Title
Recognizing Emotional Cues in Word Content versus Facial Expression: A Cross-cultural Comparison

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Yang, Yang
Möller, Corina
Lim, Steffi
et al.

Publication Date
2023

Peer reviewed

Recognizing Emotional Cues in Word Content versus Facial Expression: A Cross-cultural Comparison

Yang Yang (yang.yang@nie.edu.sg)
National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, S637616, Singapore
Corina Möller (c.moeller@mx.uni-saarland.de)
Department of Psychology, Developmental Psychology Unit, Campus Building A1.3, 66123 Saarbrücken, Germany
Steffi Lim Wan Yu (slim178@e.ntu.edu.sg)
Department of Psychology, Nanyang Technological University, 50 Nanyang Ave, S639798, Singapore
Kelly Adelia (kelly.adelia@nie.edu.sg)
National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, S637616, Singapore
Qi Wang (qiwang@cornell.edu)
College of Human Ecology, Cornell University, Ithaca, NY 14853 USA

Abstract

Cultural background may shape how people attend to different emotional cues. Emotions can be perceived from both visual and auditory channels. This cross-cultural study investigated individuals' attention to emotional cues in facial expressions and spoken words. The final sample consisted of 99 Singaporean Chinese and 81 German adults (M_{age} = 24.03 years, SD_{age} = 6.29 years). In this online study, participants completed two tasks in which they were presented with emotional facial expressions and spoken words simultaneously. They were asked to judge the pleasantness of word meanings (Word task) or facial expressions (Face task) while ignoring the other aspect. Singaporean participants' accuracies were significantly influenced by the word content while judging the pleasantness of facial expressions in the Face task. However, for German participants, there was no significant interference effect in either the Face or Word task.

Keywords: emotion recognition; facial expression; word content; culture

Introduction

Understanding others' emotions is critical for effective communication. In our increasingly interconnected world, it is important to be able to communicate our emotions with people from different cultures. The study of cultural similarities and differences in emotion processing is therefore important to facilitate inter-cultural communication. The current study examines emotion recognition based on linguistic and facial cues among Singaporean and German adults.

Culture is an ever-changing system of ways of life within a community, shaping our thoughts, behaviors, and emotions. It plays a critical role in molding the definition and experience of emotion, and the actions people take as a result of emotions (Yang & Wang, 2019). Hall (1976) proposed the concepts of *low-context* and *high-context* cultures. In high-context cultures, such as European and North American cultures, verbal content consists of more information than contextual cues, and communication is relatively more direct. However, in high-context cultures, such as Asian cultures, individuals are more interconnected, and some contextual information is enough to effectively communicate ideas, thus contextual cues play a large role in communication.

The differences between high-context and low-context cultures shape the styles of attention and cognition between Asians and Westerners (Nisbett et al., 2001; Wang, 2021). Studies with a variety of cognitive tasks have shown that individuals in Japan, a high-context culture, were more sensitive to contextual information than those in the United States, a low-context culture. For instance, in one study American and Japanese participants watched animated video clips of underwater scenes with a focal fish and background objects, Japanese made more references to the background objects and less to the focal fish when describing what they saw in the scenes than did Americans (Masuda & Nisbett, 2001). In addition, Japanese participants recognized the focal fish better if the focal fish was presented against the original background than a novel background, whereas American participants' performance was not influenced by the background. These findings suggest that Japanese were more attentive to background information than Americans. This cultural difference in attention to contextual information has been found in more abstract tasks (Kitayama et al., 2003) and more social tasks (Masuda et al., 2008). Eye-tracking studies revealed that Chinese participants fixated more on the background, while North American participants fixated more on the focal object when they were viewing pictures of naturalistic scenes (Chua et al., 2005; Lu et al., 2008). In sum, Asians tend to pay more attention to contextual cues than do Americans.

Research has also explored the role of culture in recognizing emotions from various cues. Masuda and colleagues (2008) found that Japanese participants, but not Westerners, were influenced by the surrounding people's emotions while judging a central person's emotion. Japanese participants also looked at the surrounding people more than Westerners. In addition to facial expressions, humans have been found to incorporate multiple information channels to process all available emotional cues, likely involuntarily. For instance, emotional information from facial expressions, prosody and semantics could be integrated to holistically evaluate the emotion being communicated (Paulmann & Pell, 2011). Ishii et al. (2003) found evidence that people from different cultures were differentially accustomed to attending to word meaning and tone when presented with emotionally

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

spoken words. Americans paid more attention to word meanings, while Japanese and Filipinos paid more attention to vocal tone. Tanaka et al. (2010) compared native Japanese and Dutch participants' face and voice judgements in a Stroop-like experiment, which included a face task (i.e., categorize emotion of the faces as angry or happy while ignoring the voices) and a voice task (i.e., rate the valence of the voices while ignoring the faces). Japanese participants were found to have weighted vocal cues more than the Dutch participants. Liu et al. (2015) conducted an Event-Related Potentials (ERP) study, using a face task (i.e., actively attend to facial expressions, not the voice) and voice task (i.e., actively attend to vocal stimuli, not the faces), on North American native English speakers and Mainland Chinese native Mandarin speakers. On a behavioral and neural level, the English group was more influenced by facial cues in the voice task than vocal cues in the face task, while the Chinese group yielded no significant differences between facial and vocal tasks. A recent study by Yang et al. (2021) found that European American children were more sensitive to word meanings, but less sensitive to vocal tones than Chinese children.

Notably, existing research on emotion processing have mainly used Western participants and Asian participants from a limited number of low-context and high-context countries (mainly USA, Japan, and Mainland China). Thus, it is important to extend previous research by investigating the emotion processing of individuals from a broader variety of countries. To fill this gap, the present study aimed to investigate the judgement of facial expressions and word meanings in Singapore and Germany.

In the present study, Singapore was chosen as a high-context culture with a high percentage (76.8 %) of Chinese inhabitants. Singapore is a Southeast Asian city state comparable to Hong Kong in terms of cultural background, history, and economy. Just like Hong Kong, Singapore as a former British colony was influenced by western culture, and yet Chinese traditions play a crucial role in the daily routine regarding politeness behavior and family structures. This is reflected in highly comparable patterns in the six dimensions of national culture proposed by Hofstede (2011) between Singapore (SG), Mainland China (CH) and Hong Kong (HK). In particular, all three jurisdictions show almost identical values (CH: 20; HK: 25; SG: 20) on the individualism and collectivism dimension, which corresponds to Hall's (1976) context dimension. In contrast, Germany (GER) was chosen as a low-context culture comparable to the United States (US) or the United Kingdom (UK). These countries demonstrate high values on Hofstede's individualism/collectivism dimension (US: 91; UK: 89; GER: 67), and thus, are considered as individualist/low-context cultures. Therefore, individuals of these countries prefer an explicit communication and emphasize self-actualization (Hall, 1976; Hofstede, 2011; Markus & Kitayama, 1991).

Accordingly, Germany and Singapore are suitable representatives for low-context and high-context cultures, respectively. For the present study it was intended that, besides the cultural distinction, both countries feature a comparable socioeconomic background as differences in socioeconomic status have previously been reported to influence individual' social cognitive skills (e.g., Ebert et al., 2017). As indicated by the Human Development Index (HDI), Singapore (HDI = 0.94) and Germany (HDI = 0.95) have a comparable socioeconomic status (Human Development Report, 2020). Additionally, previous cross-cultural studies majorly used English-speaking countries as high-context cultures (e.g., United States) and non-English-speaking countries as low-context cultures (e.g., Japan, China), thus language was confounded with culture in these comparisons. This study with participants from Singapore, an English high-context culture, and German, a non-English low-context culture, would provide with important information for teasing apart cultural and language effects on emotion processing. In this study, we explored Singaporeans' and Germans' sensitivity to emotional cues from facial expressions and verbal content. We expected that Singaporean participants would pay more attention to contextual information than German participants.

Method

Participants

An a priori power analysis using MorePower 6.0.4. with the following specifications was applied: One between-subjects factor with two levels (*country*: Singapore/Germany) and two repeated measures factors with two (*task*: Face task/Word task) and two (*congruency*: congruent/incongruent) levels. Alpha was set to 0.05, power to .95, and an expected effect size of $\eta^2 = 0.2$ (in accordance with previous studies reporting small to medium effect sizes, e.g., Yang et al., 2021). The effects of interest were the main effect of *country* to investigate potential differences in emotion recognition processing between Singaporean and German participants and the interaction between *country* and *congruency* for the investigation of interference/congruency effects. The analysis returned a total sample size of $N = 56$. A total of 123 Singaporean Chinese (Singaporean of Chinese ethnicity) and 96 German adults participated in the present online study. All participants were recruited through student groups, via student mailing lists or in lectures at two universities in Singapore and Germany. In the Singaporean sample, $n = 24$ participants were excluded from analyses due to having low accuracies in practice trials in at least one of the two tasks (failed more than four trials out of 12 trials, $n = 20$) or identified ethnicity other than Chinese ($n = 4$). In the German sample, $n = 15$ participants were excluded from analyses due to low accuracies in practice trials in at least one of the two tasks ($n = 11$) or due to having technical problems during participation ($n = 4$). The final sample consists of 99 Singaporean adults (63 female, 36 males) with a mean age of 23.16 ($SD = 4.84$) years, and 81 German adults (53 female, 26 male, 2 diverse) with a mean age of 25.1 ($SD = 7.60$) years. All Singaporean and German participants were native or near-to-native speakers of English and German language,

respectively. All participants gave their informed consent prior to participating. Singaporean participants received a SGD$5 voucher for compensation. German participants received course credit for participating upon request. The present study was part of a research project entitled "Singaporean children's emotion understanding and its relations to anxiety and academic achievement" which was approved by the ethic committee of Saarland University (protocol number: 20-09), Germany, and by the Nanyang Technological University (protocol number: IRB-2019-10-043), Singapore.

Procedure

Participants took part in the study via a link to the experiment developed on Gorilla experiment builder (https://app.gorilla.sc/) being used to design experiments with multimodal stimuli. Gorilla experiment builder has been shown to be a suitable platform for online research studies (Anwyl-Irvine et al., 2020; Yamamoto et al., 2021). Results from Yamamoto et al. (2021) showed that there were no statistically significant differences between participants' web data and lab data results.

After providing their informed consent, participants filled out a short demographic survey and underwent an environment check to ensure they were using laptops or computers and earphones/headphones in a quiet place, with their browser maximized. Then, participants were introduced to two bimodal emotion recognition Stroop tasks (Face task and Word task). Both tasks followed a comparable procedure: A fixation cross with a beep sound was presented in the middle of the screen for 250ms. Subsequently, an image showing either a happy or angry face (500 x 500 pixels) was displayed in the center of the screen for 1000ms. Simultaneously, a spoken word with either a pleasant or an unpleasant meaning was presented for the same duration (maximum length of the audios: 1000ms). In half of the trials, faces' and words' valence was congruent (e.g., happy face and pleasant word meaning), while in the other half of trials valence was incongruent (e.g., happy face and unpleasant word meaning). The presentation of test stimuli was followed by a screen showing two emojis (happy and frowny) on the right and left side of the screen prompting participants' responses (response slide). Participants were asked to respond as fast as possible to either the facial expressions (Face task) or the spoken words (Word task) while ignoring the other aspect of the bimodal presentation (to-be-ignored aspect), after the visual and auditory stimuli were done presenting. Responses were given by pressing two keys on the keyboard, 'D' or 'K', corresponding to pleasant and unpleasant (happy and frowny emojis). Reaction time was recorded from the onsite of the response slide which was 1000ms after the onset of the auditory and visual stimuli. Combinations of valence and keys were randomized between participants but were kept constant across both tasks. Both tasks consisted of 16 test trials each which were presented in a blocked and randomized order.

Moreover, practice phases to familiarize participants with the task were applied in both the Face and the Word task. In the Face task, a unimodal task was presented first: four happy or angry cartoon faces with neutral spoken words were played simultaneously. Subsequently, eight trials (four congruent and four incongruent) using images and spoken words comparable to the test trials were presented in a fixed order. In the Word Task, the unimodal task encompassed four neutral cartoon faces and words with pleasant or unpleasant meanings. Subsequently, eight trials (four congruent and four incongruent) were presented in a fixed order. During the practice phases, participants received visual and auditive feedback. During the test phases, no feedback was given.

Participants were randomly assigned to one of four counterbalanced conditions (1. Face $task_{first}$ / $D_{positive}$; 2. Face $task_{first}$ / $K_{positive}$; 3. Word $task_{first}$ / $D_{positive}$; 4. Word $task_{first}$ / $K_{positive}$). The experiment was programmed to ensure there was an equal number of participants assigned to each of the four conditions.

Materials

Images were gathered from the Child Affective Facial Expression set (CAFE; LoBue & Thrasher, 2015) featuring a collection of emotional facial expressions of 2- to 8-year-old children of varying ethnicities. For the present study, images of Asian American (AA) children and European American (EA) children expressing happiness and anger were selected and matched by gender and validity scores for each emotion (see Figure 1 for an example). Three further pictures had to be gathered from other sources (e.g., from the online stock photography database www.shutterstock.com). Colors and white balance of all additional pictures were adjusted in Photoshop (CS6) to guarantee maximal congruence with the CAFE stimuli. All matched pairs of images were rated by 150 Singaporean and 180 German adults confirming the congruence in validity between AA (.81) and EA pictures (.82). A final set of 32 images served as test stimuli (balanced number of images for emotions, ethnicity, and gender). A further 16 images from the CAFE set and eight cartoon faces (downloaded from the online stock photography database www.dreamstime.com) served as practice trials. Importantly, in Singapore, only AA images were used while in Germany, only EA images were applied.

A total of 28 words with pleasant and unpleasant meanings were selected from the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007) which contains a list of words commonly produced by toddlers. Moreover, four words with neutral meanings (e.g., window and table) were selected. In Singapore, the original English words were used while in the German sample words were translated to German language by a bilingual native speaker. A sample of 150 Singaporean and 180 German adults rated the degree of pleasantness of all words on a 1 (very unpleasant) to 7 (very pleasant) Likert-scale and results showed comparable values in both samples. For the present study, audio stimuli were produced using two female voices of two online voice generators (https://voicegenerator.io/ in

Singapore, https://voicemaker.in/ in Germany) reading out all words with a neutral vocal tone (pitch=0.9, speed=0.7). All audio files were downloaded as .mp3-files and had a maximum duration of one second.

Figure 1. Examples of Asian American and European American stimuli based on the Child Affective Facial Expression set (LoBue & Thrasher, 2015).

Results

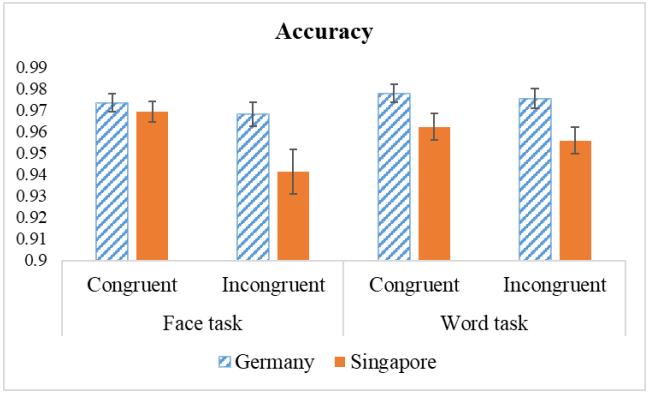
The percentage of accurate responses (accuracy) and average response time for correct trials in each task for each participant was calculated. The accuracy and response time were calculated for congruent and incongruent trials separately to examine the interference effect of the to-be-ignored aspect of the presented stimuli (congruency effect). The means of accuracy and response time are presented in Figure 2.

We first conducted a General Linear Model (GLM) with two between-subjects variables (participant's culture: Singapore vs. Germany; participant's gender: male vs. female) and two within-subject variables (task: Face vs Word; congruency: congruent vs. incongruent) on accuracy and response time, with participant's age as the covariate. The main effect of culture was significant on accuracy, $F(1,175) = 9.586$, $p = .002$. Specifically, Singaporean participants had lower accuracy than German participants overall. In terms of response time, there was a significant effect of age, $F(1,175) = 17.99$, $p < .001$, and gender, $F(1,175) = 4.38$, $p = .014$, whereby older and female participants had longer response time than younger and male participants, respectively. The interaction effect of task x culture on response time was significant, $F(1,175) = 5.57$, $p = .019$. Pairwise comparisons showed that Singaporean participants tended to have shorter response time than Germans only in the Face task, but the difference was not statistically significant, $p = .066$. The interaction effect of task x congruency was significant on accuracy, $F(1,175) = 4.14$, $p = .043$, but marginally significant on response time, $F(1,175) = 3.03$, $p = .084$. However, pairwise comparison did not show any significant difference between tasks or between congruent and incongruent trials with two cultural groups together.

Then we further explored two cultural groups separately, by conducting a GLM with two within-subjects variables (task and congruency), a between-subjects variable (gender), and a covariate (age) for each cultural group.

For Singaporeans, there was a significant effect of gender on response time, $F(1,96) = 5.53$, $p = .021$, whereby female participants had a longer response time than males in general. The effect of age on response time was marginally significant, $F(1,96) = 3.83$, $p = .053$, whereby older participants tended to respond slower than did younger participants. There was no significant within-subjects effect. Pairwise comparisons between congruent and incongruent trials showed that the congruency effect was only significant for accuracy in Face task, $p = .020$. Specifically, the accuracies for congruent trials were higher than incongruent trials for the Face task. This suggests that Singaporean participants were influenced by word meanings when they judged the faces. In addition, Singaporean participants' response time was longer in the Word task than in the Face task for both congruent and incongruent trials, ps $< .001$.

For German participants, the main effect of age on response time was significant, $F(1, 77) = 13.22$, $p < .001$, whereby older participants demonstrated longer response time than younger participants. In terms of within-subject factors, there was no significant congruency effect for any task's accuracy or response time for German participants.



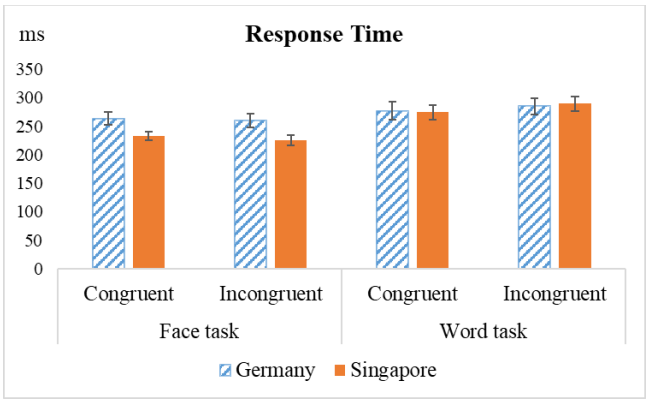


Figure 2: Accuracy and response time as a function of congruency, task, and cultural group. Error bars indicate standard errors.

Discussion

This study examined emotion processing from two types of cues in Singaporeans and Germans. We found that, compared to Germans, Singaporean participants had lower overall accuracies in the Stroop-like tasks. They were influenced by the to-be-ignored information in the Stroop-like tasks to a greater extent than Germans. Specifically, Singaporean participants had higher accuracy scores in congruent trials than incongruent trials in the Face task, whereas German participants' accuracy did not differ between congruent and incongruent trials. Regarding response time, Singaporean participants responded faster in the Face task than in the Word task. In contrast, German participants' response time did not differ between congruent and incongruent trials, or between the two tasks.

Singaporean participants had lower accuracy scores than German in these Stroop-like tasks, and the interference effect (higher performance in congruent than incongruent trials) was only found among Singaporean participants but not for the German sample. Considering Hall's (1976) theory of low-context and high-context, Singaporean Chinese may maintain traditional Chinese values and pay relatively more attention to contextual cues. Studies investigating North American, Japanese, and Chinese samples showed that Asians, Japanese and Chinese, paid more attention to contextual cues when viewing pictures with either social or non-social stimuli, compared to their North American counterparts (Chua et al., 2005; Lu et al., 2008; Masuda & Nisbett, 2001). Consistent with these findings, Singaporean Chinese were influenced by the spoken words while judging the pleasantness of facial expressions in the Stroop-like tasks. However, participants in Germany, a low-context culture, were not influenced by the to-be-ignored aspect of stimuli and able to focus on the focal aspect of stimuli in both the Face and Word task.

Interestingly, the interference effect for Singaporean participants was only significant in Face task but not in the Word task. This suggests that Singaporean participants were influenced by the word content when judging the pleasantness of the facial expressions, but not by the facial expressions when judging the verbal content. These results, to some extent, are consistent with in Tanaka and colleagues' findings in Japanese and Dutch samples that Japanese participants were more sensitive to vocal cues than Dutch participants (Tanaka et al., 2010). They argued that facial expressions might always be the focal information, whereas the voice might always serve as contextual information regardless of the task, given the importance of facial expressions. Because Asians are more sensitive to contextual information, the interference effect of voices is larger in the Face task. In contrast, German participants might be able to focus on the task-relevant stimuli (e.g., facial expressions in the Face task, spoken words in the Word task) and were not influenced by the to-be-ignored aspect of stimuli. However, results regarding the Singaporean participants' response time in the two tasks were not completely consistent with these findings. We found that Singaporean participants responded faster in the Face task than in Word task in both congruent and incongruent trials. This may simply be due to the shorter response time needed for facial stimuli than auditory stimuli which required about 1000ms to finish presenting. Further studies with multi-sensory stimuli are needed to further examine the cultural difference in the processing of multi-sensory emotional stimuli.

The main effect of age on response time was also found in both Singaporean and German samples. This is consistent with general findings in the tasks measuring response time that response time decreases rapidly in childhood and adolescence, then increases steadily through adulthood (Dykiert et al., 2012; Lahtela et al., 1985; Williams et al., 2007). This may be due to differing ability in sustained attention across lifespan (Dykiert et al., 2012).

We also found a significant gender difference in overall response time among Singaporean participants, but not among German participants. Although research findings often point to faster response time for males compared to females, more research have found how it may differ between conditions (Dykiert et al., 2012; Spierer et al., 2010). For example, Dykiert et al. (2012) found that males have faster response time in simple reaction time (SRT) where there is only one stimulus and one response, but not in choice reaction time (CRT) where there are multiple stimuli with their respective responses. The task in this study constitutes as CRT, however only German participants' response time in this study aligns with the finding but not Singaporean participants. Further research may be needed to better understand whether and how different conditions and cultural backgrounds affect gender differences in response time.

There are some limitations of the current study. First, participants completed the tasks online instead of in the lab, so it is not ensured that participants have followed the instruction strictly. It cannot be checked whether participants had headphones/earphones on during the Face task or kept their eyes on the screen during the Word task although we have emphasized these in the task instructions repeatedly. Second, the quality and the loading time for stimuli might differ across different browsers and depend on internet speed. Despite of the above-mentioned disadvantages, online studies have significant benefits allowing researchers to recruit a relatively large number of participants and collect data more efficiently.

This study provided an important empirical investigation of emotion recognition from facial expressions and verbal content in two cultural groups. It sheds light on our knowledge about the role of culture in emotion processing in different modalities.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior*

Research Methods, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12629–12633. https://doi.org/10.1073/pnas.0506162102

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Sex differences in reaction time mean and intraindividual variability across the life span. *Developmental Psychology*, *48*(5), 1262–1276. https://doi.org/10.1037/a0027550

Ebert, S., Peterson, C., Slaughter, V., & Weinert, S. (2017). Links among parents' mental state language, family socioeconomic status, and preschoolers' theory of mind development. *Cognitive Development*, *44*, 32–48. https://doi.org/10.1016/j.cogdev.2017.08.005

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *The MacArthur-Bates communicative development inventories user's guide and technical manual* (2nd ed.). Paul H. Brookes Publishing Co Inc.

Hall, E. T. (1976). *Beyond culture*. Anchor Books.

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, *2*(1). https://doi.org/10.9707/2307-0919.1014

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46. https://doi.org/10.1111/1467-9280.01416

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science*, *14*(3), 201–206. https://doi.org/10.1111/1467-9280.02432

Lahtela, K., Niemi, P., & Kuusela, V. (1985). Adult visual choice-reaction time, age, sex and preparedness: A test of Welford's problem in a large population sample. *Scandinavian Journal of Psychology*, *26*(1), 357–362. https://doi.org/10.1111/j.1467-9450.1985.tb01175.x

Liu, P., Rigoulot, S., & Pell, M. D. (2015). Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2014.11.034

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01532

Lu, Z., Daneman, M., & Reingold, E. M. (2008). *Cultural differences in cognitive processing style: Evidence from eye movements during scene processing*.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*(3), 365–381. https://doi.org/10.1037/0022-3514.94.3.365

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*(5), 922–934. https://doi.org/10.1037/0022-3514.81.5.922

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310. https://doi.org/10.1037/0033-295x.108.2.291

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*(2), 192–201. https://doi.org/10.1007/s11031-011-9206-0

Spierer, D. K., Petersen, R. A., Duffy, K., Corcoran, B. M., & Rawls-Martin, T. (2010). Gender Influence on Response Time to Sensory Stimuli. *Journal of Strength and Conditioning Research*, *24*(4), 957–963. https://doi.org/10.1519/JSC.0b013e3181c7c536

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I feel your voice. Cultural differences in the multisensory perception of emotion. *Psychological Science : A Journal of the American Psychological Society / APS*, *21*(9), 1259–1262. https://doi.org/10.1177/0956797610380698

UNDP (United Nations Development Programme). (2020). Human Development Report 2020. *UNDP (United Nations Development Programme)*. http://report2020.archive.s3-website-us-east-1.amazonaws.com/

Wang, Q. (2021). The cultural foundation of human memory. *Annual Review of Psychology*, *72*(1), 151–179. https://doi.org/10.1146/annurev-psych-070920-023638

Williams, B. R., Strauss, E. H., Hultsch, D. F., & Hunter, M. A. (2007). Reaction Time Inconsistency in a Spatial Stroop Task: Age-Related Differences Through Childhood and Adulthood. *Aging, Neuropsychology, and Cognition*, *14*(4), 417–439. https://doi.org/10.1080/13825580600584590

Yamamoto, H. W., Kawahara, M., & Tanaka, A. (2021). A Web-Based Auditory and Visual Emotion Perception Task Experiment With Children and a Comparison of Lab Data and Web Data. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.702106

Yang, Y., Wang, L., & Wang, Q. (2021). Take your word or tone for It? European American and Chinese children's attention to emotional cues in speech. *Child Development*, *92*(3), 844–852. https://doi.org/10.1111/cdev.13576

Yang, Y., & Wang, Q. (2019). Culture in emotional

development. In V. LoBue, K. Perez-Edgar, & K. Buss (Eds.), *Handbook of Emotional Development* (pp. 569–593). Springer, Cham. https://doi.org/10.1007/978-3-030-17332-6_22