**Bennett CC and S Sabanovic (2015) The effects of culture and context on perceptions of robotic facial expressions.**
***Interaction Studies*. In Press.**

# The Effects of Culture and Context on Perceptions of Robotic Facial Expressions

Casey C. Bennett[1,2] and Selma Šabanović[2]

[1]*School of Informatics and Computing*
*Indiana University*
*901 E. 10th. St.*
*Bloomington, IN, USA 47408*

[2]*Department of Informatics*
*Centerstone Research Institute*
*44 Vantage Way*
*Suite 280*
*Nashville, TN, USA 37228*

Corresponding Author:
*Casey Bennett*
*Department of Informatics*
*Centerstone Research Institute*
*365 South Park Ridge Road*
*Bloomington, IN 47401*
1.812.355.6382
cabennet@indiana.edu

## Abstract

We report two experimental studies of human perceptions of robotic facial expressions while systematically varying context effects and the cultural background of subjects (n=93). Except for Fear, East Asian and Western subjects were not significantly different in recognition rates, and, while Westerners were better at judging affect from mouth movement alone, East Asians were not any better at judging affect based on eye/brow movement alone. Moreover, context effects appeared capable of over-riding such cultural differences, most notably for Fear. The results seem to run counter to previous theories of cultural differences in facial expression based on emoticons and eye fixation patterns. We connect this to broader research in cognitive science – suggesting the findings support a dynamical systems view of social cognition as an emergent phenomenon. The results here suggest that, if we can induce appropriate context effects, it may be possible to create *culture-neutral models* of robots and affective interaction.

**Keywords:** *Human-Robot Interaction; Facial Expression; Emotion; Affective Communication; Robotic Face; Culture; Context*

# 1. Introduction

## 1.1 Background

Scientific inquiry stretching back over a century has contributed to an ongoing debate about the nature and classification of human emotions and their related facial expressions (e.g. Ekman, 2009; Nelson & Russell, 2013; Breazeal, 2003; Sosnowski et al., 2006; Pantic, 2009; Cohn, 2010). Even Charles Darwin played a role (Darwin, 1872). The main points of contention can be summarized as such: Does a basic set of universal human emotions (and their related facial expressions) exist across culture, gender, context, etc.? Moreover, are there universal facial cues associated with these expressions that we can distill out from the broader array of complex and/or idiosyncratic facial movements?

Research by Ekman and colleagues going back to the 1960's suggested that there was indeed such a basic set of universal human emotions and/or facial expressions (Ekman & Friesen, 2003; Ekman, 2009). This eventually led to the development of the Facial Action Coding System (FACS), which could be used to identify facial expressions via specific facial cues. These facial cues are referred to as *action units*, and intended to encode the movement of specific facial muscles. However, that research on the universality of emotions/expressions was challenged on multiple grounds based on the work of Russell (Russell & Fernandez-Dolz, 1997), Matsumoto (Matsumoto, 1992), and others from the 1980's onwards, using studies done with human images and confederates. Recent work over the last few years using digital avatars has further challenged the universality of basic "Ekman emotions" on the basis of variations due to culture, context, and age (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). However, that work, based heavily on visual fixation patterns, has been disputed by more recent research (see Section 1.2). In spite of these scientific controversies, sophisticated automated facial expression recognition technology has been developed over the last decade such that computers can, at least for posed Western expressions, achieve roughly 95% accuracy for identifying human facial expressions (Pantic, 2009; Cohn, 2010). Furthermore, most robotic faces with affective expression capabilities built over the last decade continue to be based on the basic Ekman emotions and their associated facial expressions (Breazeal, 2003; Sosnowski et al., 2006; Canamero & Fredslund, 2001; Bazo et al., 2010; Saldien et al., 2010; Miwa et al., 2004; Becker-Asano & Ishiguro, 2011; Bennett & Šabanović, 2013; Bennett & Šabanović, 2014). In short, the literature is full of conflicting evidence on the subject, suggesting a need for novel lines of evidence.

This paper is aimed at that need, contributing to the debate over human perception of affective facial expressions and to the application of such research in robot design through two experimental studies in which participants interacted with a previously validated minimalist face robot (MiRAE). MiRAE was designed with the aim of utilizing the minimal facial cues necessary to convey facial expressions in ways humans can perceive/understand (Bennett & Šabanović, 2013; Bennett & Šabanović, 2014). The first study investigated the effect of cultural differences in perceptions of robotic facial expressions, using three human-subject groups: Japanese (living in Japan), native East Asians (living in the United States), and Westerners (i.e. Americans). A second study evaluated the effects of context on those perceptions. Both experiments seek to understand how situational factors (e.g. context, culture) affect people's perceptions of affective facial expressions. These were part of a broader series of seven experiments involving nearly two-hundred-twenty human subjects, interacting in-person with the robot (Bennett & Šabanović, 2013; Bennett & Šabanović, 2014, Bennett et al., 2014). A novel contribution of this work is the simultaneous manipulation of participant culture and context together that allows us to analyze the effects of and interactions between both of these two factors on people's perceptions of a robot's affective expression.

This research is developed through reference to cognitive science and psychological theories, which suggest that our perceptions and modes of interaction are contextually dependent and dynamically constructed and biased by cues in our environment – culture, context, interaction partners, etc. (see Related Work and Discussion sections below). Emotions perceived in others' faces – including robots – may be an internal construct in the mind of the perceiver, based on a number of perceptual and cognitive processes.

## **1.2 Related Work**

Even if facial expressions of emotions are variable in humans, it is not precisely clear as to how or why. While certain aspects of emotional and cognitive development may be universal, researchers have shown that the specific ways in which people engage in affective interaction can vary across culturally-situated norms and context scenarios. For instance, Nisbett et al. (2001) suggested that different "cognitive styles" in Western and East Asian cultures define aspects of the environment that are worthy of attention (e.g. characteristics of the environment or of the individual) and acceptable communication patterns (e.g. implicit vs. explicit). Such cognitive differences between Western and East Asian subjects may indicate that the two groups vary in regard to their attention to the context of interaction as indicative of its affective valence. Similarly, Shore (1996) argued that “social-orientational models” in particular “provide a degree of standardization in emotional response within a community,” and designate appropriate roles/behaviors within interaction as well as culturally normative rules for displaying, perceiving, and experiencing affect (pp.62-63). Ekman, Friesen, and Izard themselves suggested a similar “Deception hypothesis” in the 1970’s to explain culturally-based affective expression encoding rules (Ekman, 1971). More recently, Elfenbein (2013) has proposed a “Dialect hypothesis” for affective communication, which posits isomorphisms between affective expressions and linguistic distributions/development.

In addition to such research with humans, researchers in recent years have also used digital avatars as stimuli for testing people’s perceptions of emotion. However, the evidence derived from these studies is subject to debate. For example, much of this recent work on cross-cultural differences is rooted in what we refer to as the “Emoticon hypothesis”. In short, this posits that since emoticons are different for specific features (e.g. eyes, mouth) between Western and Eastern/Asian styles, displays of emotions by humans between those groups must therefore be different across those features as well (for example, East Asians focus more on the eyes, and Westerners more on the mouth) (Yuki et al., 2007). Several papers in the last few years have studied visual fixation patterns as the basis for these putative cultural differences in facial expressions, arguing that the patterns support the Emoticon hypothesis (Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). However, more recent papers have provided evidence countering the use of visual fixation patterns, noting that people are engaged in a range of information-gathering activities for a variety of purposes (not simply judging affect) when looking at other faces, including determining culture, gender, confidence, sexual attraction, social referencing, etc. (Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012). Furthermore, interpretation of results from studies using digital avatars is complicated by their common use of cartoon-like facial representations that are sometimes difficult to clearly relate back the FACS and/or facial displays directly based on them.

Other recent empirical work has provided evidence for significant, culturally-variable effects due to context on facial expression recognition, using both digital avatars and human faces (Righart & de Gelder, 2008; Barrett et al., 2011; Lee et al. 2012). This literature has focused on the variable importance of context between cultures (particularly Asian and Western cultures) as an explanation for such cross-cultural differences, related to the arguments of Shore (1996) and

Nisbett et al. (2001) above. Researchers have also challenged the sole focus on facial expressions, suggesting body posture/gesture plays a significant role as well (Kleinsmith et al., 2006; de Gelder, 2009). A further complicating matter is the possible effect of variations in language and cultural connotations of emotion-label words (Perlovsky, 2009; Ruttkay, 2009). Lindquist & Gendron (2013) even suggest a dynamical systems perspective of emotion perception and word-label grounding to explain such variation.

To summarize, the debate over the universality of human emotions and facial expressions, as well as their mechanisms of display/interpretation, is complex and rife with conflicting evidence. As noted in Section 1.1, our research contributes to the production of new modes of evidence, via human-robot interaction, for systematically approaching this debate.

## 1.3 The Role of Human-Robot Interaction

There are multiple motivations for utilizing robotic faces to study the question of human display and perception of emotional expressions, both academic and pragmatic. On the one hand, robotic faces provide a three-dimensional, embodied platform that can be used as a controllable/consistent/modifiable surrogate for human images or confederates when investigating questions of human cognition and perception (Adams et al., 2000; Scasselati, 2006; Kozima et al., 2009). On the other hand, if we endeavor to add faces and facial expressions to robots in order to enhance human-robot interaction and communication, then understanding how to do so effectively is of immense importance. This is doubly true if robots are also meant to interpret human facial movements. If indeed factors like culture and context matter to human perception and performance of affective facial expressions, then future human-robot interaction design requires an empirically-based understanding of how and why.

Despite the aforementioned work using human images (Matsumoto, 1992; Russell & Fernandez-Dolz, 1997; etc.) and digital avatars (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; etc.) to investigate human facial expressions, as well as numerous papers evaluating the ability of robotic faces to display the basic Ekman emotions, limited research has been performed evaluating the purported *universality* of the Ekman-based facial expressions and facial cues using robotic faces. Becker-Asano and Ishiguro (2011) evaluated the android Geminoid-F robot across three cultural groups (Americans, Europeans, and Asians), showing clear differences across them. However, the study utilized only still, posed images of the robot distributed over the Internet, and even the Western subjects struggled to identify many of the expressions (e.g. Anger, Surprise) with high accuracy. Elsewhere, Zhang and Sharkey (2011) have evaluated the effects of context on robotic facial expression identification by humans, and Embgen et al. (2012) have conducted robotic studies on emotional body language in lieu of facial expressions.

More broadly, a number of researchers have investigated cross-cultural differences in perceptions of robots, though not necessarily for the specific purpose of affective communication (Bartneck & Okada, 2001; Bartneck et al., 2007). While many such studies agree that cultural factors influence how people perceive and behave toward robots, there is a surprising lack of agreement on the nature of these differences. A popular view among scholars is that Japanese (and possibly other Asian) subjects are more positive towards robots in general and identify them as more lifelike and animate (e.g. Geraci, 2006; Kaplan, 2004). Bartneck et al. (2007) suggest the opposite – that US participants have the most positive attitudes toward robots, particularly in terms of their willingness to interact with them on a daily basis. MacDorman et al. (2009) find more similarities than differences in how pleasant or threatening US and Japanese participants deem robots to be. Lee and Šabanović's (2014) survey study of perceptions of robots among participants in the US, South Korea, and Turkey show that, while differences among these populations exist, they are not directly correlated with broad cultural factors such as animistic or Christian beliefs, or with media portrayals of robots. These divergent results suggest that more situated contextual factors beyond broadly defined national cultures may be responsible for differential perceptions and attitudes toward interactive robotic technologies, particularly variables related to the social context of the interaction.

The studies reported here explored the effect of cultural background and environmental context on people's perceptions of affective expressions of a robotic face. We used the same robot in studies performed face-to-face with participants in the USA and in Japan, so that all subjects were able to directly interact with the robotic face, rather than only watch pre-captured images or videos of the robot in action, which is known to have drawbacks (Krumhuber et al., 2013). The studies also involved using different "cultural variants" of facial expressions to test previous research findings (see Section 3.2), as well as experiments simultaneously varying both culture and context.

## 2. Methods

### 2.1 General Overview/Subjects

Two experiments are reported in this paper. They are part of a broader series of seven experiments investigating the minimal features needed for a robotic face to communicate facial affect in a way humans could perceive and understand. The first five experiments have been previously reported (Bennett & Šabanović, 2013; Bennett & Šabanović, 2014, Bennett et al., 2014). In total, 216 human subjects participated in all the experiments, of which 93 participated in the two reported here.

Three groups of subjects were utilized: Japanese (living in Japan), native East Asians (living in the United States), and Westerners (i.e. Americans). We use the term “Westerners” here to be consistent with Jack et al. (2009) and others. The Japanese were college students recruited in Japan from a university in Yokohama. The East Asians were a mixture of Japanese, South Korean, and Chinese college students, who had lived in the United States on average for 10 months (and generally no longer than one year) and had passed an English proficiency entrance exam (TOEFL). The Westerners were all American-born college students, primarily Caucasian. The age range across all groups was approximately 18-23 (i.e. college-aged). The gender mix was roughly 50-50, with the percent male being 53.2% (Westerners), 56% (Japanese), and 46.9% (East Asians living in the U.S.). The breakdown by experiment was: 57.7% (Experiment #1A), 47.9% (Experiment #1B), and 50% (Experiment #2). Most participants came from either the computer science or psychology programs.

For the two experiments reported here, the first (#1a and #1b) examined the effects of culture, and the second (#2) examined the effects of context on the participants’ perceptions of affective expressions performed by the robotic face. Experiment #1a used a sample of Japanese subjects only (n=15). Experiment #1b used subjects from all three groups (n=48, 16 per group). Experiment #2 used samples of East Asian and Western subjects only (n=30, 15 per group). Subjects were not re-used across experiments, due to potential priming effects from repeatedly showing them facial expressions (Bennett & Šabanović, 2014). Sample sizes were determined from previously observed effect sizes (Bennett & Šabanović, 2014) with consideration for time/costs constraints.

The experiments were performed in-person through face-to-face interaction between the robot and participants at universities in the United States and in Japan. All experiments were performed in a conference room against a neutral off-white background wall. For experiments involving the digital avatar and context videos, these were shown using a laptop in the same room setup.

## 2.2 Robotic Face

The platform used here (MiRAE) is a minimalist robotic face that is capable of displaying a variety of facial expressions, previously described in (Bennett & Šabanović, 2013; Bennett & Šabanović, 2014). In a previous study, MiRAE was shown capable of producing higher, or at least comparable, identification accuracy rates (with Westerners) for all expressions as a number of other robotic faces, including Kismet (Breazeal, 2003), Eddie (Sosnowski et al. 2006), Feelix (Canamero & Fredslund, 2001), BERT (Bazo et al., 2010), and the android Geminoid-F (Becker-Asano & Ishiguro, 2011, values from Table 5 therein), as shown in Table 1 (see Bennett & Šabanović, 2014). This indicates that a minimalist robotic face such as MiRAE can provide a reliable, replicable, low-cost platform for investigating questions of affect and facial expression such as those addressed here.

**Table 1: Robot Face Comparison**

| Expression | MiRAE (n=30) | Eddie (n=24) | Kismet (n=17) | Feelix (n=86) | BERT (n=10) | Geminoid (n=71) |
|---|---|---|---|---|---|---|
| Happy | 97% | 58% | 82% | 60% | 99% | 88% |
| Sad | 100% | 58% | 82% | 70% | 100% | 80% |
| Anger | 87% | 54% | 76% | 40% | 64% | 58% |
| Fear | 43% | 42% | 47% | 16% | 44% | 9% |
| Surprise | 97% | 75% | 82% | 37% | 93% | 55% |
| Disgust | - | 58% | 71% | - | 18% | - |
| Average[1] | 85% | 57% | 74% | 45% | 80% | 58% |

Facial expression identification average accuracy for the six Ekman emotions is shown for several robotic faces (including the own used here, MiRAE). The number of subjects (n) is shown for each study as well. Appropriate citations for each are provided in text. [1]Averages do not include Disgust, since not all studies included it.

The minimalist approach for the robotic face used here is grounded in over a half-century of psychological and computer science research on emotions and facial expressions (Bennett & Šabanović, 2014). The entire premise of that work (Ekman, 2009; Nelson & Russell, 2013; Pantic, 2009; Cohn, 2010) is that people are only attending to a small number of critical moving points/lines to detect emotion in faces. This is the basis for the FACS, which dominates the emotional facial expression literature and on which many robotic faces – including androids – are based (see Section 1.1). At least within the specific task context of emotional facial expression recognition, there is evidence that many realistic aspects of the face are not necessary, and may indeed even be conflating factors (e.g. by suggesting cultural affiliation, ingroup/outgroup effects). Our previous study (Bennett & Šabanović, 2014) validated that principle in this exact robotic face, providing empirical evidence that simple moving lines work just as well for emotional expressions as more complex facial features (e.g. Kismet, see above). Other robotic research, such as Okada's Muu and Kozima's Keepon (Matsumoto et al., 2006; Kozima et al., 2009), further support such minimalism for affective interaction (not to mention Mori's work on the "Uncanny Valley" [Mori, 1970]).

Examples of MiRAE displaying various facial expressions can be seen in Figure 1. The dimensions of the robotic face are similar to an actual human face, approximately 8 inches tall by 6.5 inches wide. MiRAE also has the ability to move its neck with two degrees-of-freedom (pan and tilt), though this ability was not used in the experiments described here.

MiRAE's programming code is written as a C++/Arduino library, and easily allows facial expressions to be made with varying degrees of motion for each individual facial component (as a variable passed into the function calls). These programming libraries, along with a construction manual for MiRAE, are available from the lab website (http://r-house.soic.indiana.edu) and the first author's personal website (http://www.caseybennett.com/Research.html), in order to facilitate experimental replication.

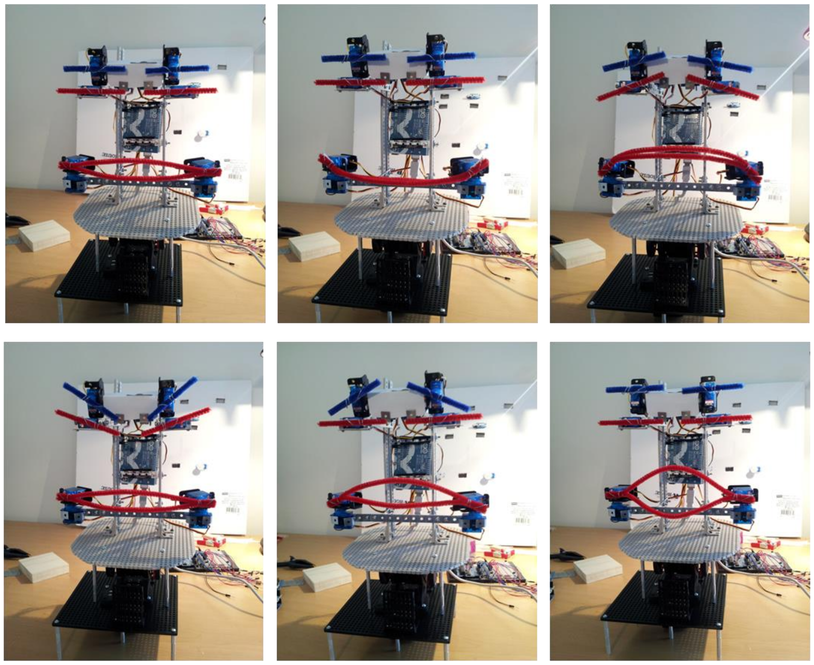
## 2.3 Experimental Design

For the two experiments reported here, the first (#1a and #1b) examined the effects of culture, and the second (#2) examined the effects of context across culture. Experiment-specific details are provided in Section 3. Here we describe the protocol and instruments used across all experiments in general.

First, we should be clear that all the experiments described here, as well the companion studies (from which some of the comparison data is derived) (Bennett & Šabanović, 2014; Bennett et al., 2014) are actually the same experiment – in terms of protocol, instruments used, and the robotic face – except for whatever independent variable was being manipulated (e.g. neck motion or added context stimuli). The only exception to this were some minor differences in the physical setup in Experiment 2 here due to the addition of the context stimuli. The robotic face was physically transported to and from Asia from the United States, so that all subjects could interact with the exact same artifact.

In all experiments, subjects observed the robotic face (and/or digital avatar, if applicable) making a randomized pre-set series of facial expressions (the six Ekman emotions, less Disgust). During each experiment, the robotic face (and/or digital avatar, if applicable) made each expression for several seconds, then returned to a neutral face. A pause of 15 seconds was provided between expressions to allow participant to fill out the FEI instrument (see next paragraph). Participants simply watched the robot, i.e. there was no interactive behavior used in these experiments. The robot (nor avatar) did not speak or make affective sounds. There were no repetitions within subjects, nor did subjects participate in multiple conditions/experiments (to avoid any "priming effect", see Section 2.1). Subjects were randomly assigned to conditions/experiments. Finally, for terminological clarity, we will use the term "eye/brow movement" to refer to the simultaneous movement of both eyes and eyebrows henceforth.

**Figure 1: MiRAE Display of Emotions**



Expression at apex of motion, without neck motion. In order (left-to-right, top-to-bottom) – Neutral, Happiness, Sadness, Anger, Fear and Surprise.

For all experiments, the same Facial Expression Identification (FEI) instrument was used as in the previous studies (Bennett & Šabanović, 2014). The FEI contains three questions. First, subjects were asked to identify the expression (Question #1) and to rate the strength of expression (Question #2). The FEI used a similar 7-option forced-choice design for Question #1 as was used in studies with Kismet, Eddie, etc. for comparability purposes (Breazeal, 2003; Sosnowski et al. 2006), although there are some issues with the forced-choice design (Nelson & Russell, 2013; Barrett et al., 2011; Fugate, 2013). The FEI also asked subjects an additional question (Question #3) for each expression, allowing (but not requiring) them to select one or more "other expressions" they thought the robot might be displaying beyond the primary one in Question #1, if desired (see [Bennett & Šabanović, 2014] for a complete description). This is the basis for the *main accuracy* (Question #1) and *other accuracy* (Question #3) in subsequent tables. The FEI is available online (in both English and Japanese) at the lab website http://r-house.soic.indiana.edu (English version: http://r-house.soic.indiana.edu/mirae/FEI_Instrument.docx).

Additionally like the previous studies, both the Godspeed (Bartneck et al., 2009) and Negative Attitudes towards Robots (NARS: Nomura & Kanda, 2003) scales were collected to evaluate user perceptions. The NARS is a commonly used metric in human-robot interaction (HRI) research, developed to measure people's attitudes towards robots *in general* and consisting of three subscales: situation of interaction, social influence of robots, and emotion in robots during interaction (Nomura et al., 2006). The NARS has often been used prior to a human-robot interaction to evaluate whether and how pre-existing attitudes affect people's behavior towards robots, as well as before and after interaction to see if the interaction itself has changed people's general attitudes toward robots. Our use of the NARS in this study was in the former sense. The Godspeed Scale was designed to gauge people's perceptions of *specific* robots and consists of five subscales: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. It is generally used in participant evaluations of robots they interact with or see; in our case we used it to measure people's perceptions of MiRAE following their interaction with it. Psychometric analyses of the NARS (Nomura et al., 2006) and Godspeed (Bartneck et al., 2009) have been previously provided, with Cronbach Alpha values consistently above 0.7. The NARS was collected prior to the interaction, the Godspeed after the interaction. Note that no significant differences in the overall NARS were found, so it will not be discussed further in this paper. For brevity, the Godspeed will only be discussed for Experiment 1a here.

All Westerner subjects and East Asian (living in the US) subjects were administered all forms, including the FEI instrument, in English. The East Asian subjects were all US university students who had passed an English proficiency entrance exam (TOEFL) prior to admission. The Japanese (living in Japan) subjects were administered the forms translated into Japanese. The 7 emotion-label options on the FEI instrument were translated into Japanese as: 怒り (*ikari* - Anger), 幸せ (*shiawase* - Happy), 悲しい (*kanashii* - Sad), 恐怖 (*kyofu* - Fear), 驚き (*odoroki* - Surprise), 嫌悪感 (*keno-kan* - Disgust), 退屈 (*taikutsu* - Bored).

## 2.4 Analysis

The analysis of data varied by experiment in accordance with the number of groups and conditions in each experiment. This included *t*-tests for Experiment #1a and ANOVAs for Experiments #1b and #2. Effect sizes are reported using Pearson's *r*. Specifics for each experiment are provided in the relevant subsections of Section 3.

Previous evaluation of statistical power suggested an *a priori* power estimate somewhere in the range 0.6 (Bennett & Šabanović, 2014), which is capable of detecting modest effects (but not smaller ones). However, since we had no basis for projecting effect sizes for most of the hypotheses reported here, it is only an estimate. Post-hoc calculations of statistical power were thus also performed, which may be informative for future experiments. For Experiment #1a, observed power was 0.64. For Experiment #1b, power was 0.98 for expression variant, and 0.32 for culture (not surprising given the small differences across cultural groups). For Experiment #2, power was 0.67 for context effects, and 0.8 for culture. Given sufficient time and money, replicating the results here with a larger study would be of great interest.

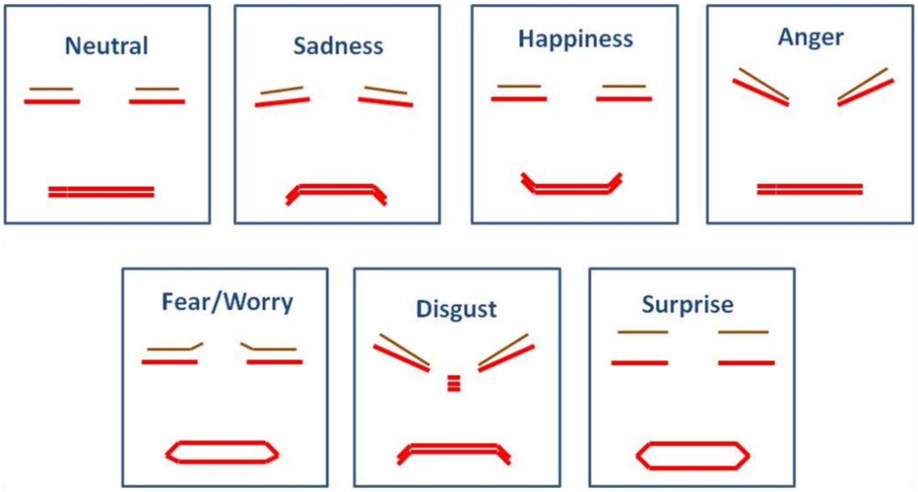
## 3. Experiments

## 3.1 Experiment 1a

### 3.1.1 Experiment 1a – Methods

**Experiment #1a** used a Japanese sample (n=15) to *replicate a previously reported study of the baseline facial expression identification results of the same robot face with Westerners* (see experiment 1 in [Bennett & Šabanović, 2014]), in order to provide a baseline comparison and ground the results of Experiment #1b. The hypothesis, based on previous research (Section 1.2), was that there would significant differences in recognition accuracy across cultures. This experiment also involved subjects observing a digital avatar designed to appear nearly identical to the embodied robotic face, as shown in Figure 2 (see [Bennett & Šabanović, 2014] for complete digital avatar description). The aim was *not* to build a state-of-the-art digital face, but to build a minimalistic avatar whose appearance and motion closely resembled the embodied robot face, for comparison purposes. The order in which subjects saw the robot and avatar (robot-1st, avatar-2nd or avatar-1st, robot-2nd) was randomized. The digital avatar was *only* used in this experiment (#1a). The experiment was exactly identical to the previous study, except for the use of Japanese subjects rather than Westerners.

**Figure 2: Schematic Facial Expressions**



For both Experiment #1a and #1b (see below), subjects observed the robotic face (and digital avatar in #1a) making a randomized pre-set series of facial expressions (the six Ekman emotions, less Disgust). As detailed in a previous paper (Bennett & Šabanović, 2014, Section 4.2), Disgust is problematic since most studies on robotic facial expressions don't actually use the Ekman "Nose Wrinkle" Disgust expression based on the FACS (e.g. Kismet [Breazeal, 2003] and Eddie [Sosnowski et al., 2006]) but rather use a contrived "Lip Twist" expression as a substitute, or do not use Disgust at all (e.g. Geminoid [Becker-Asano & Ishiguro, 2011]). To our knowledge, no robotic face has yet convincingly implemented an empirically-validated, FACS-based Disgust expression capability. In short, further work is needed.

### 3.1.2 Experiment 1a – Analysis

For Experiment #1a, we used *t*-tests (independent samples, two-tailed, equal variances not assumed) to test for differences between the original Western participants in the previously reported study (Bennett & Šabanović, 2013; Bennett & Šabanović, 2014) and the Japanese subjects evaluated in this study.

### 3.1.3 Experiment 1a - Results

Results, for both the embodied robotic face and the digital avatar, are shown in Table 2 for both the Japanese and Western subject samples (Western results reproduced from experiment 1 in [Bennett & Šabanović, 2013; Bennett & Šabanović, 2014]). A few things are notable. First, except for Fear, the identification accuracy is nearly identical for the Westerners and the Japanese, despite the fact that the facial expressions in this experiment were based on the Ekman FACS system that is purportedly biased towards Western displays of emotion (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). Fear is clearly different between the two groups (43% vs. 0%), and the Japanese clearly had trouble identifying it. However, it should be noted that – even among Westerners across an array of humanoid robotic faces (MiRAE, Kismet, Eddie, BERT, Feelix, Geminoid) – Fear is only identified on average 34% of the time (see Table 1 above) (Bennett & Šabanović, 2013; Bennett & Šabanović, 2014). *T*-tests between the two groups for overall accuracy were significantly different when including Fear ($t(43)=2.65$, p=.011, effect size=0.54), but not significant without it ($t(43)=0.53$, p=.601).

**Table 2: Experiment 1a – Main Results**

| | | **Western** | | | **Japanese** | | |
|---|---|---|---|---|---|---|---|
| | **Expression** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** |
| **Embodied** | Happy | 96.7% | 96.7% | 7.31 | 100.0% | 100.0% | 5.86 |
| | Sad | 100.0% | 100.0% | 8.30 | 86.7% | 100.0% | 7.67 |
| | Anger | 86.7% | 93.3% | 7.25 | 100.0% | 100.0% | 6.47 |
| | Fear | 43.3% | 63.3% | 6.25 | 0.0% | 6.7% | N/A |
| | Surprise | 96.7% | 100.0% | 7.96 | 93.3% | 93.3% | 5.93 |
| **Digital** | Happy | 100.0% | 100.0% | 6.93 | 100.0% | 100.0% | 4.67 |
| | Sad | 100.0% | 100.0% | 8.09 | 86.7% | 93.3% | 7.46 |
| | Anger | 100.0% | 100.0% | 7.98 | 100.0% | 100.0% | 8.07 |
| | Fear | 53.3% | 66.7% | 6.38 | 0.0% | 20.0% | N/A |
| | Surprise | 86.7% | 100.0% | 7.22 | 73.3% | 100.0% | 5.09 |

The identification results for the digital avatar followed similar patterns (significantly different with Fear, non-significant without). Strength ratings (not including Fear, since it was never identified by Japanese) were significantly different (*t*(43)=2.86, p=.008, effect size=0.41), with Westerners having higher average ratings (7.7 vs. 6.4).

Godspeed ratings were also evaluated between the two groups for the embodied robotic face. These can be seen in Table 3. Several categories were significantly different between Japanese and Westerners, with anthropomorphism and animacy being rated higher by the Japanese and perceived safety being rated higher by Westerners. It is not clear exactly why this is the case. The pattern was identical for the digital avatar (data not shown).

**Table 3: Experiment 1a – Godspeed**

| | **Western** | **Japanese** | | | |
|---|---|---|---|---|---|
| **Category** | **Embodied** | **Embodied** | ***t*-value** | **Sign.** | **Effect Size** |
| Anthropomorphism | 2.26 (.84) | 2.89 (.57) | 2.97 | 0.005* | 0.41 |
| Animacy | 2.44 (.81) | 3.24 (.58) | 3.78 | 0.001* | 0.50 |
| Likeability | 3.58 (.62) | 3.77 (.41) | 1.24 | 0.221 | |
| Perceived Intelligence | 2.86 (.81) | 3.15 (.47) | 1.49 | 0.143 | |
| Perceived Safety | 3.83 (.69) | 3.00 (.42) | 4.99 | 0.000* | 0.59 |

Mean Values for both Western and Japanese subjects are shown, with standard deviations in parentheses. T-test values are provided to the right, with statistically significant differences ($p < 0.05$) are starred with an asterisk. Effect sizes are provided for any significant differences.

## 3.2 Experiment 1b

### 3.2.1 Experiment 1b – Methods

**Experiment #1b** *evaluated two cultural variants of the baseline robotic facial expressions – an "East Asian" variant and a "Western" variant" – based on the "Emoticon hypothesis" and previous research findings that posits that East Asians focus more on the eyes and Westerners more on the mouth in interpreting facial expressions* (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). The hypothesis was that East Asians would have higher recognition accuracy for the "East Asian" variant, and the Westerners would have higher recognition accuracy for the "Western variant." In short, this resulted in the eye/brow facial feature motion being effectively turned off for the Western expressions, and the mouth facial feature motion being effectively turned off for the East Asian expressions. The exception was Anger – where the only movement in the original was in the eyes and eyebrows – which was left the same between the two variants (since there was no mouth movement to manipulate). By "effectively", we mean that the motion was set to ~10% of the original motion, so as to still be perceptible but so small as to not indicate any particular expression. Previously, we have shown that reducing the degree of motion by as much as 50% for the robot face holistically (i.e. all facial features simultaneously) had no effect on human perception of affective expression (Bennett & Šabanović, 2014). The 10% motion was in effect a small twitching motion, and was tested (for all facial features simultaneously) with several lab personnel prior to the experimental phase to verify that they conveyed no recognizable emotion/expression.

For Experiment #1b, three groups of participants were recruited each containing 16 individuals (in total, n=48) for each cultural group (see Section 2.1). Each group was randomly divided in half into two sub-groups (n=8), each of which saw only one of the variants. In other words, we had 6 sub-groups that varied by both the culture of the subjects and the facial expression variant observed.

### 3.2.2 Experiment 1b – Analysis

For Experiment #1b, we used a two-way, fixed-effects, between-subjects ANOVA to test for differences between the three cultural groups and the two cultural variants of facial expression. Post-hoc Bonferroni *t*-tests were used to determine the source of any differences.

### 3.2.3 Experiment 1b - Results

Overall results for Experiment #1b are shown in Table 4 below. Of note, we point out the similar identification patterns for Fear between the Japanese from Japan and the native East Asians living in the United States.

**Table 4: Experiment 1b – Main Results**

| | | Western | | | Japanese | | | East Asian | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Expression Variant** | **Expression** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** |
| **Western** | Happy | 100.0% | 100.0% | 7.38 | 100.0% | 100.0% | 5.63 | 100.0% | 100.0% | 7.38 |
| | Sad | 100.0% | 100.0% | 7.25 | 75.0% | 100.0% | 7.00 | 75.0% | 87.5% | 8.33 |
| | Anger | 100.0% | 100.0% | 7.25 | 87.5% | 87.5% | 6.28 | 87.5% | 87.5% | 7.71 |
| | Fear | 37.5% | 37.5% | 6.67 | 0.0% | 12.5% | N/A | 0.0% | 0.0% | N/A |
| | Surprise | 100.0% | 100.0% | 7.38 | 50.0% | 50.0% | 4.25 | 100.0% | 100.0% | 7.60 |
| **East Asian** | Happy | 50.0% | 50.0% | 5.50 | 62.5% | 87.5% | 5.00 | 50.0% | 50.0% | 7.00 |
| | Sad | 62.5% | 87.5% | 7.80 | 87.5% | 87.5% | 5.57 | 100.0% | 100.0% | 7.50 |
| | Anger | 87.5% | 87.5% | 8.00 | 100.0% | 100.0% | 7.38 | 75.0% | 75.0% | 8.33 |
| | Fear | 12.5% | 37.5% | 6.00 | 0.0% | 12.5% | N/A | 0.0% | 12.5% | N/A |
| | Surprise | 62.5% | 100.0% | 7.00 | 37.5% | 62.5% | 5.33 | 37.5% | 75.0% | 8.33 |

The results from Table 4 are succinctly summarized in Table 5. In brief, all of the cultural groups struggled to identify the East Asian expression variants (eye/brow movement only), with accuracy averaging 53.3%. Identification of the Western expression variants did vary across groups, with the Westerners having higher values, the Japanese lower values, and the East Asians living in the US somewhere in between. This pattern held even when Fear was removed, as well as Anger (which was unchanged between the variants, see Section 3.2.1). Strength ratings, however, were consistent across cultural groups for different expression variants.

**Table 5: Experiment 1b – Summary**

| | | Western | Japanese | East Asian | |
|---|---|---|---|---|---|
| **Expression Variant** | | | | | **Average** |
| **Western** | Main Accuracy | 87.5% (10.4) | 62.5% (16.6) | 72.5% (14.9) | 74.2% |
| | Strength | 7.26 (.64) | 6.01 (1.20) | 7.58 (0.97) | 7.01 |
| **East Asian** | Main Accuracy | 52.5% (18.3) | 57.5% (16.6) | 50.0% (20.9) | 53.3% |
| | Strength | 7.48 (1.64) | 5.97 (1.52) | 7.57 (1.55) | 6.95 |

Mean values are provided for each cultural group/condition, with standard deviations in parentheses.

These patterns were investigated for statistical significance via a two-way ANOVA (see Section 2.4). The results are shown in Table 6. Significant effects on accuracy were found for expression variant ($F(1,42)=17.43$, $p<.001$) but not for cultural background. The interaction effect was near significance ($F(2,42)=3.04$, $p=.058$), but not below the .05 threshold. It is possible that a larger sample size might return a significant result for the interaction effect, however. Strength ratings showed the opposite, significant variation due to cultural background, but not due to expression variant. Post-hoc test showed the significant strength differences were between the Japanese and both other groups, but not between the Westerners and East Asians living in the US.

**Table 6: Experiment 1b – ANOVA**

| | **Main Accuracy** | | **Strength Rating** | |
|---|---|---|---|---|
| | **F** | **Sign.** | **F** | **Sign.** |
| **Culture** | 1.59 | 0.216 | 6.96 | 0.002* |
| **Exp. Variant** | 17.4 | 0.000* | 0.026 | 0.873 |
| **Culture * Exp. Variant** | 3.04 | 0.058 | 0.047 | 0.954 |

F-values attaining statistical significance ($p < 0.05$) are starred with an asterisk.

To summarize the first experiment (#1a and #1b), Westerners were better at identifying robotic facial expressions from mouth movement alone than Japanese subjects (East Asians living in the US fell in between). However, none of the subject groups were significantly better at identifying facial expressions from eye/brow movement alone. Moreover, when expressions were made normally with all facial features (eyes, brows, mouth), there were no significant differences between Westerners and Japanese, except for Fear.

## 3.3 Experiment 2

### 3.3.1 Experiment 2 – Methods

For **Experiment #2**, *we evaluated the effect of the broader interaction context on participants' perceptions of the face robot's expressions*. The hypothesis, based on previous research (Section 1.2), was that context would have a larger effect on recognition accuracy for East Asians than Westerners. Subjects watched a series of videos alongside the robot-face. The videos were taken from a previous psychological study (Gross & Levenson, 1995), which validated the clips' consistent ability to elicit certain emotional responses that tie to the Ekman emotions (Happy, Sad, Anger, etc.). The same video clips were obtained in digital format and cut to length using the FRAPS software (version 3.5, http://www.fraps.com/), for the same five affective expressions as in Experiments #1a and #1b. The clips used were generally a couple minutes long, from the following (see Table 1 in [Gross & Levenson, 1995] for specific scenes/times): *When Harry Met Sally* (Happy), *Bambi* (sad), *The Shining* (Fear), *Sea of Love* (Surprise), and *Cry Freedom* (Anger). The robot face was set to automatically trigger the facial expression ("react") to match the elicited emotion of each video, at an appropriate time-point (as judged by the researchers) in the latter half of each video. Subjects were then asked to identify the expression of the robot between videos, as well as rate the strength of expression (see below). Aside from the inclusion of the video-watching, this experiment was identical to Experiments #1a and #1b in terms of protocol. As noted in Section 2.1, this experiment included two groups: Westerners and native East Asians living in the U.S. (n=30, 15 per group). Results were compared with non-context-exposed Western/Asian subjects from previous experiments (Western: n=30, Asian: n=15), with the experimental protocol being exactly the same except for the addition of context stimuli (i.e. the movie clips) during the interaction (Bennett & Šabanović, 2014; Bennett et al., 2014).

In terms of the experimental setup, the robot was placed so as to create a triadic interaction between robot, computer screen, and human subject (i.e. roughly a triangular type arrangement). Every subject was explicitly instructed prior to the experiment that the robot would "watch the video with them, and react to the video at some point, and that they should mark down the robot's reaction." A written briefing script was used by investigators to facilitate consistency.

### **3.3.2 Experiment 2 – Analysis**

For Experiment #2, we used the same ANOVA approach as in Experiment #1b (Section 3.2.2). This included a two-way, fixed-effects, between-subjects ANOVA to evaluate differences between the two cultural groups used (Westerners and East Asians living in the U.S.) and the two context exposure conditions (context-exposed vs. non-context-exposed). Post-hoc Bonferroni *t*-tests were used to determine the source of any differences.

### **3.3.3 Experiment 2 - Results**

The main results for Experiment #2 are shown in Table 7. The results show a significant increase in facial expression identification when context is supplied. This was primarily due to Fear identification, which increased from 43.3% to 100% in Westerners and from 0% to 80% in East Asians, as most of the other expressions were already in the 90-100% accuracy range without context. Of note, there was also a notable drop in identification of Happy in East Asians, which we discuss below. The results from Table 7 are summarized in Table 8.

**Table 7: Experiment 2 – Main Results**

| | | Western | | | East Asian | | |
|---|---|---|---|---|---|---|---|
| | **Expression** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** | **Main Accuracy** | **Other Accuracy** | **Strength Rating** |
| **Non-Context** | Happy | 96.7% | 96.7% | 7.31 | 100.0% | 100.0% | 5.86 |
| | Sad | 100.0% | 100.0% | 8.30 | 86.7% | 100.0% | 7.67 |
| | Anger | 86.7% | 93.3% | 7.25 | 100.0% | 100.0% | 6.47 |
| | Fear | 43.3% | 63.3% | 6.25 | 0.0% | 6.7% | N/A |
| | Surprise | 96.7% | 100.0% | 7.96 | 93.3% | 93.3% | 5.93 |
| **Context** | Happy | 93.3% | 93.3% | 5.53 | 60.0% | 80.0% | 5.67 |
| | Sad | 100.0% | 100.0% | 8.67 | 100.0% | 100.0% | 8.07 |
| | Anger | 93.3% | 100.0% | 7.50 | 93.3% | 93.3% | 7.50 |
| | Fear | 100.0% | 100.0% | 6.47 | 80.0% | 100.0% | 6.79 |
| | Surprise | 80.0% | 100.0% | 8.18 | 80.0% | 100.0% | 6.71 |

**Table 8: Experiment 2 – Summary**

| | | Western | East Asian | |
|---|---|---|---|---|
| | | | | **Average** |
| **Non-Context** | Main Accuracy | 84.0% (14.2) | 74.7% (9.2) | 80.9% |
| | Strength | 7.65 (1.36) | 6.72 (1.36) | 7.19 |
| **Context** | Main Accuracy | 92.0% (12.6) | 82.7% (16.6) | 87.3% |
| | Strength | 7.32 (1.54) | 7.25 (1.27) | 7.28 |

Mean values are provided for each cultural group/condition, with standard deviations in parentheses.

These patterns were investigated for statistical significance via a two-way ANOVA (see Section 2.4). The results are shown in Table 9. Significant effects on accuracy were found for both culture ($F(1,71)=8.02$, $p=.006$) and context ($F(1,71)=5.89$, $p=.018$). The interaction effect was not significant. There were no significant effects on strength ratings. In other words, both context and culture significantly affected facial expression perception, but context effects of similar size were present regardless of cultural background.

**Table 9: Experiment 2 – ANOVA**

| | Main Accuracy | | Strength Rating | |
|---|---|---|---|---|
| | **F** | **Sign.** | **F** | **Sign.** |
| **Culture** | 8.02 | 0.006* | 3.75 | 0.057 |
| **Context** | 5.89 | 0.018* | 0.55 | 0.463 |
| **Culture * Context** | 0.00 | 1.000 | 3.05 | 0.085 |

F-values attaining statistical significance ($p < 0.05$) are starred with an asterisk.

There were some differences across cultures, notably in the identification of Happy. Many of the East Asians identified the expression as Disgust, despite the fact that the robot expression was unchanged from previous experiments. We attribute this to the context stimuli used for that emotion (the fake orgasm scene from the film *When Harry Met Sally*), which created some discomfort and/or embarrassment in several of the East Asian participants (a few of them reported this, unsolicited, to the researcher). We also qualitatively evaluated the patterns of emotions identified as "other expression" on the FEI (Question #3, see Section 2.3) for the Westerners, which asked what if any other emotions and expression might represent beyond the primary one (data not shown for brevity). Of note, there were much higher rates of responses of Disgust for the Anger expression (80% vs. 50%, context vs. non-context) as well as higher rates of Fear for the Surprise expression (80% vs. 30%). Taken into account with the effects of context on Fear identification, these results are interesting, seeing as the robotic facial expressions themselves did not change at all.

One issue here is that many of the emotional facial expressions were already at or near 100% accuracy without context. However, a companion study to this (Bennett et al., 2014) looked at both congruent vs. incongruent context, and showed significant differences across all emotions, except (curiously) surprise. When provided incongruent context, subjects had a higher mis-recognition rate for all emotional facial expressions, revealing differences across most of them. In short, the results from Experiment #2 presented here have been partially replicated, providing further evidence for the conclusions here.

## 4. Discussion

### 4.1 General Discussion

We conducted experiments on the effects of both culture and context on perceptions of robotic facial expressions during human-robot interaction. The first set of experiments looked at the effects of culture and hypothesized culturally-variant expressions, while a second looked at the interaction of culture and context. The results are summarized below (main findings underlined).

Previous research on cultural differences in facial expressions has suggested that East Asians focus on the eyes more when viewing facial expressions in others, largely based on the "Emoticon hypothesis" and evidence from visual fixation experiments (Yuki et al., 2007; Jack et al., 2009; Koda et al., 2010; Jack et al., 2012). However, more recent research has disputed this evidence (see Introduction) (Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012). Here we investigated this hypothesis using robotic facial expressions. Our findings indicate that the issue is more complicated than those previous hypotheses might suggest. In the first experiment (#1a), we found that, except for Fear, Westerners (living in the US) and Japanese (living in Japan) were not significantly different when facial expressions were made normally (i.e. all facial features utilized). A second experiment (#1b) studied two hypothesized culturally-variant facial expressions using only mouth movement (Western) and only eye/brow movement (East Asian). We found that even though Westerners were relatively better at discerning facial expressions from mouth movement alone, Japanese were just as poor at identifying facial expressions from eye/brow movement alone, with East Asians living in the US falling somewhere in between.

These findings suggest that even if East Asians (such as Japanese) are looking at the eyes more when viewing other faces, it may be for reasons other than judging affect (as recently argued [Arizpe et al., 2012; Blais et al., 2012; Peterson & Eckstein, 2012], see Section 1.2). The results could also suggest that East Asians utilize more holistic facial feature information to judge affect in other faces. This conforms to existing research suggesting that East Asians have a more holistic cognitive style that encourages extracting meaning from relationships of multiple relevant points of attention, rather than from individual components of a scene (e.g. Nisbett et al., 2001). As for Fear, clearly current robotic facial expressions based on Ekman's FACS system appear to be ineffective for East Asians. However, we note that, even among Westerners, identification rates for Fear only average 34% across a range of humanoid robotic faces (see Section 3.1). Furthermore, Fear has been previously shown to elicit lower levels of rater agreement among research participants viewing human facial expressions, across multiple cultural groups (Biehl et al., 1997). Why this is the case remains uncertain. It is one of the most complex expressions to produce in terms of the number and control of muscles used. Its infrequency of use in daily life might also be a factor in the difficulty people have in identifying it.

The differences between Japanese and other subjects in terms of their ratings of the strength of the emotions portrayed by MiRAE can be compared to previously documented evaluations of human emotions, in which Japanese participants rated expressions as having a lower intensity than Americans (Biehl et al., 1997, pp.17). These differences in intensity might be related to the learned nature of display and decoding rules for emotional expression and to different socially normative acceptability of different expressions and levels of intensity of emotional expression in different cultures (e.g. Matsumoto, 1992). This would follow findings from previous work on identification of human emotions (e.g. Friesen, 1973), in which Japanese subjects masked negative emotions with smiles. This idea merits further study in human-robot and human-computer interaction.

As for context effects (Experiment #2), both context and culture significantly affected facial expression perception, but context effects of similar size were present regardless of cultural background. In other words, context improved recognition accuracy across cultures, and to practically the same degree. In particular, Fear – a notoriously difficult emotion to convey via robotic facial expressions – increased to nearly 100% with added context, regardless of cultural background of the subjects. These findings concur with previously reported context effects in both humans/avatars (Righart & de Gelder, 2008; Barrett et al., 2011; Lee et al. 2012) as well as robots (Zhang & Sharkey, 2011). We were also able to replicate these effects in a companion study in which we looked at the effects of both incongruent and congruent context on people's perceptions of a robots affective facial expressions which showed significant differences across all emotions, except for surprise (Bennett et al., 2014).

These findings are potentially useful for constructing robotic faces that may interact via facial expressions with different cultures, as well as for designing interactive robots or avatars that utilize facial expression identification across different cultures.

## 4.2 Implications

The results of these studies presented here have a number of potentially intriguing implications. The context effects seen in Experiment #2 seem to suggest that human subjects may be *projecting* their own internal emotions onto the facial expressions of others, including robots. Given that the context videos have been previously shown to reliably elicit certain emotions in human subjects, and the fact that the robotic facial expression stimuli were exactly the same across conditions, we arrive at such an interpretation. This concurs with other recent research findings into the role of emotion formation and cognition in human-human interaction, which may be informative for human-robot interaction.

There is evidence that such projection may in fact be a key part of such affective communication between humans. Lindquist and Gendron (2013) have proposed a “Construction hypothesis” of emotion, which is essentially a dynamical systems view of emotion perception, where language, emotion labels, and/or other context may ground our perceptions of both emotion and facial expressions. As they noted, “this leaves open the possibility, as the data reviewed here suggest, that emotions seen on other people’s face are constructed in the mind of the perceiver” (pp.70). Barrett et al. (2011) make a similar dynamical systems argument for the effects of context (including language). They also point out that context – from a human cognition standpoint – really relates to the way the brain makes predictions using visual (or other sensory) data. Recent studies provide further evidence for this explanation. Righart and de Gelder (2008) found context biases the pattern of error responses in facial expression identification of human faces. This is similar to our finding for “other expression” attribution patterns in Experiment #2 (see Section 3.3.3). Elsewhere, Lee et al. (2012) found evidence of inter-individual differences modulating the effects of context on facial expression identification.

More broadly, this relates to scholarship on the cultural aspects of social cognition and technology (e.g. Hall, 1977; Shore, 1996; Nisbett, 2001, 2003), which suggests that culturally appropriate social cues, including modes of communication, temporal interaction patterns, and expectations regarding affective display, are foundational to human sociality and that a breach of cultural norms can provide a significant barrier to successful interaction. The results here support this perspective. In previous work, Šabanović (2010, 2014) showed that various cultural models of affect, social cognition, and interaction with technology are embodied in social robot design in both explicit and implicit ways. Such culturally-situated design choices, however, generally reproduce stereotypical notions of cultural difference rather than developing technologies that can fit empirically based constructions of the cultural dynamics of social interaction. A more reflexive understanding of culture’s role in social interaction suggests a dynamic model, in which cultural models are not simply copied, but are “repeatedly assembled”: core cultural models dynamically change as they are adapted to fit contemporary circumstances (Caporael, 1997). In the development of affectively-expressive interactive technologies, this viewpoint supports the adoption of a dynamic and relational model of affect construction, which would address the situated nature of cultural expression within social interaction.

Such a dynamical systems view of emotion and affective interaction also feeds into concepts about embodied cognition and the development of robotic (and/or other artificially intelligent) interactive systems. If, as Barsalou et al. and others have suggested, higher cognition is primarily intended for the mediation of perception and action via dynamic mechanisms, then emotions are biasing factors that prime our anticipatory response systems for subsequent events (Barsalou et al., 2006; Beer, 2000). Indeed, affective communication, including facial expressions, could even be seen as a kind of *context* itself in that view. In a counter-intuitive sense, they are context created by social interaction for the explicit purpose of facilitating further social interaction. For instance, if the goal is to communicate information about food or dangers in the environment, then affective communication can provide enabling context that simplifies the need for interpretation and understanding of *future* sensory signals (including social ones) in terms of behavior/action-selection (Barsalou et al., 2006). This is an equivalent argument to Clark (2013) that we utilize social cues to “load the dice” in terms of minimizing costly prediction errors and facilitating our own cognition (see Section 3.2 therein). Or, in other words, self-structuring of sensory information into a rolling “cognitive niche” (Sterelny, 2007; Clark 2013). From another angle, this can be seen as a social-interaction-based form of cognitive scaffolding, in the vein of Gibson and visual scaffolding (Gibson, 1979). This also concurs with other recent suggestions of social cognition as an emergent phenomenon from social interaction itself (De Jaegher et al., 2010; Froese & Ziemke, 2009; Froese & Di Paolo, 2010; McGann et al. 2013). The socio-cultural and cognitive science literature both point in the same direction – that affective interaction is not necessarily about communicating some “information” about the current state of the world, but rather about biasing what we expect to experience next, both internally and externally.

Such evidence holds intriguing possibilities for robotics. If emotions perceived in others are indeed an internal construct in the mind of the perceiver based on a number of dynamic perceptual and cognitive processes, then the question exists of how we might take advantage of those processes to facilitate human-robot interaction. Facial expressions, or other direct forms of communication, may only be one piece of the puzzle. The results here suggest that, if we can induce appropriate context effects, it may be possible to create *culture-neutral models* of robots and affective interaction. Inducement of such context effects, for instance, could stem from creation of environmental conditions that correspond with certain attractor basins in human cognition. Individual-specific models could potentially be learned via machine learning methods, allowing the robot to adapt to individual people. Such an approach may be an alternative and/or potentially more effective path than direct affect communication (e.g. trying to make culturally-specific expressions or cues for every single cultural group). This is a similar concept as approaches being explored for dynamic/adaptive production of synthetic emotions in robots and intelligent agents, although from the polar opposite direction (Picard, 1997; Canamero, 2005; Asada et al. 2009; Bosse et al. 2010).

## 4.3 Limitations

There are some limitations to this study. For example, there are confounding factors we cannot rule out cross-culturally, including the effects of language. Different emotion-label words may have different cultural connotations (a.k.a. linguistic relativity), which can affect response answers (Perlovsky, 2009; Ruttkay 2009; Davies et al. 1998). Such linguistic relativity might also tie into the aforementioned view of emotions and facial expressions from a dynamical systems and embodied cognition perspective. Additionally, there are issues with the forced-choice response design – although given how common that methodology is in this area (Nelson & Russell, 2013; Barrett et al., 2011; Fugate, 2013), it becomes difficult to directly compare results to other work if other designs are utilized. Moreover, from a dynamical systems perspective, categorization is a fundamental aspect of higher cognition, as categories relate to attractor basins for otherwise continuous-valued perceptions. In that sense, it is challenging to understand or study any aspect of human cognition without categorization.

Caution should also be taken in generalizing the results seen here. There may be, for instance, tasks other than affective facial interaction where these results do not apply. Those tasks may necessitate less minimalist face/facial components for a robot, or other non-facial (i.e. bodily) cues in order to communicate information.

Other limitations include the sample size – some statistical tests here, particularly several that were near the .05 threshold for significance, might attain significance if these experiments were replicated with larger sample size. There were also some issues with the film clips used (particularly happiness, as noted in Section 3.3.3), though they were chosen because they had been previously validated to elicit certain emotional responses in a published study (Gross & Levenson, 1995). Those issues may hint at the interplay of cultural norms and context-based emotional cues. From a broader perspective, this study also leaves a number of unanswered questions that deserve further study, e.g. more deeply investigating synergistic effects between culture and context. We discuss some of these in the next section.

## 4.4 Future Directions

This work suggests a number of future directions for research. For instance, the congruence between context and facial display of emotion may have a variable effect on emotion recognition cross-culturally (Boiger & Mesquita 2012). We are currently exploring context congruence in a companion study (Bennett et al., 2014). Temporal dynamics in social cognition and interaction may also play a role. Modeling of those dynamics, in the spirit of Beer (1995), Auvray et al. (2009), and Ikegami & Suzuki (2008), may help elucidate fundamental building blocks of minimal cognition and social interaction. A study along these lines is set to begin in Japan in the summer of 2014. Moreover, exploring such interaction dynamics, both in laboratory and “robots in the wild” experiments, is warranted (Šabanović et al. 2006, MacDorman & Ishiguro 2006). We are currently studying the latter in a project that placed an interactive version of MiRAE – which could respond to the presence of people in its vicinity – into a month-long public art display to explore more naturalistic, free-form social interaction. Finally, design aspects that may affect the interaction and/or affective communication can be explored with 3D printing, allowing for rapid prototyping and testing of component design that vary in terms of shape, size, texture, range of motion, realism, etc. Understanding how certain design choices affect human-robot interaction is fundamental. We are currently working on a project involving such 3D printed robotic face design. Many other opportunities exist in this domain as well that may inform our understanding of social interaction and the artificial construction thereof.

## Acknowledgements

The authors would like to thank our collaborators in both Japan (Dr. Seite Koike, Chinatsu Nunomura) and in the United States (Kay Jessee, Marion Grant, Sam Ragsdale). We would also like to thank the anonymous reviewers of the manuscript, who made a number of invaluable suggestions. Funding was provided by NSF grant #IIS-1143712, as well as the NSF/IU IGERT summer research award in cognitive science.

## References

Adams, B., Breazeal, C., Brooks, R.A., & Scassellati, B. (2000). Humanoid robots: a new kind of tool. *IEEE Intelligent Systems and their Applications*, 15(4), 25–31.

Arizpe, J., Kravitz, D.J., Yovel, G., & Baker, C.I. (2012). Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PLoS ONE*, 7(2), e31106.

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., et al. (2009). Cognitive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development*, 1(1), 12–34.

Auvray, M., Lenay, C., & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, 27(1), 32–47.

Barrett, L.F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20(5), 286–90.

Barsalou, L.W., Breazeal, C., & Smith L.B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, 8(2), 79–91.

Bartneck, C. & Okada, M. (2001). Robotic user interfaces. *Proceedings of the Human and Computer Conference (HC2001)*, 130–140.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society*, 21(1-2), 217–30.

Bartneck, D., Kulic, E., Croft, M., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71-81.

Bazo, D., Vaidyanathan,, R., Lentz, A, & Melhuish, C. (2010). Design and testing of a hybrid expressive face for a humanoid robot. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5317–5322.

Becker-Asano, C. & Ishiguro, H. (2011). Evaluating facial displays of emotion for the android robot Geminoid F. *Proceedings of the IEEE Workshop on Affective Computational Intelligence (WACI*), 1–8.

Beer, R.D. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72(1–2), 173–215.

Beer, R.D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3), 91–99.

Bennett, C.C. & Šabanović, S. (2013). Perceptions of affective expression in a minimalist robotic face. *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 81-82.

Bennett, C.C. & Šabanović, S. (2014). Deriving minimal features for human-like facial expressions in robotic faces. *International Journal of Social Robotics*, 6(3), 367-381.

Bennett, C.C. , Šabanović, S., Fraune, M., & Shaw, K. (2014). Context congruency and robotic facial expressions: Do effects on human perceptions vary across culture? *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 465-470.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, 21(1), 3-21.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, 50(12), 2830–2838.

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1–2), 119– 155.

Boiger, M., & Mesquita, B. (2012). The construction of emotion in interactions, relationships, and cultures. *Emotion Review*, 4(3), 221–229.

Bosse, T., Pontier, M., & Treur, J. (2010). A computational model based on Gross' emotion regulation theory. *Cognitive Systems Research*, 11(3), 211-230.

Canamero, L. (2005). Emotion understanding from the perspective of autonomous robots research. *Neural Networks*, 18(4), 445-455.

Canamero, L. & Fredslund, J. (2001). I show you how I like you - can you read it in my face? *IEEE Transactions on Systems Man and Cybernetics A: Systems and Humans*, 31(5), 454 –459.

Caporael, L.R. (1997). The evolution of truly social cognition: The core configuration model. *Personality and Social Psychology Review*, 1(4), 276–298.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.

Cohn, J.F. (2010). Advances in behavioral science using automated facial image analysis and synthesis. *IEEE Signal Process Magazine*, 27(6), 128–133.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London, UK: John Murray.

Davies, I.R., Sowden, P.T., Jerrett, D.T., Jerret,t T., & Corbett, G.G. (1998). A cross-cultural study of English and Setswana speakers on a colour triads task : A test of the Sapir-Whorf hypothesis. *British Journal of Psychology*, 89(1), 1–15.

De Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535), 3475–3484.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–447.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. *Nebraska Symposium on Motivation*, 19, 207–283.

Ekman, P. (2009). Darwin's contributions to our understanding of emotional expressions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535), 3449–3451.

Ekman, P. & Friesen, W.V. (2003). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Los Altos, CA, USA: Malor Books.

Elfenbein, H.A. (2013). Nonverbal dialects and accents in facial expressions of emotion. *Emotion Review*, 5(1), 90–96.

Embgen, S., Luber, M., Becker-Asano, C., Ragni, M., Evers, V., & Arras, K.O. (2012). Robot-specific social cues in emotional body language. *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1019–1025.

Friesen, W.V. (1973) Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules. *Dissertation Abstracts International*. 33(8-B), 3976-3977.

Froese T. & Di Paolo, E.A. (2011). The enactive approach: theoretical sketches from cell to society. *Pragmatics & Cognition*. 19(1), 1-36.

Froese, T. & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, 173(3-4), 466–500.

Fugate, J.M.B. (2013). Categorical perception for emotional faces. *Emotion Review*, 5(1), 84–89.

Geraci, R. M. (2006). Spiritual robots: Religion and our scientific view of the natural world. *Theology and Science*, 4(3), 229-246.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghlin Mifflin.

Gross, J.J. & Levenson, R.W. (1995). Emotion elicitation using films. *Cognition and Emotion*, 9(1), 87–108.

Hall, E.T. (1977). *Beyond Culture*. New York, NY, USA: Anchor Books.

Ikegami, T. & Suzuki, K. (2008). From a homeostatic to a homeodynamic self. *BioSystems*. 91(2), 388–400.

Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology*, 19(18), 1543–1548.

Jack, R.E., Garrod, O.G., Yu, H., Caldara, R., & Schyns, P.G. (2012). Facial expressions of emotion are not culturally universal. *Proceeedings of the National Acadamy of Sciences USA*, 109(19), 7241–7244.

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, 1(3), 465-480.

Kleinsmith, A., De Silva, P.R., & Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers*, 18(6), 1371–89.

Kobayashi, H., Ichikawa, Y., Senda, M., & Shiiba, T. (2003). Realization of realistic and rich facial expressions by face robot. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1123-1128.

Koda, T., Ruttkay, Z., Nakagawa, Y., & Tabuchi, K. (2010). Cross-cultural study on facial regions as cues to recognize emotions of virtual agents. In: *Lecture Notes in Computer Science (Culture and Computing)*, T. Ishida, Ed. Springer, Berlin, 16–27.

Kozima, H., Michalowski, M.P., & Nakagawa, C. (2009). Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*, 1(1), 3–18.

Krumhuber, E.G., Kappas, A., & Manstead, A.S.R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, 5(1), 41–46.

Lee, T.H., Choi, J.S., & Cho, Y.S. (2012). Context modulation of facial emotion perception differed by individual difference. *PLoS ONE*, 7(3), e32987.

Lee, H. R., & Sabanović, S. (2014). Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 17–24.

Lindquist K.A. & Gendron, M. (2013). What’s in a word? Language constructs emotion perception. *Emotion Review*, 5(1), 66–71.

MacDorman K.F. & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297–337.

MacDorman, K. F., Vasudevan, S. K., & Ho, C. C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society*, 23(4), 485-510.

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, 23(1), 72–84.

Matsumoto, N., Fujii, H., & Okada, M. (2006). Minimal design for human-agent communication. *Artificial Life and Robotics*, 10(1), 49–54.

McGann. M., De Jaegher, H., & Di Paolo, E. (2013). Enaction and psychology. *Review of General Psychology*, 17(2), 203–209.

Miwa, H., Itoh, K., Matsumoto, M., Zecca, M., Takanobu, H., Rocella, S., et al. (2004). Effective emotional expressions with expression humanoid robot we-4rii: integration of humanoid robot hand rch-1. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2203-2208.

Mori, M. (1970). Bukimi no tani [The uncanny valley].  *Energy*, 7(4), 33–35. http://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley

Nelson, N.L. & Russell, J.A. (2013). Universality revisited. *Emotion Review*, 5(1), 8-15.

Nisbett, R.E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, 108(2), 291–310.

Nisbett, R.E. (2003). *The Geography of Thought: How Asians and Westerners Think Differently*. New York, NY, USA: Free Press.

Nomura, T. & Kanda, T. (2003). On proposing the concept of robot anxiety and considering measurement of it. *Proceedings of 12th IEEE International Symposium on Robot and Human interactive Communication (RO-MAN)*, 373–378.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437–454.

Pantic, M. (2009). Machine analysis of facial behaviour: naturalistic and dynamic behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535), 3505–3513.

Perlovsky, L. (2009). Language and emotions: emotional Sapir-Whorf hypothesis. *Neural Networks*, 22(5-6), 518–526.

Peterson M.F. & Eckstein, M.P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceeedings of the National Acadamy of Sciences USA*, 109(48), E3314–3323.

Picard, R.W. (1997). *Affective Computing*. Boston, MA, USA: MIT Press.

Righart, R. & de Gelder, B. (2008). Recognition of facial expressions is influenced by emotional scene gist. *Cognitive Affective & Behavavioral Neuroscience*, 8(3), 264–272.

Russell, J.A. & Fernández-Dols, J.M. (1997). *The Psychology of Facial Expression*. Cambridge, UK: Cambridge University Press.

Ruttkay, Z. (2009). Cultural dialects of real and synthetic emotional facial expressions. *AI & Society*, 24(3), 307–315.

Šabanović, S., Michalowski, M.P., & Simmons R. (2006). "Robots in the wild: Observing human-robot social interaction outside the lab." *9th IEEE International Workshop on Advanced Motion Control*, 596-601.

Šabanović, S. (2010). Emotion in robot cultures: Cultural models of affect in social robot design. *Proceedings of the Conference on Design & Emotion (D&E2010)*, Chicago, IL.

Šabanović, S. (2014). Inventing Japan’s “robotics culture”: The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science*, In Press. Available online: http://sss.sagepub.com/content/early/2014/01/17/0306312713509704.abstract

Saldien, J., Goris, K., Vanderborght, B., Vanderfaeillie, J., & Lefeber, D. 2010. Expressing emotions with the social robot Probo. *International Journal of Social Robotics*, 2(4), 377–389.

Scassellati, B. (2006). How developmental psychology and robotics complement each other. *Technical report, Massachusetts Institute of Technology, CSAIL*.

Shore, B. (1996). *Culture in Mind: Cognition, Culture, and the Problem of Meaning*. Oxford, UK: Oxford University Press.

Sosnowski, S., Bittermann, A., Kuhnlenz, K., & Buss, M. (2006). Design and evaluation of emotion-display EDDIE. *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* 3113–3118.

Sterelny, K. (2007). Social intelligence, human intelligence and niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 719-730.

Yuki, M., Maddux, W.W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, 43(2), 303–311.

Zhang, J. & Sharkey, A.J. (2011). Contextual recognition of robot emotions. In: R. Groß et al. (Ed.), *Towards Autonomous Robotic Systems* (pp.78-89). Berlin, Germany: Springer.