We focus our cultural comparisons on Western and East Asian cultures for several reasons. First, each culture is computer literate, which allows each observer to interact easily and independently (i.e., without social presence of the experimenter) with modern equipment typically used for data collection (e.g., see Sauter et al., 2010 for an illustration of the challenges using such equipment in developing countries). Second, it is well documented that Western and East Asian cultures show marked contrasts in general visual processing (Nisbett & Masuda, 2003), which plays a central role in communication – i.e., the receiving of transmitted visual information patterns (see Jack & Schyns, 2015). Finally, understanding similarities and differences in East Asian and Western social communication is critical in developing the digital economy. For all experiments, we tested all observers in the UK and used strict criteria to select Western and East Asian observers with minimal exposure to and engagement with other cultures (De Leersnyder, Mesquita, & Kim, 2011) as assessed by screening questionnaire (see *Supplemental Materials – Screening Questionnaire*). Specifically, all Western and East Asian observers had never lived in or visited a non-Western/East Asian country before, had never had any close relationships with a non-Western/non-East Asian culture individual (e.g., boy/girlfriend) or had any interest in non-Western/East Asian culture (e.g., sports or art societies). All East Asian observers were Chinese nationals of Chinese heritage, who had arrived in the UK for the first time, had a maximum UK residence of 2.5 months at the time of testing, and possessed a minimum International English Testing System (IELTS) score of 6.0 (Competent user). All Western observers were British nationals of white Caucasian ethnicity. All observers had normal or corrected-to-normal vision and were free from any lexical, reading, language (e.g., dyslexia) or emotion related atypicalities (e.g., Autism Spectrum Disorder, depression, anxiety) as per self-report, and typically from a student population. We paid each observer £6 per hour, and obtained their written informed consent. The University of Glasgow College of Science and Engineering Ethics Committee authorized the experimental protocol.

**Method**

To generate a core set of highly familiar and highly typical emotion words in each language, we proceeded in five steps.

*1. Extracting emotion words from key sources.* First, we compiled a comprehensive list of emotion words in British English and Chinese separately, sourced from established corpuses (e.g., British National Corpus – BNC), word lists (Cai & Brysbaert, 2010) and literature sources (Averill, 1983; Bedford, 2004; Davitz, 1969; de Rivera, 1977; Ho, Fu, & Ng, 2004; Li, Wang, & Fischer, 2004; Ortony & Turner, 1990; Parrott, 2001; Plutchik, 1980).

*2. Word familiarity and emotion typicality rating task.* To extract highly familiar and highly typical emotion words, we first obtained measures of familiarity and emotion typicality using a rating task with native speakers. We recruited a new group of 50 native English speakers (50 British, 25 male, mean age 21.5 years, SD 2.5 years) and 50 native Chinese speakers (50 Chinese, 26 male, mean age 24.1 years, SD 1.7 years)*.* On each experimental trial, native speakers viewed a single word in their native language and rated it according to a) familiarity on a 7-point Likert scale (‘totally unfamiliar’ to ‘highly familiar’) or b) emotion typicality on a 7-point Likert scale (‘definitely not an emotion’ to ‘definitely an emotion’) including ‘I don’t know this word’ using a Graphic User Interface (GUI). We presented all words in random order across the experiment and blocked and counterbalanced the tasks of familiarity and emotion typicality across native speakers in each group. We presented all words in lower case white font (MS Sans Serif for English; MS Song for Chinese) on a black background in the center of the screen, with all Chinese words presented in simplified form. Following the experiment, we then selected words rated as both highly familiar and highly typical of emotion (i.e., >=6 on both scales) by a large majority of native speakers (85%). See *Supplemental Materials – Table S1* and *S2* for the proportion of native speakers rating each word as highly familiar and highly typical of emotion for English and Chinese words, respectively.

*3. Commonly used emotion terms.* Finally, to relate our data to the majority of existing literature, we retained any emotion words that are widely discussed in the literature – e.g., ‘happy,’ ‘surprise,’ ‘fear,’ ‘disgust,’ ‘anger,’ ‘sad,’ ‘shame,’ ‘contempt,’ ‘pride’ and ‘embarrassment.’ Specifically, for English we retained six such emotion words – ‘surprise’, ‘disgust,’ ‘shame,’ ‘contempt,’ ‘pride,’ and ‘embarrassment.’ For Chinese, we retained five such emotion words (as determined by closest match translation provided by a professional translator) – ‘disgusted’/厌恶, ‘shame’/羞愧, ‘contempt’/鄙视, ‘pride’/骄傲 and ‘embarrassment’/尴尬.

*4. Semantic similarity of emotion words in each language.* Next, to understand the semantic relations between the emotion words in each language, we instructed a new set of 49 native English speakers (23 male, mean age 22.5 years, SD 2.2 years) and 50 native Chinese speakers (25 male; mean age 23.0 years, SD 1.9 years) to rate the semantic similarity of all word pairs in their own language. On each experimental trial, native speakers viewed a word pair (e.g., ‘happy’ and ‘fury’) and rated it according to similarity of meaning on a 7-point bipolar scale ranging from ‘very different’ to ‘very similar.’ Each native speaker rated all possible word pair combinations (excluding identical word pairs and including word order reversal) using a GUI, with words presented in random order across the experiment and positioned side-by-side in lower case white font (MS Sans Serif for English; MS Song for Chinese) on a black background in the center of the screen.

*5. Clusters of semantically similar emotion words in each language.* To identify clusters of semantically similar emotion words in each language, we first computed a semantic network of emotion words by calculating the average (mode) similarity of each word pair across native speakers in each group separately. In Figures 1 and 2, color-coded similarity matrices show the semantic relationship between all word pairs in English (Figure 1) and Chinese (Figure 2) where lighter squares indicate high similarity (e.g., in English, ‘delighted’ and ‘joy’) and darker squares indicate low similarity between word pairs (e.g., in Chinese, ‘anguish’/苦闷 and ‘pleasantly surprised’/惊喜). We then applied a graph-theoretic clustering method (Pavan & Pelillo, 2007) to each similarity matrix, revealing several clusters in each. In Figures 1 and 2, green brackets to the left of each matrix show the emotion words that form each cluster (i.e., words that are highly semantically similar) in each language.

*6. Valence, arousal and dominance of emotion words in each language.* To interpret each emotion word cluster, we obtained ratings of three main components of emotion – valence (e.g., pleasantness or unpleasantness), arousal (e.g., excited or calm) and dominance (e.g., in control or being controlled) for each word in both languages. For the English words, we extracted data from an existing database (Warriner, Kuperman, & Brysbaert, 2013). In the absence of a comparable database for Chinese words, we recruited a new set of 32 native Chinese speakers (15 male, mean age 23.3 years, SD 1.7 years) and instructed each to rate the valence, arousal and dominance of each Chinese emotion word used in the semantic network above using a similar procedure as in Warriner et al. (2013). On each experimental trial, native speakers viewed a single word in their native language and rated it according to valence, arousal or dominance on a 9-point Likert scale ranging from happy [excited/ dominated] to unhappy [calm/dominant] (depending on the dimension) using a Graphic User Interface (GUI). We blocked the dimensions of valence, arousal and dominance, with blocks presented in random order across the experiment. All Chinese words appeared in each block and presented in random order for each block. Thus, each native speaker completed 159 trials (53 words x 3 dimensions). We presented all words in white font (MS Song) simplified form on a black background in the center of the screen. To compute the average valence, arousal and dominance ratings for each Chinese word, we first normalized each native speakers responses across each dimension separately. We then computed the median valence, arousal and dominance rating across all native speakers for each word (see *Supplemental Materials – Table S3* for a full list of values). In Figures 1 and 2, color-coded circles next to each word shows the average perceived valence (red – high; blue – low), arousal (green – high; yellow – low) and dominance (magenta – high; cyan – low; see key to left of labels) as judged by native speakers in each group.

**Method**

To illustrate, consider that we aim to identify in a specific culture (e.g., East Asian) the facial movements that communicate different types of aggression (e.g., ‘fury,’ ‘rage’, ‘livid’, and ‘anger’) at different levels of intensity. On each experimental trial, a dynamic facial expression generation platform (the Generative Face Grammar - GFG; Yu et al., 2012) creates a stimulus by randomly selecting a biologically plausible combination of AUs from a core set of 42 AUs using a binomial distribution (minimum 1 AU, maximum 6 AUs, median 3 AUs). As shown in Figure 3, on this illustrative trial three AUs are selected – Upper Lid Raiser (AU5) color-coded in red, Nose Wrinkler (AU9) color-coded in green, and Upper Lip Raiser (AU10) color-coded in blue. The GFG then assigns a random movement to each AU by selecting random values for each of six temporal parameters (onset latency, acceleration, peak latency, peak amplitude deceleration, and offset latency; see labels illustrating the blue curve) from a uniform distribution. The dynamic AUs are then combined to produce a photo-realistic random facial movement, illustrated here with four snapshots. The cultural observer categorizes the stimulus according to a specific emotion (here, ‘rage’) at a given intensity (here, ‘strong’) if the face movements form a pattern that correlates with their prior knowledge (i.e. mental representation) of that facial expression of emotion at that intensity (here, ‘strong rage’). If the pattern does not correspond to any of the categories provided, the observer selects ‘other.’ In other words, as illustrated in Figure 3, this specific dynamic pattern of facial movements has elicited in this observer the response (i.e., categorical perception) ‘strong rage.’ Over many such trials, other patterns might correspond to the other emotion categories (here, ‘fury’, ‘livid’ and ‘anger’) at different levels of intensity. Following the experiment, we can therefore measure the relationship between the dynamic AUs presented on each trial and the observer’s responses, from which we can precisely identify the combinations of AUs (i.e. facial expressions) that reliably correlate with perception of different types of aggression (here, ‘fury’, ‘rage’, ‘livid,’ ‘anger’).

To model the cultural facial expressions associated with each emotion word, we proceeded in two stages involving within and between cluster categorization tasks. Each stage is detailed below.

*1. Within cluster categorization task.* First, we modeled the cultural facial expressions associated with the emotion words in each non-singleton cluster (in the English group, 6 non-singleton clusters comprising a total of 28 words; in the Chinese group, 7 non-singleton clusters comprising 48 words) using the method illustrated in Figure 3. We recruited a new set of 36 native English speakers (18 male, mean age 21.5 years, SD 2.9 years), and 32 new native Chinese speakers (16 male, mean age 22.6 years, SD 2.2 years)*.* For each observer, we presented each random facial animation on one of 8 same-race face identities (white Caucasian: 4 male, mean age 23 years, SD 4.1 years; Chinese: 4 male, mean age 22.1 years, SD 0.99 years) using the procedures described in (Yu et al., 2012) and used in (Gill, Garrod, Jack, & Schyns, 2014; Jack et al., 2014; Jack et al., 2012). Observers view the random facial animation and categorize it according to one of the emotion words comprising one of the clusters of semantically similar emotion words derived above using cluster analysis (e.g., see Figure 1 and 2), or selects ‘other.’ For example, in Figure 3, on this illustrative trial the cluster of semantically similar emotion words – response options – comprises ‘fury,’ ‘rage,’ ‘livid,’ and ‘anger.’ Observers also rate the perceived emotional intensity on a 5-point Likert scale (‘very weak’ to ‘very strong’). Each observer categorized 2400-2800 such facial animations. We presented the same response options – cluster of semantically similar emotion words – for 200 consecutive trials, and randomized the order of the emotion words clusters across the experiment. We presented each facial animation on a black background in the observers’ central visual field and displayed on a flat 17-inch panel monitor with a refresh rate of 60 Hz and resolution of 1280 × 1024. We played each animation only once for a duration of 1.25s. A chin rest ensured a constant viewing distance of 73cm, with images (average size 18.3 × 12 cm) subtending 14.29° (vertical) and 9.38° (horizontal) of visual angle, thereby reflecting the average size of a human face (Ibrahimagić-Šeper, Čelebić, Petričević, & Selimović, 2006) during typical social interaction (Hall, 1966).

*2. Between cluster categorization task.* To model the dynamic facial expressions of the remaining singleton word clusters in each culture (i.e., ‘surprise’ and ‘pride’ in the English group, and in the Chinese group ‘embarrassment’/尴尬, ‘shame’/羞愧, ‘pride’/骄傲, ‘despise’/蔑视 and ‘disgust’/厌恶), we used a between cluster categorization task. We recruited 32 new native English speakers (16 male; mean age 20.4 years, SD 2.9 years), and 32 native Chinese speakers (16 male, mean age 23 years, SD .7 years). We used the same stimulus generation and task procedure as in the *within cluster categorization task*, except the response options. On each trial, the response options comprised all singleton cluster words (e.g., in English, ‘surprise’ and ‘pride’), plus the highest frequency word from each of the non-singleton clusters, identified using the BNC for English words and Chinese National Corpus for Chinese words. Response options for the English group therefore comprised 8 emotion words – ‘surprise,’ ‘pride,’ ‘love,’ ‘fear,’ ‘hate,’ ‘anger,’ ‘sad,’ and ‘shame’ – with the Chinese response options comprising 12 emotion words ‘embarrassment’/尴尬, ‘shame’/羞愧, ‘pride’/骄傲, ‘despise’/蔑视, ‘disgust’/厌恶, ‘glad’/高兴, ‘pleasantly surprised’/惊喜, ‘surprised’/惊讶, ‘panic’/恐慌, ‘anxiety’/害怕, ‘anger’/生气, and ‘suffering’/痛苦). Each observer categorized 2400-2800 such facial animations presented on the same face identities used in the *between cluster categorization task*.

*3. Modelling dynamic facial expressions of emotion*. For both the *between* and *within cluster categorization task*, we then identified, for each culture, the dynamic face movements significantly associated with each emotion word shown in Figures 1 and 2. Specifically, we reverse correlated the observer’s categorical responses (e.g., ‘rage’) with the AUs and their temporal parameters using established model fitting procedures (Yu et al., 2012). For the *within cluster categorization task*, for each emotion word within a cluster, we pooled together all trials from all observers for each culture separately, and performed a Pearson correlation between the binary vector detailing the presence vs. absence of each AU on each trial and the corresponding binary vectors detailing the response of the observers. Consequently, we obtained for each emotion word in each culture, a 1 X 42-dimensional vector detailing the correlation coefficients for each AU. We then obtained bootstrap confidence intervals (95%, 1000 shuffled samples) for the resulting Pearson correlation coefficients, thereby producing a 1 X 42-dimensional binary vector detailing the composition of significant AUs for each emotion word in each culture. To model the dynamic components of each model, we then performed a linear regression between each of the emotional intensity response variables and the six temporal parameters for each AU. We then obtained bootstrap confidence intervals (95%, 1000 samples) for the resulting linear regression coefficients, thereby producing a 6 (temporal parameters) X 42 (AUs) matrix detailing the significant temporal parameter values for each of the significant AU derived previously. Finally, we combined the significantly correlated AUs with the temporal parameters derived from the regression coefficients, thereby an animation for each facial expression model. For the *between cluster categorization task*, for each singleton cluster word (e.g., ‘surprise’ and ‘pride’ in English), we used the same procedure as above where the set of response options (e.g., in English, ‘surprise,’ ‘pride,’ ‘love,’ ‘fear,’ ‘hate,’ ‘anger,’ ‘sad,’ and ‘shame’) comprises a single cluster*.* By combining the results of the within and between cluster categorization tasks*,* we derived 30 dynamic facial expressions of emotion for the Western group and 52 dynamic facial expressions of emotion for the East Asian group (‘sad’/悲 produced no significant Action Units).

*4. Validation of models of dynamic facial expressions of emotions.* Our method aimed to provide an accurate representation of the cultural dynamic facial expressions that communicate the emotions represented in the clusters of Figures 1 and 2. Before analyzing the resulting dynamic facial expression models (henceforth, ‘model’), we submitted each to a within-culture verification task using a new set of observers (henceforth ‘validators.’) We recruited 29 new Western white Caucasian native English speakers (14 male; mean age 20.8, SD 2.2 years), and 28 Chinese native Chinese speakers (13 male, mean age 22.9, SD 1.5) years).

On each experimental trial, validators viewed one of the emotion words presented in Figures 1 or 2 followed by a model from their own culture. Validators then indicated (using a yes/no key response) whether the emotion word accurately described the facial expression displayed. For each validator, half of the trials comprised correct (i.e., within cluster) word and facial expression pairs (e.g., ‘fury’ word + ‘livid’ facial expression) and half of the trials comprised incorrect (i.e., between cluster) word and facial expression pairs (e.g., ‘fury’ word + ‘happy’ facial expression). We presented each word for 1 second in lower case white font (Arial Unicode MS) on a black screen and played each facial expression stimulus once for 1.25s, followed again by a black screen. Observers responded using a keyboard only after the facial expression animation had finished. We displayed all models (30 Western and 52 East Asian) by mapping them onto a new set of 50 same-race identities (white Caucasian: 25 male, mean age 25 years, SD 4.2 years; Chinese: 25 male, mean age 24 years, SD 1.6 years) captured using standard procedures (Yu et al., 2012). We used standard procedure to render all stimuli (Yu et al., 2012), resulting in a total of 1500 (30 models X 50 identities) Western facial expression stimuli and 2600 (52 models X 50 identities) East Asian facial expression stimuli. Each validator completed a sub-sample of trials randomly sampled (with replacement) from the pool of culture-specific stimuli, presented in random order across the experiment. Each Western validator completed 300 such trials; East Asian validators completed 520 such trials. All validators viewed all stimuli on a black background, presented in the center of their visual field and displayed on a flat 17-inch panel monitor with a refresh rate of 60 Hz and resolution of 1280 × 1024. A chin rest ensured a constant viewing distance of 68cm, with facial expression stimuli (average size 17 × 11.5 cm) subtending 14.28° (vertical) and 9.69° (horizontal) of visual angle.

**Results**

*D-prime.* To extract the models that each culture could accurately identify, we computed the between-cluster discrimination performance (d-prime; Macmillan & Creelman, 2004) of each model in each culture. D-prime values ranged from -0.29 to 3.02 (mean 1.53, SD 0.81) in the Western group and -0.37 to 2.7 (mean 1.35, SD 0.88) in the East Asian group. Table 1 below shows all d-prime values for all Western and East Asian facial expression models.

**Method**

To explore the *latent components plus cultural accents* hypothesis, we used a multivariate data reduction technique (Non-negative Matrix Factorization – NMF; Lee & Seung, 1999) to separate culturally common and culture specific AU patterns from each modeled facial expression. Specifically, NMF performs factorization on non-negative values data to separate them into their main parts, where the parts (i.e., factors) are multivariate AU patterns. In other words, NMF applied here can identify the most common AU patterns in our data set. Each facial expression model can thus be expressed as the addition of weighted factors (i.e., common AU patterns) plus a residual (i.e., ‘accent’ AU patterns)

Model = linear\_coefficients \* Common AU patterns + Accent AU pattern (1)

NMF therefore provides an intuitive representation of each facial expression as the combination of parts (i.e., specific AU patterns) to form a whole.

*A. Common Action Unit patterns.* To extract the AU patterns common across cultures, we proceeded in two steps. First, we pooled together the 25 Western and 37 East Asian facial expression models (each represented as a 1 X 42 vector detailing the coefficients of each AU’s positive correlation with the categorization responses), resulting in a 62 (all facial expressions across cultures) X 42 (AUs) matrix of real valued correlation data.To identify the optimal (i.e., minimum) number of factors to represent the data (i.e., pooled Western and East Asian facial expressions), we applied NMF in an iterative manner (*k* = 2 to 20 factors, 1000 replicates per iteration), computed the variance explained at each iteration, and identified the minimum of the curve of variance explained as the best fit (Cattell, 1966). Our analysis revealed that four factors best represent the pooled Western and East Asian facial expressions of emotion (see *Supplemental Materials – Figure S1*). Figure 4 shows the four NMF factors (i.e., latent AU patterns) resulting from this analysis, each displayed as color-coded face maps, where red indicates a stronger AU presence and blue indicates weaker AU presence in the factor (i.e., the factor weights, normalized separately for each factor).

*B. Accent Action Unit patterns.* From the NMF factorization, we can represent each facial expression model as additive (i.e. linear) combination of the basis AU patterns (each weighted by a coefficient value detailing its contribution) to the facial expression, plus a residual (i.e., accent AU patterns). Since most facial expression models comprise one main basis AU pattern (i.e., the other three factors contribute minimally; see magenta bars in Figure 4), we can represent each model as a composition of one main basis AU pattern plus its accent (cf. Equation 1). Thus, to reveal the accent AU pattern of each facial expression model, we simply subtracted from the original real-valued facial expression model the main basis AU pattern contribution (gray-scale shapes in Figure 4 indicating contribution strength). Figure 6 shows the results using one illustrative example.