Chen, C., Hensel, L. B., Duan, Y., Ince, R., Garrod, O. G.B., Beskow, J., Jack, R. E. and Schyns, P. G. (2019) Equipping Social Robots with Culturally-Sensitive Facial Expressions of Emotion Using Data-Driven Methods. In: 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019), Lille, France, 14-18 May 2019, ISBN 9781728100890 (doi:10.1109/FG.2019.8756570).

Equipping social robots with culturally-sensitive facial expressions of emotion using data-driven methods

Chaona Chen*, Laura B. Hensel*, Yaocong Duan*, Robin Ince*, Oliver G.B. Garrod*,
Jonas Beskow†, Philippe G. Schyns* and Rachael E. Jack*

*Institute of Neuroscience and Psychology, University of Glasgow, G12 8QB, Scotland, UK
†Furhat Robotics, 11428 Stockholm, Sweden
Email: rachael.jack@glasgow.ac.uk

Abstract—Social robots must be able to generate realistic and recognizable facial expressions to engage their human users. Many social robots are equipped with standardized facial expressions of emotion that are widely considered to be universally recognized across all cultures. However, mounting evidence shows that these facial expressions are not universally recognized – for example, in East Asian cultures, they elicit significantly lower recognition than in Western cultures. Consequently, without culturally sensitive facial expressions, state-of-the-art social robots are restricted in engaging a culturally diverse range of human users, which limits their usability and global marketability. To develop culturally sensitive facial expressions, novel data-driven methods are used to model the dynamic face movements that convey basic emotions (e.g., happy, sad, anger) in any culture using cultural perception. Here, we tested whether dynamic facial expression models derived in an East Asian culture and transferred to a popular social robot enhance its performance when East Asian participants classify the displayed facial expressions and rate their human-likeness. Results show that, compared to the social robot's existing set of facial 'universal' expressions, the culturally-sensitive facial expression models are generally recognized with higher accuracy and are judged as more humanlike. We also specifically detail the dynamic face movements that produce increased recognition accuracy and judgments of human-likeness, including those that further boost the robot performance. Our results demonstrate the utility of using data-driven, methods based on social perception to derive culturally-sensitive facial expressions, which can substantially improve the performance of social robots. We anticipate that these methods will continue to inform the design of culturally-sensitive social robots and broaden their social signalling capacity, usability, and global marketability.

I. INTRODUCTION

Facial expressions are widely considered to be the universal language of emotion. Based on Darwin's ground-breaking theory on the biological origins of facial expressions of emotion [1] and Ekman's seminal cross-cultural recognition studies (e.g., [2]), several dominant theories in the field of psychology have argued that six basic emotions – happy, surprise, fear, disgust, anger and sad – are expressed and recognized in the same way across all cultures (e.g., [2-7]). To represent these universal facial expressions, the field
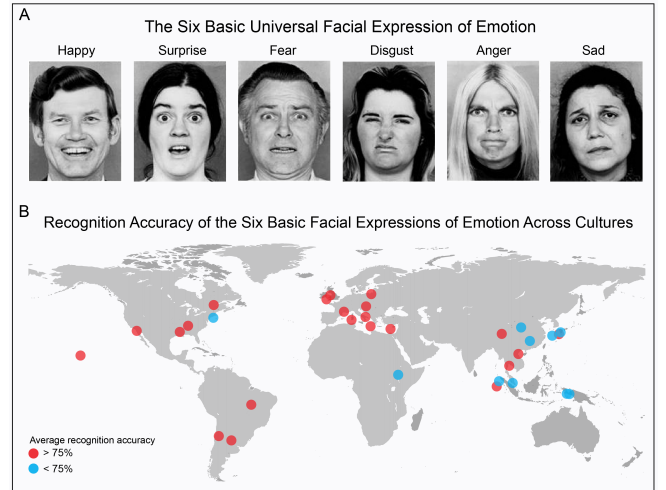
Figure 1. A. Standardized facial expressions of emotion that are widely considered to be universally recognized. B. Color-coded points show the average recognition accuracy of these facial expressions in 40 locations across the world as reported in 15 previous studies [2, 5, 6, 21-32]. Figure adapted from [33] with permission.

established a set of six standardized facial expressions (see Fig. 1A for examples), with each comprising a specific pattern of face movements called Action Units (AUs) such as Nose Wrinkler (AU9), Upper Lid Raiser (AU5) [8]. These standardized facial expression images quickly became the gold standard in research and influenced a broad range of fields including affective computing [see 9 for a review] and social robotics [10-12]. For example, state-of-the-art social robots such as Felix [13], SAYA [14] and Furhat [15, see aslo 16 for a review] generate their facial expressions based on these standardized universal Action Unit patterns. However, mounting evidence shows that these facial expressions are not recognized with similar performance across all cultures. Instead, they elicit significantly lower accuracy in a number of cultures [17, 18, see aslo reviews of 19, 20]. To illustrate, Fig. 1B shows the recognition accuracy reported in 40 locations across the world based on 15 well-known studies using these standardized facial expressions [2, 5, 6, 21-32]
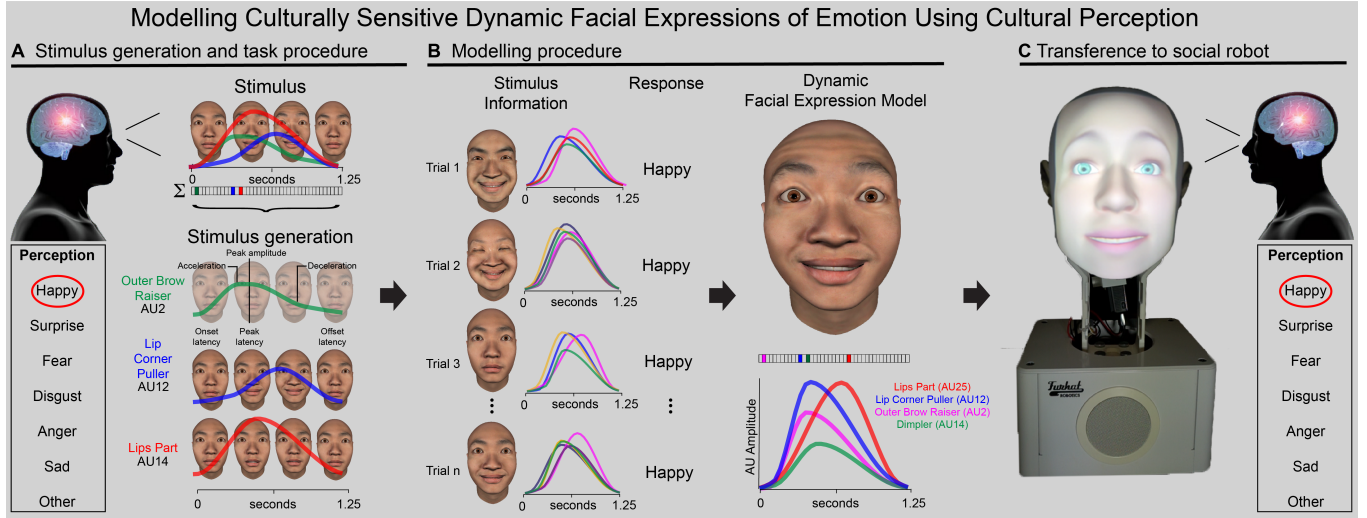
Figure 2. Data-driven, perception-based method to model culturally-sensitive dynamic facial expressions of emotion and transference to social robotics. A. Stimulus generation procedure and perceptual task. B. Facial expression modelling procedure. C. Transference of facial expression models to social robotics and cultural validation.

where red represents high accuracy (i.e., $> 75\%$) and blue represents lower accuracy (i.e., $< 75\%$) [33]. As shown by the distribution of red and blue points, the standardized facial expressions tend to be recognized primarily in Western cultures but less so in East Asian cultures. These consistent cultural differences therefore demonstrate that the facial expressions widely considered to be universal are instead more representative of the social signals of Western culture [see also reviews of 18, 34]. Indeed, such findings of substantial cultural differences are now increasingly common within the field of psychology [17, 35, 36] because the majority of psychological knowledge has been derived from Western (more specifically, Western, educated, industrialized, rich and democratic – WEIRD) [37] populations and from Western-centric theories using confirmatory methods [38]. A further limitation of these standardized facial expressions of emotion is that they are static and therefore cannot represent the naturalistic dynamics of real human facial expressions [39]. Therefore, traditional theory-driven approaches in psychology have restricted knowledge of facial expressions – that is, which dynamic face movement patterns convey these basic emotions in different cultures – which in turn has impacted related fields such as social robotics where expressive capacity is limited primarily to Western cultures.

II. RELATED WORK

To better understand facial expressions of emotion across cultures, new data-driven methods have been used to model the dynamic face movement patterns that convey the six basic emotions in different cultures [e.g., 17]. Fig. 2A-B illustrates this approach. On each experimental trial, cultural participants view a random facial animation that has been generated by a facial animation platform [40],

which randomly samples and combines a subset of facial movements (called dynamic Action Units, AUs) from a core set of 42 AUs. For example, in Fig. 2A, three AUs are selected – Outer Brow Raiser (AU2) color-coded in green, Lip Corner Puller (AU12) in blue, and Lips Part (AU14) in red. Each is activated with a random movement (see color-coded temporal activation curves for each AU; temporal parameters are labelled in the green curve). Participants view the facial animation and classify it according to one of the six emotions ('happy,' 'surprise,' 'fear,' 'disgust,' 'anger' or 'sad') and rate its intensity on a 5-point scale ('very weak' to 'very strong'). If the facial animation does not correspond to any of these emotions, participants select 'other.' Therefore, each facial animation that is classified as a particular emotion contains a face movement pattern that conveys that emotion to the participant. After many such trials, a statistical relationship is built between the dynamic AUs presented on each trial and the participant's corresponding responses (e.g., 'happy') as depicted in Fig. 2B. This procedure produces a statistically robust model of the dynamic facial expression pattern that elicits the perception of a given emotion in a participant from the culture of interest (see [40] for full details of model fitting procedure). Importantly, as these models are quantifiable representations of facial expressions, they can be directly transferred to social robotics to generate culturally-sensitive facial expressions, as illustrated in Fig. 2C. Therefore, this data-driven approach of agnostically sampling face movements and using subjective human perception to isolate the dynamic Action Unit patterns that convey different emotions is particularly suitable for exploring facial expression communication across diverse cultures [38]. Using this approach, Jack, et al. [17] modelled a set of dynamic facial expressions

of the six classic basic emotions using the cultural perception of Western and East Asian participants. Here, we aim to transfer the culturally-derived East Asian dynamic facial expression models to a popular social robot head – Furhat – and examine whether they improve performance compared to the social robot's existing 'universal' facial expressions when tested on East Asian participants.

III. METHOD

A. Transference of culturally-derived dynamic facial expression models to a social robot

To display the dynamic facial expression models on the social robot head, we first supplemented the social robot's existing facial movement vocabulary of 7 pre-set universal facial expressions of emotion (2 happy, 1 of surprise, fear, disgust, anger and sad) with a set of 42 individual dynamic AUs including all combinations (see full details of transforming the AU shape deviation data to the social robot's mesh topologies in [41]). We could then display each of the culturally-derived dynamic facial expression models of the six classic emotions (n = 30 models per emotion) on the social robot head along with the social robot's existing set of 7 facial expressions. In a first experiment, we asked a group of East Asian participants to classify these facial expressions by emotion; in a second experiment East Asian participants judged their human-likeness. For both experiments, we recruited 10 East Asian participants (10 Chinese, 5 females, mean age 23.6 years, SD = 2.12 years) with minimal exposure to and engagement with other cultures as assessed by a questionnaire (see *Supplementary Material, Screening Questionnaire* for full details). All participants had normal or corrected-to-normal vision, were free from any emotion related atypicalities (e.g. Autism Spectrum Disorder, depression), learning difficulties (e.g. dyslexia), synaesthesia, and disorders of face perception (e.g. prosopagnosia) as per self-report. All participants had a minimum International English Language Testing System (IELTS) score of 6.0 (competent user). Each participant gave written informed consent, and received a standard rate of 6 per hour for their participation. The Ethics Committee of the College of Science and Engineering, University of Glasgow provided ethical approval (Ref No: 300160186).

B. Recognition of universal facial expressions versus culturally-derived facial expressions

On each trial, participants viewed a facial animation displayed on the social robot head and classified it according to one of six emotions – happy, surprise, fear, disgust, anger or sad – in a 6-alternative forced choice task. Each participant viewed a total of 374 facial animations ([30 culturally-derived facial expression models × 6 emotions] + [7 existing universal facial expressions] × 2 repetitions) presented in random order across the experiment. We presented each
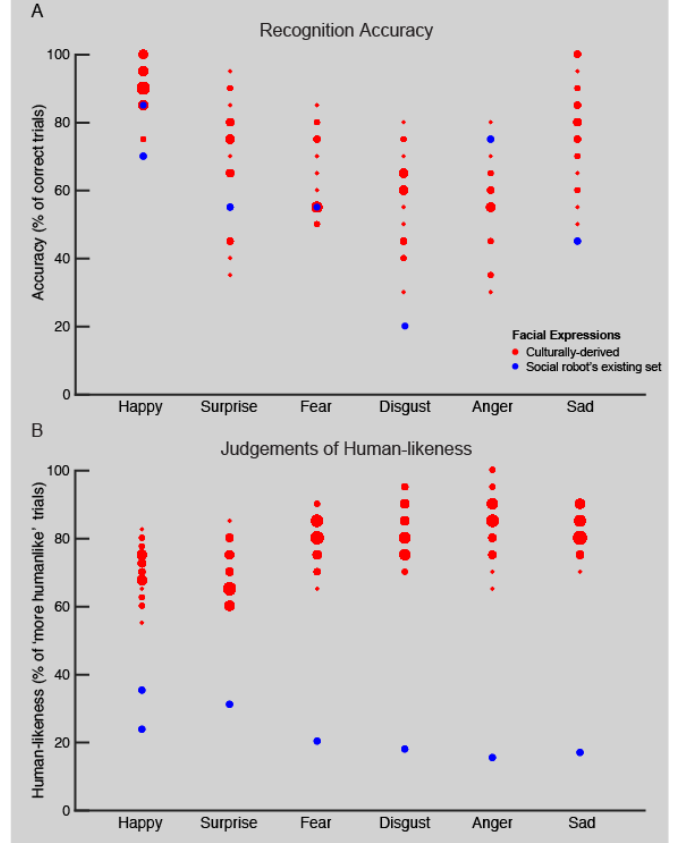

Figure 3.   A. Recognition accuracy of culturally-derived facial expressions and the social robot's existing facial expressions. B. Judgments of human-likeness.

facial animation on one of 7 available face textures ('Default,' 'Male,' 'Female,' 'Obama,' 'iRobot,' 'Gabriel,' and 'Avatar'), pseudo-randomly selected such that each texture appeared an equal number of times across the experiment for each participant. We blocked all trials by face texture and randomized the order of the blocks for each participant. We presented each facial animation once for a duration of 1.25 seconds within the participant's central visual field at a constant viewing distance of 90 cm using a chin rest. Stimuli (size 22.5 cm × 16 cm) subtended 14.25° (vertical) and 10.16° (horizontal) of visual angle, which reflects the average size of a human face [42] during natural social interaction [43]. Following each facial animation, participants responded using a Graphic User Interface (GUI) displayed on a 19-inch flat panel Dell monitor next to the robot head. We instructed participants to respond quickly and accurately. Following response, two beeps cued participants for the next trial. Participants then viewed the social robot and pressed the space bar to start next trial. We used Matlab 2016a to display the GUI and record participant responses.

To compare the recognition accuracy of the culturally-derived facial expression models with the social robot's

existing facial expressions, we computed the proportion of correct responses for each facial expression model (n = 30 per emotion) and each of the social robot's existing facial expressions (n = 7 total) by pooling the data across all trials and participants. Fig. 3A shows the results for each emotion. Red circles represent each culturally-derived facial expression model; blue represents the social robot's existing facial expressions. Circle size represents the number of facial expression models with a specific accuracy (e.g., in happy, 6 models are recognized at 95% accuracy, see the face map of each model and accuracy in *Supplemental Material, Action Unit Activations of the Culturally Derived Models by Recognition Accuracy*). As shown by the distribution of red circles in each emotion category, the majority of the culturally-derived facial expression models elicited higher recognition accuracy than the social robot's existing facial expressions with the exception of anger where only 1 model showed higher performance than the social robot's existing facial expressions.

C. Comparing judgments of humanlike-ness of universal facial expressions and culturally-derived facial expressions

Next, we compared judgments of human-likeness of the culturally-derived facial expression models and the social robot's existing facial expressions. On each trial, we presented a pair of facial expressions of the same emotion (e.g., happy) – one culturally-derived facial expression and one of the social robot's existing facial expressions displayed on the same face texture – and asked participants to choose which one looked most humanlike. We played each facial expression once for a duration of 1.25 seconds with an inter-stimulus interval (ISI) of 0.5 second and presented in a pseudo-random sequential order across the experiment. After displaying each pair of facial expressions, one beep sounded to cue participants to respond. Participants indicated which facial expression they thought was more humanlike using a GUI displayed on a 19-inch flat panel Dell monitor next to the social robot head. Following response, two beeps sounded to cue participants for the next trial. Participants viewed the social robot and pressed the space bar to start the next trial. We randomly assigned one of 7 available face textures to each emotion, blocked trials by emotion, and randomized the order of the blocks for each participant. Therefore, each participant completed a total of 420 trials ([60 pairs of happy facial expressions + 30 pairs of facial expressions for each of the other 5 emotions] × 2 pair orders). We used the same viewing conditions and equipment as in B above.

To compare judgments of human-likeness between the culturally-derived facial expression models and the social robot's existing facial expressions, we computed the proportion of times participants selected each facial expression as more humanlike by pooling trials across all trials and participants. Fig. 3B shows the results. The face maps

in *Supplemental Material, Action Unit Activations of the Culturally Derived Models by Human-Likeness Judgements* show the AUs in each model and humanlike judgements. As shown by the distribution of red points in Fig. 3B, participants consistently judged the culturally-derived facial expression models as more humanlike than the social robot's existing facial expressions.

D. Identifying the dynamic face movements associated with performance

We showed that the culturally-derived facial expression models are recognized with higher accuracy and are judged as more humanlike compared to the social robot's existing universal facial expressions. To identify which specific face movements – that is, the presence of a given AU and/or its specific dynamic properties – are associated with these improved performances, we used an information-theoretic approach based on mutual information (MI) [44, 45]. Specifically, MI quantifies the relationship between two variables – here, the presence of an AU and performance (i.e., recognition accuracy or judgments of human-likeness) of a facial expression model. High MI would indicate that an AU (e.g., Brow Lowerer, AU4) is strongly associated with performance (e.g., correct emotion classifications); low MI indicates a weak relationship. To identify, for each emotion, the AUs that are strongly associated with performance, we applied the following analysis for recognition accuracy and human-likeness separately: We computed the MI between each AU (i.e., present or absent in the culturally-derived facial expression model) and performance (e.g., correct emotion classifications) by pooling the participants' responses to the culturally-derived facial expressions displayed in *B. Recognition of universal facial expressions versus culturally-derived facial expressions*, resulting in 600 trials per emotion (30 models x 10 participants x 2 repetitions). Given that some AUs are present in 100% of the facial expression models – e.g., in happy, Lip Corner Puller (AU12) – which provides no variance to successfully compute MI, we added noise to 1% of the AU patterns by randomly allocating AU absence (i.e., 0 for absence, 1 for presence) before computing MI. We established the statistical significance of high MI values using a Monte Carlo simulation method by shuffling the participants' responses 1000 times, computing MI for each AU at each iteration, and using the random distribution of MI values to identify the AUs with MI values that are significantly higher than chance (i.e., > 95% of the distribution, uncorrected). All AUs with significantly high MI are displayed on face maps in Fig 4 for recognition accuracy (Panel A) and human-likeness (Panel B) with their Action Unit labels listed to the right of each face map in regular font.

Certain AUs could also improve performance based on their specific dynamic properties such as high amplitude, early peak latency, or fast acceleration. To identify any
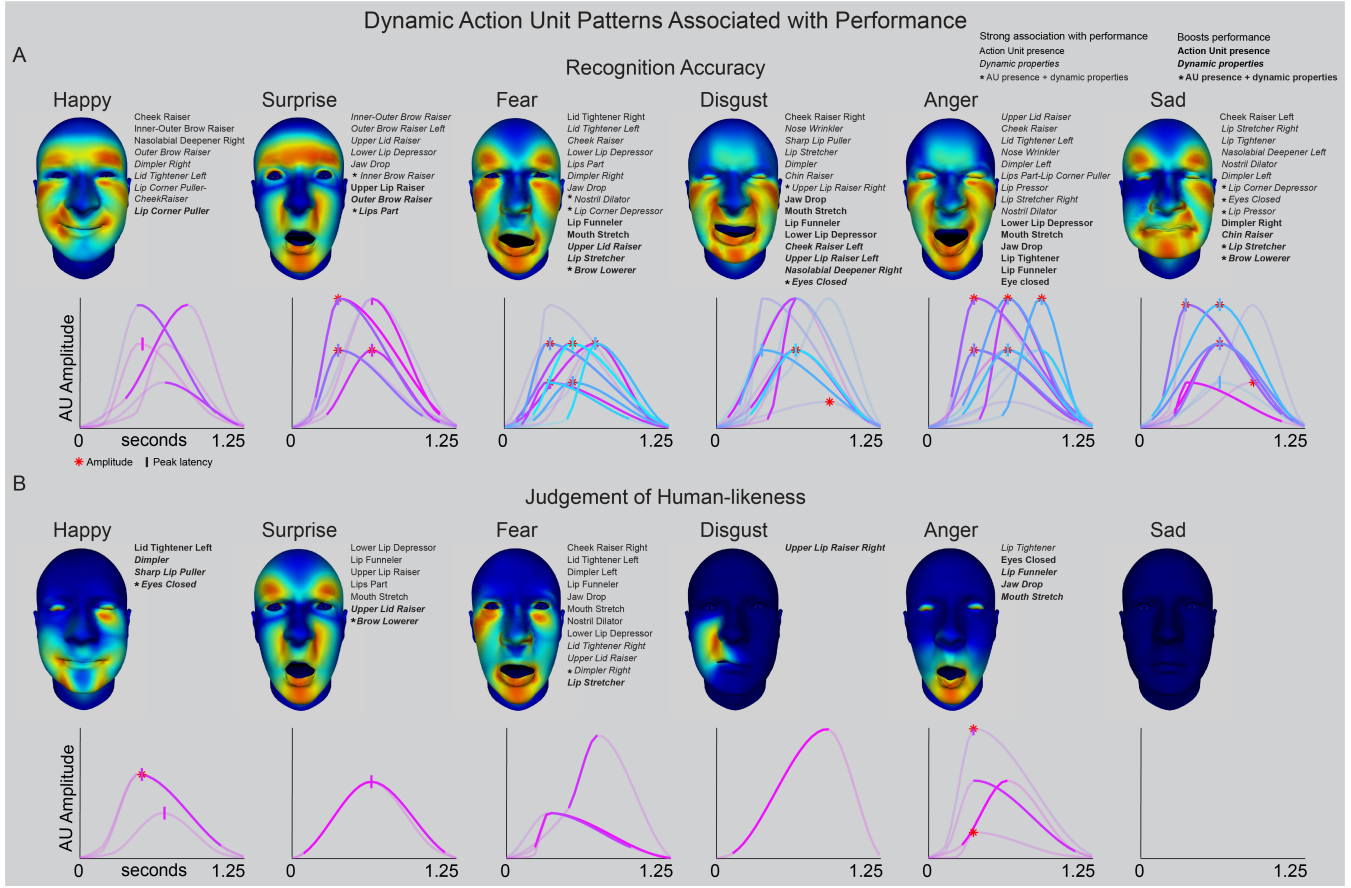
Figure 4. Dynamic Action Unit patterns associated with high recognition accuracy and judgments of human-likeness.

such AUs, we computed the MI between each dynamic property (e.g., amplitude) and performance (e.g., correct or incorrect emotion classifications). Specifically, we computed MI between performance and three levels of AU temporal parameter values (high, medium, low) for each of four temporal parameters – amplitude, peak latency, acceleration, deceleration – separately. We established statistical significance of high MI values for each and each temporal parameter using a Monte Carlo method as described above. AU dynamics with significantly high MI are displayed in the face maps in Fig 4 with their AU names listed in *italics* (see legend on top right). Next, to further specify the particular level of dynamic properties (e.g., high, medium, or low amplitude) of each of these AUs that drives high performance, we computed the frequency of each level of dynamic properties in the high-performance trials (e.g., correct emotion classifications) and identified the specific level of dynamic properties with the highest frequency. We did this for each of four temporal parameters – amplitude, peak latency, acceleration, deceleration – separately. The specific dynamic properties of each AU are displayed as activation curves below each face (significant results are

presented in saturated colour; non-significant parameters are interpolated and presented in more transparent colour). AUs with both their presence and dynamic properties associated with high performance are marked with asterisks (see legend on top right). Together, these results show that for each emotion, several specific AUs and/or their specific dynamic properties are strongly associated with recognition accuracy and with judgments of human-likeness. For example, in happy, Inner-Outer Brow Raiser (AU1-2) is strongly associated with recognition accuracy, as are the specific dynamic properties of Lip Corner Puller - Cheek Raiser (AU12-6). For perceptions of human-likeness, in fear, the dynamic properties of Upper Lid Raiser (AU5) is strongly associated with performance.

E. Dynamic Action Units that further boost performance

Above, we identified the individual AUs and the dynamic properties of AUs that are strongly associated with (and therefore important for) performance – i.e., the correct classification of emotions and for judgments of human-likeness. As shown in Fig 3, certain facial expression models elicit particularly high performance and thus comprise specific face movements that further boost performance. As

described above, these AUs could boost performance by their presence alone or by their specific dynamic properties such as high amplitude, early peak latency, or fast acceleration. To identify and characterize these specific dynamic AUs for each emotion, we first identified the AUs that are strongly associated specifically with high performance by computing the point-wise mutual information (PMI) between the presence of each AU and each of three levels of performance (i.e., 'low accuracy,' 'medium accuracy' or 'high accuracy;' and the same for human-likeness). For example, high PMI would indicate that the presence of an AU specifically drives high accuracy. We established statistical significance with the Monte Carlo method described above. These high-performance AUs are displayed in the face maps in Fig 4, with their names listed in **bold**. Next, to identify whether specific dynamic AU properties also boost performance, we conducted a similar PMI analysis applied separately to each of four temporal parameters – amplitude, peak latency, acceleration, deceleration – using three bins of parameter values for each (e.g., 'low amplitude,' 'medium amplitude' and 'high amplitude'). These AUs are displayed on the face maps in Fig 4 with their AU name in ***bold italics*** and the specific dynamic properties shown as activation curves below each face. AUs with significantly high PMI for their presence and dynamics are marked with asterisks. For example, in fear, Brow Lowerer (AU4) and Mouth Stretch (AU27) further boost recognition accuracy. In disgust, the dynamic properties of Upper Lip Raiser Right (AU10R) boost performance for judgments of human-likeness.

IV. CONCLUSIONS

Here, we transferred a set of 30 culturally-derived dynamic facial expression models to a popular social robot and compared their recognition accuracy and judgments of humanlike-ness with the social robot's existing universal facial expressions amongst a group of East Asian participants. Results show that these culturally-derived dynamic facial expression models generally outperformed the social robot's existing facial expressions on both emotion recognition accuracy and judgments of human-likeness. Further analysis of the facial expression models revealed the specific Action Units and temporal dynamic properties that drive improved performance, including those that further boost performance towards higher recognition accuracy and perceptions of human-likeness. Together, our results highlight the advantage of using culturally valid dynamic face movements to accurately signal social messages. Our results also demonstrate the power of using subjective cultural perception to model dynamic facial expressions to improve the performance of social robots within a culturally diverse global market. We anticipate that the application of data-driven approaches will further inform the design of culturally-sensitive digital agents for improved performance and usability with more diverse range of user groups.

REFERENCES

[1] C. Darwin, *The Expression of the Emotions in Man and Animals*, 3rd ed. London: Fontana Press, 1999/1872.

[2] P. Ekman, E. R. Sorenson, and W. V. Friesen, "Pan-Cultural Elements in Facial Displays of Emotion," *Science*, vol. 164, pp. 86-88, April 4, 1969 1969.

[3] P. Ekman and W. Friesen, "A new pan-cultural facial expression of emotion," *Motivation and Emotion*, vol. 10, pp. 159-168, 1986.

[4] D. Matsumoto, "American-Japanese Cultural Differences in the Recognition of Universal Facial Expressions," *Journal of Cross-Cultural Psychology*, vol. 23, pp. 72-84, 1992.

[5] M. Biehl, D. Matsumoto, P. Ekman, V. Hearn, K. Heider, T. Kudoh, et al., "Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences," *Journal of Nonverbal Behavior*, vol. 21, pp. 3-21, 1997.

[6] J. D. Boucher and G. E. Carlson, "Recognition of Facial Expression in Three Cultures," *Journal of Cross-Cultural Psychology*, vol. 11, pp. 263-280, 1980.

[7] Kimiko Shimoda, M. Argyle, and P. R. Bitti, "The intercultural recognition of emotional expressions by three national racial groups: English, Italian and Japanese," *European Journal of Social Psychology*, vol. 8, pp. 169-179, 1978.

[8] P. Ekman and W. V. Friesen, *Manual for the Facial Action Coding System*. Palo Alto, CA: Consulting Psychologists Press, 1978.

[9] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, pp. 1424-1445, 2000.

[10] N. Lazzeri, D. Mazzei, A. Greco, A. Rotesi, A. Lanat, and D. E. De Rossi, "Can a humanoid face be expressive? A psychophysiological investigation," *Frontiers in bioengineering and biotechnology*, vol. 3, p. 64, 2015.

[11] C. C. Bennett and S. Sabanovic, "Deriving minimal features for human-like facial expressions in robotic faces," *International Journal of Social Robotics*, vol. 6, pp. 367-381, 2014.

[12] E. G. Krumhuber and K. R. Scherer, "The look of fear from the eyes varies with the dynamic sequence of facial actions," *Swiss Journal of Psychology*, 2016.

[13] L. Canamero and J. Fredslund, "I show you how I like you-can you read it in my face?[robotics]," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, vol. 31, pp. 454-459, 2001.

[14] T. Hashimoto, S. Hitramatsu, T. Tsuji, and H. Kobayashi, "Development of the face robot SAYA for rich facial expressions," in *SICE-ICASE, 2006. International Joint Conference*, 2006, pp. 5423-5428.

[15] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granstrm, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive behavioural systems*, ed: Springer, 2012, pp. 114-130.

[16] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, pp. 143-166, 2003.

[17] R. E. Jack, O. G. B. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 7241-7244, 2012 2012.

[18] H. A. Elfenbein and N. Ambady, "On the universality and cultural specificity of emotion recognition: A meta-analysis," *Psychological Bulletin*, vol. 128, pp. 203-235, 2002.

[19] B. Mesquita and N. H. Frijda, "Cultural variations in emotions: a review," *Psychological Bulletin*, vol. 112, pp. 179-204, Sep 1992.

[20] J. A. Russell, "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies," *Psychological Bulletin*, vol. 115, pp. 102-41, Jan 1994.

[21] L. Ducci, L. Arcuri, and T. Sineshaw, "Emotion Recognition in Ethiopia The Effect of Familiarity with Western Culture on Accuracy of Recognition," *Journal of Cross-Cultural Psychology*, vol. 13, pp. 340-351, 1982.

[22] P. Ekman, "Universals and cultural differences in facial expressions of emotion," presented at the Nebraska Symposium on Motivation, 1972.

[23] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, et al., "Universals and cultural differences in the judgments of facial expressions of emotion," *Journal of Personality and Social Psychology*, vol. 53, pp. 712-7, Oct 1987.

[24] H. A. Elfenbein and N. Ambady, "Universals and Cultural Differences in Recognizing Emotions," *Current Directions in Psychological Science*, vol. 12, pp. 159-164, October 1, 2003 2003.

[25] H. A. Elfenbein, M. Mandal, N. Ambday, S. Harizuka, and S. Kumar, "Hemifacial differences in the in-group advantage in emotion recognition," *Cognition and Emotion*, vol. 18, pp. 613-629, 2004.

[26] Y. Huang, S. Tang, D. Helmeste, T. Shioiri, and T. Someya, "Differential judgement of static facial expressions of emotions in three cultures," *Psychiatry and clinical neurosciences*, vol. 55, pp. 479-483, 2001.

[27] R. E. Jack, C. Blais, C. Scheepers, P. G. Schyns, and R. Caldara, "Cultural confusions show that facial expressions are not universal," *Current Biology*, vol. 19, pp. 1543-1548, 2009.

[28] G. Kirouac and F. Y. Dore, "Accuracy of the judgment of facial expression of emotions as a function of sex and level of education," *Journal of Nonverbal Behavior*, vol. 9, pp. 3-7, 1985.

[29] G. Kirouac and F. Y. Dor, "Accuracy and latency of judgment of facial expressions of emotions," *Perceptual and Motor Skills,* vol. 57, pp. 683-686, 1983.

[30] D. Matsumoto and P. Ekman, "American-Japanese cultural differences in intensity ratings of facial expressions of emotion," *Motivation and Emotion*, vol. 13, pp. 143-157, 1989.

[31] F. T. McAndrew, "A cross-cultural study of recognition thresholds for facial expressions of emotion," *Journal of Cross-Cultural Psychology*, vol. 17, pp. 211-224, 1986.

[32] T. Shioiri, T. Someya, D. Helmeste, and S. W. Tang, "Misinterpretation of facial expression: A cross-cultural study," *Psychiatry and clinical neurosciences*, vol. 53, pp. 45-50, 1999.

[33] R. E. Jack, "Culture and facial expressions of emotion," *Visual Cognition*, vol. 21, pp. 1248-1286, Sep 1 2013.

[34] N. L. Nelson and J. A. Russell, "Universality revisited," *Emotion Review*, vol. 5, pp. 8-15, 2013 2013.

[35] C. Chen, C. Crivelli, O. G. B. Garrod, P. G. Schyns, J.-M. Fernndez-Dols, and R. E. Jack, "Distinct facial expressions represent pain and pleasure across cultures," *Proceedings of the National Academy of Sciences*, 2018.

[36] C. Crivelli, J. A. Russell, S. Jarillo, and J.-M. Fernndez-Dols, "The fear gasping face as a threat display in a Melanesian society," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 12403-12407, 2016.

[37] J. Henrich, S. Heine, and A. Norenzayan, "The weirdest people in the world?," *Behavioral and Brain Sciences*, vol. 33, pp. 61-83, 2010.

[38] R. E. Jack, C. Crivelli, and T. Wheatley, "Data-driven methods to diversify knowledge of human psychology," *Trends in Cognitive Sciences*, vol. 22, pp. 1-5, 2018.

[39] E. Krumhuber, A. S. Manstead, D. Cosker, D. Marshall, P. L. Rosin, and A. Kappas, "Facial dynamics as indicators of trustworthiness and cooperative behavior," *Emotion*, vol. 7, p. 730, 2007.

[40] H. Yu, O. G. B. Garrod, and P. G. Schyns, "Perception-driven facial expression synthesis," *Computers and Graphics*, vol. 36, pp. 152-162, 2012.

[41] C. Chen, O. G. Garrod, J. Zhan, J. Beskow, P. G. Schyns, and R. E. Jack, "Reverse Engineering Psychologically Valid Facial Expressions of Emotion into Social Robots," in *Automatic Face and Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, 2018, pp. 448-452.

[42] L. Ibrahimagic-Seper, A. Celebic, N. Petricevic, and E. Selimovic, "Anthropometric differences between males and females in face dimensions and dimensions of central maxillary incisors," Medicinski glasnik, vol. 3, pp. 58-62, 2006.

[43] E. Hall, The Hidden Dimension. Garden City, NY: Doubleday, 1966.

[44] T. E. Nichols and A. P. Holmes, "Nonparametric permutation tests for functional neuroimaging: a primer with examples," Human Brain Mapping, vol. 15, pp. 1-25, 2002.

[45] T. M. Cover and J. A. Thomas, Elements of information theory: John Wiley and Sons, 2012.