

Large-scale observational evidence of cross-cultural differences in facial behavior

Daniel McDuff*, Jeffrey M. Girard*, and Rana el Kaliouby

Affectiva Inc., University of Pittsburgh, and Affectiva Inc.

Self-report studies have found evidence that cultures differ in the display rules they have for facial expressions (i.e., for what is appropriate for different people at different times). However, observational studies of actual patterns of facial behavior have been rare and typically limited to the analysis of dozens of participants from two or three regions. We present the first large-scale evidence of cultural differences in observed facial behavior, including 740,984 participants from 12 countries around the world. We used an Internet-based framework to collect video data of participants in two different settings: in their homes and in market research facilities. Using computer vision algorithms designed for this data set, we measured smiling and brow furrowing expressions as participants watched television ads. Our results reveal novel findings and provide empirical evidence to support theories about cultural and gender differences in display rules. Participants from more individualist cultures displayed more brow furrowing overall, whereas smiling depended on both culture and setting. Specifically, participants from more individualist countries were more expressive in the facility setting, while participants from more collectivist countries were more expressive in the home setting. Female participants displayed more smiling and less brow furrowing than male participants overall, with the latter difference being more pronounced in more individualist countries. This is the first study to leverage advances in computer science to enable large-scale observational research that would not have been possible using traditional methods.

Keywords: facial expression, culture, gender, crowdsourcing, computer vision

Introduction

Even before Darwin (1872), writers of antiquity conjectured that facial expressions of emotion are innate and shared across cultures (Evans, 1969). However, it was not until the 20th century that researchers began to investigate this *universality thesis* empirically. In a series of now-famous studies, Ekman and Friesen found support for the universality thesis in several preliterate cultures (Ekman & Friesen, 1971; Ekman, Sorenson, & Friesen, 1969). These studies (and the many that followed) generated considerable controversy in the research community: they have been touted as conclusive evidence of the universality thesis by some researchers (e.g., Matsumoto, 1990; Oster, Daily, & Goldenthal, 1989) and have been challenged on methodological grounds by others (e.g., Russell, 1994).

Researchers on both sides of the debate, however, have

acknowledged that facial behavior is influenced by culture. Klineberg (1940) argued that different cultures possess different norms for emotional expression, including rules for how expressions should be modified in different contexts. Ekman (1977) incorporated this idea into his neuro-cultural model of emotional expression, which argues that basic emotions are linked to specific, pan-cultural body actions that are filtered through culturally-defined *display rules* that dictate when and how such actions should be modified (e.g., potentiated or attenuated).

Studies using self-report and image-rating methodologies have explored the display rules that different cultures have for facial expressions, e.g., when and for whom an expression is appropriate (e.g., Ekman et al., 1987; Matsumoto, 1990; Matsumoto, Yoo, & Fontaine, 2008). Such rules are thought to exist in order to promote the attainment of group goals and to avoid common sources of interpersonal conflict (Argyle, Henderson, Bond, Iizuka, & Contarello, 1986). However, the extent to which these rules, and other cultural factors, influence actual patterns of facial behavior is less well-understood.

Several studies have directly examined cultural differences in facial behavior using observational methodologies (and will be discussed in detail in the next section). However, few of these studies have included more than a hun-

*The first and second authors contributed equally to this study.

Correspondence concerning this article should be addressed to Daniel McDuff, 463 Waverley Oaks Road, Waltham, MA, 02452. E-mail: djmcduff@media.mit.edu or to Jeffrey M. Girard, Department of Psychology, University of Pittsburgh, Pittsburgh, PA, 15260. E-mail: j.girard@pitt.edu

dred participants from multiple cultures. This state of affairs is largely due to the high cost of collecting and annotating observational data using traditional methods. For instance, training in the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002), the current “gold standard” for facial measurement (Cohn & Ekman, 2005; Ekman & Rosenberg, 2005), can take up to six months and coding a single minute of video can take well over an hour (Ekman, 1982). It can also be costly and logistically difficult to collect observational data in diverse and naturalistic settings. As a result, previous work has been limited by constrained and atypical settings (e.g., in a laboratory or on a TV game show) in addition to small samples.

By drawing on recent advances in computer science, the current study circumvents the prohibitive cost of observational measurement and, for the first time, quantifies cross-cultural differences in facial behavior on a truly large scale. Specifically, we used the Internet to collect videos of spontaneous facial behavior from hundreds of thousands of participants in twelve different countries and two different settings (i.e., in laboratories and in their homes). We then used algorithms to analyze the facial behavior in these videos and explored the influence of culture, gender, and setting on participants’ facial behavior.

Culture, context, and gender

Culture is a shared system of meaning and information that maintains social order within a group by providing values and norms for behavior, thought, and emotion that are shared among members and transmitted across generations (Matsumoto, 2006; Matsumoto et al., 2008). While there are many ways to define and measure culture, research has settled on the individualism-collectivism dimension (Hofstede, 1980, 2001) as a key component to be considered. Cultures high in individualism view the self as independent and rely on interpersonal exchange and interaction with strangers for success, while cultures high in collectivism view the self as interdependent and rely on group solidarity and conformity to norms for success (Oyserman, Coon, & Kemmelmeier, 2002).

Differences between cultures’ priorities and values are reflected in their display rules. In the largest cross-cultural study on display rules to date (which included 5,361 participants from 32 countries), Matsumoto et al. (2008) found that individualism was positively associated with higher self-reported expressivity norms for emotions in general, and for positive emotions in particular. These effects were interpreted as evidence that emotions (and their free expression) have greater importance in cultures that are more individualistic (see also Suh, Diener, Oishi, & Triandis, 1998). It was also proposed that these effects may be related to the positive correlation between individualism and country-level differences in extraversion (Hofstede & McCrae, 2004), a per-

sonality trait that has been linked to stronger expression of positive emotion (McCrae & Costa, 1999).

In addition to this overall effect of culture, an interaction between culture and social context was also identified. Specifically, cultures’ display rules were found to differ for interactions with *in-group* members and *out-group* members. Interactions with in-group members are characterized by a sense of intimacy, familiarity, and trust that is often based on a history of shared experiences and an anticipated future, while interactions with out-group members lack these qualities (Brewer & Kramer, 1985; Messick & Mackie, 1989; Tajfel, 1982). Matsumoto et al. (2008) found that, across cultures, participants had higher expressivity norms for interactions with in-group members as compared to out-group members. Additional findings suggest that individualist display rules favor the expression of positive emotion and censor the expression of negative emotion toward out-group members, while collectivist display rules do the same for in-group members. This pattern is consistent with the idea that display rules serve to foster the building of trust and cohesion in the social contexts that a culture most relies on for success.

Other work has also identified overall effects of gender on expressivity norms, as well as gender-by-culture and gender-by-context interactions (Brody & Hall, 2008). Fischer and Manstead (2000) found that, across 37 countries, women reported more intense and longer-lasting emotions and more overtly expressed emotions than men; these differences were also more pronounced in individualistic cultures, which the authors theorized as being due to males in these cultures suppressing their expressivity in order to protect their status. Scherer, Wallbott, and Summerfield (1986) found that, in Israel and seven European countries, females reported displaying more facial reactions than males when expressing joy, sadness, fear, and anger. In a meta-analysis of 162 research reports, LaFrance, Hecht, and Paluck (2003) found that women and adolescent girls smiled more than men and adolescent boys. This effect held across all 13 included countries of origin, although the gender difference was more pronounced in some cultures than others. For instance, participants from the United States and Canada showed larger sex differences in smiling than participants from the United Kingdom. It was also found that social context moderated the gender effect. Most relevant to the current study was a finding that sex differences were significantly larger when participants were in facility settings and aware that their behavior was being observed than when they believed that their behavior was not being observed.

Previous cross-cultural observational studies

While self-reports can reveal how emotions and their expression are perceived by a culture, their results do not necessarily reflect how emotions are actually expressed in a culture. In order to quantify the influence of cultural display

rules on extant behavior, cross-cultural observational studies are required. Several such studies have been conducted; however, they have been limited in terms of sample sizes, diversity, and social context.

Friesen (1973) compared the facial behavior of 25 Japanese and 25 American university students while watching stressful films. While there were no differences between groups when participants watched the films alone, the Japanese participants were more likely than the American participants to mask their negative expressions by smiling when in the presence of an experimenter. Ekman, Friesen, and Ellsworth (1972) interpreted this result as evidence of a Japanese (i.e., collectivist) display rule that negative emotions should be masked in the presence of others in order to foster harmony, cooperation, and cohesion.

Waxer (1985) analyzed the nonverbal behavior of 30 contestants on American TV game shows and 30 contestants on Canadian TV game shows. He found that American contestants were rated as more expressive than Canadian contestants and that female contestants were rated as more expressive than male contestants. Follow-up analyses revealed that American males smiled significantly more than Canadian males. These results were interpreted as evidence that "American display rules permit (encourage) greater expressiveness of emotion than more staid Canadian display rules" (p. 118).

Matsumoto and Kupperbusch (2001) examined the behavior of 40 European American female undergraduates as they viewed positive and negative film clips either alone or in the presence of an experimenter. Personality assessments were used to split participants into groups with primarily individualist or collectivist personality traits. Despite reporting the same intensity of felt emotions, participants from these two groups differed in their expressive behavior. When in the presence of an experimenter, the "collectivist" group showed more dampening of their expressions than the "individualist" group. This dampening occurred during both films, i.e., the collectivist group expressed less positive affect during the positive film and less negative affect during the negative film.

Matsumoto and Willingham (2006) examined the facial behavior of 67 Olympic athletes immediately after winning or losing a match, when receiving their medal, and when posing on the podium. This clever design allowed them to investigate the influence of strong emotion, social context, and culture on facial behavior. Due to a small number of athletes from each country, they compared groups of athletes from three regions that have been found to differ in terms of culture: North America and Western Europe, East Asia, and other areas. Significant differences in behavior between these regions were not apparent, which the authors interpreted as evidence that the influence of culture on facial behavior, while potentially important, can be eclipsed by strong emotion and context effects or obfuscated by grouping large

numbers of countries together by region.

Matsumoto, Willingham, and Olide (2009) returned to this dataset and examined how the facial behavior of 84 Olympic athletes was influenced by their countries' demographic characteristics and cultural dimensions. They found that athletes were more expressive if they came from countries with higher affluence, population density, and individualism. With participants spanning 35 different countries, this study boasts more cultural diversity than any other cross-cultural observational study to date. However, with an average of only 2.4 participants per country, this impressive breadth of coverage came at the cost of depth.

Additional observational studies focused on infants and preschoolers, finding that Western (e.g., American) children were more facially expressive than Asian (e.g., Chinese) children in general (Camras, Bakeman, Chen, Norris, & Cain, 2006; Camras et al., 1998; Kagan et al., 1994; Kisilevsky et al., 1998). These results suggest that cultural differences in facial expressivity begin quite early in life and are related to both temperamental differences and family environmental factors. Interestingly, these studies with children tended to include more participants than the previous studies with adults. Camras et al. (2006) included 163 children from two different countries, and Kagan et al. (1994) included an impressive 433 children from three different countries. However, it is still unknown how well these effects replicate in large adult samples and across a wider range of regions.

By including relatively few total participants from just a handful of regions, the size and diversity of previous observational studies' samples have been limited. While it is well-known that studies with smaller samples have reduced "power" to detect true effects, it is less well-appreciated that the statistically significant results identified by such studies are also less likely to reflect true effects (Button et al., 2013). It is thus of critical importance to replicate previously identified results in a larger sample. It is also necessary to examine cross-cultural differences in facial behavior during more "everyday" settings than TV game shows and Olympic awards ceremonies. Everyday settings are less likely to be influenced by strong emotion and context effects that could mask cross-cultural differences and, by virtue of being more common, may be the settings in which cross-cultural differences in nonverbal behavior have the most impactful interpersonal consequences.

The Current Study

The current study aimed to address the limitations of previous work by including a very large number of participants from multiple countries and observing facial behavior during the "everyday" context of watching video advertisements. We collected videos of spontaneous facial behavior from nearly three-quarters of a million participants in twelve countries and two experimental settings (i.e., at home and in a lab

facility). We then developed algorithms to measure the activity of two important facial actions (Figure 1). “Smiles” were defined as FACS action unit 12, which pulls the lip corners toward the ears, and “brow furrows” were defined as FACS action unit 4, which pulls the eyebrows down and together to form vertical wrinkles on the inner brow. Action unit 12 corresponds to contraction of the *zygomaticus major* muscle and action unit 4 corresponds to contraction of the *corrugator supercilii*. These actions have been associated with positive and negative emotional valence, respectively, and have a long history of study using electromyography (e.g., Cacioppo, Petty, Losch, & Kim, 1986; Schwartz, Ahern, & Brown, 1979).

Based on theory and the results of previous studies, we make the following hypotheses about the influence of culture, setting, and gender on participants’ facial behavior. For the sake of clarity, separate sub-hypotheses are enumerated for smiling and brow furrowing.

- *Hypothesis 1:* Across settings and genders, participants from more individualist cultures will be more expressive than participants from more collectivist cultures. (a) Participants from more individualist cultures will show more smiling than participants from more collectivist cultures. (b) Participants from more individualist cultures will show more brow furrowing than participants from more collectivist cultures.

- *Hypothesis 2:* Across cultures and genders, participants will be more expressive in the facility (out-group) setting than in the home (in-group) setting. (a) Participants will show more smiling in the facility setting than in the home setting. (b) Participants will show more brow furrowing in the facility setting than in the home setting.

- *Hypothesis 3:* Across cultures and settings, female participants will be more expressive than male participants. (a) Female participants will show more smiling than male participants. (b) Female participants will show more brow furrowing than male participants.

- *Hypothesis 4:* Culture and setting will interact such that participants from more individualist cultures will express more positive emotion and less negative emotion in the facility setting and more negative emotion and less positive emotion in the home setting. (a) Smiling will be positively associated with individualism in the facility setting but negatively associated with individualism in the home setting. (b) Brow furrowing will be negatively associated with individualism in the facility setting but positively associated with individualism in the home setting.

- *Hypothesis 5:* Sex differences in facial behavior will be more pronounced in more individualist cultures than in more collectivist cultures. (a) The difference in smiling between female and male participants will be larger in more individualist cultures than in more collectivist cultures. (b) The difference in brow furrowing between female and male participants will be larger in more individualist cultures than in more collectivist cultures.

- *Hypothesis 6:* Sex differences in facial behavior will be more pronounced in the facility setting than in the home setting. (a) The difference in smiling between female and male participants will be larger in the facility setting than in the home setting. (b) The difference in brow furrowing between female and male participants will be larger in the facility setting than in the home setting.

Method

Data Collection

Traditional methods for collecting observational data offer researchers a great deal of control over the experimental context, but suffer from limited scalability. Typically, participants are brought into lab settings where experimenters can record information about their behavior. Alternatively, experimenters may travel to participants’ homes to achieve greater ecological validity. However, staffing, travel, and financial constraints typically limit studies to fewer than 100 participants from one or two cities. In order to collect observational data from a large number of participants around the world, our study required a novel research paradigm with substantially higher scalability. We developed a web-based platform which walked participants through the consent process before recording video from their webcams as they viewed audiovisual stimuli (McDuff, 2014).

Participants were recruited as part of market research panels and represented a wide demographic of nationalities, genders (61% female), age groups (17–65 years old), and income levels. Some participants took part in the study from their own homes, while others took part at market research facilities. Those who took part at home received an email containing a link to a webpage where they could opt-in to the study (see McDuff, 2014, for more information). Those that took part in a facility followed a similar protocol, but were in an unfamiliar setting with experimenters and other participants. In the facility setting, participants were greeted by the experimenter and then used a computer in the facility to complete the experiment. Those that took part from their own homes may have been alone or in the presence of people they knew (e.g., family or friends); informal analyses suggest that, in the majority of cases, no other people were visible on the webcam’s video recording.

At the start of each session, participants were asked for their consent to participate in the study. They were made aware that, if they chose to participate, a webcam would be used to record their facial behavior. Following this, participants viewed between one and ten randomly-ordered television ads. While they watched, videos of participants’ facial behavior were recorded at a frame rate of 14 frames per second. All videos were recorded at the same resolution of 320×240 pixels. The participants were instructed to turn lights on if the room was dark and to position themselves in

Table 1

Information on included countries' cultural dimension index and data collection

Country	Stimuli	Individualism		Home Setting		Facility Setting	
		Rank*	Index	Participants	% Female [†]	Participants	% Female [†]
United States	2,598	1	91	177,534	60	663	68
United Kingdom	688	3	89	50,323	59	674	50
Germany	369	14	67	21,386	59	1,016	47
India	1,224	23	48	279	27	206,066	78
Japan	96	24	46	3,286	86	2,801	66
Argentina	158	24	46	5,588	70	10,602	73
Russia	180	28	39	1,659	47	10,806	72
Brazil	382	29	38	6,216	62	41,670	83
Mexico	335	41	30	20,519	64	9,025	65
China	754	50	20	3,112	53	145,216	74
Peru	78	59	16	2,219	36	9,191	31
Colombia	111	63	13	1,450	53	9,684	60
Total	6,973	–	–	293,570	60	447,414	74

* The individualism rank is within the 67 countries characterized in Hofstede (2001)

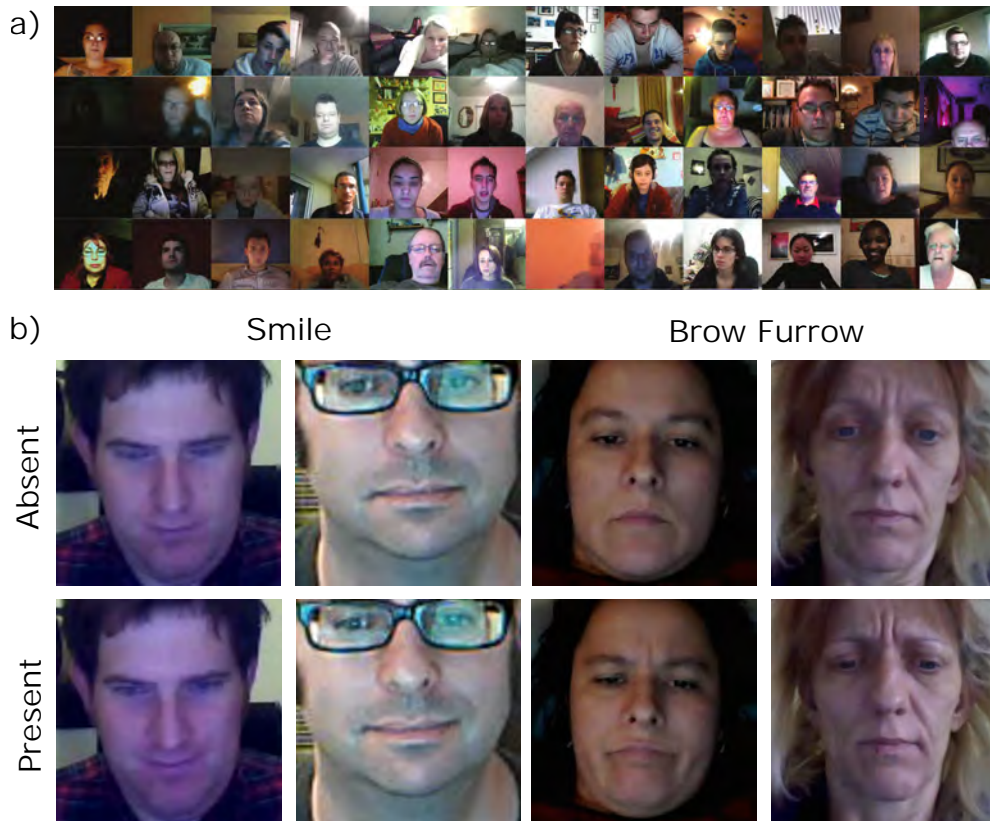
[†] The gender percentages are only based on participants who provided gender labels

Figure 1. a) Example frames from the videos collected. b) Cropped example images of the facial actions; note that other facial actions may also be present in these examples.

front of the webcam.

The recorded videos were stored on a secure server and later processed with automated facial coding algorithms. Example images shown in this paper are from participants who provided further consent to have images from their videos published. Participants were monetarily compensated at a local rate similar to what a typical 30 minute online market research survey would pay (i.e., approximately \$8 USD).

Over a period of four years, we collected 1,450,559 videos of facial responses from 740,984 participants. These numbers are excluding additional videos during which a face could not be detected (8.7%). However, in the vast majority of videos (90%), a face was tracked for more than 80% of frames, and for many videos (79%), a face was tracked for all frames. Participants were from countries that spanned the range from highly individualist to highly collectivist. Table 1 shows the number of unique participants, gender ratios, and individualism indexes for each included country.

Advertisements were selected from typical television content in each country and represented a wide cross-section of content. Using country-specific stimuli conferred several advantages over the presentation of identical media to all participants, including that it allowed us to circumvent language barriers (i.e., all participants watched videos in their countries' native languages) and to avoid confusion caused by culture-specific references. By using a large number of ads, the idiosyncratic influence of any single ad was minimized. To minimize the impact of participant age, we showed each ad to a range of age groups.

Most of the ads were aimed at the general population and were from recognizable brands (e.g., Mars, Coca-Cola, and Pepsi). A total of 6,973 different ad stimuli were used (Table 1). In all countries, ads from the following categories were selected: beverages (e.g., soda and alcohol), groceries (e.g., snacks, candy, and ice cream), personal care, home goods, services (e.g., telecoms and banking), and retail (e.g., apparel). The distribution of ad categories was similar across countries. The duration of the ads ranged from 10 to 155 seconds ($M = 45.9$, $SD = 46.6$). Most ads aimed to elicit positive emotion such as mirth or excitement. However, a range of emotions were represented including surprise and sadness.

Automated Facial Coding

To circumvent the monetary and temporal cost of manually coding facial behavior in all of the participant videos, automated algorithms were trained using a process called *supervised learning* (see McDuff, 2014). In this process, an algorithm learns from a subset of data (i.e., video frames) for which human-verified labels exist. To generate such labels, a group of 20 coders was trained using material from the FACS manual (Ekman et al., 2002). A minimum of three coders labeled each video frame for the presence or absence of smiles

and brow furrows. A fourth coder checked the agreement and quality of the labels before they were approved. This quality check was always performed by a coder who had been certified by passing the official FACS Final Test. If a label failed this quality check, the video would be relabeled by the original coders. The reliability of the manual coding was measured in terms of video-level base rates¹ using the intraclass correlation coefficient (ICC, formula [A,1]; McGraw & Wong, 1996) and in terms of frame-level agreement using the free-marginal kappa coefficient (κ_n ; Brennan & Prediger, 1981). Video-level reliability for the manual coding was $ICC = .91$ for smiles and $ICC = .91$ for brow furrows, while frame-level reliability was $\kappa_n = .81$ for smiles and $\kappa_n = .84$ for brow furrows.

A total of 80,000 video frames were used for training each of the automated classifiers. Quantitative measurements of these images (i.e., *features*) were extracted using computer vision techniques and fed to an algorithm along with the human-verified FACS labels. The algorithm then learned a high-dimensional mapping between the features and labels, which was used to generate predicted labels for novel video frames. Previous work has found that algorithms trained in this way can be highly successful at detecting spontaneous facial actions (Girard, Cohn, Jeni, Sayette, & De la Torre, 2015). Furthermore, these specific algorithms have been shown to perform well on a challenging benchmark dataset (Senechal, McDuff, & Kaliouby, 2015). Reliability analyses were used to assess how successful our algorithms were at detecting smiles and brow furrows in the current dataset, as well as how performance was impacted by area (i.e., west vs. east), setting, and gender. Such in-depth evaluation is rarely performed; however, identifying the operational parameters of a system is crucial when applying it to a heterogeneous data set (Girard & Cohn, in press).

Table 2 shows the automated facial coding algorithms' performance at detecting smiles and brow furrows, as well as how performance was impacted by various factors. These reliability analyses compare the predictions of our algorithms to the human-verified FACS labels. Because average base rates are the current study's dependent variables, the reliability (i.e., the ICC values) of these video-level predictions are the most important to consider. As a more challenging and aspirational evaluation of our algorithms, we also calculated and report the reliability of the frame-level predictions in terms of κ_n values.

As shown in Table 2, video-level reliability was $ICC = .82$ for smiles and $ICC = .65$ for brow furrows. Frame-level reliability was $\kappa_n = .74$ for smiles and $\kappa_n = .79$ for brow furrows. For smiles, video-level reliability was influenced by area, setting, and gender. Reliability was higher for participants from western countries than for participants from

¹The base rate is the proportion of time that a participant displays a facial action.

Table 2
Reliability of our facial action unit detection algorithms across different testing subsets

	Test Set Size		Smile Reliability		Brow Furrow Reliability	
	Videos	Frames	ICC	κ_n	ICC	κ_n
Full Set	1,571	1,100,000	.82	.74	.65	.79
By Area						
West	1,042	700,000	.82	.69	.66	.74
East	529	400,000	.69	.84	.53	.88
By Setting						
Facility	1,333	900,000	.83	.71	.67	.79
Home	238	200,000	.63	.85	.50	.76
By Gender						
Male	708	500,000	.90	.80	.64	.80
Female	863	600,000	.78	.69	.65	.78

Note: ICC = Intraclass correlation coefficient (for base rates); κ_n = Free marginal kappa coefficient (for video frames).

eastern countries ($p < .05$), was higher in the facility setting than in the home setting ($p < .05$), and was higher for male participants than for female participants ($p < .05$). For brow furrows, video-level reliability was influenced by area and by setting but not by gender. Reliability was higher for participants from western countries than for participants from eastern countries ($p < .05$) and it was higher in the facility setting than in the home setting ($p < .05$); reliability was not significantly different for male and female participants.

Individualism-Collectivism Indexes

To analyze the impact of culture on participants' facial behavior, we used Hofstede's (2001) individualism indexes as continuous values. These scores quantify each country's position on the individualism-collectivism dimension and are based on a long-term study of 67 different countries. These indexes have been used in a number of related studies (Matsumoto et al., 2008; Rychlowska et al., 2015) and cover a wide range of scores. Of the included countries, the United States (91) and the United Kingdom (89) had the highest individualism indexes, while Colombia (13) and Peru (16) had the lowest indexes.

Statistical Analysis

Data were analyzed using Mplus version 7.3 (Muthén & Muthén, 2012). We used multilevel models to account for the fact that participants from the same country were likely to be more similar than participants from different countries (Heck & Thomas, 2015). A separate model was constructed to predict each facial action. Setting and gender were added to the first level as uncentered effect codes, and the setting-by-gender interaction was added to this level as well (Equation 1). Each country's individualism index was added to the second level as an uncentered z -score (Equation 2). Because we had no hypothesis for it, the three-way interaction

of setting-by-gender-by-individualism was excluded in order to simplify the models and reduce the number of comparisons. In the following equations, the predicted facial action base rate for participant i from country j is a linear combination of fixed effects γ , random participant effects r , and random country effects u .

$$BASERATE_{ij} = \beta_{0j} + \beta_{1j}(SETTING_{ij}) + \beta_{2j}(GENDER_{ij}) + \beta_{3j}(SETTING_{ij})(GENDER_{ij}) + r_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(INDIVIDUALISM_j) + u_{0j} \quad (2a)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(INDIVIDUALISM_j) + u_{1j} \quad (2b)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(INDIVIDUALISM_j) + u_{2j} \quad (2c)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (2d)$$

Due to the design of the multiple market research projects contributing data to the current study, some participants were not asked to provide information about their gender (37.6%). Whether gender data was collected from a given participant was not dependent on the gender of that participant, their facial behavior, or their country of origin; it depended entirely on which project they were randomly assigned to. Thus, it is reasonable to assume that this data is missing at random (MAR) if not missing completely at random (MCAR). We used a state-of-the-art multiple imputation method to "fill in" the missing data rather than case deletion as the latter approach can lead to biased estimates when the MCAR assumption is violated and loss of power when it is not (Schafer & Graham, 2002). The Gibbs sampler (Casella & George, 1992), as implemented by Mplus, was used to generate 50 data sets with binary gender imputations. The imputation analysis included all of the variables that appeared in subsequent analyses. The imputed data sets were then provided

to Mplus, which performed the multilevel analyses on each data set using maximum likelihood estimation with robust standard errors. The parameter estimates and standard errors from each imputed data set were then combined into a single set of results.

Results

Multilevel regression models were built to estimate the impact of culture, setting, and gender on participants' facial behavior. Table 3 summarizes the regression estimates for each model. The principal intercepts (γ_{00}) shows that an average participant from a country with an average individualism score had a base rate of 5.62% for smiles and 2.84% for brow furrows. This means that, while watching the television advertisements, participants spent nearly twice as much time smiling as they did brow furrowing.

The main effect of individualism was statistically significant for brow furrowing ($\gamma_{01} = 0.56$, $p = .004$) but not for smiling ($\gamma_{01} = 0.86$, $p = .140$). Figure 2(a) shows that, controlling for setting and gender, participants from more individualist countries displayed more brow furrowing. The main effect of setting was also statistically significant for brow furrowing ($\gamma_{10} = -0.44$, $p = .021$) but not for smiling ($\gamma_{10} = 0.08$, $p = .821$). Figure 2(b) shows that, controlling for culture and gender, participants displayed more brow furrowing in the facility setting than in the home setting. The main effect of gender was statistically significant for both smiles ($\gamma_{20} = 0.73$, $p < .001$) and brow furrows ($\gamma_{20} = -0.41$, $p < .001$). Figure 2(c) shows that, controlling for culture and setting, female participants displayed more smiling and less brow furrowing than male participants.

The cross-level interaction between individualism and setting was statistically significant for smiling ($\gamma_{11} = -1.63$, $p < .001$) but not for brow furrowing ($\gamma_{11} = -0.10$, $p = .543$). Figure 2(d) that, controlling for gender, participants in more individualist cultures displayed more smiling in the facility setting, while participants in more collectivist cultures displayed more smiling at home. The cross-level interaction between individualism and gender was statistically significant for brow furrowing ($\gamma_{21} = -0.16$, $p = .049$) but not for smiling ($\gamma_{21} = -0.08$, $p = .450$). Figure 2(e) shows that the gender difference for brow furrowing was more pronounced in more individualist countries and less pronounced in more collectivist countries. The level-1 interaction between setting and gender was not statistically significant for smiling ($\gamma_{30} = -0.12$, $p = .360$) or for brow furrowing ($\gamma_{30} = 0.04$, $p = .591$) as depicted in Figure 2(f).

Discussion

Observational studies of cross-cultural differences in facial behavior have been rare and typically limited in terms of sample size, diversity, and social context. The results presented here are the first large-scale observational evidence

of cross-cultural differences in facial behavior. The current sample is orders of magnitude larger than any previous study's and includes nearly 750,000 participants from 12 countries around the world.

Hypothesis 1 was that participants from more individualist cultures would be more expressive than participants from more collectivist cultures. This hypothesis was partially supported by the results. In line with *hypothesis 1(b)*, there was a significant positive association between individualism and brow furrowing. However, *hypothesis 1(a)* was not supported in that the association between individualism and smiling was not significant. Thus, we can conclude that participants from more individualist cultures were more expressive than participants from more collectivist cultures in some ways but not in others. If we follow the example of previous research and interpret smiling as an expression of positive emotion and brow furrowing as an expression of negative emotion, then this result would suggest that more individualist cultures express negative emotion more freely than more collectivist cultures. This interpretation is consistent with the idea that self-esteem in individualist cultures is based on the ability to express the self and differentiate from others, while self-esteem in collectivist cultures is based on the ability to adjust and restrain the self in order to maintain social harmony (Markus & Kitayama, 1991).

The absence of a significant main effect of individualism on smiling is inconsistent with the results of previous studies. For instance, Matsumoto et al. (2008) found that individualism was positively associated with higher self-reported expressivity for positive emotions in particular. This divergence between self-report and observational measurement of positive emotion expression may indicate that participants are poor judges of their own behavior, perhaps due to a desire to present themselves as more positive/desirable than they really are (Gosling, John, Craik, & Robins, 1998). Matsumoto et al. (2009) found that individualism was positively associated with higher observed expressivity among Olympic athletes. However, they did not distinguish between expressions of positive and negative emotions so it is possible that this finding was driven by a difference in negative emotion expressivity alone. Alternatively, our null finding may be due to power issues; although the current study included a very large number of participants and more countries than most previous studies, there were still only 12 countries.

Hypothesis 2 was that participants would be more expressive in the facility setting than in the home setting. This hypothesis was partially supported by the results. In line with *hypothesis 2(b)*, participants showed significantly more brow furrowing in the facility setting than in the home setting. However, *hypothesis 2(a)* was not supported in that there was not a significant difference in smiling between the two settings. Thus, participants were more expressive in the facility setting than in the home setting in some ways but not in

Table 3

Multilevel regression estimates predicting each facial action's base rate percentage

	Smile (AU 12)			Brow Furrow (AU 4)		
	Estimate	SE	<i>p</i>	Estimate	SE	<i>p</i>
Principal intercepts						
Intercept (γ_{00})	5.62	0.49	<.001	2.84	0.30	<.001
Individualism (γ_{01})	0.86	0.58	.140	0.56	0.20	.004
Setting slopes						
Intercept (γ_{10})	0.08	0.36	.821	−0.44	0.19	.021
Individualism (γ_{11})	−1.63	0.47	<.001	−0.10	0.17	.543
Gender slopes						
Intercept (γ_{20})	0.73	0.12	<.001	−0.41	0.10	<.001
Individualism (γ_{21})	−0.08	0.11	.450	−0.16	0.08	.049
Setting×Gender slopes						
Intercept (γ_{30})	−0.12	0.13	.360	0.04	0.08	.591
Random components						
Level 1 (r_{ij}) Variance	144.94	28.65	<.001	119.05	21.26	<.001
Level 2 (u_0) Variance	2.87	0.94	.002	1.04	0.60	.081
Setting (u_1) Variance	1.40	0.76	.065	0.39	0.18	.030
Gender (u_2) Variance	0.12	0.08	.117	0.09	0.05	.096
Setting×Gender (u_3) Variance	0.13	0.05	.019	0.04	0.03	.160

Note: Setting: {Facility = −1, Home = 1}; Gender: {Male = −1, Female = 1}; Individualism: z-scores.

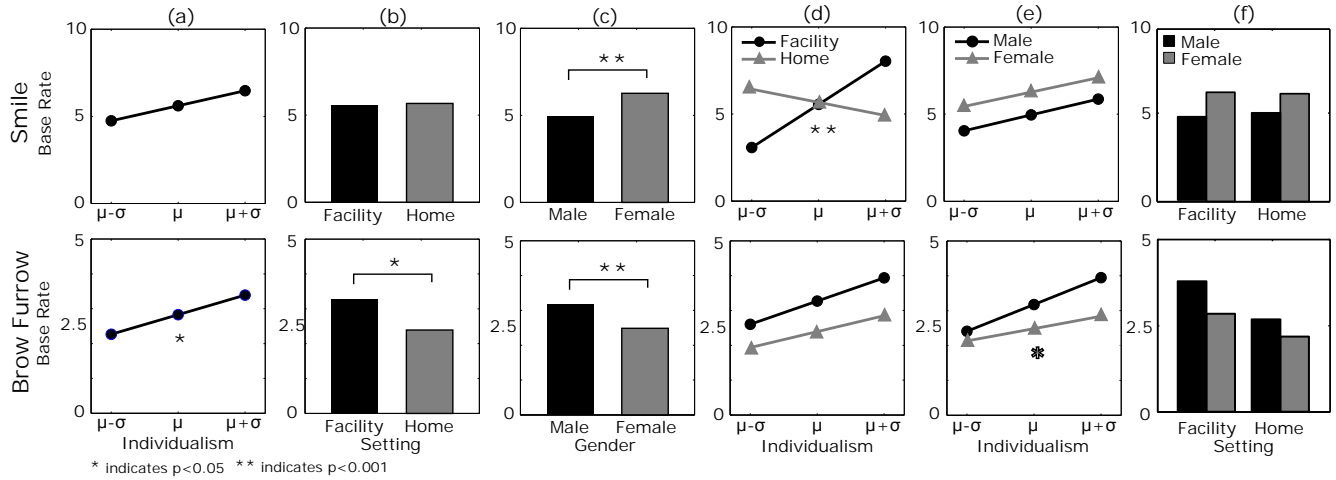


Figure 2. Multilevel regression estimates for the (a) culture, (b) setting, and (c) gender main effects. Estimates for the (d) setting by culture, (e) gender by culture, and (f) setting by gender interactions. For cultural effects, estimates for the mean individualism score and one standard deviation above and below the mean individualism score are shown.

others. If we assume that the home setting was characterized by more intimacy, familiarity, and trust than the facility setting, then we can interpret the facility and home settings as representing out-group and in-group contexts, respectively. Thus, this result may suggest a pan-cultural increase in expression of negative emotion in out-group contexts. Lowering the threshold for displaying negative affect in unfamiliar settings and around strangers may be evolutionarily adaptive

for all cultures as a way of protecting boundaries in settings with a higher likelihood of interpersonal transgression. The absence of a significant difference in smiling between settings is not consistent with previous research on “audience effects” (see Fridlund, 1991); however, much of this research used individualist samples and thus may have confounded setting main effects and culture-by-setting interactions.

Hypothesis 3 was that female participants would be more

expressive than male participants. This hypothesis was partially supported and partially contradicted by the results. In line with *hypothesis 3(a)*, and the meta-analysis of LaFrance et al. (2003), female participants showed significantly more smiling than male participants. However, contrary to *hypothesis 3(b)*, female participants showed significantly less brow furrowing than male participants. Thus, it would appear that the influence of gender on expressivity is more nuanced than a global increase in female expressivity; female participants were more expressive than male participants in some ways, but they were also less expressive in other ways. Specifically, assuming a valence-based interpretation of smiling and brow furrowing, female participants expressed more positive emotion and less negative emotion than male participants. These results are consistent with the arguments that women tend to reverse or mitigate negative emotion through displays of positive emotion (Frith & Kitzinger, 1998; Hall & Halberstadt, 1986) and that those with higher power (e.g., men) can smile or not, but those with lower power (e.g., women) are obligated to do so (Hecht & LaFrance, 1998). The results are not consistent with argument that men suppress their overall expressivity in order to maintain and protect their status (Fischer & Manstead, 2000), although a related interpretation might be proposed that men express more negative emotion in order to secure and demonstrate their status.

Hypothesis 4 was that culture and setting would interact such that participants from more individualist cultures would express more positive emotion and less negative emotion in the facility (i.e., out-group) setting and more negative emotion and less positive emotion in the home (i.e., in-group) setting. This hypothesis was partially supported by the results. In line with *hypothesis 4(a)*, the individualism-by-setting interaction was significant such that individualism was positively associated with smiling in the facility setting and negatively associated with smiling in the home setting. Figure 2(d) shows that participants from more collectivist cultures smiled more in the home setting than the facility setting, whereas participants from more individualist cultures smiled more in the facility setting than in the home setting. However, *hypothesis 4(b)* was not supported in that there was not a significant culture-by-setting interaction for brow furrowing. Thus, culture moderated the impact of setting on some facial actions but not others. Specifically, the amount that participants smiled in different settings varied greatly across cultures. These results are consistent with the argument that expressions of positive emotion are increased in the settings that a culture most relies on for success: out-group interactions for more individualist cultures and in-group interactions for more collectivist cultures. Given that there were significant main effects of culture and setting on negative emotion, the absence of a significant interaction effect may suggest that negative emotion has the same role in different cultures and is not differentially modulated based on context.

Hypothesis 5 was that sex differences in facial behavior would be more pronounced in more individualist cultures. This hypothesis was partially supported by the results. In line with *hypothesis 5(b)*, the difference in brow furrowing between female and male participants was larger in more individualist cultures and smaller in more collectivist cultures. However, *hypothesis 5(a)* was not supported in that the difference in smiling between male and female participants did not vary as a function of individualism. Thus, sex differences in expressivity were more pronounced in more individualist cultures for some facial actions but not others. Specifically, assuming a valence-based interpretation of smiling and brow furrowing, it was the sex difference in negative emotion that varied by culture. While negative emotion was expressed more commonly by male participants than by female participants overall, this sex difference was larger in more individualist cultures. If expressions of negative emotion are used to differentiate and dominate, as suggested earlier, then this interaction may be explained by the added emphasis that both males and individualist cultures place on differentiation and dominance. The lack of a significant interaction for smiling may suggest that gender norms related to the expression of positive emotion are more universal than those related to the expression of negative emotion.

Hypothesis 6 was that sex differences in facial behavior would be more pronounced in the facility (i.e., out-group) setting than in the home (i.e., in-group) setting. This hypothesis was not supported by the results. *Hypothesis 6(a)* was not supported in that the difference in smiling between male and female participants did not vary between settings. Similarly, *hypothesis 6(b)* was not supported in that the difference in brow furrowing between male and female participants did not vary between settings. The absence of a significant gender-by-setting interaction is inconsistent with previous findings that sex differences in smiling were larger when participants were in a laboratory context (LaFrance et al., 2003). Perhaps these null findings are due to the current study's experimental task (i.e., watching advertisements), which may not have activated audience effects and gender norms as much as the explicitly social tasks used in other studies.

Conclusions

The current study is the first large-scale observational investigation of cross-cultural differences in facial behavior, including a sample of 740,984 participants from 12 countries. Furthermore, we analyze facial responses in an "everyday" context (i.e., watching ads) in order to prevent cultural differences from being eclipsed by strong emotion and context effects. We present findings pertaining to the influence of culture, setting, and gender on actual facial behavior and provide convergent support for some, but not all, hypotheses derived from previous studies using self-report and small-scale

observational methodologies.

Results suggest that culture, setting, and gender each independently influence facial behavior. Participants from more individualist cultures expressed more negative emotion than participants from more collectivist cultures; participants in the out-group setting expressed more negative emotion than participants in the in-group setting; and females expressed more positive emotion and less negative emotion than males.

Additionally, results suggest that culture moderates the associations between setting and positive emotion and between gender and negative emotion. First, participants from more individualist cultures expressed more positive emotion in the out-group setting, while participants from more collectivist cultures expressed more positive emotion in the in-group setting. Second, the sex difference in expressions of negative emotion was larger in more individualist cultures and smaller in more collectivist cultures.

Significant main effects of culture and setting on positive emotion are notably absent, as these effects have been found in previous studies. We suggest that the large “cross-over” interaction effect between culture and setting explains these null findings and provides a more accurate picture than previous studies that only included a single culture or setting.

Limitations and Future Directions

The repurposing of market research data and the application of recent advances in computer science allowed us to examine important psychological questions on an unprecedented scale. However, these innovations also came with limitations.

First, although the number of participants per country was very high, the number of included countries was still modest from a statistical perspective. As such, the tests of cultural effects were likely underpowered and future work on cross-cultural differences should prioritize including participants from as many different countries as possible.

Second, although smiling and brow furrowing have a long history of use as indicators of positive and negative emotion, respectively, these relationships are not perfect one-to-one correspondences (Cacioppo & Tassinary, 1990). People smile in a variety of circumstances including some which seem unrelated to positive affect (e.g., pain, distress, and embarrassment; Ansfield, 2007). This has led some researchers to argue that smiles communicate an intention to affiliate as opposed to (or in addition to) expressing positive emotion (e.g., Hess, Adams Jr, & Kleck, 2005; Rychlowska et al., 2015). There is also some controversy over what brow furrows mean. Oster (1978) argued that “brow knitting” in infants is a signal of concentration and it has become common for researchers to interpret brow furrowing as indicative of mental effort as well. However, it is not clear that brow knitting in infants is the same expression as brow furrowing in adults. In this study, we adopt the view of Ekman (1979) that

brow furrowing expresses negative emotion (i.e., anger, fear, or sadness) or mental difficulty (e.g., uncertainty, perplexity, or doubt).

Third, our measures of facial actions were necessarily basic. To get reliable measures from the current generation of algorithms (especially in the challenging home setting), we used action unit base rates, which quantify how much time an action was present within each video. Our measures do not capture more nuanced characterizations of behavior, such as the combination of multiple facial actions, the number of discrete actions, or the intensity of such actions. We hope that future work can explore this type of data once automated algorithms advance to the level of being able to collect such measures reliably. Of particular interest would be the combination of smiling with the cheek raiser (action unit 6), which has been found to be helpful in discriminating amused and polite smiles (Ekman, Davidson, & Friesen, 1990), and the dimpler (action unit 14), which has been linked with states of contempt and depression (e.g., Girard et al., 2014).

Fourth, while our inclusion of so many different ad stimuli minimized the idiosyncratic influence of any single ad, it also made it difficult to tell exactly what participants were responding to. Although the distribution of product categories was similar across countries, and informal analyses suggest that most ads aimed to elicit mild amusement or excitement, it is possible that different cultures produce ads with different affective features or goals. Future work on a smaller scale would benefit from collecting self-reports from participants about the type and intensity of emotions that were elicited by such stimuli.

Fifth, our interpretations of the effects related to setting hinge on the assumption that participants experienced the home setting as a more in-group context and the facility setting as a more out-group context. The presence of significantly different behavior in these two settings—despite identical ad stimuli in both—suggests that there was something importantly different about the social context of these settings. We argue that the classic social psychology division of in-group/out-group contexts can be usefully (and reasonably) applied here, but we acknowledge that additional information (e.g., self-reports) would be necessary to substantiate any such claim. Future work would benefit from approaching this issue more directly than we were able to with the current data.

Sixth, characterizing the similarities and differences between cultures is a challenging task. In this work, we used Hofstede’s (2001) indexes of individualism-collectivism because they are well-established in the literature, provide a complete listing of scores for the countries included in the current study, and are highly correlated with other accepted measures of individualism-collectivism (Rychlowska et al., 2015). However, other measures may have advantages over Hofstede’s. For instance, Singelis (1994) proposed that indi-

vidualism and collectivism can be usefully modeled as separate unipolar dimensions. Future work would benefit from exploring if other measures of culture (e.g., unipolar individualism, unipolar collectivism, or power distance) also predict facial behavior.

Finally, due to the design of the current study, it was not possible to collect much information about each participant. However, future work on a smaller scale should include measures of each participant's traits and values, as there is doubtlessly a great deal of heterogeneity within each country. Such measures would also allow researchers to explore the influence of these factors on nonverbal behavior both within and across cultures.

References

- Ansfield, M. E. (2007). Smiling when distressed: When a smile is a frown turned upside down. *Personality and Social Psychology Bulletin*, 33(6), 763–775.
- Argyle, M., Henderson, M., Bond, M., Iizuka, Y., & Contarello, A. (1986). Cross-cultural variations in relationship rules. *International Journal of Psychology*, 21(1-4), 287–315.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
- Brewer, M. B., & Kramer, R. M. (1985). The psychology of intergroup attitudes and behavior. *Annual Review of Psychology*, 36(1), 219–243.
- Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 395–408). New York, NY: The Guilford Press.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., & Robinson, E. S. J. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.
- Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, 50(2), 260.
- Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *The American Psychologist*, 45(1), 16–28.
- Camras, L. A., Bakeman, R., Chen, Y., Norris, K., & Cain, T. R. (2006). Culture, ethnicity, and children's facial expressions: A study of European American, Mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion*, 6(1), 103–114.
- Camras, L. A., Oster, H., Campos, J., Campos, R., Ujiie, T., Miyake, K., . . . Meng, Z. (1998). Production of emotional facial expressions in European American, Japanese, and Chinese infants. *Developmental Psychology*, 34(4), 616–628.
- Casella, G., & George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3), 167–174.
- Cohn, J. F., & Ekman, P. (2005). Measuring facial action by manual coding, facial EMG, and automatic facial image analysis. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The new handbook of nonverbal behavior research* (pp. 9–64). New York, NY: Oxford University Press.
- Darwin, C. (1872). *The expression of emotions in man and animals* (3rd ed.). New York, NY: Oxford University.
- Ekman, P. (1977). Biological and cultural contributions to body and facial movement. In J. Blacking (Ed.), *The anthropology of the body*. London: Academic Press.
- Ekman, P. (1979). About brows: emotional and conversational signals. In M. V. Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 169–202). Cambridge: Cambridge University Press.
- Ekman, P. (1982). Methods for measuring facial action. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 45–90). Cambridge: Cambridge University Press.
- Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, 58(2), 342–353.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17, 124–129.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings*. New York, NY: Pergamon Press.
- Ekman, P., Friesen, W. V., & Hager, J. (2002). *Facial action coding system: A technique for the measurement of facial movement*. Salt Lake City, UT: Research Nexus.
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., . . . Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712–717.
- Ekman, P., & Rosenberg, E. L. (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (2nd ed.). New York, NY: Oxford University Press.
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164(3875), 86–88.
- Evans, E. C. (1969). Physiognomies in the ancient world. *Transactions of the American Philosophical Society*, 59, 1–101.
- Fischer, A. H., & Manstead, A. S. R. (2000). The relation between gender and emotions in different cultures. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 71–94). New York, NY: Cambridge University Press.
- Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, 60(2), 229–240.
- Friesen, W. V. (1973). *Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules* (Doctoral dissertation). University of California San Francisco.
- Frith, H., & Kitzinger, C. (1998). "Emotion work" as a participant re-source: A feminist analysis of young women's talk-in-interaction. *Sociology*, 32, 299–320.
- Girard, J. M., & Cohn, J. F. (in press). A primer on observational measurement. *Assessment*, 1–10. doi:

- 10.1177/1073191116635807
- Girard, J. M., Cohn, J. F., Jeni, L. A., Sayette, M. A., & De la Torre, F. (2015). Spontaneous facial expression in unscripted social interactions can be measured automatically. *Behavior Research Methods*, 47(4), 1136–1147.
- Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., Hammal, Z., & Rosenwald, D. P. (2014). Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image and Vision Computing*, 32(10), 641–647.
- Gosling, S. D., John, O. E., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, 74(5), 1337–1349.
- Hall, J. A., & Halberstadt, A. G. (1986). Smiling and gazing. In J. S. Hyde & M. C. Inn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 136–185). Baltimore, MD: Johns Hopkins University Press.
- Hecht, M. A., & LaFrance, M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin*, 24(12), 1332–1342.
- Heck, R. H., & Thomas, S. L. (2015). *An introduction to multi-level modeling techniques: MLM and SEM approaches using Mplus* (3rd ed.). New York, NY: Routledge.
- Hess, U., Adams Jr, R. B., & Kleck, R. E. (2005). Who may frown and who should smile? Dominance, affiliation, and the display of happiness and anger. *Cognition and Emotion*, 19(4), 515–536.
- Hofstede, G. H. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage Publications.
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Sage Publications.
- Hofstede, G. H., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research*, 38(1), 52–88.
- Kagan, J., Arcus, D., Snidman, N., Feng, W. Y., Hendler, J., & Greene, S. (1994). Reactivity in infants: A cross-national comparison. *Developmental Psychology*, 30(3), 342–345.
- Kisilevsky, B. S., Hains, S. M. J., Lee, K., Muir, D. W., Xu, F., Fu, G., ... Yang, R. L. (1998). The still-face effect in Chinese and Canadian 3- to 6-month-old infants. *Developmental psychology*, 34(4), 629–639.
- Klineberg, O. (1940). *Social psychology*. New York, NY: Holt.
- LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin*, 129(2), 305–334.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253.
- Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion*, 14(3), 195–214.
- Matsumoto, D. (2006). Culture and nonverbal behavior. *Handbook of nonverbal communication*, 219–236.
- Matsumoto, D., & Kupperbusch, C. (2001). Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. *Asian Journal of Social Psychology*, 4(2), 113–131.
- Matsumoto, D., & Willingham, B. (2006). The thrill of victory and the agony of defeat: spontaneous expressions of medal winners of the 2004 athens olympic games. *Journal of personality and social psychology*, 91(3), 568.
- Matsumoto, D., Willingham, B., & Olide, A. (2009). Sequential dynamics and culturally-moderated facial expressions of emotion. *Psychological Science*, 20(10), 1269–1274.
- Matsumoto, D., Yoo, S. H., & Fontaine, J. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology*, 39(1), 55–74.
- McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In L. A. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). New York, NY: Guilford Press.
- McDuff, D. (2014). *Crowdsourcing affective responses for predicting media effectiveness* (Doctoral dissertation). Massachusetts Institute of Technology.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46.
- Messick, D. M., & Mackie, D. M. (1989). Intergroup relations. *Annual Review of Psychology*.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Oster, H. (1978). Facial expression and affect development. In M. Lewis & L. A. Rosenblum (Eds.), *The development of affect*. New York, NY: Plenum Press.
- Oster, H., Daily, L., & Goldenthal, P. (1989). Processing facial affect. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 101–161). Amsterdam: Elsevier.
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128(1), 3.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102.
- Rychlowska, M., Miyamoto, Y., Matsumoto, D., Hess, U., Gilboa-Schechtman, E., Kamble, S., ... Niedenthal, P. M. (2015). Heterogeneity of long-history migration explains cultural differences in reports of emotional expressivity and the functions of smiles. *Proceedings of the National Academy of Sciences*, 112(19), E2429–E2436.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177.
- Scherer, K. R., Wallbott, H. F., & Summerfield, A. B. (1986). *Experiencing emotion: A cross-cultural study*. New York, NY: Cambridge University Press.
- Schwartz, G. E., Ahern, G. L., & Brown, S. L. (1979). Lateralized facial muscle response to positive and negative emotional stimuli. *Psychophysiology*, 16(6), 561–571.
- Senechal, T., McDuff, D., & Kaliouby, R. (2015). Facial action unit detection using active learning and an efficient non-linear

- kernel approximation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 10–18).
- Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, 20(5), 580–591.
- Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology*, 74(2), 482–493.
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1–39.
- Waxer, P. H. (1985). Video ethology: Television as a data base for cross-cultural studies in nonverbal displays. *Journal of Nonverbal Behavior*, 9(2), 111–120.