

Interdependence modulates the brain response to word–voice incongruity

Keiko Ishii,¹ Yuki Kobayashi,² and Shinobu Kitayama³

¹Center for Experimental Research in Social Sciences, Hokkaido University, N10W7, Kita-ku, Sapporo 060-0810, Japan, ²Center for Evolutionary Cognitive Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-Ku, Tokyo 153-8902, Japan, and ³Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043, USA

Social orientation (interdependence as opposed to independence) has been suggested as a major cultural dimension. In the present work, we used a specific stimulus-locked component of electroencephalogram and found, for the first time, that the perceiver's social orientation modulates the brain response to incongruity of word meaning to attendant vocal tone. Participants judged the verbal meaning of emotionally spoken emotional words while ignoring the vocal tone. As predicted, there was a greater negative event-related potential between 450 and 900 ms after the stimulus onset when the verbal content was incongruous with the background vocal tone, relative to when the two were congruous. Of importance, this incongruity-based late negativity was larger when participants were unobtrusively exposed to schematic human faces while listening to the stimulus words and for females than for males. Moreover, this late negativity was reliably predicted by chronic social orientation for females, but not for males and in the face condition, but not in the no-face control condition. Implications for future work on culture are discussed.

Keywords: independence versus interdependence; event-related potential; face; attention; and vocal tone

Imagine someone tells you that you are great, but she does so with a harsh tone of voice. Unless you are completely oblivious to vocal tone, you may be immediately puzzled by the incongruity between the verbal meaning and the vocal tone—for vocal tone and verbal meaning of a spoken message are generally expected to be consistent with each other (Grice, 1974). We investigated this phenomenon by observing electric activities of the brain as a function of the perceiver's social orientation (i.e. interdependence as opposed to independence). By social orientation we mean one's preparedness to engage in social relations. Interdependent individuals are said to be higher in social orientation than independent individuals. Here we report, for the first time, that the perceiver's social orientation increases the likelihood of spontaneously attending to background vocal tones, thereby detecting the incongruity of the tones vis-à-vis the focal verbal meanings of spoken words.

In a series of experiments Ishii, Kitayama and colleagues (Kitayama and Ishii, 2002; Ishii *et al.*, 2003) investigated this issue by adopting a modified version of the Stroop interference paradigm (Stroop, 1935). Participants listened

to either positive or negative words, one at a time. These words were spoken in either positive or negative vocal tones. Even though participants were explicitly instructed to ignore the vocal tone, it took them longer to judge the pleasantness of each word if the vocal tone was incongruous than when it was congruous. This shows that people often fail to filter out vocal tone even when they try to ignore it. The vocal tone, as a consequence, causes a response interference when it is incongruous with the focal word meaning.

Several studies in social and personality psychology have shown that vocal tone reveals the speaker's relational attitudes (Zuckerman *et al.*, 1982; Ambady *et al.*, 1996). Accordingly, as one's social orientation increases, the person may be expected to spontaneously allocate more attention to vocal tone, resulting in a greater sensitivity to any incongruities between word and voice. Consistent of this analysis, the vocal Stroop effect is larger when participants are reminded of their social rejection experience (Pickett *et al.*, 2004), when they are subliminally exposed to relationship words (e.g. friendship, unity) and, thus, an interdependent orientation is unobtrusively primed (Gray *et al.*, 2007), or when they are induced to feel sadness, which is typically accompanied by an affiliation motive (Gray *et al.*, 2007).

As may be predicted by the hypothesis that Asians are more interdependent and thus more socially oriented than North Americans (Markus and Kitayama, 1991; Kitayama *et al.*, 2007), Asians including Japanese and Filipinos exhibit a stronger vocal Stroop effect than North

Received 6 January 2009; Accepted 19 October 2009

Advance Access publication 2 December 2009

This work was supported by Center for Evolutionary Cognitive Sciences at the University of Tokyo. It was also supported by a grant from National Science Foundation (BCS 0717982). This article was completed while the third author was a Fellow at the Center for the Advanced Study in Behavioral Sciences (Stanford, CA). The authors thank Emre Demiralp, Jun'ichi Katayama, and Jinkyung Na for their comments on a draft of this article.

Correspondence should be addressed to: Keiko Ishii, Center for Experimental Research in Social Sciences, Hokkaido University, N10W7, Kita-ku, Sapporo 060-0810, Japan. E-mail: ishii@let.hokudai.ac.jp

Correspondence may also be addressed to Shinobu Kitayama, Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043, USA. E-mail: kitayama@umich.edu

Americans do (Kitayama and Ishii, 2002; Ishii *et al.*, 2003). Furthermore, consistent with the notion that independence is relatively less emphasized in Catholic denominations of Christianity than in Protestant denominations, Sanchez-Burks (2002) show that Americans with Catholic backgrounds are more sensitive to vocal tone, as shown by a larger vocal Stroop effect, relative to their Protestant counterparts, especially when they engage in informal social interactions. These cross-cultural findings affirm the claim that social orientation (i.e. interdependence as opposed to independence) is an active element of culture.

So far, the measure of choice in all the studies reviewed above is response time. However, response time is necessarily marked by an overt response the participant makes to the stimulus and, as a consequence, it may be expected to be influenced by a number of (largely unknown) factors that might come into play between the time when attention is spontaneously deployed to the vocal tone of an impinging stimulus word and the time when a response is eventually made on the verbal meaning of the stimulus word. For example, the size of a Stroop-type interference effect might depend on one's ability to manage competing responses. It is theoretically crucial, then, to measure spontaneous attention to vocal tone, on line, when it actually occurs rather than waiting until the delivery of an overt response to the stimulus word.

Electrophysiological responses of the brain may provide a powerful alternative to response time as a measure of spontaneous attention to vocal tone. Numerous studies have demonstrated that one reliable neurological marker of the detection of semantic incongruity is a negative event-related potential (ERP) component, called the N400, which typically occurs relatively late in the processing (although typically earlier than any manual responses), approximately 400 ms after the stimulus onset (Kutas and Hillyard, 1980; Kutas and Federmeier, 2000).

In a vocal Stroop paradigm participants are asked to judge the pleasantness of the verbal meaning of a spoken word while ignoring the attendant vocal tone. To detect any semantic incongruity between vocal tone and verbal meaning, then, the vocal tone will have to be spontaneously attended to. Otherwise, the person would not notice any existing incongruity. Once detected, this inconsistency, by and in itself, might give rise to N400. Moreover, there may be a general canon of language use that prescribes consistency (Grice, 1974). Thus, the detection of an incongruity between word and voice may in turn violate the canon of language use, thereby producing N400. For these reasons, N400 can serve as a highly proximal measure of spontaneous attention to vocal tone. Whenever vocal tone is spontaneously attended to and, thus, an existing mismatch between vocal tone and verbal meaning has been detected, N400 responses may be expected.

As compared with the traditional response time measure, the N400 measure is far less likely to be contaminated

by various subsequent processes, including response competition that might come into play after the detection of a mismatch between vocal tone and verbal meaning. We thus anticipated that the current ERP measure would tap the actual process of interest (i.e. spontaneous attention to vocal tone) much more directly and, thus, it will be a powerful tool for testing the present hypothesis.

Recent studies by Schirmer and colleagues used a vocal Stroop procedure and observed a strong negativity between 350 and 600 ms after the stimulus onset if vocal tone is incongruous in pleasantness with the focal word meaning, compared with when the two channels of information are congruous (Schirmer *et al.*, 2002, 2006; Schirmer and Kotz, 2003). Importantly, this effect was observed only for female participants. This finding is consistent with the present hypothesis because women can be assumed to be more socially oriented than men (Hall, 1978). If the present analysis were correct, however, the likelihood of detecting word-voice incongruity and, thus, showing an incongruity-based late negativity should increase as a function of other factors that contribute to social orientation. Here we examined two such factors.

First, we expected that one's social orientation should increase when the person is exposed to human faces, insofar as the exposure to faces is often associated with social interaction. Humans, including infants, are known to be exquisitely sensitive to face-like stimuli (Goren *et al.*, 1975; Johnson *et al.*, 1991). Furthermore, mere exposure to faces or face-like stimuli seems sufficient to produce, albeit unconsciously, outside of any conscious recognition, an impression of 'watching eyes', thereby making social norms and expectations more salient. In support of this hypothesis, mere exposure to face-like stimuli causes important changes in cognitive dissonance (Kitayama *et al.*, 2004). Furthermore, it tends to increase altruistic behaviors (Haley and Fessler, 2005; Bateson *et al.*, 2006; Rigdon *et al.*, 2009). Drawing on this literature, we expected that mere exposure to schematic faces would increase one's social orientation. In particular, we anticipated that this exposure would lead the perceiver to believe that the impinging situation might potentially be 'social' and 'public'. We thus predicted that an incongruity-based late negativity would be greater when participants are exposed to faces while listening to stimulus words than when they are not.

Second, to directly test the prediction that more socially oriented (i.e. interdependent rather than independent) individuals would more readily detect the word-voice incongruity, we measured each individual's chronic social orientation. At present, there are many measures of independence and interdependence and it is not obvious which measure(s) might be suitable in measuring chronic levels of social orientation. We reasoned that chronic social orientation would be best assessed in terms of the degree to which people involve themselves, consistently across multiple situations, in interdependent or engaging

(as opposed to independent or disengaging) social interactions. These two modes of social interaction are signaled by the types of emotions that are experienced. Interdependent social relations are associated with socially engaging emotions such as 'friendly feelings', 'feelings of connection', 'shame' and 'guilt', whereas independent social relations are associated with socially disengaging emotions such as 'pride in the self', 'self-esteem', 'anger' and 'frustration' (Kitayama *et al.*, 2006). We thus measured the relative intensity of experiencing these two types of emotions across multiple social situations (Kitayama and Park, 2007).

METHODS

Forty-seven Japanese young adults (24 males and 23 females) were tested individually. They listened to a number of emotional words that were spoken in different emotional vocal tones and reported, by pressing one of two appropriately labeled mouse buttons, whether the verbal meaning of each word was pleasant or unpleasant while ignoring the vocal tone. Both accuracy and speed were emphasized. Ten practice trials were followed by 64 experimental trials, with a block consisting of 32 utterance stimuli (see below) repeated twice. Within each block, the order of the 32 stimuli was randomized for each participant. Response time was measured from the onset of the stimulus word. Four participants showed unusually high rates of artifacts or error responses (greater than 25% when the two types of trials were combined). These four participants were excluded from the following analyses. The data from the remaining 43 participants (20 males and 23 females) were analyzed. Approximately half of participants (10 males and 12 females) were randomly assigned to a face condition, with the remaining participants assigned to a no-face control condition.

At the start of each trial, an instruction was presented on the computer screen, which asked the participants to use two relevant mouse buttons to indicate their judgments. In the face condition, when this instruction appeared, two schematic faces were unobtrusively presented to illustrate the two response options (see the top of Figure 1 for the instruction and the faces used). In the control condition the same instruction appeared without any schematic faces. In both conditions, when the participants were ready to proceed, they pressed either one of the two mouse buttons. Immediately afterward, a stimulus word was presented through a speaker. The participants then reported their judgment by pressing one of the two mouse buttons. There was a 1500-ms interval between trials.

Electroencephalogram (EEG) was recorded from 19 Ag/AgCl electrodes (anterior [FP1, FP2, F7, F3, FZ, F4, F8], central [T7, C3, CZ, C4, T8] and posterior [P7, P3, PZ, P4, P8, O1, O2]) mounted in an elastic cap (Quikcap, NeuroScan) according to the International 10–20 system. To control for horizontal and vertical eye movements, a bipolar electroencephalogram (EOG) was also recorded

using four electrodes. All the EEG and EOG channels were digitized at a 500 Hz sampling rate using a Neuroscan Synamp2s amplifier with a band-pass between DC and 100 Hz.¹ Recordings were referenced to the electrode located between Cz and CPz and then re-referenced offline to the average of the left and right mastoids. Electrode impedance was kept below 10 kOhm. ERP averages were computed with a 200 ms baseline and a 900 ms ERP time window. In the ERP analysis 10.7% of the trials were rejected because of eye blinks or movement artifacts (the EOG rejection criterion was 100 μ V). Only correctly answered trials were averaged. The overall error rate was 2.5%, which was distributed equally across the conditions. Grand averages were smoothed with an 8-Hz low-pass filter for illustration purposes.

A set of eight positive (e.g. 'arigatai [grateful]' and 'ataakai [warm]') and eight negative Japanese words (e.g. 'surai [bitter]' and 'zurui [sly]') used by Ishii and her colleagues (2003) in their behavioral study on the vocal Stroop effect were adopted. The words were spoken in either positive (i.e. smooth and round) or negative tones (i.e. harsh and constricted), resulting in a set of 32 utterances. The stimuli were produced by two male and two female speakers. The perceived extremity of emotional vocal tone (assessed with low-pass filtered stimuli) was no different as a function of word meaning. The length of the stimuli varied somewhat with the means of 1113, 1028, 1116 and 784 ms for the positively spoken positive words, positively spoken negative words, negatively spoken positive words, and negatively spoken negative words, respectively. After completing the meaning judgment task, participants were asked to fill out an Implicit Social Orientation Questionnaire (ISOQ) (Kitayama and Park, 2007).

In the ISOQ, participants were presented with 10 mundane social situations (e.g. 'having a positive interaction with friends') and were asked to remember the most recent occasion in which each of the 10 situations happened to them. Participants reported the extent to which they experienced each of 10 different emotions in the situation (1: not at all, 6: very strongly). The emotions were classified in terms of two dimensions of social orientation (socially engaging versus socially disengaging) and valence (positive versus negative). Thus, there were (1) *socially engaging positive emotions* (e.g. friendly feelings), (2) *socially engaging negative emotions* (e.g. shame), (3) *socially disengaging positive emotions* (e.g. pride in self), and (4) *socially disengaging negative emotions* (e.g. anger). In addition, *general positive emotions* (elated, happy, and calm) and *general negative emotions* (unhappy) were also included.

We computed the mean intensities for the four types of emotions defined by valence (positive versus negative) and social orientation (engaging versus disengaging) for each of the 10 situations. We then determined whether each

¹EEG and EOG amplified offline using a 30 Hz low-pass filter were also analyzed. The results were in line with the analysis reported in the main text.

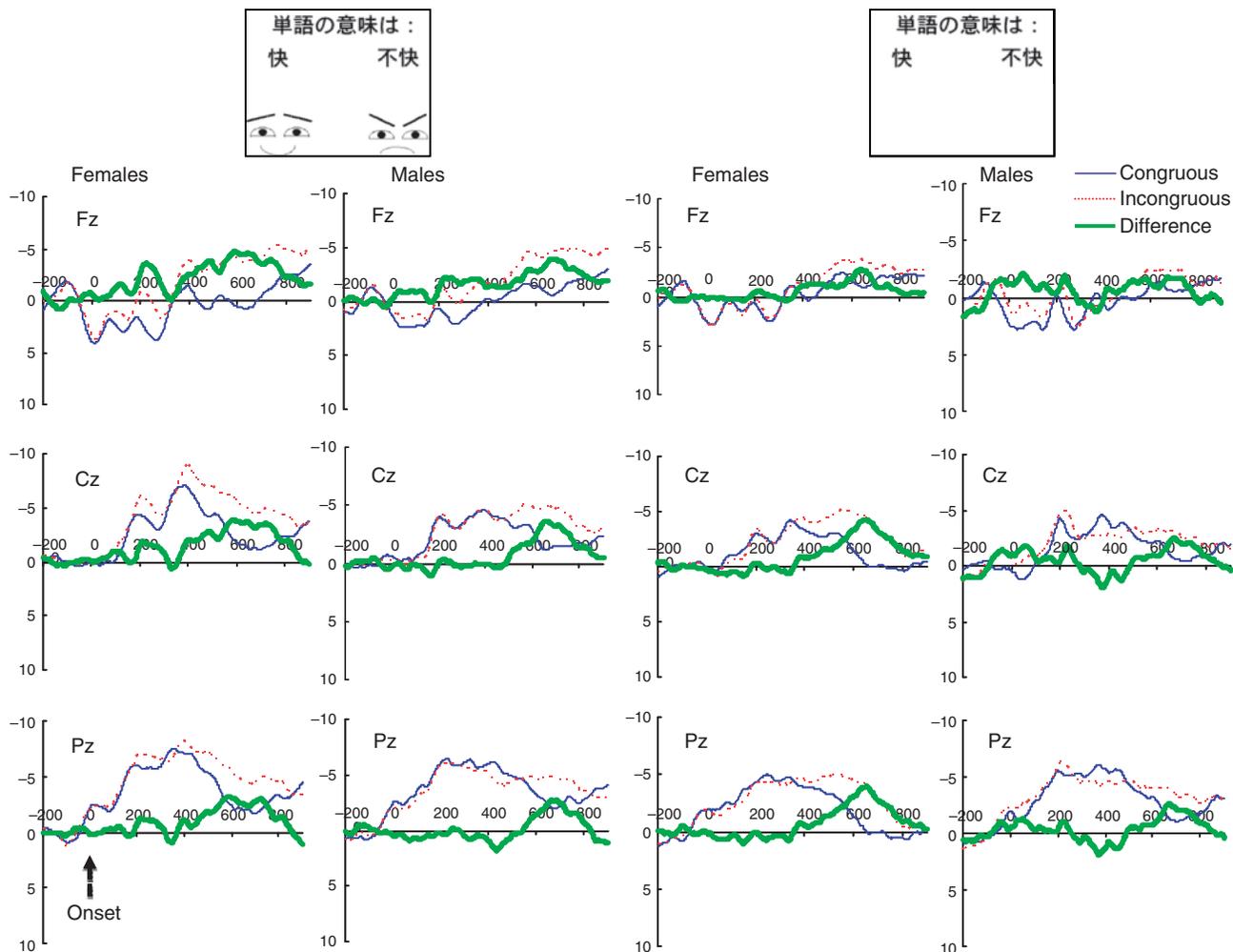


Fig. 1 Wave forms in the face condition and the control condition by male and female participants. Wave forms for congruous and incongruous utterances and their differences (incongruity – congruity) are shown in blue, red and green, respectively. The incongruity related late negativity was greater in the face condition than in the control condition. This effect was statistically significant in the frontal midline location (Fz) during the 600–750 ms window. Moreover, the incongruity-based late negativity was also greater for females than for males. This effect was statistically significant in the central and posterior midline locations (Cz and Pz) in the 450–600 ms window. The instruction reads: 'What is the meaning of the word? (単語の意味は), Pleasant (快), Unpleasant (不快)'.

situation is generally positive or negative for each participant by comparing the mean intensities for general positive and negative emotions. Using only the engaging and the disengaging emotions that were matched in valence to each situation, we computed the difference in the reported intensity scores between the engaging emotions and the disengaging emotions. We then averaged the differences over the 10 situations for each participant to yield an index of interpersonal orientation (see Kitayama and Park, 2007 for details). This index takes positive scores if, across the 10 situations, disengaging emotions were more strongly experienced than engaging emotions and negative scores if the reverse is the case. In the present sample, the mean score was $-.04$. Females were somewhat more interdependent than were males (-0.15 versus 0.08), although the difference was not significant ($P > 0.15$). There was no difference in the mean score between face and control conditions ($F < 1$).

RESULTS

Behavioral data

Response times for the correct responses were averaged for each participant and submitted to an analysis of variance (ANOVA), which showed a significant interaction between word meaning and vocal tone, $F(1, 39) = 118.68$, $P < 0.0001$, $\eta_p^2 = 0.75$. Replicating previous studies by Ishii, Kitayama and colleagues (Kitayama and Ishii, 2002; Ishii et al., 2003), response time was significantly slower when word meaning and vocal tone were incongruous than when they were congruous. A four-way interaction involving word meaning, vocal tone, face condition, and participant's gender was also significant, $F(1, 39) = 4.50$, $P < 0.05$, $\eta_p^2 = 0.10$. All means are summarized in Table 1. To examine the nature of this interaction, we subtracted the mean response time (ms) in the congruous condition from the mean response time (ms) in the incongruous condition for

Table 1 Mean response time (ms) in the word meaning judgment

	Word meaning			
	Pleasant		Unpleasant	
	Vocal tone			
	Pleasant	Unpleasant	Pleasant	Unpleasant
Face/Female				
<i>M</i>	1308	1343	1207	1118
<i>s.d.</i>	(177)	(180)	(173)	(149)
Face/Male				
<i>M</i>	1224	1321	1174	1068
<i>s.d.</i>	(128)	(159)	(126)	(111)
Control/Female				
<i>M</i>	1205	1282	1163	1095
<i>s.d.</i>	(160)	(156)	(166)	(133)
Control/Male				
<i>M</i>	1338	1374	1252	1178
<i>s.d.</i>	(224)	(214)	(206)	(168)

each participant to yield an index of the interference by vocal tone. The mean interference score was greater in the face condition than in the control condition, as predicted, for male participants (102.5 versus 54.2), $t(39) = 2.47$, $P < 0.05$, $d = 0.92$. Female participants showed a moderate degree of interference regardless of the face conditions (62.0 versus 70.7), $t(39) = -.48$, $P > 0.20$. Controlling for word length did not cause any differences in the general pattern.²

ERP data

We inspected wave forms for each of the four conditions defined by face and gender. In Figure 1, wave forms from midline areas are displayed. Wave forms from the incongruous condition are indicated by dotted lines whereas those from the congruous condition are indicated by solid lines. The difference between the two is showed with thick solid lines. Wave forms at lateral areas are given in Supplementary Figure. As can be seen, wave forms are very similar at both midline and lateral areas. In line with previous work on N400 (Hagoort and van Berkum, 2007), however, the effects of the experimental variables were stronger at the midline areas than in the lateral areas. We therefore focus on the midline areas hereafter.³

A series of ANOVAs were performed on the mean amplitudes at Fz, Cz and Pz electrode sites for each of six 150-ms time windows between 0 and 900 ms. All ANOVAs included two dummy-coded categorical variables

(gender [male versus female], face [present versus absent]), and one continuous variable (chronic social orientation) as between-subjects factors. Social orientation was centered. As within-subjects factors, both word meaning and vocal tone were included. All interactions were tested.

Incongruence effects are indicated by significant interactions that include word meaning and vocal tone. The pertinent interactions that proved statistically significant are highlighted in Table 2. The vocal tone \times word meaning interaction was observed quite consistently at all the electrode sites after 450 ms post-stimulus. All the interactions here indicated incongruence-based negativity. That is to say, a more pronounced negativity was observed when word meaning and vocal tone were incongruous than when the two were congruous. As noted below, these interactions were qualified by the face manipulation, gender, and chronic social orientation.

An incongruence effect was also evident as early as 150-ms post-stimulus at the Fz electrode site. Curiously, this latency was much shorter than word length. We thus suspect that this early negativity might be due to certain confounding factors in our stimuli. One possibility is that this negativity is responding to incongruity of certain [unknown] stimulus features to generic lexical or phonemic expectations (Frishkoff *et al.*, 2004). This early negativity might also be capturing a spontaneous brain response designed to inhibit the activation of incongruous semantic meaning (Jodo and Kayama, 1992). Yet, because we presented the same set of words eight times, we cannot exclude another possibility, namely, that (at least some) participants correctly guessed the semantic meaning of a word just by listening to the first fraction of it when the word was repeated. Regardless of the precise reasons for the early negativity associated with word-voice incongruity, it is important to note that during this earlier time-window, the negativity was not moderated by any of the three indices of social orientation.

Our prediction implies that the vocal tone \times word meaning interaction should be qualified by face, gender, and chronic social orientation. The predicted interaction involving face (vocal tone \times word meaning \times face) was observed at 600–750 ms at the Fz electrode site, $F(1, 35) = 6.41$, $P < 0.05$, $\eta_p^2 = 0.15$. Figure 2 displays the mean negativity as a function of vocal tone and word meaning for each of the two face conditions. In support of the present analysis, an incongruity-based late negativity was more pronounced in the face condition than in the control condition. More specifically, the vocal tone \times word meaning interaction was significant in both the face condition and the no-face control condition, $F(1, 18) = 31.94$, $P < 0.0001$, $\eta_p^2 = 0.64$ and $F(1, 17) = 14.33$, $P < 0.005$, $\eta_p^2 = 0.46$, respectively. In the no-face control condition, when vocal tone was negative, the negativity was stronger for positive words than for negative words, $t(17) = 4.81$, $P < 0.01$, $d = 1.48$, but no reliable effect of incongruence was found when vocal tone was positive, $t(17) = 1.69$, *n.s.* The pattern was similar in the

²We also computed correlations between the interference index of the behavioral measure (i.e. response time) and the index of chronic social orientation. The correlation was not significant in either of the conditions (face: $r = -0.04$, control: $r = 0.30$, $P > 0.20$).

³It should be noted that there was a distinct negativity at the frontal electrode, shortly before stimulus presentation. This potential, known as readiness potential, reflects preparedness for a motor response (Deecke *et al.*, 1976)—in the present case for pressing the mouse button to initiate each trial.

Table 2 Significant effects on the mean amplitudes of 150-ms time windows from 0 to 900 ms

	0–150	150–300	300–450	450–600	600–750	750–900
Fz						
Tone × Meaning		6.87		25.84	44.11	
Tone × Meaning × Face					6.41	
Tone × Meaning × Gender						
Tone × Meaning × SO × Face				4.93		
Tone × Meaning × SO × Gender						
Cz						
Tone × Meaning				12.57	31.73	
Tone × Meaning × Face						
Tone × Meaning × Gender				4.87		
Tone × Meaning × SO × Face						
Tone × Meaning × SO × Gender						
Pz						
Tone × Meaning				9.85	26.61	
Tone × Meaning × Face						
Tone × Meaning × Gender				6.33		
Tone × Meaning × SO × Face						
Tone × Meaning × SO × Gender					6.84	4.50

Light grey box: $P < .05$, Dark grey box: $P < .01$.

F-value was included in each box of a time window where a significant effect was found.

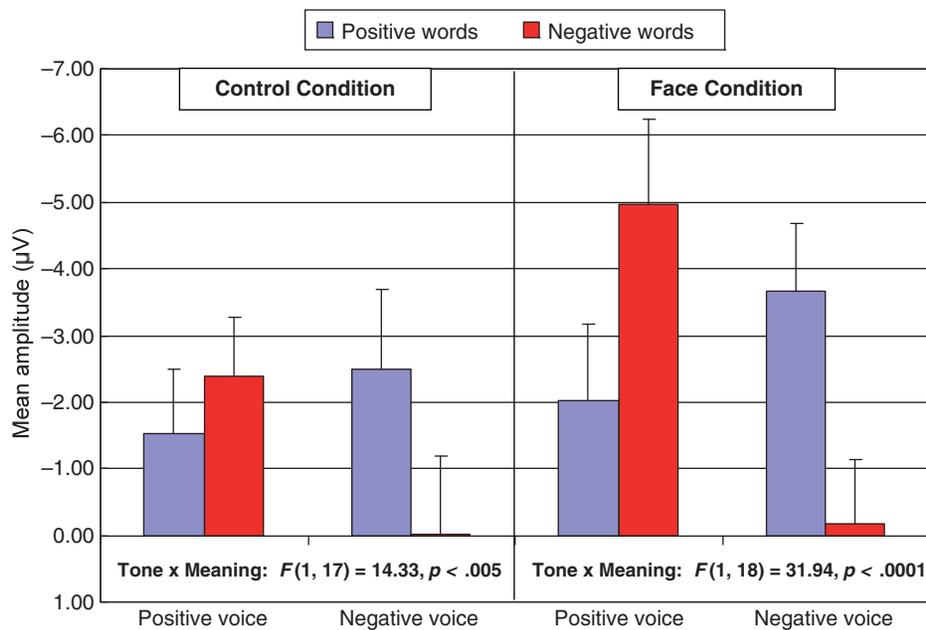


Fig. 2 Mean amplitudes of the late negativity in the face and the control conditions: amplitude was assessed at the Fz site, in the 600–750 ms time window.

face condition. However, as illustrated in Figure 2, it was distinctly more pronounced in the face condition, as predicted. The incongruity effect was highly significant both when the vocal tone was positive, $t(18) = 3.86$, $P < 0.01$, $d = 1.16$, and when the vocal tone was negative; $t(18) = 4.61$, $P < 0.01$, $d = 1.39$.

From a different angle, it is evident that positive tones caused a greater incongruity-based late negativity in the face condition than in the no-face condition, whereas

negative tones resulted in a significant difference between the positive word condition and the negative word condition both in the face condition and in the no-face condition. It appears, then, that people are attentive to negative, potentially threatening vocal signals even in seemingly non-social situations (i.e. in the no-face condition), but they become attentive to positive, potentially comforting signals only in ostensibly social situations (i.e. in the face condition). This possibility must be pursued in future work.

The predicted interaction involving gender (vocal tone × word meaning × gender) proved significant at 450–600 ms at the central and posterior electrode sites. At the Cz electrode site, the vocal tone × word meaning × gender interaction proved significant, $F(1, 35) = 4.87, P < 0.05, \eta_p^2 = 0.12$. As can be seen in Figure 3, the vocal tone × word meaning interaction was highly significant for female participants, $F(1, 19) = 14.04, P < 0.005, \eta_p^2 = 0.42$. The incongruity effect was significant for both positive vocal tones and

negative tones, $t(19) = 2.34, P < 0.05, d = 0.69$ and $t(19) = 3.46, P < 0.01, d = 1.02$, respectively. For male participants, however, the vocal tone × word meaning interaction was negligible.

The vocal tone × word meaning × gender interaction also proved significant at the Pz electrode site during the 450–600 ms time period, $F(1, 35) = 6.33, P < 0.05, \eta_p^2 = 0.15$. As illustrated in Figure 4, the data pattern was very similar to the one observed above at the central

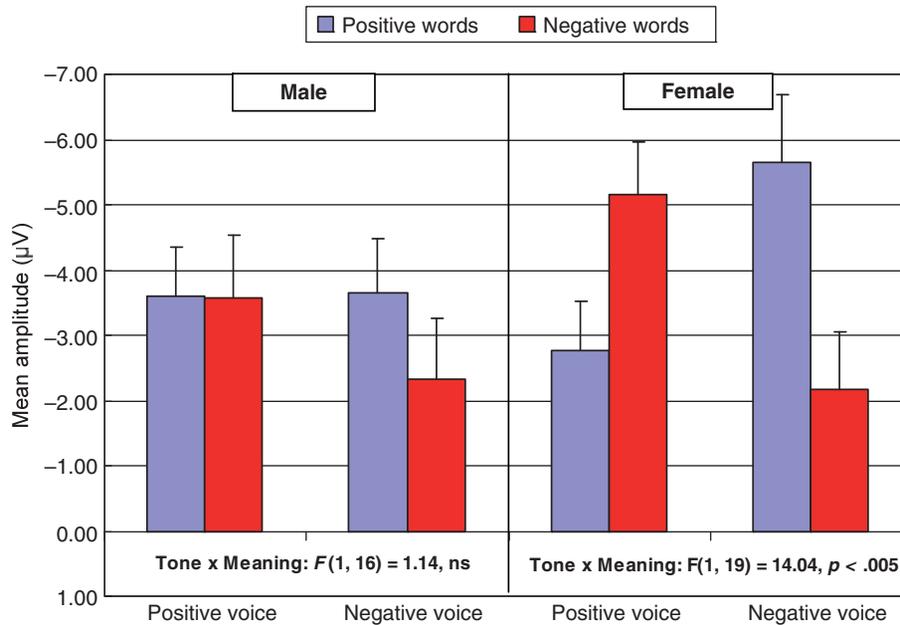


Fig. 3 Mean amplitudes of the late negativity for female and male participants: amplitude was assessed at the Cz site in the 450–600 ms time window.

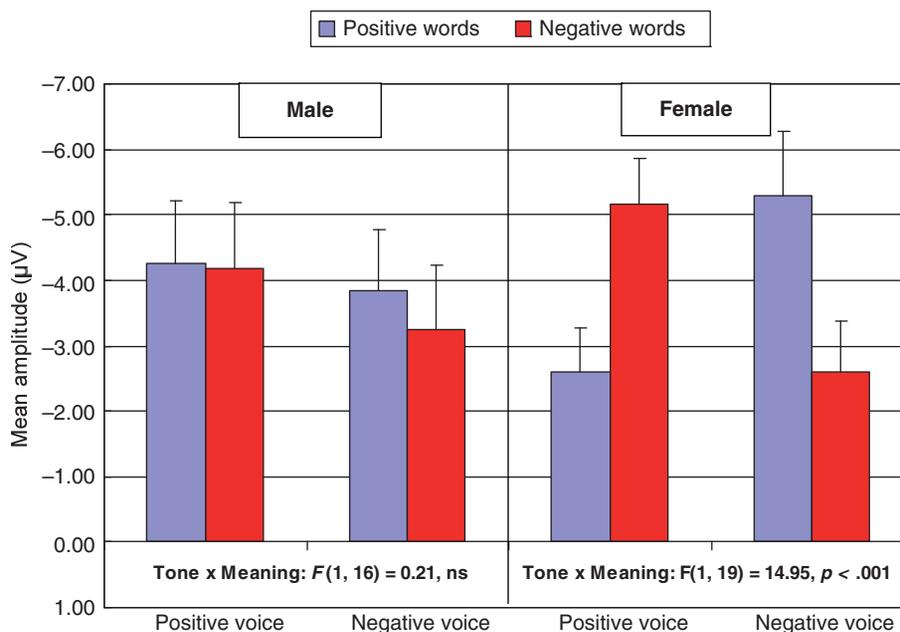


Fig. 4 Mean amplitudes of the late-negativity for female and male participants: amplitude was assessed at the Pz site in the 450–600 ms time window.

electrodes. The incongruity-based negativity was significant for female participants. $F(1, 19) = 14.95, P < 0.001, \eta_p^2 = 0.44$. This was the case for both positive and negative vocal tones, $t(19) = 2.74, P < 0.05, d = 0.81$ and $t(19) = 2.92, P < 0.01, d = 0.86$, respectively. As before, the interaction was negligible for male participants ($F < 1$). Together, the two interaction patterns summarized in Figures 3 and 4 replicate previous work by Schirmer and colleagues (Schirmer *et al.*, 2002, 2006; Schirmer and Kotz, 2003).

Recall the face effect was apparent when the vocal tones were positive, but not when they were negative. In contrast, the gender effect was clearly observed regardless of the valence of the vocal tones: Women were spontaneously more attentive to vocal tones than men were regardless of whether the tones were positive or negative.

Another predicted interaction involved vocal tone, word meaning, and chronic social orientation. Yet, the vocal tone \times word meaning \times chronic social orientation interaction was never significant in any of the sites throughout the time course. Instead, significant effects involving chronic social orientation showed up in conjunction with either face or gender. First, a vocal tone \times word meaning \times chronic social orientation \times face interaction was significant at Fz during the 450–600 ms period, $F(1, 35) = 4.93, P < 0.05$. To examine specific patterns, we first subtracted the negativity in the congruous cases from the negativity in the incongruous cases to obtain a measure of an incongruity-based negativity. This index was then plotted as a function of chronic social orientation in each of the two face conditions. Keep in mind that the current index of chronic social orientation takes more negative values as one becomes more interdependent (rather than independent). Further, incongruity-based negativity is indexed by negative amplitude differences.

As shown in Figure 5, incongruity-based negativity was predicted, albeit marginally, by chronic social orientation

in the face condition ($r = 0.39, P < 0.10$). This relationship was somewhat reversed in the no-face control condition ($r = -.29, ns$).

Next, a vocal tone \times word meaning \times gender \times chronic social orientation interaction was significant at Pz during the 600–750 ms period, $F(1, 35) = 6.84, P < 0.02$. As before the average negativity of the congruous cases was subtracted from the average negativity of the incongruous cases to yield a measure of incongruity-based negativity, which was then plotted as a function of chronic social orientation for both males and females. As shown in Figure 6, incongruity-based negativity was reliably predicted by chronic social orientation for female participants ($r = 0.42, P < 0.05$). In contrast, there was no such relationship for male participants ($r = -.28, ns$).

Finally, the same interaction effect (vocal tone \times word meaning \times gender \times chronic social orientation) was also significant at the Pz site during the subsequent time period (750–900 ms), $F(1, 35) = 4.50, P < 0.05$. The measure of incongruity-based negativity is plotted as a function of social orientation for both males and females. As shown in Figure 6, incongruity-based negativity was reliably predicted by chronic social orientation for female participants ($r = 0.44, P < 0.05$), but not for male participants ($r = -.03, ns$).

Overall, the data suggest that chronic social orientation does not modulate the late negativity in and by itself, but it does so in conjunction with one or the other of the remaining two indicators of social orientation we tested, namely, face and gender. First, in the face condition, the incongruity-based late negativity increased as a function of chronic social orientation. Curiously, this effect disappeared in the no-face control condition. Second, the incongruity-based late negativity also increased as a function of chronic social orientation for females, but not for males.

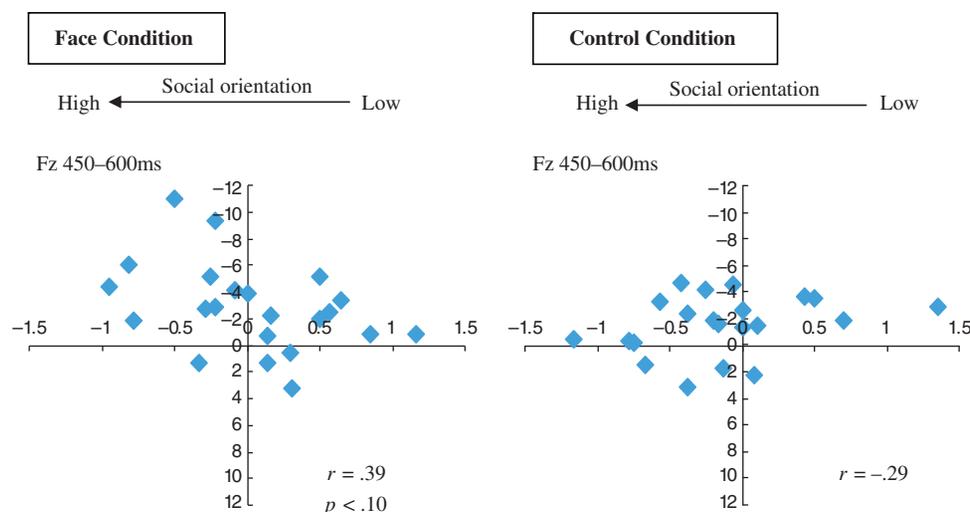


Fig. 5 Scatter plots of the incongruity-based negativity index (the amplitude of negativity for incongruous cases—the amplitude of negativity for congruous cases) as a function of chronic social orientation in the face and the no-face control conditions at the Fz site in the 450–600 ms time window.

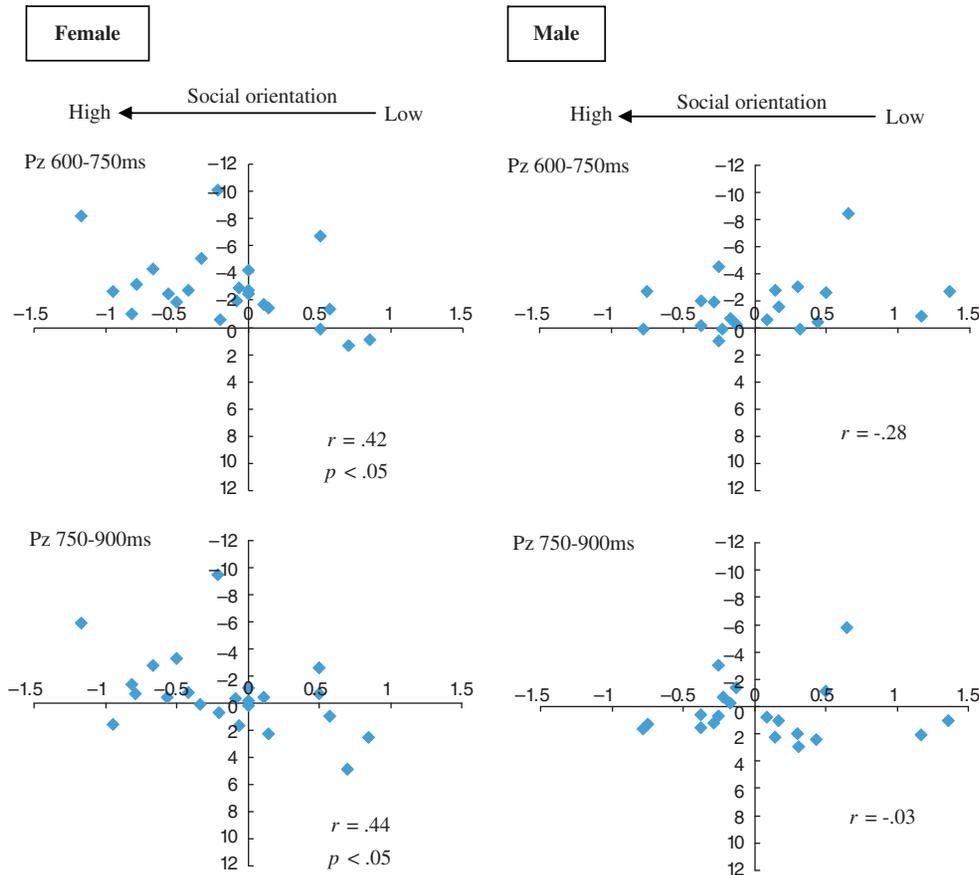


Fig. 6 Scatter plots of the incongruity-based negativity index (the amplitude of negativity for incongruous cases—the amplitude of negativity for congruous cases) as a function of chronic social orientation for female and male participants at the Pz site in the 600–750 ms and 750–900 time windows.

Why is the incongruity-based late negativity modulated by chronic social orientation under certain conditions, but not in certain others? We assessed chronic social orientation in terms of a tendency to be attuned to and oriented toward other individuals across multiple social situations. One conjecture, then, is that chronic social orientation, as assessed this way, may become an active element of behavioral regulation only in situations that are construed as 'social' or 'public' as opposed to 'non-social' or 'private'. When the situation is perceived as non-social and private, the same personal disposition may become irrelevant to and, thus, inert in regulating behavior. The social or public nature of the situation could be signaled by 'watching faces'. Moreover, females may be more likely than males to construe the same situation as potentially 'social' and 'public'. This explains why chronic social orientation increased the likelihood of detecting word-voice incongruity only for either those exposed to the 'watching faces' or for females.

According to this interpretation, there are two component processes that constitute chronic social orientation. One is the degree to which a given situation is construed as 'social' and 'public' as opposed to 'non-social' and 'private', and the

other refers to the extent to which one becomes socially engaged (versus disengaged) if the situation is construed as 'social' and 'public'. What this means is that social orientation is not a unified construct. To the contrary, it is constituted by at least two component processes. This analysis must be further examined in future work.

DISCUSSION

The present work offers the first evidence that the brain response to word-voice incongruity varies in highly systematic fashion as a function of three different indicators of social orientation (i.e. interdependence as opposed to independence). First, the incongruity-based late (N400-like) negativity was more pronounced when participants were unobtrusively exposed to human schematic faces. Evidently, this effect was more pronounced when the vocal tones were positive (rather than negative). Second, we replicated earlier studies (Schirmer *et al.*, 2002, 2006; Schirmer and Kotz, 2003) and showed that the incongruity-based late negativity is greater for females than for males. The gender effect was equally strong regardless of the valence of the vocal tones. Finally, the incongruity-based late negativity tended to increase as a function of

chronic social orientation in the frontal and posterior locations. Interestingly, this effect occurred for females, but not for males and, moreover, it tended to occur in the face condition, but not in the no-face control condition. We speculated that chronic social orientation becomes relevant only when the impinging situation is construed as social and public. The incidental exposure to face might increase the likelihood of such a construal. Likewise, females might be more likely than males to make such a construal.

The incongruity-based negativity was moderated by the three indicators of social orientation 450 ms after the stimulus onset and some of these moderation effects appear to persist for 500 ms (see Table 2). This finding may imply that even though the detection of incongruity is prompted by spontaneous attention to vocal tone, the moderation effect by social orientation is mediated by late-occurring, intentional attention. That is to say, social orientation may intensify intentional attention to vocal tone (i.e. a social cue), which in turn may enhance the salience of the incongruity between word meaning and vocal tone, leading to a larger N400 response.

Curiously, the behavioral measure (response time) yielded less clear results. The predicted effect of face was observed for males, but not females. Moreover, there was no effect of chronic social orientation in the behavioral measure (see footnote 2). Whereas response time is necessarily crude as a measure of the underlying component process that is involved, the brain measure may be tapping this process with a greater precision. That is to say, our ERP measure may be tapping the detection of incongruity between verbal meaning and vocal tone, whereas our response time measure may be assessing a much more down-stream outcome of incongruity between word and voice, i.e. response competition. It should not come as any surprise, then, that (1) the patterns of results are not aligned perfectly between the two measures and, moreover, that (2) the predictions regarding the detection of word-voice incongruity are borne out much more clearly with the ERP measure than with the response time measure.

One limitation of the present work is that we tested only Japanese participants. Future work should directly look at cross-cultural differences. Given the behavioral evidence of Ishii, Kitayama, and colleagues (Kitayama and Ishii, 2002; Ishii et al., 2003), the brain response to word-voice incongruity should also be greater for those engaging in interdependent cultures (e.g. Asian cultures) than for those engaging in independent cultures (e.g. North America and Western Europe).

One important challenge the field of cultural psychology is facing today is to determine the extent to which culture is inscribed into the brain. Nevertheless, it has become increasingly clear that culture is likely to influence some significant brain pathways involved in self, cognition, emotion, and motivation (e.g. Chiao and Ambady, 2007; Han and Northoff, 2008; Kitayama and Park, 2009; Fiske, 2009).

The emerging view of culture as a fundamental force in shaping these brain mechanisms implies that to an important extent the brain must be seen as an open system whose functions can change over time as it interacts with the surrounding socio-cultural environment. We thus believe that the understanding of brain processes would be greatly enhanced by testing them as a function of a variety of socio-cultural variables. The study reported here is a small step in this direction.

REFERENCES

- Ambady, N., Koo, J., Lee, F., Rosenthal, R. (1996). More than words: Linguistic and nonlinguistic politeness in two cultures. *Journal of Personality and Social Psychology*, 70, 996–1011.
- Bateson, M., Nettle, D., Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2, 412–414.
- Chiao, J. Y., Ambady, N. (2007). Cultural neuroscience: parsing universality and diversity across levels of analysis. In: Kitayama, S., Cohen, D., editors. *Handbook of Cultural Psychology*. NY: Guilford Press, pp. 237–54.
- Deecke, L., Grözinger, B., Kornhuber, H. H. (1976). Voluntary finger movement in man: cerebral potentials and theory. *Biological Cybernetics*, 23, 99–119.
- Fiske, S. T. (2009). Cultural processes. In: Berntson, G.G., Cacioppo, J.T., editors. *Handbook of Neuroscience for the Behavioral Sciences*. New York: Wiley, pp. 985–1004.
- Frishkoff, G. A., Tucker, D. M., Davey, C., Scherg, M. (2004). Frontal and posterior sources of event-related potentials in semantic comprehension. *Cognitive Brain Research*, 20, 329–54.
- Goren, C. C., Sarty, M., Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56, 544–9.
- Gray, H. M., Ambady, N., Ishii, K., Kitayama, S. (2007). *When Misery Loves Company: Effects of Sad Mood on Affiliation Goals and Attention to Relational Cues*. Tufts University, Unpublished manuscript.
- Grice, P. (1974). Method in philosophical psychology: from the banal to bizarre. *Proceedings and Addresses of the American Philosophical Association*, 48, 23–53.
- Hagoort, P., van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 801–11.
- Haley, K. J., Fessler, D. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous dictator game. *Evolution and Human Behavior*, 26, 245–56.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845–57.
- Han, S., Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: a transcultural neuroimaging approach. *Nature Neuroscience*, 9, 646–54.
- Ishii, K., Reyes, J. A., Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, 14, 39–46.
- Jodo, E., Kayama, Y. (1992). Relation of a negative ERP component to response-inhibition in a go/no-go task. *Electroencephalography and clinical neurophysiology*, 82, 477–82.
- Johnson, M. H., Dziurawiec, S., Ellis, H., Morton, J. (1991). Newborns preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40, 1–19.
- Kitayama, S., Duffy, S., Uchida, Y. (2007). Self as cultural mode of being. In: Kitayama, S., Cohen, D., editors. *Handbook of Cultural Psychology*. NY: Guilford Press, pp. 136–74.
- Kitayama, S., Ishii, K. (2002). Word and voice: spontaneous attention to emotional utterances in two languages. *Cognition and Emotion*, 16, 29–60.
- Kitayama, S., Mesquita, B., Karasawa, M. (2006). Cultural affordances and emotional experience: socially engaging and disengaging emotions in

- Japan and the United States. *Journal of Personality and Social Psychology*, 91, 890–903.
- Kitayama, S., Park, H. (2007). Cultural shaping of self, emotion, and well-being: how does it work? *Social and Personality Psychology Compass*, 1, 202–22.
- Kitayama, S., Park, J. (2009). *The social self, the social brain: a perspective of cultural neuroscience*, University of Michigan, unpublished manuscript.
- Kitayama, S., Snibbe, A.C., Markus, H.R., Suzuki, T. (2004). Is there any 'free' choice?: self and dissonance in two cultures. *Psychological Science*, 15, 527–33.
- Kutas, M., Federmeirer, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463–70.
- Kutas, M., Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203–5.
- Markus, H. R., Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–53.
- Pickett, C. L., Gardner, W. L., Knowles, M. (2004). Getting a cue: the need to belong and enhanced sensitivity to social cues. *Personality and Social Psychology Bulletin*, 30, 1095–107.
- Rigdon, M. L., Ishii, K., Watabe, M., Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30, 358–67.
- Sanchez-Burks, J. (2002). Protestant relational ideology and (in)attention to relational cues in work settings. *Journal of Personality and Social Psychology*, 83, 919–29.
- Schirmer, A., Kotz, S. A. (2003). ERP evidence for a gender-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, 15, 1135–48.
- Schirmer, A., Kotz, S. A., Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, 14, 228–33.
- Schirmer, A., Lui, M., Maess, B., Escoffier, N., Chan, M., Penney, T. (2006). Task and sex modulate the brain response to emotional incongruity in Asian Listeners. *Emotion*, 6, 406–17.
- Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, 18, 643–62.
- Zuckerman, M., Amidon, M. D., Bishop, S. E., Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology*, 32, 347–57.