the increased physiological cost of emotionally suppressive behavior (Gross and Levenson, 1993; Butler et al.,

2003). A possible dissolution of this dispute might be found in cultural adaptation effects. For instance, an experiment conducted by Butler et al. (2009) has shown that emotion-expressive behavior led to an increase in blood pressure among Asian Americans but to a decrease among European Americans. Similarly, a recent study from the field of neuroscience has reported that, after viewing unpleasant pictures, Asian Americans showed a much faster decrease of the brain potentials related to emotional processing than US Americans (Murata et al., 2013). As more and more similar findings emerge for various tasks and contexts (e.g., Shen et al., 2004; Zhou and Bishop, 2012), the bulk of the empirical evidence argues for reduced physiological activity in East Asians as they, consciously and sub-consciously, regulate their emotional displays—an effect that might turn out quite differently for members of other cultures.

The Current Study

Introducing common findings from cross-cultural psychology to the research field of affective HCI, the current study set out to scrutinize how Chinese and German users would react to the impression of being "unmasked" by a computer-based form of emotional recognition. For a comprehensive examination of this response, we investigated both participants' physiological arousal as well as their subjective affinity to AFER software following a (supposedly) automatic reading of their facial emotions.

The consulted literature provided us with two reasonable assumptions how the chosen cultural samples might differ in their reaction to the presented technology. On the one hand, it seemed highly likely that Chinese participants, as members of a collectivism-oriented culture, would perceive the AFER procedure as an unpleasant experience, considering that the recognition of facial micro-expressions serves as a substantial—and, compared to human interactions, unprecedented—form of "losing face." In consequence, we found it logical to assume that individuals with this cultural background would show a more pronounced arousal reaction (meaning either a stronger or longer increase of physiological activity—or both). On the other hand, we pondered that the stronger sense of caution against humanlike technology in the West (e.g., Kaplan, 2004; MacDorman et al., 2009; Nomura et al., 2015) could just as well lead to more anxiety among German individuals, who might consider affective computers as a threat to human uniqueness. In consequence, we decided to juxtapose these contradicting hypotheses:

H1a: The physiological arousal measured after feedback from AFER software will be more intense among Chinese participants.

H1b: The physiological arousal measured after feedback from AFER software will be more intense among German participants.

Apart from physiological effects, we were also interested in participants' subjective evaluation of the presented technology. In our interpretation, the arguments that had led to our first set of hypotheses applied just as well to this research focus: If a cultural group would perceive the emotionally aware computer as discomforting and arousing, it seemed highly likely that they would also report a lower affinity to the system in question. As such, we again formulated a set of two conflicting assumptions, matching H1a and H1b:

H2a: The subjective evaluation of emotion recognition software will be less favorable among Chinese participants.

H2b: The subjective evaluation of emotion recognition software will be less favorable among German participants.

Compensating for the exploratory nature of our two-fold hypotheses, we included an additional measure to help us understand why one assumption might overrule the other: Participants' attribution of human likeness to the presented AFER system. Doing so, we strived to find out whether our cultural groups differed in their perceptions of AFER (non-)artificiality—and whether these attributions served as a mediator between culture and the eventual affinity to the presented technology.

RQ1: Do Chinese and German participants differ in terms of the human likeness they ascribe to AFER technology?

RQ2: Is participants' AFER affinity mediated by these human likeness attributions?

Apart from potential main effects of culture, we were also interested if users' reactions were to turn out differently as soon as the digital system indicated a violation of—instead of a match with—cultural display rules from the inter-human context. For interactions with human strangers, East Asians have been shown to favor socially engaging emotions (even negative ones such as shame), while Westerners focus on more ego-focused and exclusively positive expressions (e.g., Eid and Diener, 2001; Matsumoto et al., 2008a; Boiger et al., 2013). Considering this, we were curious to find out whether AFER systems would count as "just another stranger," toward whom the discussed display norms would be fully in play. If so, the impression of having shown pride in front of the computer should emerge as unpleasant for Chinese but not for German participants; the latter should dislike reports of ashamed facial expressions instead.

H3: If emotion recognition software reports facial displays that contradict culture-specific display rules, participants will show more physiological arousal.

H4: If emotion recognition software reports facial displays that contradict culture-specific display rules, participants will evaluate it less favorably.

Our interest in comparing different display rule conditions, however, heralded some methodological challenges. After some initial deliberation, we quickly disregarded the idea of artificially inducing specific facial expressions, because we were highly skeptical of the validity and reliability of this approach—especially since our study was supposed to revolve around rather subtle displays of emotion (e.g., expressions of pride). Of course, the alternative of waiting for participants to express a specific emotion without purposely stimulating it seemed just as impractical. An elimination of the identified problems eventually occurred in the form of a more deceptive approach. By providing

participants with a fully standardized, faux result instead of a genuine reading of their facial expressions, we found a unique solution to manipulate display rule violations as needed. As an additional merit, this procedure allowed us to standardize the alleged intensity of emotional expressivity across participants—which would have been impossible otherwise. At the same time, we now had to contain the risk of participants doubting the provided results. For this purpose, our study was adjusted in two ways: First, during the initial presentation of the AFER system, we repeatedly emphasized that the technology would focus on so-called micro-expressions, which are hardly discernible to the human eye; secondly, we decided to have participants fill in a demanding cognitive test during the alleged analysis, thereby distracting them from a conscious monitoring of their own face.

METHODS

Participants

Following an a-priori analysis of required sample size with *G*Power* software (Faul et al., 2007), we recruited 100 students from two ethnic groups at a German university: 51 participants who self-identified as Chinese (24 female, 27 male) and 49 participants who self-identified as German (39 female, 10 male). To control for possible acculturation effects among the Chinese participants who temporarily lived as exchange students in Germany, we formulated two inclusion criteria: (a) having spent one's youth in Mainland China, Taiwan, or Hong Kong, and (b) speaking Chinese (e.g., Mandarin) as current main language. According to previous research (e.g., Guan, 2007; Yu and Wang, 2011), the community of Chinese exchange students in Germany remains highly connected to their home culture, yet extremely isolated within the host society. This has been confirmed by reports from our participants and colleagues, so that we consider our sample a valid reference for the experiences of young Chinese students. Moreover, the fact that our final sample consisted exclusively of individuals from Mainland China—without any students from Taiwan or Hong Kong—lends further support to the homogeneity of this experimental group.

Following a manipulation check, a total of three participants (all from the German sample) had to be excluded from further analysis, as they had doubted the authenticity of the presented stimuli. Additionally, the data of seven participants (3 Chinese, 4 German) could not be used due to technical difficulties in their physiological measurements. Lastly, one Chinese participant had to quit the experiment early on account of his previously undisclosed color blindness. Therefore, our final sample included 47 Chinese (age $M = 26.1$ years, $SD = 2.62$) and 42 German students (age $M = 22.6$ years, $SD = 3.93$), for a total of 89 participants. As a compensation for his or her time, each participant could choose between €5 or credits for our university's mandatory "experimental participation" course, in which students are required to participate in 20 experiments (available from a large catalog of different research projects).
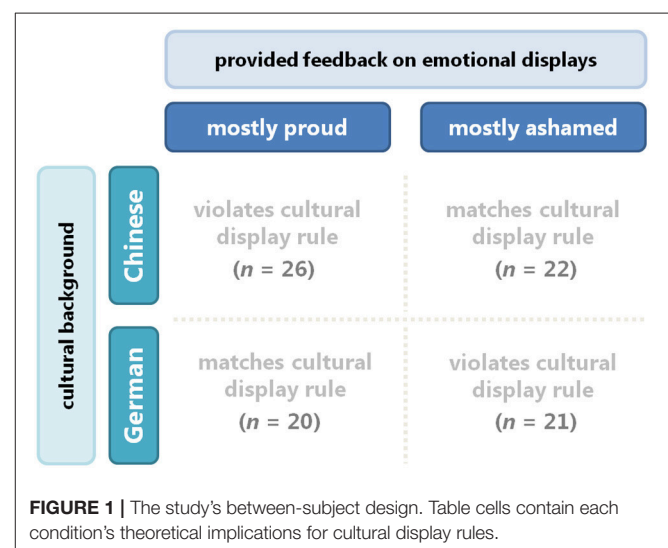
Procedure

At the beginning of the experiment, participants were told that the current study served to explore cultural differences in facial micro-expressions during a cognitive task. Pointing out the video camera and desktop PC in our laboratory, we explained that the experiment would involve state-of-the-art AFER software, which could "monitor [the user's] face for spontaneous muscle movements"—including "even the tiniest contractions"—and "calculate a summary of the emotional displays during any given task." As we were not interested in participants' actual facial displays, but only in their reaction to a manipulated recollection of it, we prepared two standardized result sheets, with one claiming that the software had recognized pride and the other indicating shame in the user's micro-expressions. This resulted in a 2 × 2 between-subject design as illustrated in **Figure 1**. Participants were assigned to one of the two feedback conditions by means of a block randomization procedure.

After they had filled out an informed consent form, we equipped participants with an unobtrusive physiological monitoring wristband, recording their heart rate for the remainder of the experiment. Following a 1-min baseline measurement (which was conducted while sitting in silence), we then provided the materials for a short intelligence test and instructed participants to keep their head directed toward the video camera for the duration of the task. Under the deceptive impression that their facial displays were being monitored, participants filled in the test for a timed duration of 3 min. Subsequently, they were instructed to access the analysis' alleged result on a private screen—a method that we chose not only to prevent potential feelings of humiliation in front of the study conductor, but also to put all focus on the concept of being analyzed in a "non-human" context.

As soon as participants finished reading the provided results, they filled in a short questionnaire on their acceptance of the software, as well as a few control variables. Lastly, we took down each participant's e-mail address to provide them with an extensive explanation of the study's goals, its deceptive design, and our findings. Those who did not want to enter their e-mail address were debriefed directly and kindly asked to keep the

FIGURE 1 | The study's between-subject design. Table cells contain each condition's theoretical implications for cultural display rules.

experiment's true nature a secret until all recruited students had taken part.

Stimulus Design

As materials for the brief intelligence test, we used self-created matrix completion tasks modeled after the widely-used Raven Progressive Matrices (Raven et al., 2003). In this type of test, participants have to choose one of eight options as the missing tile of a 3 × 3 symbol matrix. Since the performance in matrix tasks does not depend on language or factual knowledge, they are considered a culture-faire method of testing; even though the actual performance was not relevant to our hypotheses, we therefore chose this type of cognitive test for our cross-cultural experiment.

For the believable introduction of an authentic AFER system, we used the recognition software *MultiSense*, which is freely available as part of the Virtual Human Toolkit (Hartholt et al., 2013). Among other functions, the software includes face tracking and a basic form of expressivity analysis, visualizing the results via various computer windows (e.g., camera feed with automatically aligned 3D grid). Although we chose not to record any actual facial data from our participants, we briefly showed them a live stream of their face within the MultiSense environment at the beginning of the experiment in order to foster our deception of genuine facial recognition. **Figure 2** depicts the visualization setup as it was presented to the participants on the experimenter's swiveling computer screen. Doing so, we always made sure to show the interface for only a couple of seconds and from a small distance, thus obscuring its technological specifics.

Furthermore, to mitigate any privacy concerns, we explicitly stated that the software "worked in real-time" and would not need to save recordings of the participant's face at any time.

Apart from the introduction of the software itself, the authentic presentation of its manipulated results was absolutely crucial for the success of the experiment. As such, we prepared a step-by-step procedure to convey the analysis' outcome in a believable way. Firstly, we composed a web-based result sheet using both HTML and JavaScript code (see **Figure 3**), which was able to dynamically display the metadata of each appointment, including the participant's individual number, cultural group, and time of measurement. As center part of the sheet, we prepared a table with fictional parameters for eight affective states (anger, fear, sadness, happiness, surprise, trust, shame, and pride) in the style of existing AFER frameworks (Doerrfeld, 2015). Most importantly, one of the table's rows was colored in bright red, highlighting the according emotion as most prevalent feeling during the intelligence test. Depending on experimental condition, this highlight was set on either *pride* or *shame*, with all other emotions fixed at medium levels. To explicitly direct participants' attention toward the relevant part of the sheet, the study conductor (who was sitting several feet away) was instructed to always tell them to "focus on the scores in red, which indicate the predominant emotion during the experiment." Lastly, we compiled two additional graphs and added them to our sheet, further increasing the saliency of the relevant facial expressions. The completed result page was then translated from German into Simplified Chinese, with back-translations ensuring the similarity of both versions.
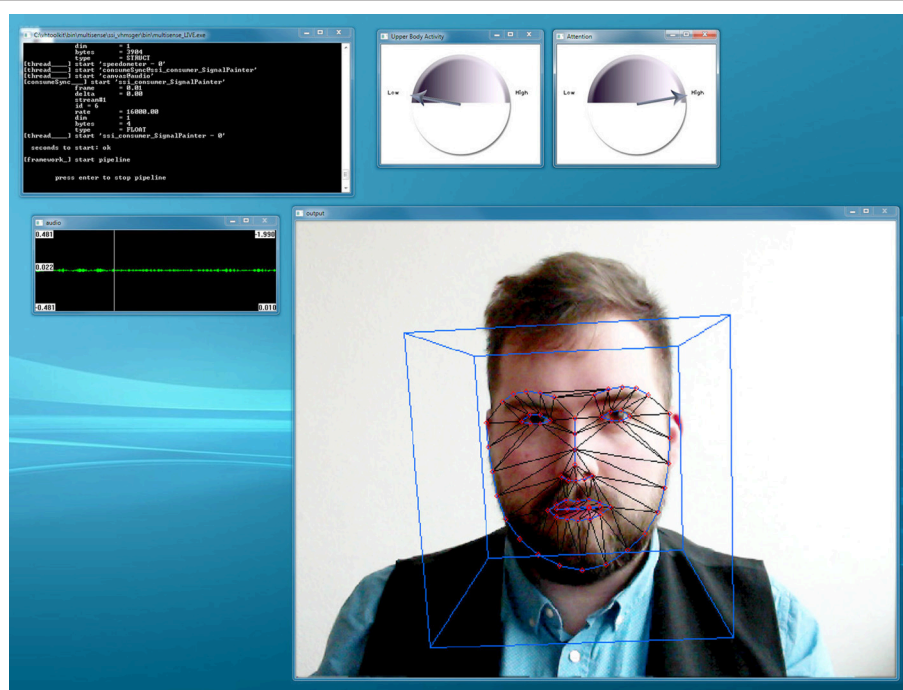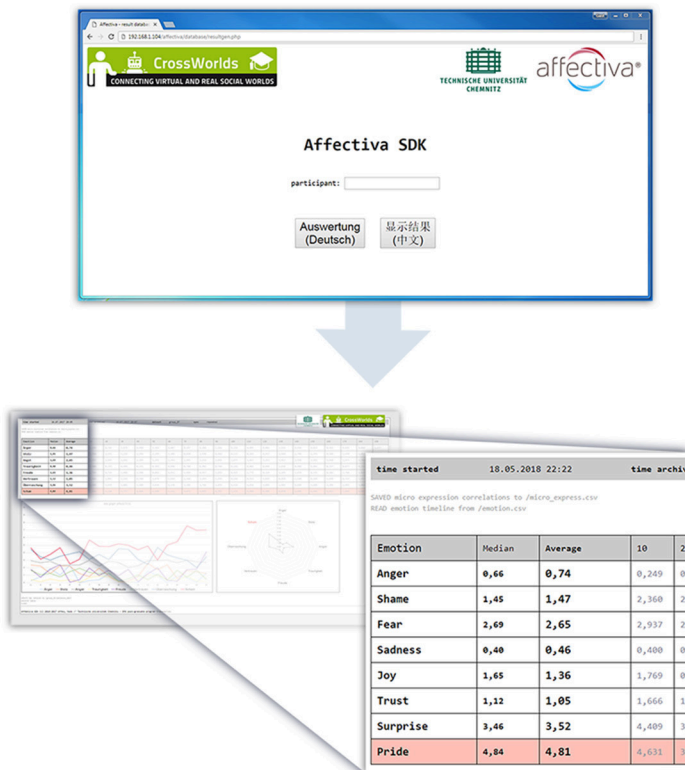

FIGURE 2 | MultiSense visualization used for the deceptive narrative of emotion recognition. The main window (bottom right) shows a live feed of the user's face, with a 3D grid automatically fitted to relevant facial points (demonstrated by a study conductor in this image).

FIGURE 3 | Procedure to convey the deceptive result sheet after the experimental task. **(A)** The login screen hosted on a local web server. **(B)** Result sheet with fictitious data. The allegedly most prevalent emotion is displayed in red—either shame or pride, depending on condition.

Measures

Cardiovascular Activity

We used an Empatica E4 physiological measurement wristband (Empatica Inc., 2017) to achieve unobtrusive monitoring of our participants' heart rate. Although the Empatica E4 is able to measure heart rate and heart rate variability with a frequency of 1 Hz (one data point per second), we averaged the values of 60 s to achieve a meaningful reduction of data. During the experiment, the study conductor marked the exact second in which participants opened the manipulated AFER result sheet as time of stimulus onset. For a theoretically coherent analysis, however, we always added 5 s to this time to account for a basic cognitive processing of the presented stimuli. Since the cardiovascular system is known to respond more slowly to stressors than other physiological indicators, we aligned our procedure with previous studies, which have examined autonomic arousal and recovery during the first few minutes after stimulus onset (e.g., Brosschot and Thayer, 2003; Roberts et al., 2008; Boer, 2016). As a result, the following measures were obtained for all participants: (1) a 1-minute baseline of resting heart rate, starting shortly after the wristband had been equipped; (2) the average heart rate during the first minute after stimulus presentation, starting with a delay of 5 s; (3) the average

heart rate during the subsequent second minute after stimulus presentation.

Attractiveness and Human Likeness of the Presented Technology

Participants rated their subjective impression of the presented AFER procedure using the technology-related attractiveness index by Ho and MacDorman (2010), which captures the affective response toward a technological stimulus. Although the questionnaire's five semantic differentials (e.g., "repulsive—agreeable," "messy—sleek"; rated on a 7-point scale) were originally designed to address visual features in relation to the "uncanny valley" phenomenon, we found them suitable for an evaluation of our more abstract stimuli as well; in fact, the authors suggest that their index basically pinpoints participants' reaction on an evolutionary avoidance-approach continuum, which strongly correlates to perceptions of interpersonal warmth. Due to our study's cross-cultural nature, we translated the original English items into German and Simplified Chinese and used back-translations by native speakers to ensure semantic equivalence. Both translations proved to be of acceptable to high internal consistency (Chinese version, $\alpha = 0.70$; German version, $\alpha = 0.81$). To establish measurement invariance

between both versions, we conducted a series of increasingly restrictive confirmatory factor analyses (CFAs), which is a common procedure to test measures for configural, metric, and scalar invariance. Doing so, partial scalar invariance could be established for our translated attractiveness indices. **Table 1** gives an overview of the conducted model comparisons.

For the exploratory investigation of human likeness attributions to the AFER technology, we used Ho and MacDorman's *human likeness* index Ho and MacDorman (2010), which assesses the amount of animacy and human nature ascribed to a technology (with artificiality and synthetic nature as other endpoints of the spectrum). Whereas the original version of the measure consists of six semantic differentials (e.g., "human-made—humanlike," "artificial—lifelike"), we excluded two items that did not apply to our disembodied scenario, namely "biological movement—mechanical movement" and "without definitive lifespan—mortal." The resulting four item scale was again translated, with internal consistency turning out a bit lower for the Chinese ($\alpha = 0.63$) than for the German version ($\alpha = 0.79$). In our interpretation, this concurs with the reviewed literature as it suggests a more complex understanding of human likeness in Chinese tradition. Nevertheless, multi-group CFA testing for measurement invariance again indicated partial scalar invariance between our two translations (see **Table 1**) so that we still included the measure in the exploratory part of our study.

Manipulation Checks

By design, the current study did not focus on participants' real emotional displays; quite the opposite, we aimed at convincing them of a standardized result. Although various efforts were expended to facilitate this goal (e.g., emphasizing the role of micro-expressions, distracting participants from their face, and keeping the reported affect at a moderate level), we also decided

to include some form of measurement to assess the success of our deception. Specifically, we asked participants to rate the accuracy of the software's analysis on a self-developed two item scale ($\alpha = 0.84$)—which was then used to identify the cases where had to assume a discrepancy between the provided feedback and a person's self-perception. As a conservative rule, all participants who had filled in the lowest score (1 out of 5) on one or both items were completely excluded from our study. Eventually, this was the case for three participants from the German sub-sample, who reported frequent participation in psychological studies and, as such, might have been especially wary of potential deceptions.

To gain additional insight into our manipulation's validity, we inquired participants to rate their own performance in the matrix completion task on a 5-point scale. By connecting these ratings to the actual test results, we were able to assess whether participants could actually estimate their accomplishment in a realistic manner—which would have disrupted our manipulation. Fortunately, our analysis showed that both groups had very little insight into their true performance.

RESULTS

The measured scores and physiological data from our final sample can be obtained from the **Data Sheet 1** in the Supplementary Material.

Manipulation Checks
Software Accuracy
To check the evaluations of *software accuracy* for significant group differences, we calculated a two-way analysis of variance (ANOVA) with the between-subject factors culture and type of feedback. The procedure yielded no significant main effect for the latter, $F_{(1, 85)} = 0.81$, $p = 0.81$, and no significant interaction

TABLE 1 | Multi-group confirmatory factorial analyses to check translated scales for measurement invariance.

Scale	Model	Constrained to be equal across groups	χ^2 ($\Delta \chi^2$)	*df* (Δdf)	(Δp)	CFI (ΔCFI)	RMSEA (ΔRMSEA)
Attractiveness	**M1** Configural invariance	Factor structure	16.128	10	–	0.947	0.117
	M2 Metric invariance	Factor structure, loadings	(2.814)	(4)	(0.59)	(0.010)	(0.028)
	M3a Scalar invariance	Factor structure, loadings, intercepts	(19.784)	(4)	(<0.01**)	(0.136)	(0.072)
	M3b Partial scalar invariance (except items 4 and 5)	Factor structure, loadings, intercepts	(1.609)	(2)	(0.45)	(0.003)	(0.009)
	M4 Structural invariance	Factor structure, loadings, intercepts, means	(0.083)	(1)	(0.77)	(0.008)	(0.011)
Human likeness	**M1** Configural invariance	Factor structure	3.058	4	–	1.000	0.000
	M2 Metric invariance	Factor structure, loadings	(1.179)	(3)	(0.76)	(0.000)	(0.000)
	M3a Scalar invariance	Factor structure, loadings, intercepts	(9.609)	(3)	(<0.05*)	(0.055)	(0.093)
	M3b Partial scalar invariance (except item 4)	Factor structure, loadings, intercepts	(4.712)	(2)	(0.09)	(0.000)	(0.000)
	M4 Structural invariance	Factor structure, loadings, intercepts, means	(7.329)	(1)	(<0.01**)	(0.090)	(0.119)

According to Chen (2007), $\Delta CFI < 0.010$ and $\Delta RMSEA < 0.015$ indicate that the invariance assumption also holds for the more restricted model. $^*p < 0.05$, $^{**}p < 0.01$.

between factors, $F_{(1, 85)} = 2.51$, $p = 0.11$. Accordingly, we note that both reports of "proud" and "ashamed" facial displays were seen as moderately accurate, regardless of cultural background. However, we did observe a significant main effect of culture, $F_{(1, 85)} = 5.43$, $p = 0.02$, $\eta_p^2 = 0.06$; Chinese participants generally ascribed higher accuracy to the presented software ($M = 3.49$, $SD = 0.84$) than German participants ($M = 3.08$, $SD = 0.80$). While this might somewhat reflect the stronger skepticism of the native-speaking students at our university—who typically take part in more psychological studies than exchange students— our finding could also just emphasize the stronger belief in technological prowess among the Chinese. In any case, we report that both group means manifested slightly above the scale's midpoint, so that we still deem our manipulation acceptably successful.

Actual and Perceived Test Performance

A two-way ANOVA with participants' actual test results as a dependent variable revealed no significant differences between Chinese and German participants, $F_{(1, 85)} = 0.32$, $p = 0.57$. On average, German students solved $M = 5.76$ test items correctly ($SD = 1.76$), closely matching the performance of the Chinese students ($M = 5.60$ correct answers, $SD = 1.54$). However, German participants facilitated their scores with a higher number of total answers ($M_{GER} = 9.83$, $M_{CN} = 8.17$), including more wrong answers ($M_{GER} = 4.07$, $M_{CN} = 2.57$). We further investigated whether the two groups receiving different AFER feedback had, by chance, produced significantly different test results, which may have been problematic for the believability of the deception. However, this was not the case, $F_{(1, 85)} = 2.20$, $p = 0.14$.

In terms of participants' own perception of their task performance, another two-way ANOVA with the factors culture and type of feedback revealed a strong main effect of culture, $F_{(1, 85)} = 11.36$, $p < 0.01$, $\eta_p^2 = 0.12$. Indeed, our data show that Chinese participants considered their performance significantly better ($M = 3.64$, $SD = 0.82$) than German participants ($M = 3.05$, $SD = 0.83$). While there was no notable main effect for type of feedback, an interaction between both factors emerged as marginally significant, $F_{(1, 85)} = 4.14$, $p = 0.05$, albeit with a very small effect size of $\eta_p^2 = 0.03$. Examining our data pattern, we observed that only Chinese participants rated their performance significantly higher if the software had indicated "proud" ($M =$

3.92, $SD = 0.64$) instead of "ashamed" facial displays ($M = 3.32$, $SD = 0.89$); for the German participants, self-assessment in both "proud" ($M = 3.00$, $SD = 0.97$) and "ashamed" conditions ($M = 3.09$, $SD = 0.68$) turned out rather similar. On account of the effect's marginal size and significance, however, we advise to interpret this finding with caution.

As our final, but probably most crucial manipulation check, we conducted two separate linear regression analyses to find out if the real test result predicted the self-assessment from participants of both groups. This investigation did not result in significant regression equations, neither for Chinese [$F_{(1, 46)} = 1.14$, $p = 0.29$] nor for German [$F_{(1, 41)} = 0.34$, $p = 0.56$] students. Hence, we argue that participants had relatively little insight if their performance had been good or bad, which certainly supported our manipulation of alleged micro-expressions.

Cardiovascular Activity

Table 2 shows the means and standard deviations for participants' heart rate during the baseline measurement, as well as the 2 min after stimulus presentation. For reasons of clarity, the table also contains the relative differences between subsequent data points. As an additional illustration, the graphs in **Figure 4** depict the average heart rate changes in the four experimental groups, compared to their respective baseline values.

To check the acquired data for statistically meaningful differences in physiological activity, we focused on the heart rate changes during the first (initial arousal) and second minute (sustained arousal) with two separate analyses of covariance. In both ANCOVA procedures, we controlled for participants' age and gender by entering them as covariates, as these variables have been shown to profoundly influence cardiovascular (re-)activity (e.g., Carroll et al., 2000).

A 2 (culture) × 2 (type of feedback) ANCOVA with the mean heart rate change between baseline and first minute post-stimulus as dependant variable resulted in no significant main effects for culture [$F_{(1, 83)} = 0.05$, $p = 0.81$], type of feedback [$F_{(1, 83)} = 1.39$, $p = 0.24$], or interaction between both [$F_{(1, 83)} = 0.06$, $p = 0.80$]. In terms of participants' initial cardiovascular response, we therefore conclude that there was no meaningful difference between the selected cultural groups—or that the effect was too small to be detected by the test power achieved with our sample size. The same applies to our assumptions about

TABLE 2 | Descriptive statistics for heart rates and relative heart rate changes between the three measuring points.

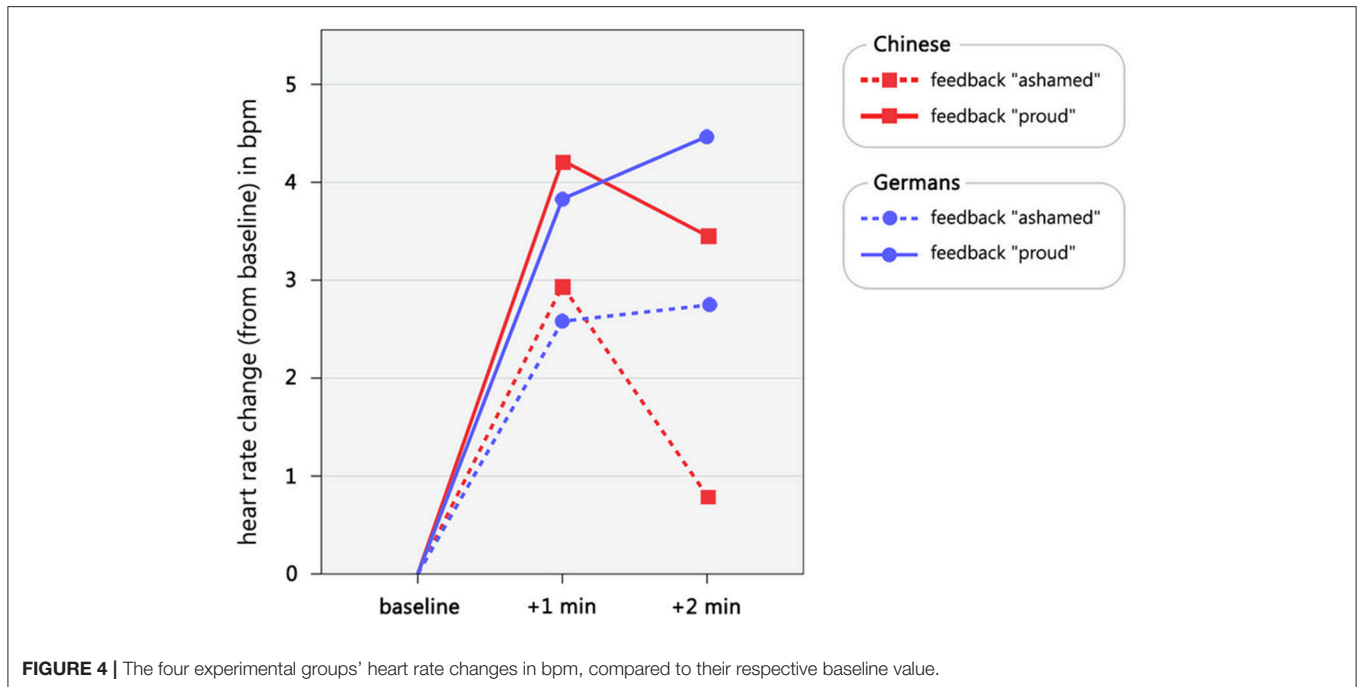| | | Heart rate in beats per minute (bpm) | | | | |
| | | Baseline | 1st minute | | 2nd minute | |
Culture	Feedback	M (SD)	M (SD)	Δ to prev.	M (SD)	Δ to prev.
Chinese	Ashamed	79.35 (11.89)	82.29 (12.88)	+ 3.7%	80.15 (12.97)	−2.6%
	Proud	77.95 (7.77)	82.17 (12.00)	+ 5.4%	81.40 (11.07)	−0.9%
	Total	78.60 (9.82)	82.23 (12.28)	+ 4.6%	80.82 (11.88)	−1.7%
German	Ashamed	76.48 (8.14)	79.06 (9.00)	+ 3.4%	79.22 (7.64)	+ 0.2%
	Proud	76.86 (8.78)	80.69 (9.08)	+ 5.0%	81.33 (8.88)	+ 0.8%
	Total	76.66 (8.35)	79.84 (8.97)	+ 4.1%	80.23 (8.22)	+ 0.5%

FIGURE 4 | The four experimental groups' heart rate changes in bpm, compared to their respective baseline value.

the type of given feedback. Although the data of the Chinese participants indeed suggest more initial arousal if the AFER result had claimed micro-expressions of pride ($M = +4.22$ bpm, $SD = 8.56$) instead of shame ($M = +2.94$ bpm, $SD = 7.50$), the observed effect missed the threshold of statistical significance and should therefore be interpreted cautiously. For the German participants, the influence of the given feedback was even smaller.

However, using the heart rate change between the first and second minute post-stimulus as a dependent variable, another ANCOVA yielded a significant main effect of culture, $F_{(1, 83)} = 6.26$, $p = 0.01$, with a moderate effect size of $\eta_p^2 = 0.07$. At the same time, no significant effects were found for type of feedback [$F_{(1, 83)} = 1.22$, $p = 0.27$] or factor interaction [$F_{(1, 83)} = 0.03$, $p = 0.87$]. Thus, in terms of "sustained" arousal, German participants remained more agitated after the computerized recognition procedure, even showing yet another increase in heart rate ($M = +0.39$ bpm, $SD = 5.53$). The Chinese group, on the other hand, reduced their heart rate by $M = -1.41$ bpm ($SD = 3.98$) during the second minute after stimulus presentation.

In summary, our findings offer first evidence in favor of hypothesis H1b over H1a: Even though their initial response was not stronger *per se*, we argue that the *longer* arousal in the German sub-sample does constitute a more intense physiological reaction to the AFER feedback. At the same time, hypothesis H3 could not be confirmed by our data, as the manipulated feedback had no noteworthy effect on the arousal of our participants (if potential type II errors are dismissed).

Subjective Impression of the Software

Table 3 gives an overview of the scores yielded from this study's self-report measures. To check both ratings for statistically

TABLE 3 | Descriptive statistics for self-report measures.

Culture	Feedback	Human likeness M (SD)	Attractiveness M (SD)
Chinese	Ashamed	2.82 (0.81)	3.62 (0.54)
	Proud	3.07 (0.71)	3.78 (0.52)
	Total	2.95 (0.76)	3.71 (0.53)
German	Ashamed	2.51 (0.90)	3.63 (0.59)
	Proud	2.29 (0.78)	3.45 (0.53)
	Total	2.40 (0.84)	3.54 (0.56)

Both scales range from 1 to 5.

significant group differences, we conducted a pair of 2 × 2 ANOVAs, again using culture and type of feedback as between-subject factors.

The first ANOVA focusing on participants' attractiveness ratings did not uncover any significant effects, neither main effects for culture [$F_{(1, 85)} = 1.962$, $p = 0.17$] or type of feedback [$F_{(1, 85)} = 0.01$, $p = 0.96$], nor an interaction between both [$F_{(1, 85)} = 2.19$, $p = 0.14$]. As we found almost identical attractiveness ratings in all four groups, our results supported neither hypothesis H2a nor H2b: The cultural groups did not differ in their subjective liking of the presented stimuli. As alleged violations of human-to-human display rules hardly influenced this as well, we further reject hypothesis H4.

Focusing on our second subjective measure, a two-way analysis of variance with human likeness as the dependent variable resulted in no significant main effect for type of feedback, $F_{(1, 85)} = 0.01$, $p = 0.94$, and no interaction between feedback and culture, $F_{(1, 85)} = 1.95$, $p = 0.17$. However, we

now obtained a significant main effect of culture $F_{(1, 85)} = 10.24$, $p < 0.01$, with a very large effect size of $\eta_p^2 = 0.11$. As expected from the literature review, Chinese participants showed a notably higher attribution of humanness to the affective software ($M = 2.95$, $SD = 0.76$) than German participants ($M = 2.40$, $SD = 0.84$), therefore giving a positive answer to RQ1.

Mediation Analysis

Our exploration of culture as a quasi-experimental macro-level predictor had not resulted in a clear empirical effect on technology attractiveness. However, to address our question if human uniqueness concepts might actually stand as a mediator between these two constructs (RQ2), we proceeded to a mediation analysis using the PROCESS macro for IBM SPSS (Hayes, 2013), using five thousand iterations for bootstrap confidence intervals. Indeed, the procedure uncovered a significant indirect effect from culture over human likeness attributions to attractiveness ratings, b = −0.09, SE = 0.05, 95% CI [−0.23, −0.01], with the mediator accounting for roughly half of the total effect, $P_M = 0.55$. **Figure 5** gives an overview of all obtained standardized regression coefficients in the mediation analysis; due to the dummy coding of our cultural groups (0 = "Chinese", 1 = "German"), the negative coefficient indicates higher outcomes among the Chinese sample. In particular, our data show that having a Chinese cultural background significantly predicted higher human likeness attributions, which in turn predicted higher attractiveness ratings. Apart from this indirect effect, the direct relationship between culture and technology attractiveness remained insignificant. RQ2 can therefore be answered positively.
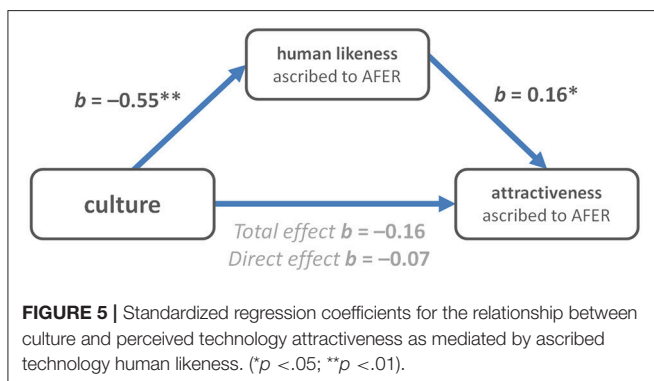
DISCUSSION

To investigate if emotion-sensitive forms of human-computer interaction are influenced by sociocultural factors, we introduced Chinese and German participants to a facial recognition system and prepared its results to either match or violate traditional norms for emotional expressions toward strangers. Measuring cardiovascular parameters as well as subjective impressions of the AFER

software, we addressed both the subconscious and conscious processing of this new, digitally mediated form of "becoming unmasked."

Based on an interdisciplinary review of literature, two contrasting assumptions emerged on how Chinese and German individuals might compare in their experience of automatic facial emotion analyses. Specifically, we contemplated that either the Western preference for candid facial expressions or the East Asian tradition to accept "human-like" qualities in non-human entities as something completely natural would eventually tip the scales in favor of one of the two groups. Our results suggest the latter. Despite comparable increases of cardiovascular activity immediately after the presentation of the AFER results, Chinese participants soon returned to a notably lower heart rate, whereas German participants lingered in a state of sustained arousal. Accordingly, we report a response pattern that matches findings by Matsumoto et al. (2009), suggesting that the initial response to a stimulus might be universal before slightly delayed "cultural influences kick in" (p. 1273). Considering that complex artificiality is often seen as a threat in Western cultures (Kaplan, 2004; Złotowski et al., 2017), we argue that the arousal of the German students might indeed indicate some form of anxiety, i.e., an autonomic reflection of their subliminal wariness toward the novel technology. At the same time, it has to be noted that changes in cardiovascular activity do not allow a direct interpretation of emotional quality: Being aroused by a stimulus could just as well signal curiosity or positive excitement. Similarly, we deem it possible that the quicker regulation of Chinese participants' heart rate is heavily influenced by their proficiency in autonomic emotional suppression, which has been demonstrated by previous experiments (e.g., Shen et al., 2004; Zhou and Bishop, 2012). With these limitations in mind, we suggest that the reported main effect of culture on the physiological arousal evoked by AFER is interpreted conservatively.

At the same time, our investigation of human likeness evaluations and their revealed role as a mediator between culture and the final AFER affinity clearly support the suggested interpretation of our results. Whereas cultural background showed no isolated effect on the attractiveness participants ascribed to the emotionally aware system, we found that our groups differed greatly in how "animate" and "human-like" they considered the presented computer, which in turn predicted the final attractiveness ratings. In our explanation, this implies that culture as a macro-level container for many confounding variables may not suffice to completely explain views on technology—yet be essential in forming people's basic philosophy and worldview (e.g., the idea of what constitutes a human-natured entity), which then interacts with other individual dispositions to form actual attitudes and behaviors. For HCI developers, this builds toward an unambiguous recommendation: Only by tapping into both cross-cultural and individual-level forms of research, they might eventually settle on the right amount of "human" that customers from different backgrounds would like to rediscover in their technology.

FIGURE 5 | Standardized regression coefficients for the relationship between culture and perceived technology attractiveness as mediated by ascribed technology human likeness. (*$p < .05$; **$p < .01$).

Figure content:
human likeness ascribed to AFER
b = −0.55**
b = 0.16*
culture
attractiveness ascribed to AFER
Total effect b = −0.16
Direct effect b = −0.07

Lastly, we report that different types of feedback by the emotion recognition system had very little impact on participants' experience in our scenario. Empirically, it did not matter whether the computer claimed displays of pride or shame as the most prevalent facial expression, neither for participants' arousal nor for their liking of the software—and regardless of cultural background. Providing an interpretation of these findings, we suggest that the confidential reveal in our experiment (the AFER results were conveyed on a private screen) must have weakened the perceived importance of traditional display rules by a great extent. Since a recognition system that is used privately has no bearing on one's social standing or any in-group coherence, people might be completely indifferent about culturally desirable expressions in such a setting. Based on this argument, however, we would expect much stronger effects once the technology was used to trigger meaningful real-life consequences. Considering that psychological counseling, e-learning, and job assessments are all being targeted as application fields for AFER, we expect various use cases in which a much stronger need for "appropriate" facial displays will arise—may it be a virtual classroom or a automatic job selection procedure. As such, we strongly believe that, even though they remained insignificant in our single-user experiment, traditional display rules will instantly find new relevance once emotionally aware systems turn into actual mediators of interdependence, social standing, or financial success.

LIMITATIONS

Our results are limited in their generalizability due to the use of a convenience sample that consisted exclusively of students ranging between 18 and 37 years. Similarly, we note a slightly uneven distribution of female and male participants, especially in the German group. Although we tried to control for these factors (e.g., by including them as covariates or additional predictors in our statistical tests), different results might emerge if other samples, for instance children or elderly participants, were to experience the provided scenario. This also applies to participants' level of education: Since our student sample represents only a small part of the socio-economic spectrum, we consider it highly likely that other findings would be obtained with participants pursuing other occupations.

In regard to the cultural comparison conducted in this study, we note that both the Chinese and the German group consisted of individuals from different parts of the respective country, which potentially underestimates regional influences. However, in light of the long cultural distance between both nations, we still think that our study achieved an insightful juxtaposition of cultural differences in the perception and experience of affective technology. Nevertheless, we think that future studies might benefit from explicitly asking participants about their reliance on cultural values. Qualitative methods such as semi-structured interviews might be particularly useful in this regard, promising a less ambiguous understanding of how users' socialization has contributed to their conceptualization of emotional appropriateness, preference for collectivistic values, and, consequently, their reaction to AFER analyses.

CONCLUSION

From an early age, most people are socialized to adjust their emotional output as soon as they interact with other perceptive entities—that is, typically, with other humans. Yet, due to breakthroughs in AFER technology, computers are now also able to read emotions from the human face, thereby entering the world of emotional communication as an exciting new intermediary.

At first, we pondered the possibility that common display rules might simply be transferred one-to-one to the interaction with emotion-sensitive technology. However, based on our empirical efforts, we have come to the conclusion that AFER systems do not provoke the same concerns about "appropriate" facial displays that are common among humans—at least as long as they possess only limited influence on other outcomes. In isolated interactions with "benevolent" affective computers, people simply have no incentive to be anxious about a certain kind of result. For the near future, however, we predict that real-life applications of AFER will hardly stay as inconsequential or innocuous as our experimental scenario. Clearly, the technology has not been designed as a "single-player gimmick" but as a method of collecting data in numerous contexts. In current times, it seems all but far-fetched to envision autonomous cars, medical robots, or automatic job interview systems whose emotional perceptiveness all but determines crucial outcomes for the humans involved; and the establishment of such procedures will surely have users turn back to their human-to-human norms of behavior. If our study is any indication, it will depend both on cultural and individual factors whether society appreciates this development—or faces it with an underlying anxiety.

ETHICS STATEMENT

This study was conducted in accordance with all institutional regulations and the ethical guidelines of the German Psychological Society (2016). Formal approval by a special ethics committee is not required for psychological research in Germany, as long as research objectives do not involve issues regulated by law. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The provided consent form included the voluntariness and the anonymity of the obtained data. Only if participants had thoroughly read and signed the provided form, we continued with the experiment. Due to the partially deceptive nature of the experimental procedure, participants also received a full written debriefing at the end of the study.

AUTHOR CONTRIBUTIONS

J-PS study conception and design, literature review, creation of stimulus materials, data collection, statistical analysis, and

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdigh. 2018.00018/full#supplementary-material

Data Sheet 1 | The collected data of the final study sample.

REFERENCES

Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., and Holley, A. (2002). Influence of affective and cognitive judgments on autonomic parameters during inhalation of pleasant and unpleasant odors in humans. *Neurosci. Lett.* 319, 162–166. doi: 10.1016/S0304-3940(01) 02572-1

Boer, D. P. (2016). *The Wiley Handbook on the Theories, Assessment and Treatment of Sexual Offending*. Hoboken, NJ: John Wiley & Sons.

Boiger, M., Mesquita, B., Uchida, Y., and Feldman Barrett, L. (2013). Condoned or condemned: the situational affordance of anger and shame in Japan and the US. *Pers. Soc. Psychol. Bull.* 39, 540–553. doi: 10.1177/0146167213478201

Brody, L. R. (1997). Gender and emotion: beyond stereotypes. *J. Soc. Issues* 53, 369–394. doi: 10.1111/j.1540-4560.1997.tb02448.x

Brosschot, J. F., and Thayer, J. F. (2003). Heart rate response is longer after negative emotions than after positive emotions. *Int. J. Psychophysiol.* 50, 181–187. doi: 10.1016/S0167-8760(03)00146-6

Butler, E. A., Egloff, B., Wilhelm, F. H., Smith, N. C., Erickson, E. A., and Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion* 3, 48–67. doi: 10.1037/1528-3542.3.1.48

Butler, E. A., Lee, T. L., and Gross, J. J. (2009). Does expressing your emotions raise or lower your blood pressure? The answer depends on cultural context. *J. Cross Cult. Psychol.* 40, 510–517. doi: 10.1177/0022022109332845

Camras, L. (1985). "Socialization of affect communication," in *The Socialization of Emotions*, eds M. Lewis and C. Saarni (New York, NY: Plenum Press), 141–160.

Carroll, D., Harrison, L. K., Johnston, D. W., Ford, G., Hunt, K., Der, G., et al. (2000). Cardiovascular reactions to psychological stress: the influence of demographic variables. *J. Epidemiol. Commun. Health* 54, 876–877. doi: 10.1136/jech.54.11.876

Chen, F. F. (2007). Sensitivity of Goodness of Fit indexes to lack of measurement invariance. *Struct. Equat. Model. Multidiscipl. J.* 14, 464–504. doi: 10.1080/10705510701301834

Chen, M., Zhang, Y., Li, Y., Mao, S., and Leung, V. C. M. (2015). EMC: Emotion-aware mobile cloud computing in 5G. *IEEE Netw.* 29, 32–38. doi: 10.1109/MNET.2015.7064900

Chen, X., Chen, H., Li, D., and Wang, L. (2009). Early childhood behavioral inhibition and social and school adjustment in Chinese children: a 5-year longitudinal study. *Child Dev.* 80, 1692–1704. doi: 10.1111/j.1467-8624.2009.01362.x

Davis, E., Greenberger, E., Charles, S., Chen, C., Zhao, L., and Dong, Q. (2012). Emotion experience and regulation in China and the United States: how do culture and gender shape emotion responding? *Int. J. Psychol.* 47, 230–239. doi: 10.1080/00207594.2011.626043

Derntl, B., Habel, U., Robinson, S., Windischberger, C., Kryspin-Exner, I., Gur, R. C., et al. (2012). Culture but not gender modulates amygdala activities during explicit emotion recognition. *BMC Neurosci.* 13:54. doi: 10.1186/1471-2202-13-54

Dickerson, S. S., and Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol. Bull.* 130, 355–391. doi: 10.1037/0033-2909.130.3.355

Doerrfeld, B. (2015). *20+ Emotion Recognition APIs That Will Leave You Impressed, and Concerned [Blog post]*. Available online at: http://nordicapis. com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/

Dong, P., Huang, X. I., and Wyer, R. S. (2013). The illusion of saving face: how people symbolically cope with embarrassment. *Psychol. Sci.* 24, 2005–2012. doi: 10.1177/0956797613482946

Eid, M., and Diener, E. (2001). Norms for experiencing emotions in different cultures: inter- and intranational differences. *J. Pers. Soc. Psychol.* 81, 869–885. doi: 10.1037/0022-3514.81.5.869

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98. doi: 10.1515/semi.1969.1.1.49

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

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

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/BF031 93146

Ford, B. Q., and Mauss, I. B. (2015). Culture and emotion regulation. *Curr. Opin. Psychol.* 3, 1–5. doi: 10.1016/j.copsyc.2014.12.004

German Psychological Society (2016). *Berufsethische Richtlinien des BDP und der DGPs.* Available online at: https://www.dgps.de/index.php?id=85

Gray, K., and Wegner, D. M. (2012). Feeling robots and human zombies: mind perception and the uncanny valley. *Cognition* 125, 125–130. doi: 10.1016/j.cognition.2012.06.007

Gross, J. J., and Levenson, R. W. (1993). Emotional suppression: physiology, self-report, and expressive behavior. *J. Pers. Soc. Psychol.* 64, 970–986. doi: 10.1037/0022-3514.64.6.970

Guan, H. (2007). *Anpassung und Integration der Chinesischen Studierenden in Deutschland.* Doctoral dissertation. Available online at: http://webdoc.sub. gwdg.de/ebook/dissts/Bremen/Guan2007.pdf

Gunes, H., and Hung, H. (2016). Is automatic facial expression recognition of emotions coming to a dead end? The rise of the new kids on the block. *Image Vision Comput.* 55, 6–8. doi: 10.1016/j.imavis.2016.03.013

Haring, K. S., Mougenot, C., Ono, F., and Watanabe, K. (2014). Cultural differences in perception and attitude towards robots. *Int. J. Affect. Eng.* 13, 149–157. doi: 10.5057/ijae.13.149

Hartholt, A., Traum, D., Marsella, S. C., Shapiro, A., Stratou, G., Leuski, A., et al. (2013). "All together now: introducing the virtual human toolkit," in *Proceedings of the 13th International Conference on Intelligent Virtual Agents*, eds R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira (Berlin; Heidelberg: Springer), 361–368.

Hayes, A. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis.* New York, NY: The Guilford Press.

Ho, C.-C., and MacDorman, K. F. (2010). Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. *Comput. Hum. Behav.* 26, 1508–1518. doi: 10.1016/j.chb.2010. 05.015

Ho, D. Y. F. (1976). On the concept of face. *Am. J. Soc.* 81, 867–884. doi: 10.1086/226145

Ho, D. Y. F. (1986). "Chinese patterns of socialization: a critical review," in *The Psychology of the Chinese People*, ed M. H. Bond (Hong Kong: Oxford University Press), 1–37.

Ho, D. Y. F., Fu, W., and Ng, S. M. (2004). Guilt, shame and embarrassment: revelations of face and self. *Cult. Psychol.* 10, 64–84. doi: 10.1177/1354067X04044166

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work Related Values*. Beverly Hills, CA: Sage.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Thousand Oaks, CA: Sage Publications.

Hwang, H., and Matsumoto, D. (2015). "Evidence for the universality of facial expressions of emotion," in *Understanding Facial Expressions in Communication*, eds M. Mandal and A. Awasthi (New Delhi: Springer), 41–56.

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *Int. J. Humanoid Robot.* 1, 465–480. doi: 10.1142/S0219843604000289

Kazuhiko, K. (2017). *An Introduction to Yokai Culture: Monsters, Ghosts, and Outsiders in Japanese History*. Tokyo: Japan Publishing Industry Foundation for Culture.

Kim, H. S., and Sherman, D. K. (2007). "Express yourself": culture and the effect of self-expression on choice. *J. Pers. Soc. Psychol.* 92, 1–11. doi: 10.1037/0022-3514.92.1.1

King, A. Y. C., and Bond, M. H. (1985). "The Confucian paradigm of man: a sociological view," in *Chinese Culture and Mental Health*, eds W. S. Tseng and D. Y. H. Wu (San Diego, CA: Academic Press), 29–45.

Kitano, N. (2007). "Animism, Rinri, modernization; the base of Japanese robotics," *Paper presented at the 2007 IEEE International Conference on Robotics and Automation (ICRA'07),* Rome. Available online at: http://www.roboethics.org/icra2007/contributions/KITANO%20Animism%20Rinri%20Modernization%20the%20Base%20of%20Japanese%20Robo.pdf

Kitayama, S., Markus, H. R., and Kurokawa, M. (2000). Culture, emotion, and well-being: good feelings in Japan and the United States. *Cogn. Emot.* 14, 93–124. doi: 10.1080/026999300379003

Kitayama, S., Mesquita, B., and Karasawa, M. (2006). Cultural affordances and emotional experience: socially engaging and disengaging emotions in Japan and the United States. *J. Pers. Soc. Psychol.* 91, 890–903. doi: 10.1037/0022-3514.91.5.890

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* 84, 394–421. doi: 10.1016/j.biopsycho.2010.03.010

Kring, A. M., and Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *J. Pers. Soc. Psychol.* 74, 686–703. doi: 10.1037/0022-3514.74.3.686

Levenson, R. W. (2003). "Autonomic specificity and emotion," in *Handbook of Affective Sciences*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 212–224.

Li, D., Rau, P. L. P., and Li, Y. (2010). A cross-cultural study: effect of robot appearance and task. *Int. J.Soc. Robot.* 2, 175–186. doi: 10.1007/s12369-010-0056-9

MacDorman, K. F., Vasudevan, S. K., and Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI Soc.* 23, 485–510. doi: 10.1007/s00146-008-0181-2

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. (2006). Are cultural differences in emotion regulation mediated by personality traits? *J. Cross Cult. Psychol.* 37, 421–437. doi: 10.1177/0022022106288478

Matsumoto, D., and Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motiv. Emot.* 13, 143–157. doi: 10.1007/BF00992959

Matsumoto, D., LeRoux, J. A., Bernhard, R., and Gray, H. (2004). Personality and behavioral correlates of intercultural adjustment potential. *Int. J. Intercult. Relat.* 28, 281–309. doi: 10.1016/j.ijintrel.2004.06.002

Matsumoto, D., Willingham, B., and Olide, A. (2009). Sequential dynamics of culturally moderated facial expressions of emotion. *Psychol. Sci.* 20, 1269–1274. doi: 10.1111/j.1467-9280.2009.02438.x

Matsumoto, D., Yoo, S. H., Fontaine, J., Anguas-Wong, A. M., Arriola, M., Ataca, B., et al. (2008a). Mapping expressive differences around the world: the relationship between emotional display rules and Individualism v.collectivism. *J. Cross Cult. Psychol.* 39, 55–74. doi: 10.1177/0022022107311854

Matsumoto, D., Yoo, S. H., and Nakagawa, S. (2008b). Culture, emotion regulation, and adjustment. *J. Pers. Soc. Psychol.* 94, 925–937. doi: 10.1037/0022-3514.94.6.925

Mendes, W. B. (2016). "Emotion and the autonomic nervous system," in *Handbook of Emotions*, eds L. F. Barrett, M. Lewis, and J. Haviland-Jones (New York, NY: Guilford Publications), 166–181.

Michaud, F., Pirjanian, P., Audet, J., and Létourneau, D. (2000). "Artificial emotion and social robotics," in *Distributed Autonomous Robotic Systems 4*, eds L. E. Parker, G. Bekey, and J. Barhen (Tokyo: Springer), 121–130.

Moran, C. M., Diefendorff, J. M., and Greguras, G. J. (2013). Understanding emotional display rules at work and outside of work: the effects of country and gender. *Motiv. Emot.* 37, 323–334. doi: 10.1007/s11031-012-9301-x

Murata, A., Moser, J. S., and Kitayama, S. (2013). Culture shapes electrocortical responses during emotion suppression. *Soc. Cogn. Affect. Neurosci.* 8, 595–601. doi: 10.1093/scan/nss036

Nomura, T., Syrdal, D. S., and Dautenhahn, K. (2015). "Differences on social acceptance of humanoid robots between Japan and the UK," in *Proceedings of the 4th International Symposium on New Frontiers in Human-Robot Interaction,* eds M. Salem, P. Baxter and K. Dautenhahn (Canterbury: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour), 115–120.

Oker, A., Prigent, E., Courgeon, M., Eyharabide, V., Urbach, M., Bazin, N., et al. (2015). How and why affective and reactive virtual agents will bring new insights on social cognitive disorders in schizophrenia? An illustration with a virtual card game paradigm. *Front. Hum. Neurosci.* 9:133. doi: 10.3389/fnhum.2015.00133

Ottaviani, C., Shapiro, D., Davydov, D. M., Goldstein, I. B., and Mills, P. J. (2009). The autonomic phenotype of rumination. *Int. J. Psychophysiol.* 72, 267–275. doi: 10.1016/j.ijpsycho.2008.12.014

Oyserman, D., Coon, H. M., and Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychol. Bull.* 128, 3–72. doi: 10.1037/0033-2909.128.1.3

Parker, R., S., Haytko, D. L., and Hermans, C. M. (2009). Individualism and collectivism: reconsidering old assumptions. *J. Int. Bus. Res.* 8:127.

Raven, J., Raven, J. C., and Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio, TX: Harcourt Assessment.

Roberts, N. A., Levenson, R. W., and Gross, J. J. (2008). Cardiovascular costs of emotion suppression cross ethnic lines. *Int. J. Psychophysiol.* 80, 82–87. doi: 10.1016/j.ijpsycho.2008.06.003

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

Seo, Y. S. (2011). Individualism, collectivism, and client expression of different emotions: their relations to perceived counselor effectiveness. *Asia Pac. Educ. Rev.* 12, 251–261. doi: 10.1007/s12564-010-9136-7

Shen, B.-J., Stroud, L. R., and Niaura, R. (2004). Ethnic Differences in cardiovasular responses to laboratory stress: a comparison between Asian and White Americans. *Int. J. Behav. Med.* 11, 181–186. doi: 10.1207/s15327558ijbm1103_7

Soto, J. A., Perez, C. R., Kim, Y.-H., Lee, E. A., and Minnick, M. R. (2011). Is expressive suppression always associated with poorer psychological functioning? A cross-cultural comparison between European Americans and Hong Kong Chinese. *Emotion* 11, 1450–1455. doi: 10.1037/a0023340

Stein, J.-P., and Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160, 43–50. doi: 10.1016/j.cognition.2016.12.010

Takahashi, K., Ohara, N., Antonucci, T. C., and Akiyama, H. (2002). Commonalities and differences in close relationships among the Americans and Japanese: a comparison by the individualism/collectivism concept. *Int. J. Behav. Dev.* 26, 453–465. doi: 10.1080/01650250143000418

Triandis, H. C., and Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *J. Pers. Soc. Psychol.* 1, 118–128. doi: 10.1037/0022-3514.74.1.118

Underwood, M. K., Coie, J. D., and Herbsman, C. R. (1992). Display rules for anger and aggression in school-age childen. *Child Dev.* 63, 366–380. doi: 10.2307/1131485

Vess, M., Arndt, J., and Cox, C. R. (2012). Faith and nature: the effect of death-relevant cognitions on the relationship between religious fundamentalism and connectedness to nature. *Soc. Psychol. Pers. Sci.* 3, 333–340. doi: 10.1177/1948550611420303

Wang, Y., Liu, X., and Cui, L. (2012). Influential factors of emotional display rules in Chinese adolescents. *Psychology* 3, 504–506. doi: 10.4236/psych.2012.36072

Wass, S., De Barbaro, K., and Clackson, K. (2016). Learning and the autonomic nervous system: understanding interactions between stress, concentration and learning during early childhood. *Front. Neurosci.* doi: 10.3389/conf.fnins.2016.92.00015

Wegner, D. M., and Gray, K. (2016). *The Mind Club: Who Thinks, What Feels, and Why It Matters*. New York, NY: Viking.

Yu, W., and Wang, W. (2011). An investigation into the acculturation strategies of Chinese students in Germany. *Intercult. Commun. Stud.* 20, 190–210.

Zhou, T., and Bishop, G. D. (2012). Culture moderates the cardiovascular consequences of anger regulation strategy. *Int. J. Psychophysiol.* 86, 291–298. doi: 10.1016/j.ijpsycho.2012.10.010

Złotowski, J., Yogeeswaran, K., and Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *Int. J. Hum.Comput. Stud.* 100, 48–54. doi: 10.1016/j.ijhcs.2016.12.008

Zuckerman, M., Klorman, R., Larrance, D. T., and Spiegel, N. H. (1981). Facial, autonomic, and subjective components of emotion: the facial feedback hypothesis versus the externalizer–internalizer distinction. *J. Pers. Soc. Psychol.* 41, 929–944. doi: 10.1037/0022-3514.41.5.929