Neurocultural evidence that ideal affect match promotes giving

BoKyung Park, Elizabeth Blevins, Brian Knutson, and Jeanne L. Tsai

Department of Psychology, Stanford University, Stanford, CA 94305, USA

Correspondence should be addressed to Jeanne L. Tsai, Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305 USA.
E-mail: jltsai@stanford.edu.

Abstract

Why do people give to strangers? We propose that people trust and give more to those whose emotional expressions match how they ideally want to feel ("ideal affect match"). European Americans and Koreans played multiple trials of the Dictator Game with recipients who varied in emotional expression (excited, calm), race (White, Asian) and sex (male, female). Consistent with their culture's valued affect, European Americans trusted and gave more to excited than calm recipients, whereas Koreans trusted and gave more to calm than excited recipients. These findings held regardless of recipient race and sex. We then used fMRI to probe potential affective and mentalizing mechanisms. Increased activity in the nucleus accumbens (associated with reward anticipation) predicted giving, as did decreased activity in the right temporo-parietal junction (rTPJ; associated with reduced belief prediction error). Ideal affect match decreased rTPJ activity, suggesting that people may trust and give more to strangers whom they perceive to share their affective values.

Key words: dictator game; giving; culture; emotion; fMRI

Introduction

People sometimes give to others whom they do not know and may never meet again. Researchers have elegantly captured this phenomenon in an experimental task called the "Dictator Game," in which individuals receive money to distribute between themselves and others ("recipients") (Forsythe *et al.*, 1994; Kahneman *et al.*, 1986). In contrast to pure self-interest, across studies, ~60% of individuals choose to give at least some of their money to strangers (Engel, 2011). Although minimizing concerns about reciprocity or reputation can decrease giving somewhat (Franzen and Pointner, 2012), as many as 40% of people still give something, suggesting that other factors drive giving. Here, we propose that people may give more to strangers whom they perceive to share their affective values (or "ideal affect"), since the perception of shared values enhances trust. Because cultures differ in ideal affect, this may account for cultural differences in giving.

Ideal affect match

Although most people report wanting to feel good in general, they vary in terms of which specific types of positive states they ideally want to feel (their "ideal affect"). Ideal affect can be distinguished from how people actually feel (their "actual affect") and varies across cultures (Tsai, Knutson, & Fung, 2006; Tsai, 2007; Tsai *et al.*, 2007a, b). For example, European Americans typically value high-arousal positive states (HAP) such as excitement and enthusiasm more, and low-arousal positive states (LAP) such as peacefulness and serenity less, than East Asians (Tsai *et al.*, 2006). These differences are reflected in popular media content found in children's storybooks, Facebook pages and leaders' official photos (Huang and Park, 2013; Tsai *et al.*, 2007, 2016). Ideal affect can also influence social perceptions and actions. For instance, the more people value HAP, the more likely they are to choose physicians who promote "exciting" *vs* "calming" lifestyles, in part because they trust them more (Sims

et al., 2014). Furthermore, European Americans prefer excited *vs* calm physicians more than do Hong Kong Chinese (Sims *et al.* manuscript under review). Variation in ideal affect does not depend on people's actual affective experience.

Ideal affect might also influence people's willingness to give to others. People might interpret others' emotional expressions as signals of shared affective values, which could promote trust. Since people give more to those they perceive as trustworthy (Van't Wout and Sanfey, 2008; Fehrler and Przepiorka, 2013), we propose that people may give more to strangers whose expressions match their ideal affect ("ideal affect match"). Thus, the more people value HAP, the more they should give to excited *vs* calm recipients, but the more people value LAP, the more they should give to calm *vs* excited recipients.

This study has the potential to fill several gaps in the literature on giving. First, although previous studies indicate that people give more to smiling than non-smiling recipients (Tidd and Lockard, 1978; Scharlemann *et al.*, 2001; Krumhuber *et al.*, 2007; Genevsky *et al.*, 2013; Tortosa *et al.*, 2013), research has not distinguished among different positive expressions (e.g., open, toothy "excited" smiles *vs* closed "calm" smiles). Second, while previous research suggests that people give more to recipients who appear more trustworthy (Fehrler and Przepiorka, 2013; Van't Wout and Sanfey, 2008), it remains unclear which specific emotional signals increase perceived trustworthiness. Finally, although large-scale studies suggest that giving varies across cultures (Henrich *et al.*, 2005), the specific mechanisms underlying cultural differences in giving remain unclear. For instance, several theorists have proposed that individuals should give more to others from the same ingroup (Fong and Luttmer, 2007). Empirical findings, however, are mixed when ingroup is defined by race, national identity, experimentally induced membership or institutional affiliation (Buchan *et al.*, 2006; Ferraro and Cummings, 2007; Suzuki *et al.*, 2007). Here, we instead propose that culture might influence giving through perceived ideal affect match, above and beyond matches in less malleable characteristics like race or sex. Specifically, members of cultures that value HAP should give more to excited *vs* calm recipients, whereas members of cultures that value LAP should give more to calm *vs* excited recipients.

Neurobehavioral mechanisms of ideal affect match

Ideal affect match could enhance giving in multiple ways. First, ideal affect match could induce positive affective experience, which might enhance perceived trustworthiness and giving. Indeed, increased brain activity in circuits associated with positive arousal and reward (such as the nucleus accumbens [NAcc]) can promote giving across a number of charitable and lending contexts (Moll *et al.*, 2006; Harbaugh *et al.*, 2007; Hare *et al.*, 2010; Genevsky *et al.*, 2013; Genevsky and Knutson, 2015). Thus, even in the minimal social context of a Dictator Game, NAcc activity may promote giving, particularly if potential recipients' emotional expressions match individuals' ideal affect. Indeed, consistent with cultural differences in ideal affect, Chinese individuals show reduced ventral striatal (including NAcc) activity in response to excited (*vs* calm) faces compared with their European American counterparts, and ventral striatal activity during face viewing correlated with choices to view excited *vs* calm faces months after scanning (Park *et al.*, 2016).

Second, ideal affect match may influence giving through a mentalizing mechanism. The specific relationship between mentalizing and giving, however, is unclear. For instance,

increases in right temporo-parietal junction [rTPJ] activity have been associated with belief prediction error, or the perception of others' incongruent mental states (Saxe and Wexler, 2005; Cloutier *et al.*, 2011; Koster-Hale and Saxe, 2013; Mende-Siedlecki *et al.*, 2013). When forced to choose between allocating to the self *vs* others, increases in rTPJ activity have been associated with more generous choices (Hutcherson *et al.*, 2015), particularly when they override selfish tendencies (Strombach *et al.*, 2015). Thus, ideal affect match may promote giving through increased rTPJ activity.

However, there may be circumstances when the dominant response is to give (e.g. when people expect to give, as when making charitable donations). Under these circumstances, perceiving incongruent mental states may actually interfere with and inhibit giving. Thus, ideal affect match may promote giving through decreased rTPJ activity. Assuming an underlying tendency to give in the Dictator Game (Rand *et al.*, 2014), we predicted that ideal affect match would decrease perceived differences in beliefs and associated rTPJ activity, which would promote giving.

Thus, to examine the effects of ideal affect match on giving as well as underlying neural mechanisms, we conducted two studies during which participants played multiple rounds of a Dictator Game with recipients who varied in their positive emotional expressions.

Study 1: Ideal affect match and giving in the U.S. and Korea

In Study 1, we recruited European Americans and Koreans, and we hypothesized that: (1) European Americans would value HAP *vs* LAP more than their Korean counterparts; (2) European Americans would trust and offer more money to excited *vs* calm recipients than their Korean counterparts; and (3) these differences in offers would be mediated by ideal affect match, which would also drive perceived trustworthiness.

Materials and methods

Participants

We recruited 101 European American students (61.4% female) from Stanford University and 65 Korean students (58.5% female) from Seoul National University for a study on "personality and decision-making." European Americans were born and raised in the U.S., had parents who were born and raised in the U.S., and had grandparents who were born and raised in North America or Western Europe. Koreans were born and raised in Korea and had parents and grandparents born and raised in Korea. We excluded seven participants who did not vary their responses across trials (two European American, five Koreans; see Study 2 for rationale), resulting in a final sample of 99 European Americans and 60 Koreans. Results did not change when analyses included the entire sample. All procedures were approved by the Stanford Institutional Review Board.

The groups differed in age (European American Mean = 19.42, SD = 1.96; Korean Mean = 23.00, SD = 2.57; $t(157) = -9.88$, $P < 0.001$ and family annual income (1 = less than \$10,000, 2 = \$10 001–\$20 000, 3 = \$20 001–\$30 000, 4 = \$30 001–\$40 000, 5 = \$40 001–\$50 000, 6 = \$50 001–\$75 000, 7 = \$75 001–\$100 000, 8 = over \$100 000); European American Mean = 7.24, SD = 0.82; Korean Mean = 5.03, SD = 1.94, $t(157) = 9.97$, $P < 0.001$. However, results did not change when age and family income were entered as

covariates in our analyses, and therefore, these differences will not be discussed further.

Facial stimuli

Using the Facegen Modeller program (http://facegen.com), we created faces that varied by expression (excited, calm, neutral), race (White, Asian) and sex (male, female), resulting in 12 recipient groups (see Supplementary Materials Section S1 for Facegen parameters). Neutral faces were included as fillers (see Supplementary Materials Section S2 for results with neutral faces). For each recipient group, we generated two different identities, resulting in 24 different faces. We counterbalanced whether participants saw a particular recipient's excited or calm expression as well as the order in which excited *vs* calm recipients were presented.

Instruments

To assess participants' actual affect and ideal affect, we administered the Affect Valuation Index (AVI) (Tsai *et al.*, 2006). Participants rated how often they *actually felt* and how often they *ideally wanted to feel* 39 different affective states over the course of a typical week, using a 5-point scale 1 ("never") to 5 ("all the time"). These states sampled each octant of the affective circumplex as well as other emotional states: elated, enthusiastic, euphoric, excited, content, happy, satisfied, peaceful, calm, relaxed, serene, fearful, hostile, nervous, sad, unhappy, ashamed, disgusted, stressed, guilty, contemptuous, fatigued, angry, lonely, dull, sleepy, sluggish, astonished, surprised, strong, aroused, rested, energetic, quiet, still, idle, passive, inactive, no emotion) (Watson and Tellegen, 1985; Larsen and Diener, 1991; Barrett and Russell, 1999). We created actual HAP (alpha: European Americans = 0.81; Koreans = 0.69) and ideal HAP (alpha: European Americans = 0.66; Koreans = 0.60) aggregates by averaging ratings of actual and ideal "enthusiastic," "excited," "elated" and "euphoric," respectively, and actual LAP (alpha: European Americans = 0.84; Koreans = 0.73) and ideal LAP (alpha: European Americans = 0.72; Koreans = 0.57) by averaging "peaceful" and "serene," respectively. Although the LAP aggregate usually also includes "calm" and "relaxed," these terms were not reliable for the European American participants, and therefore, were not included in the LAP aggregate in this study (but were included in Study 2).

Dictator game

Participants played 24 trials of the Dictator Game, each with a different recipient. In each trial, participants were given an endowment of $6 or $14, shown a putative picture of the recipients' avatar face (Figure 1A), and then given an opportunity to offer from $0 to $6 (or $14) in 1-dollar increments to that recipient.

Procedure

Participants were recruited on-line. At the beginning of the study, participants were told that they would be randomly assigned to one of two roles in a game, the "proposer" or the "recipient," and would play the game with different players. Proposers would receive a certain amount of money, which they could then offer to the recipients. Recipients had no choice but to accept the offer.

To explain why participants would see computer-generated faces, before playing the game, participants were told they would see their partners' avatars during the game, as would their partners, and therefore, they needed to choose their own avatar. Participants were then asked to choose an avatar that they identified with most among 12 faces that varied by expression, race and sex, but that differed from the faces used in the actual game (see Supplementary Materials Section S3 for participants' avatar selection).

Participants were then told that they had been assigned to be the "proposer" and would play 24 trials of the game with ostensibly 24 different recipients (described above). Participants were informed that one of the trials of the game would be randomly chosen, and the amount of money that they kept for themselves on that trial would be added to their participation compensation. Thus, participants understood that their decisions to give on any given trial could have real monetary consequences.

After completing the task, participants viewed a subset (12 out of 24) of recipients' avatars again, and rated the extent to which each recipient was trustworthy, friendly, assertive, dominant, intelligent and financially needy, along with other filler items, using a scale from 1 = 'Not at all', to 7 = 'Extremely.' We included these traits to assess whether our findings were specific to trustworthiness. Afterwards, participants completed the AVI.

All study materials were translated and back-translated into Korean by two independent translators.

Data analyses and results

Hypothesis 1: Cultural differences in ideal affect

We conducted a repeated-measures ANOVA, with ideal affect as a within-subjects factor and participants' culture as a between-subject factor. Because participants' sex did not change our results, we collapsed across participants' sex to increase statistical power. Actual HAP and LAP were treated as covariates in the model, to control for overlap between actual and ideal affect. As predicted, there was a significant Ideal Affect × Participant Culture interaction, $F(1,152) = 19.33$, $P < 0.001$, partial $\eta^2 = 0.11$, showing that European Americans valued HAP more, $P = 0.001$, 95% CI = [0.18, 0.67] and LAP less, $P = 0.052$, 95% CI = [−0.57, 0.00] than Koreans. Within cultures, European Americans valued HAP ($M = 3.91$, SE = 0.06) more than LAP ($M = 3.73$, SE = 0.07), $P = 0.027$, 95% CI = [0.02, 0.34], whereas Koreans valued LAP ($M = 4.01$, SE 0= 0.10) more than HAP ($M = 3.48$, SE = 0.09), $P < 0.001$, 95% CI = [−0.75, −0.30] (Figure 1B).

Hypothesis 2: Cultural differences in offers and perceived trustworthiness

We divided the amount participants offered in each trial by the amount of endowment in that trial ($6 or $14) to compute an 'offer ratio' made to each recipient. We conducted a 2 Participant Culture (European American, Korean) × 2 Recipient Expression (excited, calm) × 2 Recipient Race (White, Asian) × 2 Recipient Sex (Male, Female) × 2 Amount of Endowment ($14, $6) repeated-measures ANOVA on the ratio of offer. Participant Culture was a between-subjects factor, while Recipient Expression, Recipient Race, Recipient Sex and Amount of Endowment were within-subjects factors. We collapsed across participant sex since it did not alter the results.

Overall, European Americans ($M = 0.53$, SE = 0.02) offered significantly more than Koreans ($M = 0.34$, SE = 0.03), $F(1,153) = 27.53$,
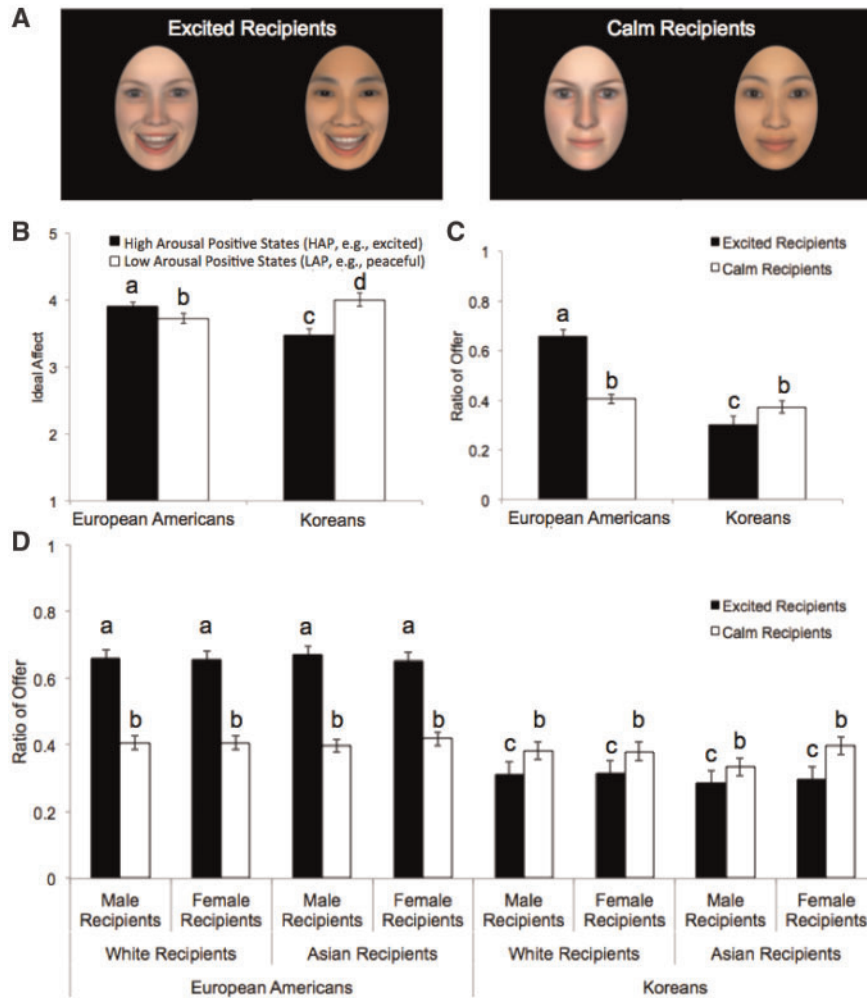
Fig. 1. Ideal affect match influences giving (Study 1). (A) *Facial stimuli*: Stimuli varied by emotional expression (excited, calm, neutral), race (White, Asian) and sex (male, female). Neutral expressions were treated as fillers, and are shown in Supplementary Materials Section S1. (B) *Cultural differences in ideal affect*: European Americans valued HAP more than LAP, Koreans valued LAP more than HAP, and European Americans valued HAP more and LAP less than Koreans. Ps $\leq$ 0.05, *except b vs d P* = 0.052. (C) *Ratio of offers*: European Americans offered more (regardless of endowment) to excited than calm recipients; Koreans offered more to calm than excited recipients. European Americans offered more to excited recipients than did Koreans. Ps < 0.001. (D) *Ratio of offers divided by recipients' expression, race, and sex*: Cultural differences in offers to excited *vs* calm recipients held across recipient race and sex. Error bars represent standard errors (S.E.). Ps < 0.05. In all panels, different letters indicate significant differences at indicated P-levels.

$P < 0.001$, partial $\eta^2 = 0.15$, 95% CI = [0.12, 0.27] (Figure 1C). Participants gave more to excited ($M = 0.48$, SE = 0.02) than calm recipients ($M = 0.39$, SE = 0.02), $F(1,153) = 58.45$, $P < 0.001$, partial $\eta^2 = 0.28$, 95% CI = [0.07, 0.11]. However, these main effects were qualified by a significant Participant Culture by Recipient Expression interaction, $F(1,153) = 186.95$, $P < 0.001$, partial $\eta^2 = 0.55$. As predicted, European Americans offered significantly more to excited than to calm recipients, $P < 0.001$, 95% CI = [0.22, 0.28], while Koreans offered significantly more to calm than excited recipients, $P < 0.001$, 95% CI = [−0.11, −0.03]. Also consistent with observed cultural differences in ideal affect, European Americans offered significantly more to excited recipients ($M = 0.66$, SE = 0.03) than did Koreans ($M = 0.30$, SE = 0.04), $P < 0.001$, 95% CI = [0.27, 0.45]. Contrary to predictions, there were no significant cultural differences in offers to calm recipients, $P = 0.296$, 95% CI = [−0.03, 0.10] (European American $M = 0.41$, SE = 0.02; Korean $M = 0.37$, SE = 0.03). When offers were mean-deviated to control for the main effect of culture, however, Koreans did offer more to calm recipients than did European Americans (see Supplementary Materials Section

S4A). These findings held across recipient race, sex and amount of endowment, all Ps > 0.40 (Figure 1D).

Showing a similar pattern, a 2 Participant Culture (European American, Korean) × 2 Recipient Expression (excited, calm) × 2 Recipient Race (White, Asian) × 2 Recipient Sex (Male, Female) repeated-measures ANOVA on perceived trustworthiness revealed that European Americans ($M = 5.39$, SE = 0.06) rated recipients overall as more trustworthy than did Koreans ($M = 3.87$, SE = 0.08), $F(1,149) = 221.87$, $P < 0.001$, partial $\eta^2 = 0.60$. Participants rated excited recipients ($M = 4.83$, SE = 0.07) as more trustworthy than calm recipients ($M = 4.42$, SE = 0.06), $F(1,149) = 23.16$, $P < 0.001$, partial $\eta^2 = 0.14$.

As predicted, however, these main effects were qualified by a significant Participant Culture × Recipient Expression interaction, $F(1,149) = 266.92$, $P < 0.001$, partial $\eta^2 = 0.64$. European Americans rated recipients as more trustworthy than did Koreans, $P < 0.001$, 95% CI = [2.63, 3.18]; European Americans rated excited recipients as more trustworthy ($M = 6.28$, SE = 0.09) than calm recipients ($M = 4.49$, SE = 0.08), $P < 0.001$, 95% CI = [1.59, 2.01]; and Koreans rated calm recipients as more

trustworthy ($M = 4.36$, $SE = 0.10$) than excited recipients ($M = 3.38$, $SE = 0.11$, $P < 0.001$, 95% CI = [-1.24, -0.72]. Although there were no cultural differences in the trustworthiness ratings of calm recipients, $P = 0.300$, when trustworthiness ratings were mean-deviated, Koreans rated calm recipients as more trustworthy than did European Americans (see Supplementary Materials Section S4B). These findings held across recipient race and sex.

Hypothesis 3: Cultural differences in offers mediated by ideal affect match and trustworthiness ratings

To test meditational hypotheses, we created four difference scores by subtracting: (1) participants' ratings of trustworthiness of calm recipients from those of excited recipients, collapsing across recipient race and sex, (2) participants' mean ratio of offers to calm recipients from those to excited recipients, collapsing across recipient race and sex, (3) participants' ideal LAP from their ideal HAP or 'Participant Ideal Affect (HAP-LAP)' and (4) participants' actual LAP from their actual HAP or 'Participant Actual Affect (HAP-LAP).' We then conducted a serial mediation analysis (bootstrapped $n = 10\,000$) using the "process" macro (Hayes, 2012).

Participant Culture ($+1 = $EA, $-1 = $KR) significantly predicted Participant Ideal Affect (HAP-LAP) controlling for Participant Actual Affect (HAP-LAP), $B = 0.40$, $SE = 0.07$, $ß = 0.42$, $t = 5.36$, $P < 0.001$, 95% CI = [0.25, 0.55], indicating that European Americans valued HAP *vs* LAP more than Koreans (Figure 2; path a1). Second, Participant Ideal Affect (HAP-LAP) significantly predicted higher trustworthiness ratings of excited *vs* calm recipients, $B = 0.32$, $SE = 0.14$, $ß = 0.18$, $t = 2.22$, $P = 0.028$, 95% CI = [0.04, 0.60] (path a3). The more participants valued HAP *vs* LAP, the more they perceived excited recipients versus calm recipients as trustworthy. Third, trustworthiness ratings of excited *vs* calm recipients significantly predicted mean ratio of offers to excited *vs* calm recipients (path b2). The more participants trusted excited *vs* calm recipients, the more they offered to excited *vs* calm recipients, $B = 0.07$, $SE = 0.01$, $ß = 0.52$, $t = 7.06$, $P < 0.001$, 95% CI = [0.05, 0.08]. The significant effect of Participant Culture on differences in offer to excited *vs* calm recipients (path c), $B = 0.16$, $SE = 0.01$, $ß = 0.71$, $t = 11.40$, $P < 0.001$, 95% CI = [0.13, 0.18], was reduced after entering participants' ideal affect and trustworthiness ratings in the model (path c'), $B = .06$, $SE = 0.02$, $ß = 0.29$, $t = 3.83$, $P < 0.001$, 95% CI = [0.03, 0.10]. The indirect effect through ideal affect and trustworthiness ratings was significant, Standardized Effect = 0.04, $SE = 0.02$, 95% CI = [0.01, 0.09] (Figure 2). Thus, as predicted,

cultural differences in offers to excited *vs* calm recipients were mediated by differences in ideal affect and trustworthiness ratings.

When substituting Participant Actual Affect (HAP-LAP) for Participant Ideal Affect (HAP-LAP), the same model was not significant, Standardized Effect = -0.0001, $SE = 0.0037$, 95% CI = [-0.0095, 0.0069], suggesting that the ideal affect match, not actual affect match, contributed to the observed effects. Moreover, none of the other trait ratings significantly mediated the effect of ideal affect match on offer (see Supplementary Materials Section S5).

We reproduced this pattern of findings in another study (32 European Americans, 55 Koreans) using more realistic facial stimuli (i.e. real faces that were imported into FaceGen). European Americans offered more to realistic excited recipients than realistic calm recipients compared with Koreans, and these cultural differences in offers were mediated by differences in perceived trustworthiness (see Supplementary Materials Section S6).

Study 1: Discussion

As predicted, European Americans offered more to excited *vs* calm recipients relative to their Korean counterparts. This difference was selectively mediated by cultural differences in ideal affect and the perceived trustworthiness of excited versus calm recipients, and held across recipient race and sex.

To examine the underlying neural mechanisms by which ideal affect influenced decisions to give, we conducted another study in which participants played the Dictator Game in the scanner. We predicted that people would give more to recipients whose emotional expressions matched their ideal affect, and with whom they experienced either greater liking (indexed by greater NAcc activity) and/or greater similarity in beliefs and values (indexed by lesser rTPJ).

Study 2: Neural mechanisms underlying the influence of ideal affect match on giving

We predicted: (1) participants would trust and give more to recipients whose emotional expressions matched their ideal affect; (2) the more participants experienced positive arousal (associated with increased NAcc activity) and/or perceived similar beliefs and feelings (associated with reduced rTPJ activity), the more money they would offer to recipients; (3) when participants saw expressive faces that matched their ideal affect, they would experience more positive arousal or reward (increasing
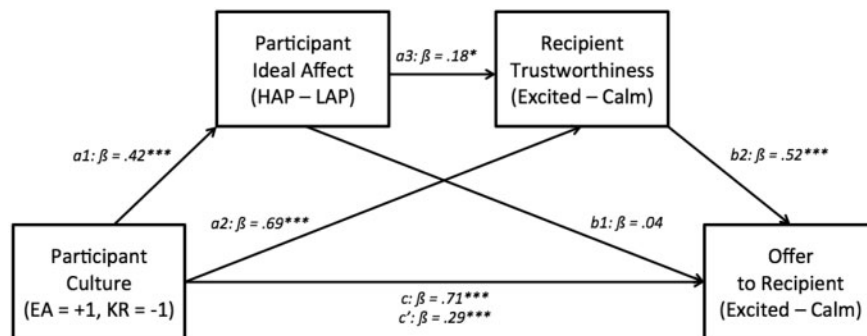

Fig. 2. Ideal affect match mediates cultural differences in giving (Study 1). Compared with Koreans, European Americans valued HAP *vs* LAP more, which predicted greater trust in excited *vs* calm recipients, which predicted higher offers to excited *vs* calm recipients. We report standardized *ß* values and unstandardized culture codes (European American = +1, Korean = −1) for ease of presentation. Additional significant indirect effects are available upon request. *P < 0.05, ***P < 0.001.

activity in NAcc), and/or they would perceive more similar thoughts and feelings (reducing activity in rTPJ); and (4) ideal affect match would increase giving by increasing trust, as well as increasing NAcc activity, and/or decreasing rTPJ activity.

Method

Participants

We recruited 30 healthy, right-handed European American and 24 Korean students (18–30-years-old) from San Francisco Bay Area universities for a study of "personality and decision-making." Although we ideally sought to compare European Americans living in the US with Koreans living in Korea as in Study 1, to control for scanner characteristics, we recruited European American and Korean participants living in the US. Korean participants spent on average 20.53 months (SD = 17.16) in US. Because the number of months spent in US did not alter the results, we omitted this variable from further analyses to maximize statistical power. Participant numbers were based on the required sample size (15 per group), estimated by mean ratio of offers from European Americans and Koreans in Study 1, to achieve desired power = 0.80 and alpha = 0.001 (https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html).

Eight participants were excluded from further analyses due to excessive head movement (>2 mm) (five participants); an anatomical abnormality (one participant); and inconsistency in the study procedure (two participants). In addition, 10 participants who varied their responses during fewer than 10% of trials were excluded due to insufficient variation in behavioral responses to examine neural predictors of giving. Analyses were conducted on the remaining 18 European Americans and 18 Koreans.

Because this study aimed to examine the influence of ideal affect on giving, which varies both within and across cultures, and cultural group differences in ideal affect did not differ significantly (possibly due to self-selection, since Koreans who chose to study in the U.S. and participate in an fMRI study may have valued HAP more), we collapsed across cultural groups in our final analyses (total N = 36; 16 females), and treated ideal affect match rather than cultural grouping as the independent variable. All procedures were approved by the Stanford Institutional Review Board.

Facial stimuli

For the fMRI study, we created faces that varied in expression (no smile [neutral], low-intensity smile [calm], moderate-intensity smile [moderately excited], high-intensity smile [excited]), race (White, Asian) and sex (male, female), resulting in 16 different recipient groups. For each recipient group we created three different faces, resulting in 48 faces in total (Supplementary Materials Section S1). Since our hypotheses focused on excited *vs* calm expressions, we binned the four levels of expressions into 'calm' (by aggregating "no smile" and 'low-intensity' smiles) and 'excited' (by aggregating 'moderate-intensity' and 'high-intensity' smiles) categories as in Park *et al.* (2016). However, analyses with the four levels of expression yielded similar results (Supplementary Materials Section S7).

Instruments

To assess actual and ideal affect, participants completed the same version of the AVI used in Study 1. We created actual HAP (alpha = 0.56) and ideal HAP (alpha = 0.74) aggregates by averaging ratings of actual and ideal "elated," "enthusiastic," "euphoric" and "excited," respectively, and actual LAP (alpha = 0.86) and ideal LAP (alpha = 0.85) aggregates by averaging ratings of actual and ideal "peaceful," "calm," "relaxed" and "serene," respectively. As mentioned above, there were no significant cultural differences in ideal HAP or LAP (European Americans: ideal HAP M = 3.23, SE = 0.15, ideal LAP M = 3.64, SE = 0.18; Koreans: ideal HAP M = 3.35, SE = 0.15, ideal LAP M = 3.92, SE = 0.18, F(1,32) = 0.38, P = 0.544).

Dictator game

As in Study 1, participants were presented with one facial stimulus per trial. Participants were endowed with $12 at the onset of each trial, and made offers in increments of $3, ranging from $0 to $12. We did not vary the absolute endowment across trials since this did not influence the results in the previous studies described above. To maintain participants' attention, $0 appeared on the left or on the right of the scale pseudo-randomly. Each recipient face was presented twice, resulting in a total of 96 trials. The presentation order of recipient faces was pseudo-randomized and fixed for all participants to mitigate against systematic carry-over effects in fMRI activity time course analyses.

Each trial began with a recipient face presented for 4 s. All facial stimuli (i.e. face and keyhole) occupied 640 × 640 pixels and were presented in the center of a black screen on a 47″ LCD display, with a screen resolution of 1920 × 1080 p, and then projected to a 17.78 × 6.35 cm mirror with viewing distance of 15 cm from the eyes. Next, a scale appeared for 4 s, during which participants were prompted to make their offer to the recipient. If participants did not make their offer within this window, the trial ended and was counted as 'missed.' Trials ended with a jittered fixation interval (2–6 s, mean = 4 s) (Figure 3A). Equal numbers of each inter-trial interval were evenly distributed across conditions in a pseudorandomized order.

Procedure

Prior to scanning, participants were told that they would play a game with other players whom they did not know, but whose avatar faces they would see. Approximately 1 week before their scan (M = 6.28 days before the scan, SD = 6.91), participants selected their own avatar and completed the AVI in an online survey.

Before scanning, participants were instructed about the game, and told that they had been assigned the role of proposer. Participants were informed that one of the 96 trials would be randomly chosen, and they would receive the amount they kept for themselves on the selected trial. They were also informed that the amount of money they offered to the recipient would go into a 'recipient pool' that included offers from previous proposers. Participants were told that at the end of the study, they would play one trial as a recipient and receive a randomly selected offer that previous proposers made, which they did.

Participants played four practice trials of the game with different faces and then entered a 3.0-T General Electric Discovery MR750 scanner outfitted with a 32-channel head coil. Once inside the scanner, participants completed 96 trials of the dictator game (total time = 19 min 32 s) while functional scans were acquired. Forty-six slices of gradient echo T2* weighted echo-planar images (EPI) provided whole brain coverage (axial acquisition from inferior to superior; interleaved EPI; repetition time, 2 s; echo time, 25 ms; flip angle, 77°; in-plane resolution and
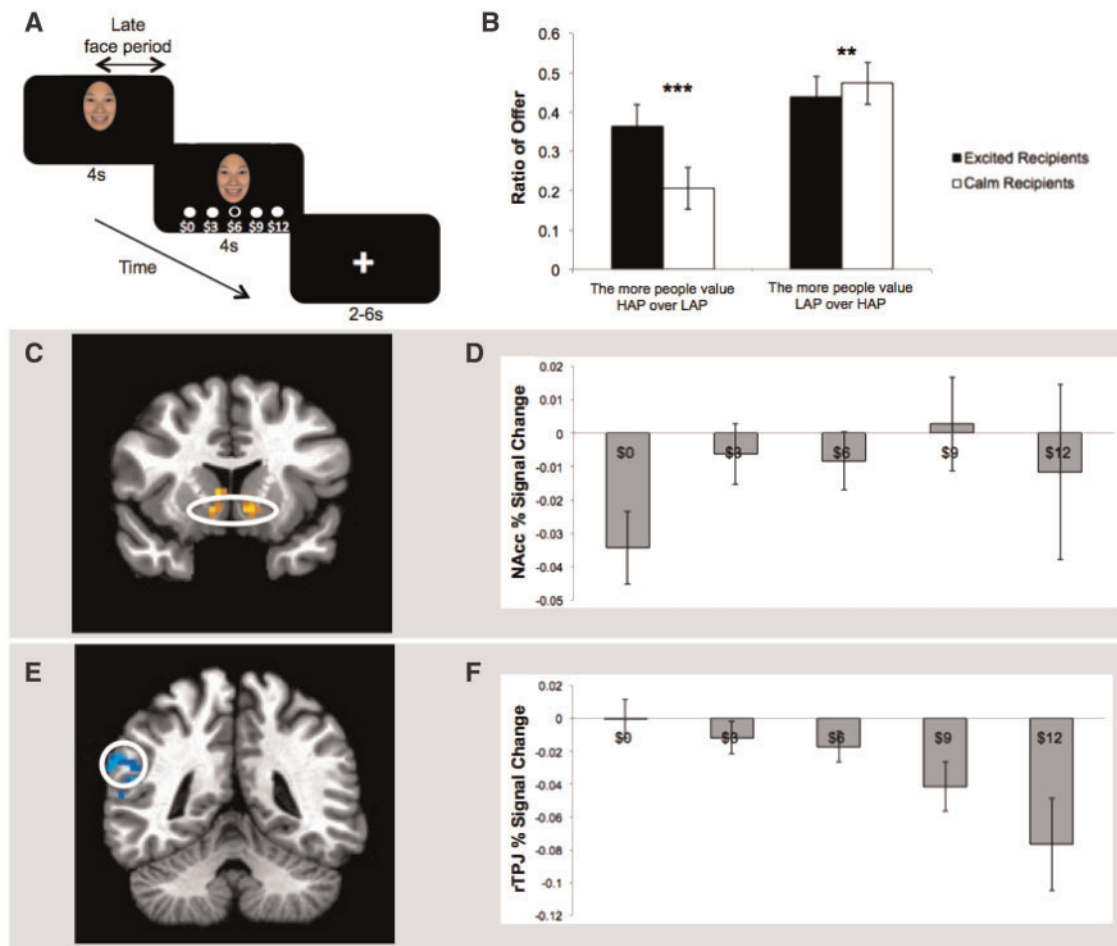
Fig. 3. Neural predictors of giving (Study 2). (A) *Representative trial of Dictator Game*: Participants viewed the recipient's face (4 s); a scale appeared underneath the face so that participants could indicate how much to offer the recipient (4 s); followed by a jittered fixation cross (2–6 s, average 4 s). Neuroimaging analyses focused on the "late face period," or the latter half (2 s) of face viewing. (B) *Ratio of offers*: Participants who valued HAP over LAP made higher offers to excited *vs* calm recipients; whereas participants who valued LAP over HAP made higher offers to calm *vs* excited recipients. Analyses were run on continuous difference (Participant Ideal Affect[HAP-LAP]) scores and categorized offers, but the ratios of offers estimated at ± 2SD of Participant Ideal Affect[HAP-LAP] are depicted for ease of visualization. (C–D) Increased bilateral NAcc activity predicted greater subsequent giving (mean % signal change of NAcc preceding each offer amount is depicted for ease of visualization). (E–F) Decreased rTPJ activity predicted greater subsequent giving (mean % signal change of rTPJ preceding each amount of offer is depicted for ease of visualization). Warmer colors indicate positive associations; cooler colors indicate negative associations. Thresholded at $P < 0.005$ uncorrected, cluster ≥ 11 continuous voxels, $P < 0.05$ corrected. Error bars represent standard errors (S.E.). ***$P < 0.001$, **$P < 0.01$.

thickness, 2.9 mm; field of view = [232, 232]; acquisition matrix = [80, 80]; no gap between slices). After the dictator game, whole-brain T1-weighted structural scans were acquired (repetition time, 7.2 ms; echo time, 2.8 ms; flip angle, 12°; in-plane resolution and thickness, 0.9 mm; field of view = [255.55, 230.0]; acquisition matrix = [256, 256]) as participants rested.

Immediately after scanning, participants moved to a nearby testing room, viewed the same recipient avatars they had seen in the scanner, and rated recipients' trustworthiness, friendliness, intelligence, assertiveness, dominance, financial neediness and physical attractiveness along with a number of filler items. Again, other than trustworthiness, these traits did not mediate the effects of ideal affect match on giving, and therefore will not be discussed further. Afterwards, participants answered demographic questions, and an open-ended question about their understanding of the task. Based on content analyses of their responses, participants appeared to interpret the task as related to giving rather than keeping or withholding money, since 40.63% of participants mentioned "giving,"

"sharing" or "empathy" but 0% mentioned "keeping" or "withholding" money.

Finally, participants were debriefed and received their compensation. All materials, instructions and measurements were translated and back-translated into Korean by two independent translators. All procedures were approved by the Stanford Institutional Review Board.

Data analyses and results

Hypothesis 1: Does ideal affect match increase offers and perceived trustworthiness?

We predicted that the more participants valued HAP over LAP, the more they would trust and give to excited over calm recipients, as in Study 1. To test these hypotheses, we made two difference scores subtracting: (1) participants' ideal LAP from their ideal HAP (Participant Ideal Affect [HAP-LAP]), and (2) participants' actual LAP from their actual HAP (Participant Actual Affect [HAP-LAP]). We implemented a Participant Ideal Affect

(HAP-LAP) × Recipient Expression (excited, calm) cumulative mixed model on the offer amount, controlling for Participant Actual Affect (HAP-LAP) and treating participants as random effects. Overall, the more participants valued HAP over LAP, the less they offered to recipients, Estimate = −0.52, SE = 0.26, z = −2.02, P = 0.044. Moreover, participants offered more to excited than calm recipients overall, Estimate = 0.43, SE = 0.04, z = 10.71, P < 0.001.

As predicted, however, these effects were qualified by a significant Participant Ideal Affect (HAP-LAP) by Recipient Expression interaction, Estimate = 0.30, SE = 0.04, z = 7.08, P < 0.001. The more participants valued HAP over LAP (at +2SD of Participant Ideal Affect [HAP-LAP]), the more they offered to excited *vs* calm recipients, Estimate = 0.76, SE = 0.08, z = 9.85, P < 0.001 (Figure 3B). Conversely, the more participants valued LAP over HAP (at −2SD of Participant Ideal Affect [HAP-LAP]), the less they offered to excited *vs* calm recipients, Estimate = −0.19, SE = 0.07, z = −2.65, P = 0.008 (Figure 3B). As in Study 1, these effects did not vary by recipients' race or sex or participants' sex. We ran the same analyses substituting Participant Ideal Affect (HAP-LAP) with Participant Actual Affect (HAP-LAP), which revealed no significant main effects or interactions involving Participant Actual Affect (HAP-LAP).

Similarly, a Participant Ideal Affect (HAP-LAP) × Recipient Expression (excited, calm) linear mixed-effects model (using lme function in the R nlme package) on trustworthiness ratings, controlling for Participant Actual Affect (HAP-LAP), with participants treated as random effects, revealed participants rated excited *vs* calm recipients as more trustworthy overall, Effect = 0.26, SE = 0.02, t(3386) = 10.46, P < 0.001. This main effect was qualified, however, by the predicted significant Participant Ideal Affect (HAP-LAP) by Recipient Expression interaction, Effect = 0.16, SE = 0.03, t(3386) = 5.80, P < 0.001, bootstrapped (n = 10 000) 95% CI = [0.10, 0.21]. The more participants valued HAP over LAP (at +2SD of Participant Ideal Affect [HAP-LAP]), the more trustworthy they perceived excited *vs* calm recipients, Estimate = 0.43, SE = 0.05, t(3386) = 9.08, P < 0.001, bootstrapped (n = 10 000) 95% CI = [0.34, 0.52], whereas the more participants valued LAP over HAP (at −2SD of Participant Ideal Affect [HAP-LAP]), the less trustworthy they perceived excited *vs* calm recipients, Estimate = −0.06, SE = 0.05, t(3386) = −1.28, P = 0.200, bootstrapped (n = 10 000) 95% CI = [−0.15, 0.03], although this latter simple effect was not significant. These effects held across recipient race and sex. We ran the same analyses substituting Participant Ideal Affect (HAP-LAP) with Participant Actual Affect (HAP-LAP), but this yielded a different pattern of results (see Supplementary Materials Section S8).

In sum, consistent with findings from Study 1, the more participants valued HAP over LAP, the more they trusted and gave money to excited *vs* calm recipients.

fMRI data

Whole-brain analyses were conducted using Analysis of Functional Neural Images (AFNI; AFNI_16.2.06 version) software (Cox, 1996). The first six scans before the task were omitted to compensate for magnet stabilization. All other images were submitted to slice timing correction (using the first slice as reference), motion correction (using the third volume as a reference and Fourier interpolation), spatial smoothing (using a 3D isotropic Gaussian kernel of a 4-mm full width at half maximum), normalization to average percent signal change and high-pass filtering (omitting frequencies <0.01 Hz, as described in Wu *et al.*, 2014). Analyses focused on the second volume

acquisition, or the latter half (2 s) of the face viewing period (the "late face period," see Figure 3A), during which participants had seen facial stimuli and were making their decision, but were not yet able to view the scale or indicate their offer. This allowed us to control for specific motor responses, as well as capture the most psychologically relevant period of the choice task (as in Genevsky *et al.*, 2013).

We constructed a general linear model (GLM, ordinary least-squares regression) including five orthogonal regressors of interest. The first regressor weighted the second volume acquisition of each trial ('Late Face Period') when facial stimuli were presented but before participants indicated their offer. Four additional regressors of interest contrasted different aspects of the faces during this period: (1) Recipient Race (White = +1, Asian = −1), (2) Recipient Sex (male = +1, female = −1), (3) Recipient Expression (excited = +1, calm = −1) and (4) the interaction of recipient expression and recipient race. Eight regressors of no interest were also included: one sampling white matter activity, one sampling cerebrospinal fluid activity (to minimize the influence of physiological confounds, since activity in this region correlates with heart rate; Chang and Glover, 2009), and six modeling head movement. Before entry into the model, regressors of interest were convolved with a canonical gamma variate hemodynamic response kernel to fit the expected hemodynamic delay (Cohen, 1997). Linear regression *t*-statistic maps were converted to Z-scores, coregistered with structural maps, spatially normalized by warping to Montreal Neurological Institute space (linear to colin27T1_seg template), and resampled as 2.9 mm cubic voxels.

Regression analyses were conducted with AFNI program 3dRegAna to examine how neural responses to excited *vs* calm faces changed as a function of Participant Ideal Affect (HAP-LAP), controlling for Participant Actual Affect (HAP-LAP). This regression map was initially voxelwise thresholded (at P < 0.005) and then cluster thresholded (cluster size ≥11 continuous 2.9 mm³ voxels) to yield corrected maps for detecting whole-brain activity (P < 0.05 corrected, derived with 10 000 Monte Carlo iterations using AFNI program 3dClustSim implanted in AFNI_16.2.06 version).

In order to visualize neural activity associated with giving decisions during the late face period, we constructed another GLM including two orthogonal regressors of interest. The first regressor represented the 'Late Face Period' of each trial. The second regressor modeled the 'Offer', parametrically weighted according to the offer that subjects subsequently made ($0 = −2, $3 = −1, $6 = 0, $9 = +1, $12 = +2). Regressors of no interest and additional processes were the same as those described above. To compare brain activity before different offer amounts, we ran a one-sample *t*-test on the 'Offer' regressor coefficients. This one-sample *t*-test map was submitted to the same voxelwise and cluster thresholds.

Volume-of-interest (VOI) analyses were then conducted to confirm and clarify findings from the whole-brain analyses. Spherical VOIs (8 mm diameter) based on regions associated with anticipatory affect (Knutson and Greer, 2008) and charitable giving (Genevsky *et al.*, 2013) in the bilateral NAcc (Talairach coordinates: ±10, 12, −2), and regions associated with mentalizing in the right TPJ (MNI coordinates, 39, −63, 21; from Hutcherson *et al.*, 2015, Supplementary Table S4), were constructed. VOIs were defined based on findings from these independent studies in order to assure prediction (and thereby avoid "double-dipping"). Use of raw percent signal change data is necessary to make trial-to-trial predictions of choice, so percentage signal change during the late face period (2 s before the

offer prompt onset) was averaged within each VOI, averaged across bilateral VOIs, and then extracted as activity time courses. Sampling was lagged by 4 s to account for the hemodynamic lag to peak as in previous research (Knutson *et al.*, 2005; Knutson *et al.*, 2007).

NAcc and rTPJ timecourse data were submitted to a cumulative link mixed-effect model as independent variables, with participants as random effects, to examine whether brain activity could predict the amount of subsequent offers. We also conducted a Participant Ideal Affect (HAP-LAP) × Recipient Expression (excited, calm) linear mixed-effects model, controlling for Participant Actual Affect (HAP-LAP), with individual subjects as random effects, respectively, to test if ideal affect match modulated NAcc and rTPJ activity. A full linear mixed-effects model, including Participant Ideal Affect (HAP-LAP) × Recipient Expression (excited, calm) × Recipient Race (White, Asian) × Recipient Sex (male, female) on timecourse data yielded the same results.

Hypothesis 2: Does enhanced NAcc activity and/or diminished rTPJ activity predict greater offers?

As predicted, whole brain analyses revealed that participants showed enhanced NAcc activity (Figure 3C and D) but decreased rTPJ activity (Figure 3E and F) before making higher offers to recipients (Table 1). VOI analyses further revealed that activity in these areas significantly predicted offer on a trial-to-trial basis. Consistent with the whole brain results, VOI analyses specifically revealed that enhanced NAcc activity, Estimate = 0.26, SE = 0.12, z = 2.27, P = 0.023, and diminished rTPJ activity, Estimate = −0.26, SE = 0.11, z = −2.42, P = 0.016, during the late face period predicted higher subsequent offers to recipients. These effects were not modulated by the interaction between Participant Ideal Affect (HAP-LAP) and Recipient Expression, indicating that the effects of NAcc and rTPJ on subsequent offers held across participants as well as recipients. VOI analyses of other coordinates for the rTPJ (available upon request) as well as other areas associated with mentalizing such as the dorsal medial prefrontal cortex (DMPFC; Talairach coordinates: 0, 50, 20; from Van Overwalle and Baetens, 2009) yielded weaker but similar results (see Supplementary Materials Section S9).

Hypothesis 3: Does ideal affect match increase NAcc activity and decrease rTPJ activity?

Contrary to prediction, whole brain analyses (Table 2) revealed that NAcc activity during the late face period was *not* significantly modulated by ideal affect match (Figure 4A and B). NAcc activity during the early face period (i.e. the first 2 s of face viewing period), however, was significantly modulated by ideal affect match (see Supplementary Materials Section S10), consistent with our previous work indicating that ideal affect match can increase NAcc activity prior to face evaluation (Park *et al.*, 2016). Thus, ideal affect match influenced participants' positive response to the targets at the beginning but not the end of the face viewing period. VOI analyses similarly revealed no significant main effects or interactions involving ideal affect for NAcc activity during the late face period, all Ps > 0.210 (Figure 4A and B).

As predicted, whole brain analyses revealed that rTPJ activity during the late face viewing period was significantly modulated by ideal affect match (Table 2, Figure 4C). Specifically, when participants viewed recipients whose emotional expression matched their ideal affect, rTPJ activity decreased. VOI

Table 1. Whole-brain analyses of neural activity during anticipation of giving *vs* not giving

Region	x	y	z	Peak Z	Voxels
Face period					
L Cuneus	−1	−72	17	−8.18	13801
L Medial Frontal Gyrus	−4	1	52	7.10	1691
R Cingulate Gyrus	4	−25	40	−6.47	1342
R Fusiform Gyrus	39	−45	−18	6.72	379
R Insula	30	21	6	5.75	325
L Inferior Frontal Gyrus	−45	4	32	4.88	197
R Precuneus	28	−57	40	5.46	163
R Superior Frontal Gyrus	22	15	46	−5.09	153
L Insula	−28	24	9	5.63	93
R Middle Temporal Gyrus	42	7	−32	−4.72	92
L Lingual Gyrus	−22	−86	0	4.42	84
L Fusiform Gyrus	−36	−45	−15	5.53	73
L Middle Frontal Gyrus	−22	15	55	−3.39	52
L Thalamus	−4	−28	−3	4.47	49
R Paraphippocampal Gyrus/R Amygdala	19	−8	−12	6.02	45
L Parahippocampal Gyrus	−19	−40	9	4.43	35
L Middle Frontal Gyrus	−36	4	46	−4.40	32
L Medial Frontal Gyrus	−13	15	−18	−4.43	29
R Anterior Cingulate	4	4	−6	4.08	29
L Parahippocampal Gyrus	−30	−8	−26	4.80	28
R Parahippocampal Gyrus	30	−5	−26	4.18	23
L Superior Frontal Gyrus	−30	56	−6	−4.25	19
L Superior Frontal Gyrus	−19	62	9	−3.63	18
R Caudate	19	−34	17	4.43	18
L Middle Frontal Gyrus	−36	39	−9	−3.83	15
L Cingulate Gyrus	−22	7	26	−3.37	14
R Superior Temporal Gyrus	33	7	−18	4.25	11
Offer					
L Lingual Gyrus	−10	−74	−3	−3.79	65
L Superior Frontal Gyrus	−19	21	49	3.90	39
R Posterior Cingulate	16	−60	14	−4.01	23
L Caudate/L NAcc	−10	10	−3	3.84	21
R Supramarginal Gyrus/RTPJ	62	−43	26	−3.75	20
L Middle Occipital Gyrus	−45	−72	9	−3.51	18
R Precentral Gyrus	59	−2	11	−3.74	18
R Caudate/R NAcc	10	10	−3	3.56	17
R Superior Temporal Gyrus	51	−45	17	−3.35	14
L Lingual Gyrus	−10	−60	3	−3.28	13
R Middle Temporal Gyrus	59	−22	−9	−3.30	12
L Posterior Cingulate	−13	−66	17	−3.39	12

n = 36; voxel-wise P < 0.005 uncorrected, cluster corrected P < 0.05, minimum cluster size = 112.9 × 2.9 × 2.9 mm continuous voxels; *x* = right, *y* = anterior, *z* = superior Talairach coordinates, bold indicates activation of predicted volumes of interest.

analyses also revealed a significant interaction of Participant Ideal Affect (HAP-LAP) and Recipient Expression (excited, calm), Estimate = −0.01, SE = 0.01, *t*(3398) = −2.13, P = 0.033, bootstrapped (*n* = 10 000) 95% CI = [−0.03, −0.001] on rTPJ activity. The more participants valued HAP over LAP (at +2SD of Participant Ideal Affect [HAP-LAP]), the *less* rTPJ activity occurred in response to excited versus calm recipients, although this result was marginally significant, Estimate = −0.02, SE = 0.01, *t*(3398) = −1.80, P = 0.072, bootstrapped (*n* = 10 000) 95% CI = [−0.04, 0.00]. The more participants valued LAP over HAP (at −2SD of Participant Ideal Affect [HAP-LAP]), the *more* rTPJ activity occurred in response to excited versus calm recipients, Estimate = 0.02, SE = 0.01, *t*(3398) = 2.02, P = 0.044, bootstrapped (*n* = 10 000) 95% CI = [0.0004, 0.05] (Figure 4D).
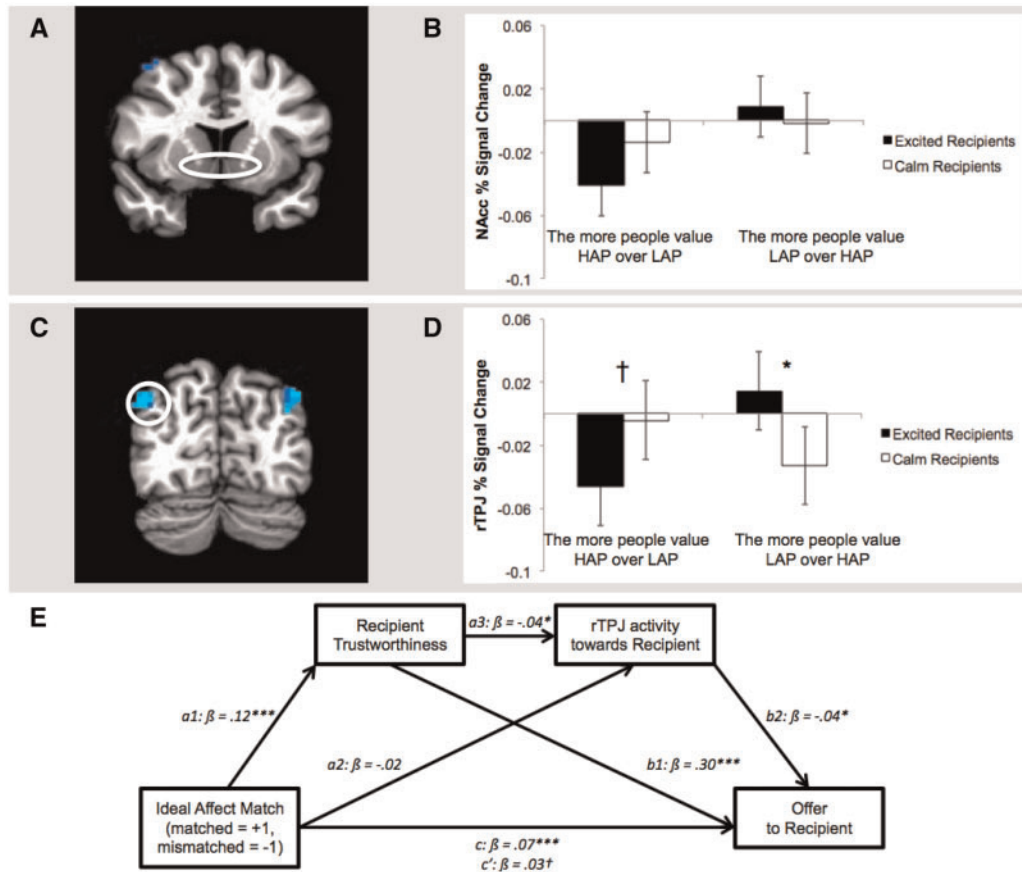
Fig. 4. Neural activity mediates influence of ideal affect match on giving (Study 2). (A–B) Ideal affect did not influence NAcc activity, (C–D) but did influence rTPJ activity during viewing of excited *vs* calm recipients (late face viewing period). Participants who valued HAP over LAP showed lower rTPJ activity in response to excited *vs* calm recipients; however, participants who valued LAP over HAP showed higher rTPJ activity in response to excited *vs* calm recipients [analyses were run on continuous difference (Participant Ideal Affect [HAP–LAP]) scores, but % signal change of rTPJ estimated at ± 2SD of Participant Ideal Affect (HAP-LAP) is depicted for ease of visualization]. Warmer colors indicate positive associations; cooler colors indicate negative associations. Thresholded at $P < 0.005$ uncorrected, cluster ≥ 11 continuous voxels, $P < 0.05$ corrected. (E) *Serial mediation:* A bootstrapped ($n = 10\,000$) serial mediation analysis indicated that participants rated recipients whose emotional expression matched their ideal affect as more trustworthy, which predicted decreased rTPJ activity in response to those recipients, which predicted greater offers to recipients. Ideal affect match refers to the match between participants' ideal affect (based on median split) and recipients' expression (excited, calm). We present standardized β values with unstandardized ideal affect match index (Match = +1, Mismatch = −1) for ease of interpretation. Error bars represent standard errors (S.E.). ***$P < 0.001$, *$P < 0.05$, †$P < 0.10$.

Table 2. Whole brain analyses revealing interactions between participant ideal affect and recipient expression

Region	x	y	z	Peak Z	Voxels
Recipient expression (excited *vs* calm)					
R Middle Occipital Gyrus	28	−89	0	6.58	323
L Lingual Gyrus	−19	−92	−6	7.81	272
R Superior Temporal Gyrus	45	−34	6	5.12	20
L Middle Frontal Gyrus	−22	1	55	−3.80	13
R Insula	45	−14	6	3.89	12
Participant Ideal Affect (HAP−LAP) X Recipient Expression(Excited *vs* Calm)					
L Cingulate Gyrus	−4	−43	40	−5.30	65
L Precuneus/L TPJ	−42	−72	38	−4.29	41
R Middle Frontal Gyrus	33	13	55	−4.90	41
R Precuneus/R TPJ	42	−72	35	−5.08	28
R Superior Frontal Gyrus	16	53	14	−3.93	12
L Superior Frontal Gyrus	−16	56	17	−3.41	11

$n = 36$; voxel = wise $P < 0.005$ uncorrected, cluster corrected $P < 0.05$, minimum cluster size = 11 2.9 × 2.9 × 2.9 mm continuous voxels; x = right, y = anterior, z = superior Talairach coordinates, bold indicates activation of predicted volumes of interest.

Analyses conducted on the four levels of expression yielded the same results (Supplementary Materials Section S7). Similar analyses substituting Participant Ideal Affect (HAP-LAP) with Participant Actual Affect (HAP-LAP), however, revealed no significant main effects or interactions involving actual affect for either NAcc or rTPJ activity.

In sum, ideal affect match altered rTPJ but not NAcc activity during the late face viewing period.

Hypothesis 4: Does brain activity mediate the influence of ideal affect match on giving?

To test mediation, we divided the participants into two groups based on their ideal affect. We computed the median value of Participant Ideal Affect (HAP-LAP), Median = −0.50, and coded participants with scores above the median as the Value HAP group ($n = 16$), and coded participants with scores below the median as the Value LAP group ($n = 20$). To account for the interaction between participants' ideal affect and recipients' expressions, we multiplied the group codes (Value HAP = +1, Value LAP = −1) by Recipient Expression (Excited = +1, Calm = −1) to calculate an "ideal affect match index," which represented whether participants viewed recipients that matched their ideal affect (i.e. participants in Value HAP [*vs* LAP] group viewed an excited [*vs* calm] recipient) or did not match their ideal affect (i.e. participants in Value HAP [*vs* LAP] group viewed a calm [*vs* excited] recipient). We also mean-deviated trustworthiness ratings to account for individual differences in response style. We then conducted a bootstrapped serial mediation analysis ($n = 10\,000$) on trial-to-trial data to examine whether ideal affect match increased participants' trustworthiness ratings, which in turn reduced rTPJ activity, resulting in greater offers (Figure 4E).

As predicted, and consistent with the findings of Study 1, ideal affect match predicted trustworthiness ratings, $B = 0.15$, SE = 0.02, ß = 0.12, $t = 7.10$, $P < 0.001$, 95% CI = [0.11, 0.19], indicating that participants rated faces that matched their ideal affect as more trustworthy (path a1). In turn, controlling for ideal affect match, trustworthiness ratings predicted decreased rTPJ activity, $B = −0.01$, SE = 0.004, ß = −0.04, $t = −2.05$, $P = 0.041$, 95% CI = [−0.02, −0.0004], indicating that the more trustworthy that participants rated a particular recipient, the less rTPJ activity occurred in response to viewing that recipient (path a3). Moreover, decreased rTPJ activity predicted greater subsequent offers, $B = −0.03$, SE = 0.01, ß = −0.04, $t = −2.26$, $P = 0.024$, 95% CI = [−0.06, −0.004], controlling for ideal affect match and trustworthiness ratings, supporting our hypothesis that decreased rTPJ activity could promote increased offers (path b2). The significant effect of Ideal Affect Match on offer to excited *vs* calm recipients (path c), $B = 0.02$, SE = 0.005, ß = 0.07, $t = 4.01$, $P < 0.001$, 95% CI = [0.01, 0.03], was reduced after entering perceived trustworthiness and rTPJ activity in the model (path c'), $B = 0.01$, SE = 0.004, ß = 0.03, $t = 1.94$, $P = 0.053$, 95% CI = [−0.0001, 0.02]. The indirect effect of Ideal Affect Match -> Trustworthiness -> rTPJ activity -> Offer was marginally significant, Standardized Effect = 0.0002, SE = 0.0001, 95% CI = [0.0000, 0.0005], perhaps due to reduced variance associated with converting the continuous ideal affect measure into a dichotomous one (results of other indirect effects are available upon request). None of the other traits were correlated with rTPJ activity (Supplementary Materials, Section S11).

Study 2: Discussion

In summary, neuroimaging results suggested that the more participants found recipients' faces rewarding (consistent with increased NAcc activity), and the more they appeared to share their thoughts or feelings (consistent with decreased rTPJ activity), the more they gave to potential recipients. Participants also showed reduced rTPJ activity in response to faces whose expressions matched their ideal affect. Moreover, participants gave more to recipients whose expressions matched their ideal affect because they trusted them more and showed reduced rTPJ activity.

Contrary to hypotheses, NAcc activity during the late face period (right before the decision to give), was *not* modulated by ideal affect match, and did not mediate the influence of ideal affect match on participants' choices to give. Thus, although ideal affect match initially evoked rapid responses associated with reward, our data suggest that the influence of ideal affect match did not critically rely on those feelings in this study.

General discussion

Why do people give to strangers? In two studies, we found that individuals were more likely to trust and give to recipients whose emotional expressions reflected their ideal affect (or showed "ideal affect match"). This held for both European Americans and Koreans, despite the fact that they differed in valued affective states. Consistent with cultural variation in ideal affect, European Americans trusted and gave more to excited *vs* calm recipients, whereas Koreans trusted and gave more to calm *vs* excited recipients. Moreover, only recipients' emotional expression, but not their race or sex, influenced participants' choices to give. Neuroimaging data further revealed that ideal affect match decreased activity in the rTPJ, a circuit associated with perceiving that others hold different thoughts and feelings. Decreased rTPJ activity was associated with increased attributions of trustworthiness, and ultimately, enhanced giving. Thus, people may have trusted and given more to strangers whose emotional expressions matched their ideal affect because they inferred shared thoughts and feelings.

Consistent with the notion that experienced as well as ideal positive affect can increase giving, NAcc activity during the late face viewing period also predicted increased giving overall (Genevsky *et al.*, 2013; Genevsky and Knutson, 2015). These findings therefore support the notion that positive arousal can promote choices to give on a trial-to-trial basis. Ideal affect match, however, did not influence NAcc activity immediately prior to choice, suggesting that ideal affect match did not influence giving through this channel.

Decreased rTPJ activity also predicted increased giving. A growing literature has implicated rTPJ activity in perceiving that others have different thoughts or feelings, or "mentalizing" (Saxe and Wexler, 2005; Zaki and Ochsner, 2012). Although increased rTPJ activity has also been associated with enhanced prosocial behavior, including giving to others (Hare *et al.*, 2010; Telzer *et al.*, 2011; Hutcherson *et al.*, 2015; Strombach *et al.*, 2015), the current findings suggest that in some instances, increased rTPJ activity may instead decrease prosocial behavior. Consistent with this possibility, a recent study found that increased activity in mentalizing circuits boosted negative judgments of in-group members (Hughes *et al.*, 2016). Similarly, in certain contexts of resource allocation, increased rTPJ activity might hinder or delay a pre-potent tendency to give (Rand *et al.*, 2014). Future studies might explore whether framing choices in terms of taking versus giving alters the relationship between rTPJ activity and resource allocation to the self *vs* others.

Future research should also examine exactly how perceptions of shared beliefs and feelings increase giving. One possibility is that signals of shared beliefs and feelings are easier to

process. Supporting this notion, participants made faster choices to give (i.e. shorter reaction times) when presented with recipients whose expressions matched their ideal affect (Supplementary Materials Section S12).

These findings contribute several novel insights. First, they clarify that affect can play multiple roles in choices to give. While previous research suggests that actual affect can promote giving (Genevsky *et al.*, 2013; Genevsky and Knutson, 2015), the current findings further demonstrate that ideal affect match can also distinctly promote giving. But in the case of ideal affect match, the emotional values of the giver and the emotional expression of the recipient must jointly interact to increase giving. Indeed, in the current studies, individual differences in actual affect (and its interaction with the emotional expression of the recipient) did *not* predict giving.

Second, the findings raise the intriguing possibility that ideal affect match may provide a strong signal of in-group identity, above and beyond more commonly studied but less malleable factors including race and sex. Indeed, previous research suggests that individuals are better able to recognize the emotions of ingroup members (Elfenbein and Ambady, 2002), and ideal affect may provide one specific source of this "ingroup bias." Individuals clearly offered more money to others whose emotional expression matched their ideal affect—regardless of race or sex. These results extend previous work in which people give more to others who explicitly share religious values (Saroglou, 2006; Preston and Ritter, 2013) to others who implicitly share affective values. Further, the link between ideal affect match and giving was mediated by increased perceptions of trustworthiness (an index of ingroup membership), but not other traits (e.g. friendliness, intelligence, or dominance). Third, the current findings suggest that emotional signals of trust may have been difficult to identify in previous work in part because they vary as a function of people's ideal affect. Finally, because cultures differ in what kinds of affect they value, these findings suggest a mechanism through which culture may shape giving in the guise of ideal affect match. This study thus contributes to an increasing trend in which scientists use neuroscientific tools to explore cultural processes (Kitayama and Uskul, 2011; Han *et al.*, 2013).

Limitations and future directions

Despite their novel contributions, these studies had some limitations that raise questions for future research. First, while the task's features mimic those used by many online charities that solicit donations online, the minimal nature of the presentation format may have emphasized characteristics related to the manipulated dimensions of the facial stimuli used. Nonetheless, despite systematic variation not only in emotional expressions but also sexual and racial facial characteristics, the interaction effect of participants' ideal affect by target emotional expression was not modulated by these other factors. Further, similar behavioral patterns were observed with more naturalistic facial stimuli, although future research might incorporate stimuli actually found in charitable appeals online (e.g. Genevsky and Knutson, 2015). Second, the current task focused on unilateral transactions (e.g. the Dictator Game), while other giving scenarios include opportunities for reciprocity. Thus, future studies might profitably examine whether ideal affect match predicts giving in reciprocal exchanges, both cooperative and competitive (e.g. the Trust Game; Bó, 2005). Third, emerging work suggests that deliberation may reduce giving under certain circumstances (Rand *et al.*, 2014). The current findings

conversely suggest that ideal affect match may lower deliberative hurdles that delay and can ultimately diminish giving. Future studies may more directly examine the generality and limits of this account. Fourth, as mentioned above, while the task was framed as and interpreted to be a game about giving, other frames related to "taking" or "keeping" might be interesting to examine, even controlling for incentive structure (e.g. Knutson *et al.*, 2008). Fifth and finally, while ideal affect match may promote giving (and other prosocial behaviors), ideal affect mismatches may conversely diminish giving toward others. Future work could explore how to minimize potential problems related to ideal affect mismatches in the classroom, corporation, and clinic.

Despite these limitations, our findings suggest that at both cultural and individual levels of analysis, people trust and give more to strangers whose emotional expressions match their ideal affect, in part because those emotional expressions signal shared values.

References

Barrett, L.F., Russell, J.A. (1999). The structure of current affect: controversies and emerging consensus. *Current Directions in Psychological Science*, **8**(1), 10–4.

Bó, P.D. (2005). Cooperation under the Shadow of the Future: experimental evidence from in nitely repeated games. *American Economic Review*, **95**(5), 1591–604.

Buchan, N.R., Johnson, E.J., Croson, R.T.A. (2006). Let's get personal: an international examination of the influence of communication, culture and social distance on other regarding preferences. *Journal of Economic Behavior and Organization*, **60**(3), 373–98.

Chang, C., Glover, G.H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *NeuroImage*, **47**(4), 1448–59.

Cloutier, J., Gabrieli, J.D.E., O'Young, D., Ambady, N. (2011). An fMRI study of violations of social expectations: when people are not who we expect them to be. *NeuroImage*, **57**(2), 583–8.

Cohen, M.S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, **6**(2), 93–103.

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, **29**(3), 162–73.

Elfenbein, H.A., Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, **128**(2), 203–35.

Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, **14**(4), 583–610.

Fehrler, S., Przepiorka, W. (2013). Charitable giving as a signal of trustworthiness: disentangling the signaling benefits of altruistic acts. *Evolution and Human Behavior*, **34**(2), 139–45.

Ferraro, P.J., Cummings, R.G. (2007). Cultural diversity, discrimination, and economic outcomes: an experimental analysis. *Economic Inquiry*, **45**(2), 217–32.

Fong, C.M., Luttmer, E.F.P. (2009). What determines giving to Hurricane Katrina victims? Experimental evidence on income, race, and fairness. *American Economic Journal: Applied Economics*, **1**(2), 64–87.

Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, **6**(3), 347–69.

Franzen, A., Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior and Organization*, **81**(1), 74–81.

Genevsky, A., Knutson, B. (2015). Neural affective mechanisms predict market-level microlending. *Psychological Science*, **26**(9), 1411–22.

Genevsky, A., Västfjäll, D., Slovic, P., Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: affect shifts preferences for giving. *Journal of Neuroscience*, **33**(43), 17188–96.

Han, S., Northoff, G., Vogeley, K., Wexler, B.E., Kitayama, S., Varnum, M.E.W. (2013). A cultural neuroscience approach to the biosocial nature of the human brain. *Annual Review of Psychology*, **64**, 335–59.

Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, **316**(5831), 1622–5.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, **30**(2), 583–90.

Hayes, A.F. (2012). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. The Guildford Press.

Henrich, J., Boyd, R., Bowles, S., *et al.* (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, **28**(6), 795–815.

Huang, C., Park, D. (2013). Cultural influences on Facebook photographs. *International Journal of Psychology*, **48**(3), 334–43.

Hughes, B.L., Zaki, J., Ambady, N. (2017). Trusting outgroup, but not ingroup members, requires control: neural and behavioral evidence. *Social Cognitive Affective Neuroscience*, **12**(3), 372–81.

Hutcherson, C.A., Bushong, B., Rangel, A. (2015). A neurocomputational model of altruistic choice and its implications. *Neuron*, **87**(2), 451–62.

Kahneman, D., Knetsch, J.L., Thaler, R. (1986). Fairness as a constraint on profit seeking: entitlements in the market. *The American Economic Review*, **76**(4), 728–41.

Kitayama, S., Uskul, A.K. (2011). Culture, mind, and the brain: current evidence and future directions. *Annual Review of Psychology*, **62**, 419–49.

Knutson, B., Greer, S.M. (2008). Anticipatory affect: neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society of London B*, **363**(1511), 3771–86.

Knutson, B., Rick, S., Wimmer, G.E., Prelec, D., Loewenstein, G. (2007). Neural predictors of purchases. *Neuron*, **53**(1), 147–56.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience*, **25**(19), 4806–12.

Knutson, B., Wimmer, G.E., Rick, S., Hollon, N.G., Prelec, D., Loewenstein, G. (2008). Neural antecedents of the endowment effect. *Neuron*, **58**(5), 814–22.

Koster-Hale, J., Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron*, **79**(5), 836–48.

Krumhuber, E., Manstead, A.S.R., Cosker, D., Marshall, D., Rosin, P.L., Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, **7**(4), 730–5.

Larsen, R.J., Diener, E. (1991). *Promises and Problems with the Circumplex Model of Emotion*. (Clark M. S., Ed.). Newbury Park, CA: Sage.

Mende-Siedlecki, P., Cai, Y., Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, **8**(6), 623–31.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, **103**(42), 15623–8.

Park, B., Tsai, J.L., Chim, L., Blevins, E., Knutson, B. (2016). Neural evidence for cultural differences in the valuation of positive facial expressions. *Social Cognitive and Affective Neuroscience*, **11**(2), 243–52.

Preston, J.L., Ritter, R.S. (2013). Different effects of religion and God on prosociality with the ingroup and outgroup. *Personality & Social Psychology Bulletin*, **39**(11), 1471–83.

Rand, D.G., Peysakhovich, A., Kraft-Todd, G.T., *et al.* (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, **5**(3677), 1–12.

Saroglou, V. (2006). Religion's role in prosocial behavior: myth or reality? *Psychology of Religion Newsletter*, **31**(2), 1–8.

Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, **43**(10), 1391–9.

Scharlemann, J.P., Eckel, C.C., Kacelnik, A., Wilson, R.K. (2001). The value of a smile: game theory with a human face. *Journal of Economic Psychology*, **22**(5), 617–40.

Sims, T., Tsai, J.L., Koopmann-Holm, B., Thomas, E.A.C., Goldstein, M.K. (2014). Choosing a physician depends on how you want to feel: the role of ideal affect in health-related decision making. *Emotion*, **14**(1), 187–92.

Strombach, T., Weber, B., Hangebrauk, Z., *et al.* (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences*, **112**(5), 1619–24.

Suzuki, N., Konno, Y., Yamagishi, T. (2007). In-group bias in trusting behavior: a choice of allocator experiment with minimal groups. *Japanese Journal of Psychology*, **78**(1), 17–24. (In Japanese with an English abstract).

Telzer, E.H., Masten, C.L., Berkman, E.T., Lieberman, M.D., Fuligni, A.J. (2011). Neural regions associated with self control and mentalizing are recruited during prosocial behaviors towards the family. *NeuroImage*, **58**(1), 242–9.

Tidd, K.L., Lockard, J.S. (1978). Monetary significance of the affiliative smile: a case for reciprocal altruism. *Bulletin of the Psychonomic Society*, **11**(6), 344–6.

Tortosa, M.I., Strizhko, T., Capizzi, M., Ruz, M. (2013). Interpersonal effects of emotion in a multi-round Trust Game. *Psicologica*, **34**(2), 179–98.

Tsai, J.L. (2007). Ideal affect: cultural causes and behavioral consequences. *Perspectives on Psychological Science*, **2**(3), 242–59.

Tsai, J.L., Knutson, B., Fung, H.H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology*, **90**(2), 288–307.

Tsai, J.L., Louie, J.Y., Chen, E.E., Uchida, Y. (2007). Learning what feelings to desire: socialization of ideal affect through children's storybooks. *Personality and Social Psychology Bulletin*, **33**(1), 17–30.

Tsai, J.L., Miao, F.F., Seppala, E. (2007). Good feelings in christianity and buddhism: religious differences in ideal affect. *Personality and Social Psychology Bulletin*, **33**(3), 409–21.

Tsai, J.L., Miao, F.F., Seppala, E., Fung, H.H., Yeung, D.Y. (2007). Influence and adjustment goals: Sources of cultural differences in ideal affect. *Journal of Personality and Social Psychology*, **92**(6), 1102–17.

Tsai, J.L., Ang, J.Y.Z., Blevins, E., *et al.* (2016). Leaders' smiles reflect cultural differences in ideal affect. *Emotion*, **16**(2), 183–95.

Van't Wout, M., Sanfey, A.G. (2008). Friend or foe: the effect of implicit trustworthiness judgments in social decision-making. *Cognition*, **108**(3), 796–803.

Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, **48**(3), 564–84.

Watson, D., Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, **98**(2), 219–35.

Wu, C.C., Samanez-Larkin, G.R., Katovich, K., Knutson, B. (2014). Affective traits link to reliable neural markers of incentive anticipation. *NeuroImage*, **84**, 279–89.

Zaki, J., Ochsner, K.N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience*, **15**(5), 675–80.