

Context Congruency and Robotic Facial Expressions: Do Effects on Human Perceptions Vary across Culture?

Casey C. Bennett, Selma Šabanović, Marlena R. Fraune, and Kate Shaw

Abstract— We performed an experimental study (n=48) of the effects of context congruency on human perceptions of robotic facial expressions across cultures (Western and East Asian individuals). We found that context congruency had a significant effect on human perceptions, and that this effect varied by the emotional valence of the context and facial expression. Moreover, these effects occurred regardless of the cultural background of the participants. In short, there were predictable patterns in the effects of congruent/incongruent environmental context on perceptions of robot affect across Western and East Asian individuals. We argue that these findings fit with a dynamical systems view of social cognition as an emergent phenomenon. Taking advantage of such context effects may ease the constraints for developing culturally-specific affective cues in human-robot interaction, opening the possibility to create *culture-neutral models of robots and affective interaction*.

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

I. INTRODUCTION

A fundamental question for human-robot interaction (HRI) is whether – and to what degree – variables *external to the robot* affect the perceptions a human user has of what the robot is communicating. This includes affective interaction [1]. For instance, environmental context (due to music, lighting, etc.) is known to elicit resonant emotions in people [2]. Certain colors of light elicit happiness, certain sounds evoke fear, certain scenes evoke surprise, and so forth. Moreover, the effects of such environmental context may vary depending on the characteristics of the person, e.g. their cultural background.

*Research supported by NSF grant #IIS-1143712.

Casey Bennett is with the School of Informatics and Computing (SOIC) at Indiana University and the Dept. of Informatics at Centerstone Research Institute, Bloomington, IN, USA (phone: 812-355-6382; e-mail: cabennet@indiana.edu). *Member IEEE*

Selma Šabanović is with the School of Informatics and Computing (SOIC) at Indiana University, Bloomington, IN, USA (e-mail: selmas@indiana.edu). *Member IEEE*

Marlena Fraune is with the Cognitive Science Program at Indiana University, Bloomington, IN, USA (e-mail: mfraune@indiana.edu).

Kate Shaw is with the School of Informatics and Computing (SOIC) at Indiana University, Bloomington, IN, USA (e-mail: shawkaa@imail.iu.edu).

Many researchers have explored affective communication by robots, such as facial expressions [3-8] and other less explicit emotional cues [9,10]. As in interaction among humans, context effects can play a role in how people perceive such cues performed by a robot.

In previous work studying robotic facial expressions in HRI, we have empirically shown that context effects of similar size were present regardless of the participant's cultural background [11]. Providing context known to elicit matching emotions significantly improved human recognition of the robotic facial expressions over non-context experiments, even though the facial expressions were exactly the same in both conditions. The results suggested a form of *projection*. Emotions perceived in the faces of others – including robots – appeared to be an internal construct in the mind of the perceiver, based on a number of perceptual and cognitive processes [11,12]. This was equally true across human subjects from Western and Asian cultural backgrounds.

In that previous study, the provided context was always congruent with the robotic facial expression, i.e. the emotion elicited by the context was the same as the emotion communicated by the robot's facial expression. A separate, but related, question is what would happen if context congruency was varied – if the emotion expressed by the context was sometimes congruent, sometimes incongruent, with the robotic expressions [13,14]. In this study, we empirically explore *whether the effects of context congruency on human perceptions of robotic facial expressions vary across culture*.

II. RELATED WORK

This work is informed by previous research on emotions, facial expressions, and robotic faces, which we review here along with scholarship on the interplay of both culture and context on affective interaction.

A. Emotion, Facial Expressions, and Robotic Faces

In the section, we provide a brief overview (for brevity) of emotion and facial expressions, and their use in robotic faces. We have provided a more extensive overview of the scientific literature on robotic facial expressions and human emotion in previous papers [11,15,16].

The scientific study of emotions in humans has a long and venerable history going back nearly two centuries [17]. Over the last half century, scholarly debate has focused on emotional facial expressions and how to classify them [18-21]. A principle question is whether a basic set of universal human emotions (and their related facial expressions) exist across culture, gender, context, etc.? The study of facial expressions of emotion has evolved into two major camps during this time period: 1) Ekman et al., who argue for 6-7 “basic” *categorical* emotional expressions that are universal across cultures [18], and 2) Russell et al., who argue that facial expressions are emergent states from a *continuous*, multi-dimensional space of affect (circumplex model), typically defined by three principle axes: valence, arousal, and stance (Fig.1) [19,22]. Valence, which relates to the positivity/negativity of the emotion/expression, is of particular interest in the present study.

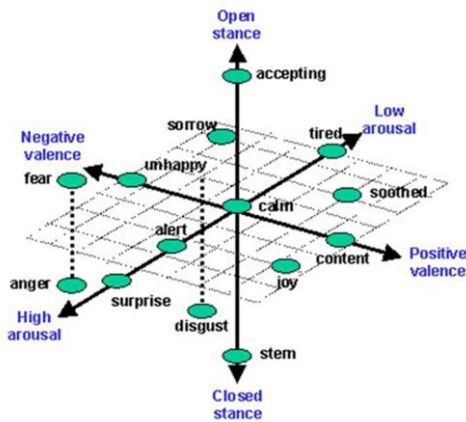


Figure 1. 3-Dimensional Affect Space (From [3])

Various robotic faces have been constructed over the last decade that integrate aspects of both Ekman and Russell’s theoretical approaches (e.g. [3-8]).

B. Culture and Affective Interaction

Numerous theories about the role of culture in affective interaction, including facial expressions, exist. One primary theory is the “Emoticon hypothesis”, which posits cultural differences in facial expressions based on differences in emoticons between Western and East Asian cultures (e.g. East Asians focus more on the eyes, and Westerners more on the mouth) [23]. A number of papers have studied visual fixation patterns as the basis for these putative differences in recent years, related to that hypothesis [24,25]. However, more recent studies have provided evidence countering the use of such 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 [26-28]. Recent work in human-robot interaction has provided empirical evidence that also runs counter to this hypothesis [11]. In short, the empirical basis at this point for the Emoticon hypothesis is tenuous at best.

Broader socio-cultural research has examined the possibility of different “cognitive styles” in affective interaction across cultures that prescribe salient features of an individual’s environment and appropriate modes of

communication [29]. Culturally variable “social-orientational models” may designate appropriate roles/behaviors within interaction as well as culturally-normative rules for displaying, perceiving, and experiencing affect [30]. Along similar lines, Ekman, Friesen, and Izard themselves suggested a “Deception hypothesis” in the 1970’s to explain culturally-based affective expression encoding rules [31]. More recently, Elfenbein has proposed a “Dialect hypothesis” for affective communication, which posits isomorphisms between affective expressions and linguistic distributions/development [32].

C. Context Congruency and Culture

An ongoing debate in recent years is whether cultural differences influence the role context plays in affective interaction, including perceptions of facial expressions [12,33,34]. Across cultures, context is considered important for discerning emotions, with evidence suggesting that without context cues emotion recognition decreases [12,35,36]. For example, Western participants (from the Netherlands) displayed faster reaction times to correctly identify emotions when a background image invoked an emotion congruent with displayed facial expressions. This trend varied by the valence of the expression [33]. Recent work has shown the importance of context in perceptions of robotic facial expressions across cultures as well [11,37].

There is some research suggesting that people in Eastern Asian cultures pay greater attention to context than do Westerners. This has been shown on neutral tasks such as describing the contents of a fishbowl [38] and on tasks of detecting emotion in faces [39,40]. In particular, the effects of context congruency varied across culture, having a greater effect on East Asians than Westerners.

However, these findings on cultural variability are subject to debate. They mainly involve looking at pictures of static faces and images on a computer screen, not direct interaction with a physically embodied human or robotic face. Thus, an open question is whether the effects of context congruency may vary across cultures in face-to-face interaction with a robot. We empirically explore that question here.

III. METHODS

A. General Overview/Subjects

This paper reports on a single experiment involving robotic facial expression recognition, in which we systematically varied the congruency of the environmental context in respect to the affective facial expressions made by the robot and the cultural background of human subjects.

Two groups of subjects participated in the study: 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. and others [24]. The East Asians were a mixture of Japanese, South Korean, and Chinese college students, who had lived in the United States on average for 6 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 gender mix was 58.3% females. Subjects were college age (18-25 years old). Results with subjects outside this age/gender composition, of course, may vary from those seen here. Most participants came from either the computer science or psychology programs.

A total of 48 subjects were recruited (n=48), 24 each for the two cultural groups. There were three experimental conditions (see Section 3.C), resulting in n=8 for each condition for each cultural group. Sample sizes were based on estimated effect sizes from previous studies [11,16].

B. 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 [15,16]. In previous studies, 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 [3, Eddie [4], Feelix [5], BERT [6], and the android Geminoid-F [8], as shown in Table I (see [16]).

TABLE I. EXPERIMENTAL CONDITIONS

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	85%	57%	74%	45%	80%	58%

Examples of MiRAE displaying various facial expressions can be seen in Fig. 2. 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).

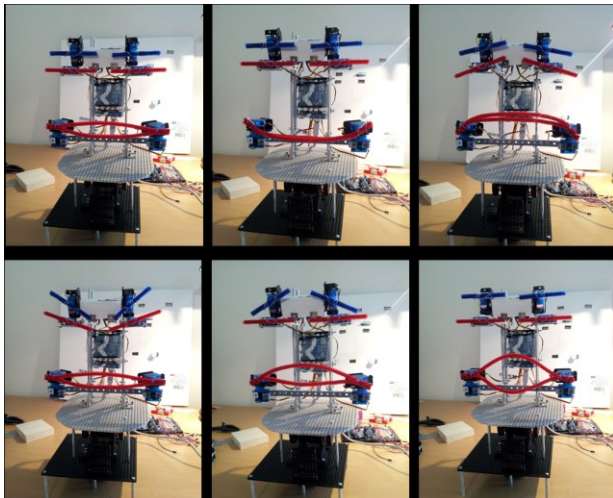


Figure 2. Expression at apex of motion, without neck motion. In order (left-to-right, top-to-bottom) – Neutral, Happiness, Sadness, Anger, Fear and Surprise.

C. Experimental Design

The experiment took place in the R-House HRI Lab at Indiana University, Bloomington. The experiment design was the same as in previously reported experiments (experiment #2 in [11]), except that the context congruency was varied in this case. After giving informed consent, subjects were asked to watch a series of videos alongside the robot-face. The videos were taken from a previously validated psychological study [2], which verified the clips’ ability to consistently elicit certain emotional responses that tie to the Ekman emotions (e.g. Happy, Sad, Anger). 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 used in previous experiments: Happy, Sad, Surprise, Fear, Anger [11,15,16].

The clips used were generally a couple minutes long, excerpted from the following films (see Table 1 in [2] 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 either match (*congruent*) or not match (*incongruent*) the elicited emotion of each video, depending on the experimental condition (see below). Expressions were triggered 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 to rate the strength of expression (see below). Results were compared with non-context-exposed subjects from previous studies [11].

The goal was to evaluate the effects of context congruency on human perceptions of robotic facial expressions. However, such effects may depend on the *degree* of incongruency, i.e. how similar the elicited emotion of the context is to the emotion of the facial expression. In this study, we define similarity based on emotional valence, which is a primary component of emotion classification systems (see Section II.A). Previous studies have also suggested that the effects of context congruency may vary by valence [33]. In order to account for similarity as a conflating factor, three experimental conditions were used, in which we “switched” certain expressions so that they were incongruent with the context (Table II). Other expressions were left as congruent with the context. Each expression was shown only once for each subject, to avoid priming effects [16]. For Condition 1, positive-valence emotional expressions (Happy, Surprise) were switched with each other. For Condition 2, negative-valence emotional expressions were switched (Sad, Fear, Anger). For Condition 3, we switched expressions *across* valence, so that positive-valence expressions were shown with negative-valence context, and vice versa (Fear was left congruent as a control).

TABLE II. EXPERIMENTAL CONDITIONS

Context	Expression Shown		
	Positive Switch	Negative Switch	Cross Switch
Happy	Surprise	-	Sad
Sad	-	Anger	Happy
Anger	-	Fear	Surprise
Fear	-	Sad	-
Surprise	Happy	-	Anger

** Entries with a dash were unchanged (i.e. context and facial expression were congruent)

For all experiments, the same Facial Expression Identification (FEI) instrument was used as in the previous studies [11,15,16]. 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 [3,4] (although there are some issues with the forced-choice design, see [12,19,41]). 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 [16] for a complete description). The FEI is available online at the lab website (<http://r-house.soic.indiana.edu>) or the first author’s personal website. Like previous studies [11,16], both the Godspeed and NARS scales were collected, but are not discussed here for brevity.

D. Analysis

The analysis of the data consisted of two separate parts in order to answer two primary questions: 1) whether the effects of context congruency on perceptions of robotic facial expressions varied by culture, and 2) whether such effects depended on the similarity of emotional valence of the facial expressions and the context.

For the first question, we used a two-way, fixed-effects, *within-subjects* ANOVA to test for differences between congruent and incongruent context across the two cultural groups. Repeated measures for each subject were the recognition accuracies of facial expressions for congruent context and for incongruent context.

For the second question, we used a two-way, fixed-effects, *between-subjects* ANOVA to test for differences in recognition accuracy between the three different conditions across the two cultural groups. The three conditions varied by which expressions were incongruent with the context, based on emotional valence (see Section III.C). Post-hoc Bonferroni *t*-tests were used to determine the source of any differences across the conditions.

IV. RESULTS

This section is broken into two parts which each address one of the primary questions of the paper (see Section II.D).

A. Effects of Context Congruency across Cultures

One primary question was whether the effects of context congruency on perceptions of robotic facial expressions varied by culture. A summary of facial expression

recognition accuracy rates by context congruency and culture is shown in Table III. The “None” context values were taken from previously reported studies [16].

TABLE III. RECOGNITION ACCURACY BY CONTEXT CONGRUENCY AND CULTURE

Context	Western	East Asian
None	84.0%	74.7%
Congruent	93.8%	87.5%
Incongruent	51.4%	45.8%

As Table III shows, congruent context produced facial expression recognition rates that were nearly twice that of incongruent context, regardless of culture. Moreover, incongruent context significantly reduced recognition rates over providing no context at all. Meanwhile, congruent context increased facial expression recognition rates by about 10-12% over no context, which replicates previous findings [11]. These patterns occurred regardless of the cultural background of the subjects.

The patterns were investigated for significance via a two-way, *within-subjects* ANOVA (see Section III.D). The results are shown in Table IV. Significant effects on accuracy were found for context congruency ($p < .001$) but not for cultural background. The interaction effect was not significant.

TABLE IV. CONTEXT CONGRUENCY BY CULTURE - ANOVA

	Main Accuracy	
	F	Sign.
Culture	0.639	0.428
Congruency	46.38	0.000*
Culture * Congruency	0.02	0.883

**Significant values ($p < .05$) marked with an asterisk

A breakdown of the recognition rates for each facial expression by cultural group is provided in Table V. There are some interesting patterns of note, although caution is warranted when sub-dividing the experimental data to that degree. In general, recognition rates for all expressions were reduced when context was incongruent, except for Surprise (curiously). Surprise recognition actually increased slightly. This was true regardless of whether Surprise was switched with expressions of similar (positive) or opposite (negative) valence. The reason for this is unclear. The result warrants further experimental investigation.

TABLE V. INDIVIDUAL FACIAL EXPRESSIONS AND CONTEXT CONGRUENCY

Context	Expression	Western	East Asian
Congruent	Happy	100.0%	100.0%
	Sad	100.0%	87.5%
	Anger	100.0%	87.5%
	Fear	93.8%	87.5%
	Surprise	75.0%	75.0%
Incongruent	Happy	50.0%	43.8%
	Sad	37.5%	37.5%
	Anger	50.0%	25.0%
	Fear	0.0%	25.0%
	Surprise	93.8%	87.5%

In summary, context congruency had a significant influence in human perceptions of robotic facial expressions. Providing incongruent context was worse than providing no context at all. There were no significant differences in context effects due to culture.

B. Effects of Emotional Valence Similarity between Facial Expression and Context

A second primary question was whether the effects of context congruency depended on the similarity of emotional valence of the facial expressions and the context. In order to test this, we had three experimental conditions that varied by emotional valence similarity (see Section III.C). A summary of facial expression recognition rates by condition and culture is shown in Table VI.

TABLE VI. RECOGNITION ACCURACY BY CONDITION AND CULTURE

Condition	Context	Western	East Asian
Positive Switch	Congruent	100.0%	87.5%
	Incongruent	75.0%	62.5%
Negative Switch	Congruent	87.5%	87.5%
	Incongruent	16.7%	37.5%
Cross Switch	Congruent	87.5%	87.5%
	Incongruent	65.6%	46.9%

As can be seen in the table, congruent context produced fairly stable recognition rates across conditions and culture. However, incongruent context produced recognition rates that varied significantly depending on the condition. The pattern (positive < cross < negative) was the same for both cultural groups, although the specific values differed. In short, perceptions of negative-valence emotional expressions (Sad, Fear, Anger) were more heavily affected by context congruency than their positive-valence counterparts. Switching across valence fell somewhere in between.

The patterns were investigated for significance via a two-way, *between-subjects* ANOVA (see Section III.D). The results are shown in Table VII. Significant effects were found for condition ($p < .001$) – i.e. emotional valence similarity – but not for cultural background. The interaction effect was not significant. Post-hoc Bonferroni *t*-tests revealed the significant differences were between the negative switch condition (Condition 2) and the other two.

TABLE VII. CONDITION BY CULTURE - ANOVA

	Main Accuracy	
	F	Sign.
Culture	0.503	0.482
Condition	9.577	0.000*
Culture * Condition	1.94	0.157

**Significant values ($p < .05$) marked with an asterisk

In summary, the similarity between emotional valence of the facial expressions and the context had a significant effect on in human perceptions of robotic facial expressions. In particular, the effects were significantly larger for facial expressions of negative-valence emotions. These results suggest that there may be predictable patterns in the effects of congruent/incongruent environmental context on perceptions of robot affect, regardless of culture.

V. DISCUSSION

We performed an empirical study ($n=48$) investigating the effects of context congruency on perceptions of robotic facial expressions across cultures. There were two key findings. First, context congruency had a significant effect on human perceptions of robotic facial expressions. This effect occurred regardless of culture, and was even of similar size. Providing incongruent context was worse than providing no context at all. Second, the similarity of emotional valence between the context and the facial expressions played a significant role, whereas negative-valence emotions were more affected by context congruency. Again, this effect occurred regardless of culture.

The results suggest that there may be predictable patterns in the effects of environmental context on perceptions of robot affect, regardless of culture. These patterns are shaped by the congruence/incongruence of the context, as well as its emotional valence. As has been suggested previously, these patterns fit the notion that emotions which humans perceive 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 [11,12,42]. In other words, humans appear to be *projecting* their own internal emotions. More broadly, this fits into recent cognitive science research on social cognition as an emergent phenomenon [43-45].

From a robot design standpoint, understanding these sorts of phenomena holds great potential to enhance socially interactive robots and human-robot interaction. Inducement of certain external context effects (see Section I) may allow us to shape the interaction without necessarily redesigning the robot itself. Moreover, given the predictable patterns of context effects, such an approach may allow us to produce *culture-neutral models* of robots and affective interaction [11]. In other words, taking advantage of context in the dynamical process of perception formation may ease the constraints for developing culturally-specific affective cues in human-robot interaction. The goal is still to design robots in culturally relevant ways, but such an approach allows us to do so in a more flexible manner [46, 47]. In short, it may not make sense to design robots *in toto* for specific cultures (especially since culture itself is dynamic and constantly in

flux), but rather to design robots that are sensitive and adaptive to particular cultural factors.

ACKNOWLEDGMENT

The authors would like to thank our collaborators in both Japan (Dr. Seite Koike, Chinatsu Nunomura) and in the United States (Kay Jessee, Amyra Asamoah). Funding was provided by NSF grant #IIS-1143712, as well as the NSF/IU IGERT summer research award in cognitive science.

REFERENCES

- [1] R.W. Picard, *Affective Computing*. Boston, MA, USA: MIT Press, 1997.
- [2] J.J. Gross and R.W. Levenson, "Emotion elicitation using films," *Cognition and Emotion*, vol.9, no.1, pp.87–108, 1995.
- [3] C. Breazeal, "Emotion and sociable humanoid robots," *Int J Hum Comput Stud*, vol. 59, no. 1–2, pp.119– 155, 2003.
- [4] S. Sosnowski, A. Bittermann, K. Kuhlentz, and M. Buss, "Design and evaluation of emotion-display EDDIE," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006, pp.3113–3118.
- [5] L. Canamero and J. Fredslund, "I show you how I like you - can you read it in my face?" *IEEE Trans Syst Man Cybern A Syst Hum*, vol.31, no.5, pp.454–459, 2001.
- [6] D. Bazo, R. Vaidyanathan, A. Lentz, and C. Melhuish, "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)*, Taipei, Taiwan, 2010, pp.5317–5322.
- [7] J. Saldien, K. Goris, B. Vanderborght, J. Vanderfaellie, and D. Lefeber, "Expressing emotions with the social robot Probo," *Int J of Social Robotics*, vol.2, no.4, pp.377–389.
- [8] C. Becker-Asano and H. Ishiguro, "Evaluating facial displays of emotion for the android robot Geminoid F," *Proceedings of the IEEE Workshop on Affective Computational Intelligence (WACI)*, 2011, pp.1–8.
- [9] N. Matsumoto, H. Fujii, and M. Okada, "Minimal design for human-agent communication," *Artificial Life and Robotics*, vol.10, no.1, pp.49–54, 2006.
- [10] H. Kozima, M.P. Michalowski, and C. Nakagawa, "Keepon: A playful robot for research, therapy, and entertainment." *Int J of Social Robotics*, vol.1, no.1, pp.3–18, 2009.
- [11] C.C. Bennett and S. Šabanović, "The effects of culture and context on perceptions of robotic facial expressions," *Interaction Studies*, In Submission, 2014. Available from: http://www.caseybennett.com/uploads/HRI_CultureContext_Paper_Final.pdf
- [12] L.F. Barrett, B. Mesquita, and M. Gendron, "Context in emotion perception." *Current Directions in Psychological Science*, vol.20, no.5, pp.286–290, 2011.
- [13] M.E. Kret, J.J. Stekelenburg, K. Roelofs, and B. de Gelder, "Perception of face and body expressions using electromyography, pupillometry, and gaze measures," *Frontiers in Psychology*, vol.4, pp.28, 2013.
- [14] M.E. Kret and B. de Gelder, "Social context influences recognition of bodily expressions," *Exp Brain Res*, vol.203, no.1, pp.169–180, 2010.
- [15] C.C. Bennett and S. Šabanović, "Perceptions of affective expression in a minimalist robotic face," *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 2013, pp.81–82.
- [16] C.C. Bennett and S. Šabanović, "Deriving minimal features for human-like facial expressions in robotic faces," *Int J of Social Robotics*, In Press, 2014. Available from: http://www.caseybennett.com/uploads/MiRAE_Paper_Final.pdf
- [17] C. Darwin, *The Expression of the Emotions in Man and Animals*. London, UK: John Murray, 1872.
- [18] P. Ekman, "Darwin's contributions to our understanding of emotional expressions," *Philos Trans R Soc Lond B Biol Sci*, vol.364, no.1535, pp.3449–3451, 2009.
- [19] N.L. Nelson and J.A. Russell, "Universality Revisited," *Emotion Review*, vol.5, no.1, pp.8–15, 2013.
- [20] M. Pantic, "Machine analysis of facial behaviour: naturalistic and dynamic behavior," *Philos Trans R Soc Lond B Biol Sci*, vol.364, no.1535, pp.3505–3513, 2009.
- [21] J.F. Cohn, "Advances in behavioral science using automated facial image analysis and synthesis," *IEEE Signal Process Mag*, vol.27, no.6, pp.128–133, 2010.
- [22] D.J. Schiano, S.M. Ehrlich, K. Rahardja, and K. Sheridan, "Face to interface: facial affect in (hu)man and machine," *SIGCHI conference on Human Factors in Computing Systems*, 2000, pp.193–200.
- [23] M. Yuki, W.W. Maddux, and T. Masuda, "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*, vol.43, no.2, pp.303–311, 2007.
- [24] R.E. Jack, C. Blais, C. Scheepers, P.G. Schyns, and R. Caldara, "Cultural confusions show that facial expressions are not universal," *Curr Biol*, vol.19, no.18, pp.1543–1548, 2009.
- [25] T. Koda, Z. Ruttkay, Y. Nakagawa, and K. Tabuchi, "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. Berlin: Springer, 2010, pp.16–27.
- [26] J. Arizpe, D.J. Kravitz, G. Yovel, and C.I. Baker, "Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use," *PLoS ONE*, vol.7, no.2, pp.e31106, 2012.
- [27] C. Blais, C. Roy, D. Fiset, M. Arguin, and F. Gosselin, "The eyes are not the window to basic emotions," *Neuropsychologia*, vol.50, no.12, pp.2830–2838, 2012.
- [28] M.F. Peterson and M.P. Eckstein, "Looking just below the eyes is optimal across face recognition tasks," *Proc Natl Acad Sci USA*, vol.109, no.48, pp.E3314–3323, 2012.
- [29] R.E. Nisbett, K. Peng, I. Choi, and A. Norenzayan, "Culture and systems of thought: Holistic versus analytic cognition," *Psychol Rev*, vol.108, no.2, pp.291–310, 2001.
- [30] B. Shore, *Culture in Mind: Cognition, Culture, and the Problem of Meaning*. Oxford, UK: Oxford University Press, 1996.
- [31] P. Ekman, "Universals and cultural differences in facial expressions of emotion," *Nebraska Symposium on Motivation*, vol.19, pp.207–283, 1971.
- [32] H.A. Elfenbein, "Nonverbal dialects and accents in facial expressions of emotion," *Emotion Review*, vol.5, no.1, pp.90–96, 2013.
- [33] R. Righart and B. de Gelder, "Recognition of facial expressions is influenced by emotional scene gist," *Cognitive Affective & Behavioral Neuroscience*, vol.8, no.3, pp.264–272, 2008.
- [34] T.H. Lee, J.S. Choi, and Y.S. Cho, "Context modulation of facial emotion perception differed by individual difference," *PLoS ONE*, vol.7, no.3, pp.e32987, 2012.
- [35] L.F. Barrett and E.A. Kensing, "Context is routinely encoded during emotion perception," *Psychol Sci*, vol.21, no.4, pp.595–599, 2010.
- [36] S. Kitayama, B. Mesquita, and M. Karasawa, "Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States," *J Pers Soc Psychol*, vol.91, no.5, pp.890–903, 2006.
- [37] J. Zhang and A.J. Sharkey, "Contextual recognition of robot emotions." In: *Towards Autonomous Robotic Systems*. R. Groß, L. Alboul, C. Melhuish, M. Witkowski, T.J. Prescott, and J. Penders, Eds. Berlin: Springer, 2011, pp.78–89.
- [38] T. Masuda and R.E. Nisbett, "Attending holistically versus analytically: comparing the context sensitivity of Japanese and Americans," *J Pers Soc Psychol* vol.81, no.5, pp.922, 2001.
- [39] S.G. Ko, T.H. Lee, H.Y. Yoon, J.H. Kwon, and M. Mather, "How does context affect assessments of facial emotion? The role of culture and age." *Psychological Aging*, vol.26, no.1, pp.48–59, 2011.
- [40] T. Masuda, P.C. Ellsworth, B. Mesquita, J. Leu, S. Tanida, and E. Van de Veerndonk, "Placing the face in context: cultural differences in the perception of facial emotion," *J Pers Soc Psychol*, vol.94, no.3, pp.365–381, 2008.
- [41] J.M.B. Fugate, "Categorical perception for emotional faces," *Emotion Review*, vol.5, no.1, pp.84–89, 2013.

- [42] K.A. Lindquist and M. Gendron, "What's in a word? Language constructs emotion perception." *Emotion Review*, vol.5, no.1, pp.66–71, 2013.
- [43] L.W. Barsalou, C. Breazeal, and L.B. Smith, "Cognition as coordinated non-cognition," *Cogn Process*, vol. 8, no.2, pp.79–91, 2007.
- [44] H. De Jaegher and E.A. Di Paolo, "Participatory sense-making: An enactive approach to social cognition," *Phenomenology in Cognitive Science*, vol.6, no.4, pp.485-507, 2007.
- [45] H. De Jaegher, E.A. Di Paolo, and S. Gallagher, "Can social interaction constitute social cognition?" *Trends in Cognitive Science*, vol.14, no.10, pp.441-447, 2010.
- [46] S. Šabanović, C.C. Bennett, and H. Lee, "Towards culturally robust robots: A critical social perspective on robotics and culture," *ACM/IEEE HRI 2014 Workshop on Culture Aware Robotics (CARS)*, In Press, 2014. Available from: http://www.caseybennett.com/uploads/Culture_Aware_Robotics.pdf
- [47] M. Rehm, N. Bee, B. Endrass, M. Wissner, and E. André, "Too close for comfort? Adapting to the user's cultural background," *ACM Proceedings of the International Workshop on Human-Centered Multimedia*, 2007, pp.85-94