

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Guilt in Social Interaction: How Partner Perception Shapes Emotional Responses

Permalink

<https://escholarship.org/uc/item/5sj15084>

ISBN

9798186387596

Author

Zhao, Shangcheng

Publication Date

2026-04-02

Supplemental Material

<https://escholarship.org/uc/item/5sj15084#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

Guilt in Social Interaction: How Partner Perception Shapes Emotional Responses

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of the Arts in Psychology and Brain Science

by

Shangcheng Zhao

Committee in charge:

Professor Hongbo Yu, Chair

Professor David Pietraszewski

Professor Nils Karl Reimer

June 2026

The thesis of Shangcheng Zhao is approved.

David Pietraszewski

Nils Karl Reimer

Hongbo Yu, Committee Chair

March 2026

ABSTRACT

Guilt in Social Interaction: How Partner Perception Shapes Emotional Responses

by

Shangcheng Zhao

This study examines how partner perceptions, more specifically, expectation of partner's emotion influence guilt-related emotional responses in interpersonal contexts. Using interactive transgression tasks across two studies (N=121 and N=185), we demonstrate that individuals' emotional responses closely track their expectations of partners' emotions, with this pattern persisting across American and Chinese samples. Inter-subject representational similarity analysis revealed associations between general social trait judgments style and context-specific emotion expectations. Cultural differences emerged in avoidance behaviors, with American participants showing increased avoidance when anticipating partner anger, while Chinese participants maintained low avoidance regardless of expected anger. These findings highlight the interpersonal nature of self-conscious emotions and the importance of studying emotions within authentic social contexts.

1 Introduction

As a self-conscious emotion that involves complex self-evaluation and cognitive judgment, guilt plays a significant role in social interactions and exerts broad societal influence on motivation, decision-making, and interpersonal relationships (Baumeister et al., 1994; Leith and Baumeister, 1998; Tangney and Fischer, 1995). A comprehensive understanding of guilt requires examining both endogenous higher-order cognitive processes—including self-reflection and social comparison (Tracy et al., 2013)—and exogenous climate such as social norms (De Hooze et al., 2007) and interpersonal dynamics (Yu et al., 2014).

However, previous research on guilt has largely relied on retrospective methods, where participants reflect on past experiences to report their emotional responses in an isolated, non-interactive context (Tangney et al., 1992). However, these retrospective approaches miss essential elements of self-conscious emotions, particularly their dynamic and interpersonal nature (Baumeister et al., 1994). Recent findings offer compelling evidence for the interpersonal nature of self-conscious emotion, particularly from an evolutionary perspective. Sznycer (2019) argues that self-conscious emotions have evolved to address adaptive challenges related to social valuation. They function as signals that promote actions yielding positive evaluations while also warning against potential devaluation and relational crises. For instance, pride reinforces behaviors that enhance others' perceptions of us, shame helps limit the information that could lead to devaluation, and guilt signals situations where we insufficiently consider a valuable social partner's welfare. Although self-conscious emotions can emerge in isolated, non-interactive contexts, fundamentally, they are still produced from our evolutionary history as social beings. Even when an individual feels guilty after an unnoticed misbehavior, such as breaking a vase, that emotion is still modulated by the broader social climate, including cultural factors (Su and Hynie, 2019; Sznycer et al., 2018; Wong and Tsai, 2007).

Since the key drive of self-conscious emotion is interpersonal, it is important to put them back into interpersonal contexts (Yu et al., 2024). One key finding in interpersonal research is that social partners play a critical role in emotion generation (Finn et al., 2020). In a dyadic

interaction, holding one person's behavior constant while varying the partner's response reveals that the partner's reaction substantially shapes the emotion experienced by the first person. Gao et al. (2021) demonstrated this hypothesis using an interactive transgression task, where they assessed the emotion reported by the transgressor in response to victims' forgiving or blaming attitudes. They found that unexpected forgiveness intensified guilt, whereas unexpected blame reduced guilt and elicited anger. Thus, even when the transgressor's behavior remains constant, the recipient's emotional reaction can substantially influence the transgressor's emotional experience.

Moreover, partner's emotional reaction might also play a role on the motivational consequences following transgression. Previous research has identified two opposing motivational directions—approach and avoidance—that can emerge following blameworthy behavior. However, although theories suggest that guilt motivates approach behaviors and active relationship repair, empirical support for these distinctions remains mixed. A review of early evidence reveals that only a handful of studies have directly tested this motivational distinction, and the results are often ambiguous (Schmader and Lickel, 2006; Tangney and Fischer, 1995; Wicker et al., 1983; Tangney and Fischer, 1995; Frijda et al., 1989; Roseman et al., 1994; Scherer and Wallbott, 1994). In fact, certain studies even indicate an opposite pattern; for example, individuals with high guilt proneness have shown a greater tendency to avoid interdependent partnerships in a cooperation game (Wiltermuth and Cohen, 2014), and one study employing interactive paradigms and physiological measures found that individuals avoided eye contact when feeling guilty, with forced eye contact linked to heightened skin conductance, emotional arousal, and psychological discomfort (Yu et al., 2017). Here, we sought to further clarify the link between guilt and motivational tendencies by not only creating an interpersonal context and assessing both tendencies directly, but also taking the partner's emotions into account. We aimed to test whether the partner-transgressor dynamic would influence the transgressor's behavioral choices following wrongdoing.

Given the importance of the role a partner plays, another crucial factor to investigate is

how social perception—the way individuals interpret, analyze, and remember information about others—shapes their own emotions. Social trait judgments, which represent consistent patterns of perceived intentions and capabilities (Fiske et al., 2007), are pervasive in everyday social life, as humans spontaneously form impressions of others upon minimal exposure (Baron et al., 2011; Willis and Todorov, 2006). For example, individuals can even judge whether someone is trustworthy or intelligent based solely on facial appearance (Lin et al., 2021; Oosterhof and Todorov, 2008; van 't Wout and Sanfey, 2008). More importantly, these judgments influence consequential decisions in the real world (Ahler et al., 2017; Blair et al., 2004; Lenz and Lawson, 2011; Lin et al., 2017, 2018; Martin, 1978; Todorov et al., 2005). Given this background, it is essential to explore how social trait perceptions impact emotion generation in interpersonal contexts.

The current study intends to extend previous findings on several aspects: using interpersonal paradigm, we measure guilt not only in the real time but also naturally take social partner's contribution into account. Moreover, we will evaluate the influence of social trait judgments on emotion generation. We hypothesis that: (a) individuals' expectation of their partner's reaction will influence their own emotional experiences. (b) individuals' social trait judgment style indicates their stationary emotion reaction. (c) guilt elicit approach and/or avoidance tendencies, while expectations of partners' emotions would differentially modulate the motivational directions.

In addition, drawing on the adaptationist view of self-conscious emotion (Sznycer, 2019, 2022), we hypothesized that these associations would generalize across cultures. To test this hypothesis, in Study 2, we implemented our model in two countries that differ markedly in individualism and collectivism (China and the United States) (Hofstede, 1978).

2 Study 1

2.1 Method

2.1.1 Participants

We recruited 171 participants from the Prolific platform. We only included participants who had English fluency, normal or corrected-to-normal vision, an education level above high school, and a Prolific approval rate greater than 95%. Fifty participants were excluded, leading to a sample of 121 participants (46 females; age = 26.75 ± 7.46) for further analysis. Among the excluded participants, seven did not complete the interpersonal transgression task (see below), thirty-eight did not pass the comprehension check of the interpersonal transgression task, five missed at least an entire condition of the interpersonal transgression task. None of the participants reported a diagnosis of psychopathological disease. All participants provided written informed consent using procedures approved by the Institutional Review Boards of West Virginia University (Protocol #2012188080) and University of California Santa Barbara (Protocol #2-22-0520).

2.1.2 Interpersonal transgression task

In this online task, participants were paired with another hypothetical participant (hereafter, “partner”). On each trial (Figure 1(a)), the participant and the partner saw an array of dots (about 20) displayed on the screen for a short interval (1.5 s). The participant and, ostensibly, the partner indicated whether the number of the dots was larger or smaller than a reference number (e.g., 20). Afterward, their performance was displayed on the screen. If one or both of them responded incorrectly, the partner had to watch an aversive image (i.e., unpleasant outcome), which was selected from the International Affective Picture System (IAPS) (Lang et al., 1997). By manipulating the participant’s and the partner’s performance (i.e., correct/incorrect feedback), we were able to manipulate the participant’s responsibility in causing unpleasant outcomes to the partner. To make sure that the number of trials was balanced across condi-

tions (i.e., both-correct, partner-incorrect, self-incorrect, both-incorrect), unbeknownst to the participants, the outcome feedback was predetermined. There were 12 for each of the possible conditions. The order of the conditions was randomized across participants. Previous studies have demonstrated the validity of this task in inducing different levels of perceived guilt, negative self-conscious emotions, and compensatory behaviors (Gao et al., 2018; Li et al., 2020; Yu et al., 2014, 2017). At the end of trials in which the partner viewed an aversive image, participants answered one of the following questions: (1) How guilty did you feel about the partner’s unpleasant outcome? (2) How shameful did you feel? (3) How long would you be willing to view the image yourself to reduce your partner’s exposure (i.e., a measure of compensation)? (4) To what extent do you think your partner feels angry? (5) . . . disappointed? (6) . . . resentful? Each question was randomly presented twice in each condition (excluding the both-correct condition), and participants indicated their responses on analog scales.

2.1.3 Social trait judgment task

After an interval, participants could choose to proceed to a face evaluation task in which they evaluate facial images across multiple dimensions. To avoid contamination between social trait judgments and guilt assessments, we use a completely independent set of stimuli that are unrelated to the prior social interaction. This design allows us to link the general pattern of individuals’ social trait judgments to their intrinsic emotional responses when interact unfamiliar others.

We used photos of celebrities from the CelebA dataset. A total of 500 photos of 50 identities were selected, resulting in 10 photos for each identity. The 50 identities included both sexes (33 male) and multiple races (40 identities were Caucasian, 9 identities were African American, and 1 identity was biracial). All photos were taken in naturalistic contexts, with faces displaying various expressions, head poses, and gaze directions, surrounded by various backgrounds and lighting. Participants provided judgments of social traits on a 7-point Likert scale. The social traits include warm, critical, competent, practical, feminine, strong, youthful,

and charismatic. We selected these eight social traits to represent the comprehensive dimensions of trait judgments from faces, which was found in prior research (Lin et al., 2021). The 500 images were divided into 10 modules, with each module containing one face image for each of the 50 identities. Participants rated all the faces on all the eight social traits (rated in blocks) in each module.

During data cleaning, we applied the following three exclusion criteria: (1) Trial-wise exclusion: we excluded trials with reaction times shorter than 100 ms or longer than 5000 ms. (2) Block/trait-wise exclusion: we excluded the entire block per module if more than 30% of the trials were excluded from the block per (1) above, or if there were fewer than 3 different rating values in the block (this suggests that the participant may not have used the rating scale properly). (3) Module-wise exclusion: we excluded a module if more than 3 blocks were excluded from the module per (2) above.

2.1.4 Analytical plan

To explore the association between participants' emotions, partners' expected emotion, and social trait judgment, we used methods from two folds. First, we tested the association on a shared level by calculating Pearson correlation. Next, we moved from shared response to idiosyncratic response. Here, we used intersubject representational similarity analysis (IS-RSA), a novel data analysis approach (Finn et al., 2020). To this end, we first constructed representation dissimilarity matrices (RDMs) based on partners' expected emotions and social trait judgments, respectively. Specifically, each RDM consists of the manhattan distance in the measure between all possible pairs of participants, which is calculated by summing up the absolute difference between two participants' ratings on each item. We then calculate the spearman correlation between the upper triangle of RDMs. A positive correlation coefficient would indicate that two participants who are similar in their social trait judgments would also be similar in their expectation of partners' emotion. This analysis would reveal a different aspect of individual differences compared to the standard correlation method described above.

To correct the significance level and account for possible inflation stemming from the high dimensionality of the RDMs, we conducted a permutation test. A null distribution of correlation coefficient was built with 1000 iterations to statistically compare the observed coefficient with those generated from surrogate data. In each iteration, we shuffled the participant labels, recalculated the social trait judgment RDM, and recalculated the correlation coefficient based on this surrogate RDM. We then compared the actual coefficients to the null distribution.

2.2 Result

2.2.1 Guilt is modulated by responsibility

We first examined participants' performance in the interpersonal transgression task. We focused on comparing the self-incorrect condition and the both-incorrect condition because the participants' performance feedback was the same in these two conditions. Therefore, participants' perception of their own performance should not have any impact on the comparison. The only difference between the two conditions was the participants' responsibility in causing the unpleasant outcome to the partner—while in the self-incorrect condition the participant was fully responsible, in the both-incorrect condition the partner and the participant shared the responsibility, thereby diluting it (Yu et al., 2014, 2017).

As a sanity check, we showed that, overall, participants reported experiencing significantly more guilt in the self-incorrect condition compared to the incorrect condition (Figure 1(b)), paired t -test, $t(120) = 9.16$, $p < .001$, $d = 0.83$, 95% CI [14.23, 22.08], and significantly more shame, $t(120) = 6.05$, $p < .001$, $d = 0.55$, 95% CI [6.35, 12.52]. In addition, they exhibited a stronger tendency for compensation, $t(120) = 3.55$, $p < .001$, $d = 0.32$, 95% CI [0.21, 0.75].

Regarding participants' expectation of their partners' emotions, the data revealed that partners were perceived as significantly more angry, $t(120) = 7.70$, $p < .001$, $d = 0.70$, 95% CI [11.18, 18.93], more disappointed, $t(120) = 8.22$, $p < .001$, $d = 0.75$, 95% CI [12.47, 20.39], and resentful, $t(120) = 7.63$, $p < .001$, $d = 0.69$, 95% CI [11.13, 18.92], in the self-incorrect

condition compared to the both-incorrect condition. These results suggest that the experimental design successfully manipulated participants' responsibility in causing unpleasant outcomes to a social partner, and resulted in an impact on their expectation of partner's emotion, their own emotion, and behavioral responses.

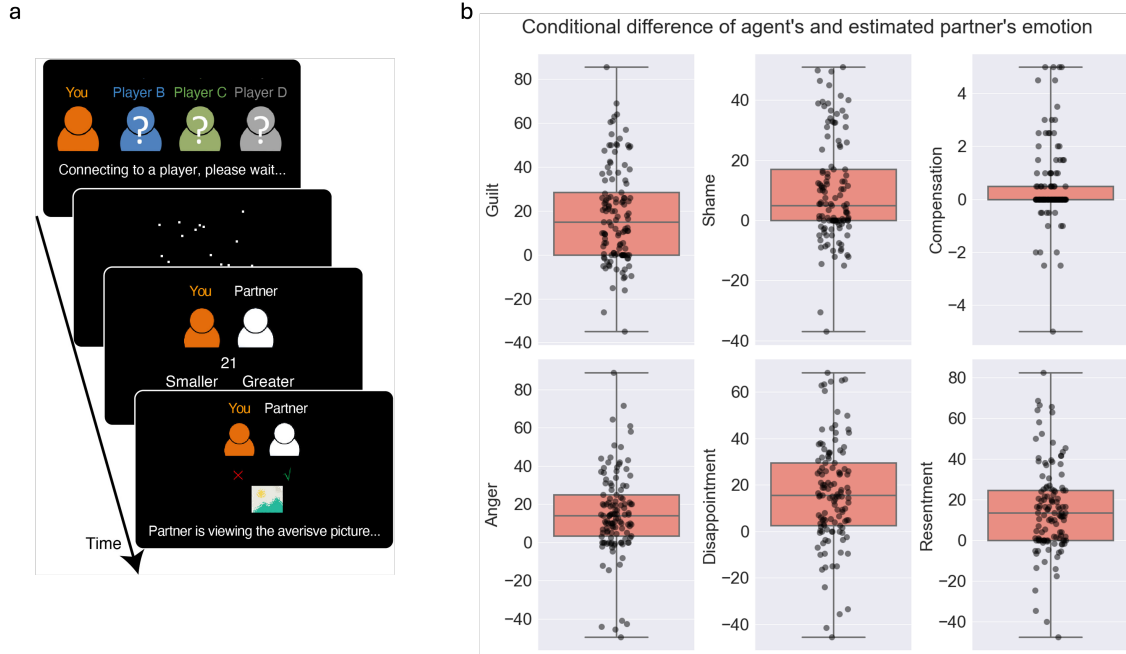


Figure 1: Procedure of the interpersonal transgression task and emotional measures. (a) Procedure of the guilt induction task. At the beginning of each trial, the participant was randomly paired with a player (i.e., Partner). The participant and the Partner then completed a dot estimation task, where they estimated the number of dots on the screen and compared it with a reference number (21 in this case). The participant's and the Partner's performance was then presented. Failure of either of them would lead to an unpleasant outcome for the Partner—viewing an aversive picture for 10 s. On different trials, participants were asked to report their own feelings, their expectation on partner's feelings, and willingness to share the Partner's unpleasant outcome as a hypothetical measure of compensation. (b) Participant's emotions and expected partner emotion. The box plot showed the difference between self-incorrect and both-incorrect condition, indicating the magnitude that modulated by higher responsibility of partner's suffering. Whiskers indicate the full range of the data. The top and bottom of the box indicate the lower and upper quantiles, respectively. The central line indicates the medium. Individual data points were shown as overlays.

2.2.2 Associations between self-emotion and other's emotion

Based on the significant differences observed in self-reported emotion and compensation between the self-incorrect and both-incorrect conditions, we defined a new variable representing guilt-related measures by subtracting the self-report value in the both-incorrect condition from that in the self-incorrect condition. These difference scores were then used in subsequent analyses.

We examined whether participants' emotional and behavioral responses were associated with their expectation of partners' emotional reactions. Pearson correlations were computed between self-reported conditional differences of guilt, shame, and the tendency to compensate and expectation of partners' anger, disappointment, and resentment (Table 1). For self-reported guilt, significant positive correlations emerged with partners' anger, $r(119) = 0.34$, $p < .001$, disappointment, $r(119) = 0.29$, $p = .001$, and resentment, $r(119) = 0.20$, $p = .029$. In contrast, while shame was significantly correlated with partners' anger, $r(119) = 0.26$, $p = .004$, and disappointment, $r(119) = 0.23$, $p = .011$, it was not significantly related to resentment, $r(119) = 0.08$, $p = .399$. Moreover, the tendency to compensate was significantly associated only with partners' resentment, $r(119) = 0.20$, $p = .027$, but not with emotional measures (anger, $r(119) = 0.12$, $p = .207$; disappointment, $r(119) = 0.06$, $p = .537$).

Table 1: Pearson correlations between self-reported emotion and expectation of partners' emotions in Study 1 (N = 121)

Self-reported measure	Anger	Disappointment	Resentment
Guilt	0.34 *** ($p < .001$)	0.29 ** ($p = .001$)	0.20* ($p = .029$)
Shame	0.26 ** ($p = .004$)	0.23* ($p = .011$)	0.08 ($p = .399$)
Compensation	0.12 ($p = .207$)	0.06 ($p = .537$)	0.20* ($p = .027$)

Note: *** $p < .001$, ** $p < .01$, * $p < .05$. All correlations are Pearson correlation coefficients with $df = 119$.

These findings reveal distinct patterns across self-conscious emotions and behaviors.

Guilt was robustly associated with multiple facets of the partner's negative response—anger, disappointment, and resentment—whereas shame showed more selective associations, relating to emotional devaluation (anger and disappointment) but not to anticipated behavioral retaliation (resentment). Compensatory behavior, in turn, appeared particularly tied to expectations of the partner's retaliatory intentions rather than their affective states.

2.2.3 Associations between trait rating and partner's emotion

In subsequent analyses, we tested the hypothesis that general social trait judgments are associated with context-specific expectations of a partner's emotion. First, simple Pearson correlations were computed between social trait judgments and participants' ratings of partners' anger, disappointment, and resentment. However, none of these correlations reached significance ($ps > .05$; see Supplementary Material).

Next, we move on to test the idiosyncrasy of social trait judgment and emotion expectation. The IS-RSA analysis revealed a significant positive association between social trait judgments and expectation of partners' emotion (permutation test with 1,000 iterations, $r = .07$, corrected $p < .001$, Figure 2). This result indicates that similarity in social trait judgment style is closely linked to a corresponding pattern of mentalizing partners' emotional states. Together, these findings support an association between trait perception and partner response expectation at inter-subject level.

2.3 Discussion

In this study, we observed a striking association between individuals' emotional responses and their expectation of a partner's emotions. Moreover, social trait judgments style were linked to expected partner's emotion in an inter-subject manner.

Interestingly, individuals' emotion appears to relate to different facets of a partner's affect. For instance, individuals' guilt was associated with partners' anger, disappointment, and resentment, whereas shame was linked only to anger and disappointment. This selective

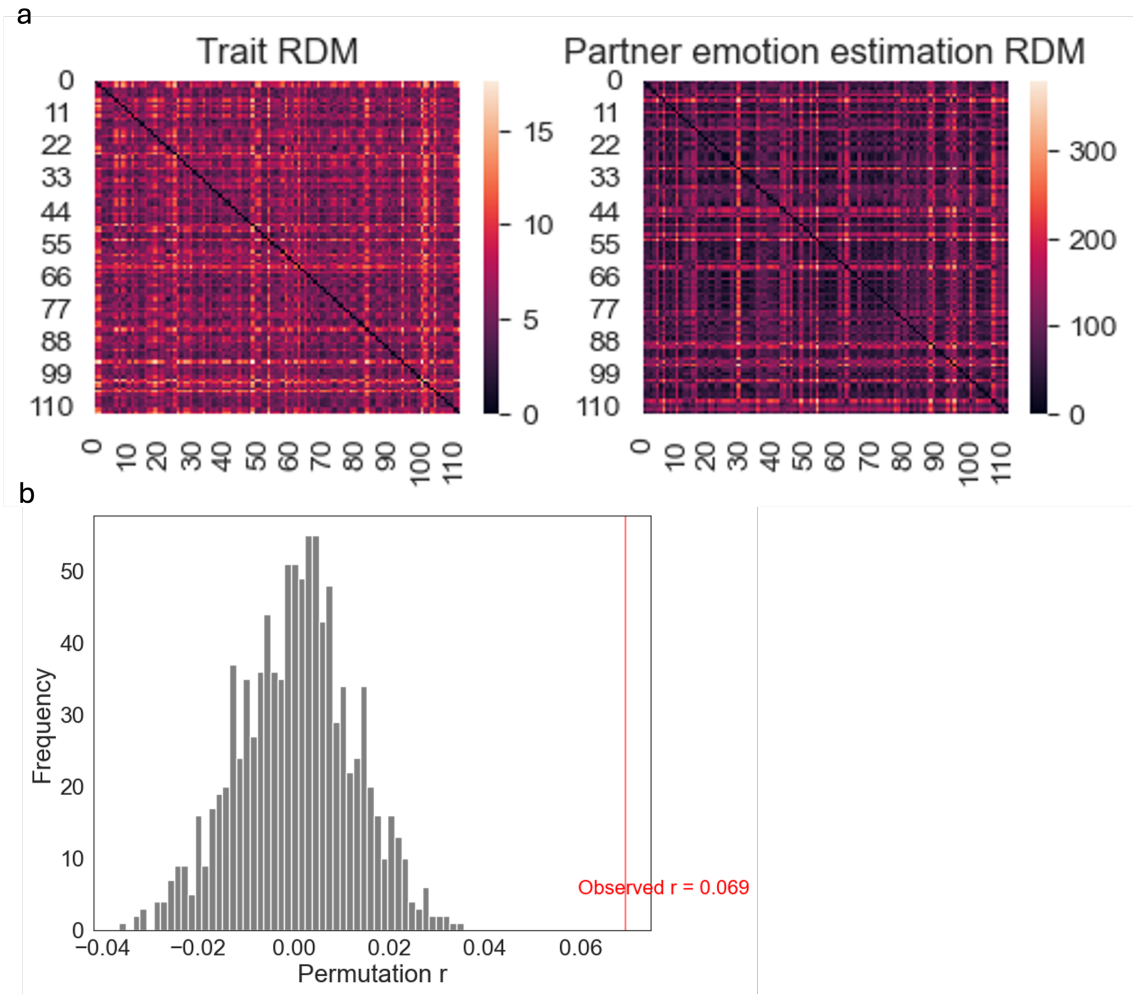


Figure 2: Inter-subject representational similarity analysis (IS-RSA). (a) Representational Dissimilarity Matrices (RDMs). On the left side we showed the social trait judgments RDM, while the left side showed the expectation of a partner's emotion RDM. Each cell represents the manhattan distance from one participant's to the other participant's rating/judgment style across all items. Darker color indicates higher similarity, while lighter color indicates larger distance. (b) Result of permutation test. The gray bars showed the null distribution, while the red line showed the observed correlation coefficient between trait and emotion RDMs.

association suggests that guilt and shame serve distinct functions as social signals that promote adaptive behavior (Sznycer, 2019). Research indicates that shame is engineered to counter social devaluation. Across cultures, shame closely tracks potential audience devaluation (Sznycer et al., 2016). From an attributional perspective, shame is often triggered by events that conflict with one's identity goals, especially when these events are attributed to uncontrollable aspects of the self (Tangney, 1990; Tracy et al., 2013). Consequently, shame functions either in a preventive manner by discouraging actions that might lead to further devaluation (de Hooge et al., 2008), and/or, once an event has occurred, instead of initiating remedy, it encourages one to downplay self-evaluation by de-emphasizing personal welfare (Gilbert, 2000). In contrast, guilt signals a failure to sufficiently value a significant other (Baumeister et al., 1994; Sznycer et al., 2016; Tooby and Cosmides, 2008) and is typically linked to a more controllable and specific mismatch with one's identity goals (Tangney, 1990; Tracy and Robins, 2004). When feeling guilty, individuals are more motivated to repair relationships and reaffirm their valuation of others (Baumeister et al., 1994; Gilbert, 1997; Sznycer et al., 2016; Tooby and Cosmides, 2008; Tracy and Robins, 2004). Our findings mirrors the theoretical root of guilt and shame. More specifically, shame is selectively activated in proportion to the magnitude of partners' negative emotions to prevent future devaluation, while guilt not only tracks partner emotion but is also closely related to the anticipation of resentment, underscoring the importance of relationship repair.

We also observed a strong correlation between individuals behavioral intentions and their expectation of a partner's hostile actions. When anticipating punishment, individuals were more motivated to take steps to appease perceived hostility, as evidenced by a stronger intention to compensate and reduce the partner's suffering.

Furthermore, we found an association between social trait judgments and the expectation of partners' emotions, implying an interaction between general trait impressions and situational evaluation. It is important to emphasize that the association we observed is not a straightforward, linear relationship in which higher social trait judgments simply result in either

enhanced or diminished emotional expectation. Instead, the effect manifests inter-subjectively: individuals who share similar social trait judgment styles tend to exhibit similar patterns of emotion prediction. Since these social trait judgments were gathered in a separate, minimal-information context that reflects an individual's general perception style, this relationship offers valuable insight into how context-free social perceptions can influence context-specific emotional processes. This finding underscores the complexity of the relationship, suggesting that nonlinearity—patterns that cannot be captured by simple correlational analyses, may be present.

3 Study 2: Cross-cultural uniformity/difference

Although Study 1 provides evidence for the associations between emotional and behavioral processes within an interactive dyad, several questions remain. First, the motivational consequences of guilt warrant further investigation. Specifically, we seek to explore how guilt relates to two opposing behavioral motivations—approach and avoidance—and how these motivations link to expectations of partners' emotions. We assess whether expectation of partner forgiveness and anger differentially drive approach and avoidance tendencies, respectively. This interpersonal perspective enhances the ecological validity of our research and provides a more accurate reflection of real-world social interactions.

Moreover, to verify the generalizability of our model, we tested the associations observed in Study 1 across different cultures. According to evolutionary adaptation theory (Sznycer, 2019), guilt is calibrated to encode mismatches in social valuation. Hence, we expect to observe similar patterns across cultures. We recruited participants from an individualistic culture (the U.S.) and a collectivistic culture (China), and compared results across both samples.

In summary, Study 2 extended the paradigm from Study 1 by recruiting culturally diverse samples to test cross-cultural generalizability. We investigated how guilt drives distinct motivational consequences and how these relate to expectations of partners' emotions. We hypothesized that: (a) consistent with Study 1, individuals' expectations of their partners'

reactions would influence their own guilt; (b) social trait judgment styles would be associated with expectations of partners' emotions; (c) guilt would be associated with both approach and avoidance tendencies, albeit with different magnitudes; (d) expectations of partners' emotions would differentially modulate approach and avoidance tendencies; and (e) the results would be generalized across cultures.

3.1 Method

3.1.1 Participants

We recruited 125 participants from an internal research platform (SONA) in University of California, Santa Barbara (UCSB). All the participants are UCSB students. Forty-four participants were excluded because of incomplete response and didn't meet data cleaning criteria (see following sections for more details), leading to a sample of 81 participants for further analysis (Gender: 82.72% females, 13.58% males, 3.70% others; Age: 19.50 ± 1.41 ; Race and ethnicity: 25% White, 20% Asian, 16% Hispanic or Latino, 39% Other).

We recruited 120 participants from a Chinese online experimental platform Naodao. Eight participants were excluded using the same criteria, leading to a sample of 104 participants (62.5% females, 37.5% males; age = 23.59 ± 3.70 ; all Asian) for further analysis.

None of the participants reported a diagnosis of psychopathological disease, including prosopagnosia. All participants provided written informed consent using procedures approved by the Institutional Review Boards of University of California Santa Barbara (Protocol #2-24-0341).

3.1.2 Social Trait judgement task

The experimental procedure is similar to Study 1. A few changes were made to broaden the span of the trait space and improve the cultural variety of facial images.

Participants will see neutral, frontal facial pictures sourced from three publicly available

face databases (Chelnokova et al., 2014; DeBruine and Jones, 2017; Ma et al., 2015). The facial images pool include Asian, Black and White faces, and faces from both genders. To ensure a comprehensive representation of facial traits, we employed maximum variation sampling based on an independent set of facial rating data. We started with 100 images, and selected 50 faces with the highest variation across the first four trait dimensions (Lin et al., 2021). This approach is aimed to maximize the diversity of facial features presented to participants, thereby ensuring representative and robust results.

During the social trait judgment task, participants evaluated trustworthiness, friendliness, criticalness, and competence of each face on a 1-7 continuous scale. We chose these specific traits on the warm/criticalness dimension of the facial trait space due to the significant association between perception of criticalness and self-reported guilt in Study 1 ($r(119) = .26$, $p = .004$). Moreover, competence is included to incorporate the warmth-competence dimensions of trait judgments from faces, drawing from a well-established framework (Fiske et al., 2007).

The order of the trait dimensions was randomized across participants to control for order effects. Each face will be presented for 0.5 seconds before the trait rating scale appears. There is no time limit for the trait rating, but participants will be encouraged to respond as quickly and accurately as possible, while swift (< 0.2 seconds) and delayed (> 10 seconds) responses will be warned.

For Naodao experiment, all the procedures are the same as the SONA sample. Instructions were translated into Mandarin Chinese by two bi-lingual speaker. Backward translation was conducted to make sure the meaning stayed consistent.

During data cleaning, we applied the following three exclusion criteria: (1) We excluded trials with reaction times shorter than 100 ms or longer than 5000 ms, aligning with Study 1. (2) Participants that failed more than 20% of the attention check ('Press key # to indicate that you are paying attention') were excluded.

3.1.3 Interpersonal transgression task

The overall procedure of the transgression task remains unchanged compared to Study 1. Several modifications were implemented to improve the ecological validity and deepen the understanding of different aspects of the guilt-related emotional process:

First, after getting the feedback, participants were prompted to answer one of following questions: (1) how guilty they were for the partner's unpleasant outcome (0-100 continuous scale) and (2) how long they would be willing to watch the aversive picture themselves, which could reduce the time the partner had to watch the aversive picture (i.e., a measure of compensation) (0-5 continuous scale); (3) the extent to which they were inclined to approach their partner for an apology or, conversely, to avoid interacting with them (0-100 continuous scale) (4) the likelihood of their partner forgiving them, as opposed to holding resentment towards them (0-100 continuous scale). Notably, question (3) comprised two separate items assessing approach and avoidance tendencies within the same trial. Each question was randomly presented three times in each condition (excluding the both-correct condition), and participants indicated their responses on analog scales.

Moreover, to intensify participants deliberation when compensating, an aversive image was shown to them at the end of the experiment with a presentation duration according to their average compensation choices.

There were 4 trials for the condition that both participant and their partner answer correctly ('both right'), and 12 trials for each of the rest possible conditions. To make sure participants are paying attention during the task, we included two attention check trials with apparent answers of the dot estimation task. In the attention check trials, the number of dots largely deviates from 20. Participants who fail to answer correctly will be warned to focus on the task. During data cleaning, we excluded participants that have less than 80% accuracy in attention check trials.

Participants have 5 seconds to make judgement about the number of dots in each trial. If they fail to respond within the time limit, they will be warned to respond faster, while the

trial will be marked as incomplete. After they complete the rest of the experiment, they will automatically move to an extra loop for the incomplete trials, which gives them one more opportunity to respond to the incomplete trials.

3.2 Result

3.2.1 Guilt is modulated by responsibility

As a sanity check and to confirm the pattern we observed in Study 1, we first tested the conditional difference between all guilt-related measures (Figure 3). Results from paired t -tests indicated a similar pattern across both cultures. In the U.S. sample, participants reported significantly more guilt in the self-incorrect condition than in the both-incorrect conditions, $t(80) = -6.25, p < .001, d = 0.69, 95\% \text{ CI } [8.63, 16.69]$. Participants also showed a marginally significant tendency for greater compensation, $t(80) = -1.93, p = .057, d = 0.21, 95\% \text{ CI } [-0.01, 0.38]$. Furthermore, they were significantly more willing to apologize, $t(80) = -4.47, p < .001, d = 0.50, 95\% \text{ CI } [4.50, 11.73]$, and to avoid further interaction in the self-incorrect condition, $t(80) = -2.67, p = .009, d = 0.30, 95\% \text{ CI } [1.50, 10.27]$, compared to the both-incorrect conditions.

Regarding participants' expectation of their partners' emotions, the data revealed that partners were perceived as significantly less likely to forgive, $t(80) = 4.03, p < .001, d = -0.45, 95\% \text{ CI } [-12.38, -4.19]$, and significantly more angry, $t(80) = -4.12, p < .001, d = 0.46, 95\% \text{ CI } [4.22, 12.09]$, in the self-incorrect condition relative to the incorrect conditions.

In the Chinese sample, participants reported experiencing more guilt in the self-incorrect condition compared to both incorrect conditions, $t(103) = -9.24, p < .001, d = 0.91, 95\% \text{ CI } [13.91, 21.51]$; a stronger tendency for compensation, $t(103) = -5.46, p < .001, d = 0.54, 95\% \text{ CI } [0.27, 0.57]$; and were more willing to apologize, $t(103) = -7.72, p < .001, d = 0.76, 95\% \text{ CI } [9.78, 16.54]$. However, there was no significant conditional difference for avoidance, $t(103) = -0.75, p = .452, d = 0.07, 95\% \text{ CI } [-1.59, 3.53]$.

As for participants' evaluation of partners' emotion, they reported that their partner

would be less likely to forgive, $t(103) = 8.29, p < .001, d = -0.81, 95\% \text{ CI } [-16.69, -10.24]$, and more angry at them, $t(103) = -6.89, p < .001, d = 0.68, 95\% \text{ CI } [9.70, 17.55]$, in the self-incorrect condition compared to both incorrect conditions.

Overall, the results replicated Study 1's findings and established a shared emotional response pattern across cultures. However, compared to the U.S. sample, the Chinese sample exhibited generally lower avoidance tendencies, which were not modulated by responsibility.

3.2.2 Associations between self- and other's emotion

We tested the correlation between self- and expected partner's emotion. Both the U.S. and Chinese samples revealed a pattern consistent with Study 1, indicating that all self emotions are closely related to expected partner's emotion (Figure 4). In the U.S. sample, perceiving a partner as more forgiving was associated with lower guilt ($r(79) = -.34, p = .002$) and lower compensation ($r(79) = -.27, p = .013$). Conversely, when participants believed that their partner would be angry, they reported higher guilt ($r(79) = .48, p < .001$) and tended to compensate more ($r(79) = .36, p = .001$). The Chinese sample demonstrated a similar pattern: a more forgiving partner was associated with lower guilt ($r(102) = -.38, p < .001$) and lower compensation ($r(102) = -.23, p = .017$), whereas perceiving a partner as angry was linked to higher guilt ($r(102) = .55, p < .001$) and higher compensation ($r(102) = .30, p = .002$).

3.2.3 Association between guilt and motivational tendencies

We further investigate the link between guilt and motivational tendencies. First, we explore whether guilt elevate the magnitude of approach and avoidance. Results showed that guilt is positively correlated to both approach and avoidance tendencies in both the U.S. sample (approach: $r(79) = .50, p < .001$, avoidance: $r(79) = .24, p = .032$) and the Chinese sample (approach: $r(102) = .69, p < .001$, avoidance: $r(102) = .25, p = .011$). To further examine the magnitude of the association, we conducted partial correlation test, as well as Steiger's Z-test to formally examine the strength of correlation. We observed that,

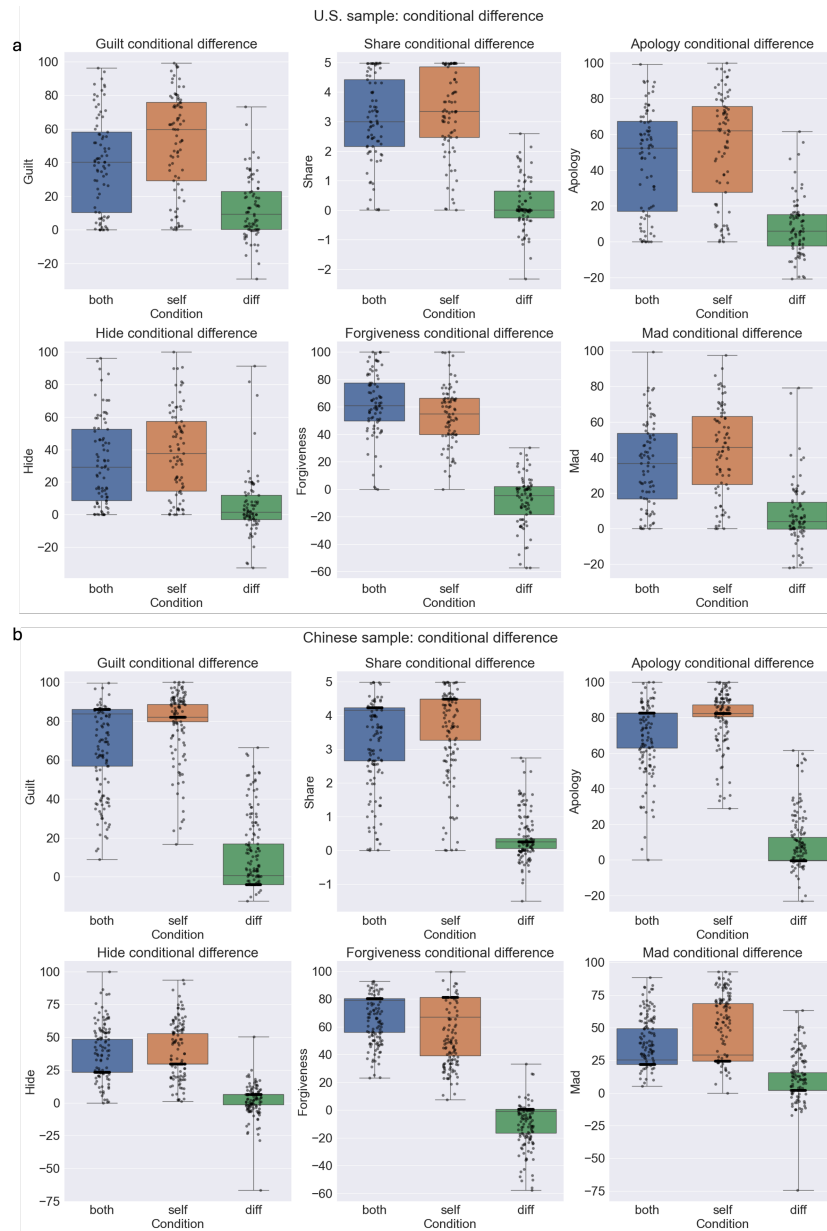


Figure 3: Conditional differences between self-emotions and expected partner emotions. “Both” denotes the baseline condition in which both the participant and their partner answered incorrectly, resulting in lower participant responsibility for the partner’s punishment. “Self” refers to the experimental condition in which only the participant answered incorrectly while the partner answered correctly, resulting in higher responsibility as the partner suffered due to the participant’s error. “Diff” represents the difference between the two conditions (Self – Both). Self-emotions include guilt and compensation (share); self-motivations encompass approach (apology) and avoidance (hide); expected partner emotions are reflected in ratings of forgiveness and anger (mad). (a) Results from the U.S. sample; (b) Results from the Chinese sample.

in both U.S. and Chinese sample, guilt is more strongly related to approach tendency (U.S.: $\text{partial}r(79) = 0.47, p < 0.001$, Chinese: $\text{partial}r(102) = 0.68, p < 0.001$) than avoidance (U.S.: $\text{partial}r(79) = 0.14, p = 0.218$, Chinese: $\text{partial}r(102) = 0.19, p = 0.051$) tendency, which this difference is statistically significant (Steiger Z test, U.S.: $z = 4.58, p < 0.001$, Chinese: $z = 2.15, p < 0.031$).

3.2.4 Partner's emotion modulate motivational tendencies

Next, we tested whether and how expected partner's emotion is associated with motivational tendencies. In the U.S. sample, participants who perceive partners as more forgiving reported less inclination to apologize ($r(79) = -.39, p < .001$) and less avoidance ($r(79) = -.23, p = .039$), while participants perceiving partners as more angry exhibited a higher tendency to apologize ($r(79) = .50, p < .001$) and avoid ($r(79) = .39, p < .001$). In the Chinese sample, the correlation between partner's forgiveness and less inclination to apologize ($r(102) = -.42, p < .001$) and avoid ($r(102) = -.26, p = .009$) remains. Partner's anger is only associated with greater tendency to apologize ($r(102) = .51, p < .001$), but no significant link between partner's anger and participants' inclination to avoid is found ($r(102) = .07, p = .492$). This cultural deviation is further confirmed by a linear regression model, with avoidance as dependent variable and expected partners' emotions, culture, as well as their interaction terms as regressors. The model results a significant interaction effect between culture and expected partner's anger ($B = 0.52 \pm 0.15, 95\% \text{ CI } [0.22, 0.82]; t = 3.41, p < 0.001$).

3.2.5 Associations between trait rating and partner's emotion

In subsequent analyses, we tested the hypothesis that general social trait judgments are associated with context-specific expectations of a partner's emotion. In both the U.S. and the Chinese sample, simple Pearson correlations didn't reveal a significant association between social trait judgments and partners' anger and forgiveness. ($ps > .05$; see Table 2 and Table 3 in Supplementary Material).

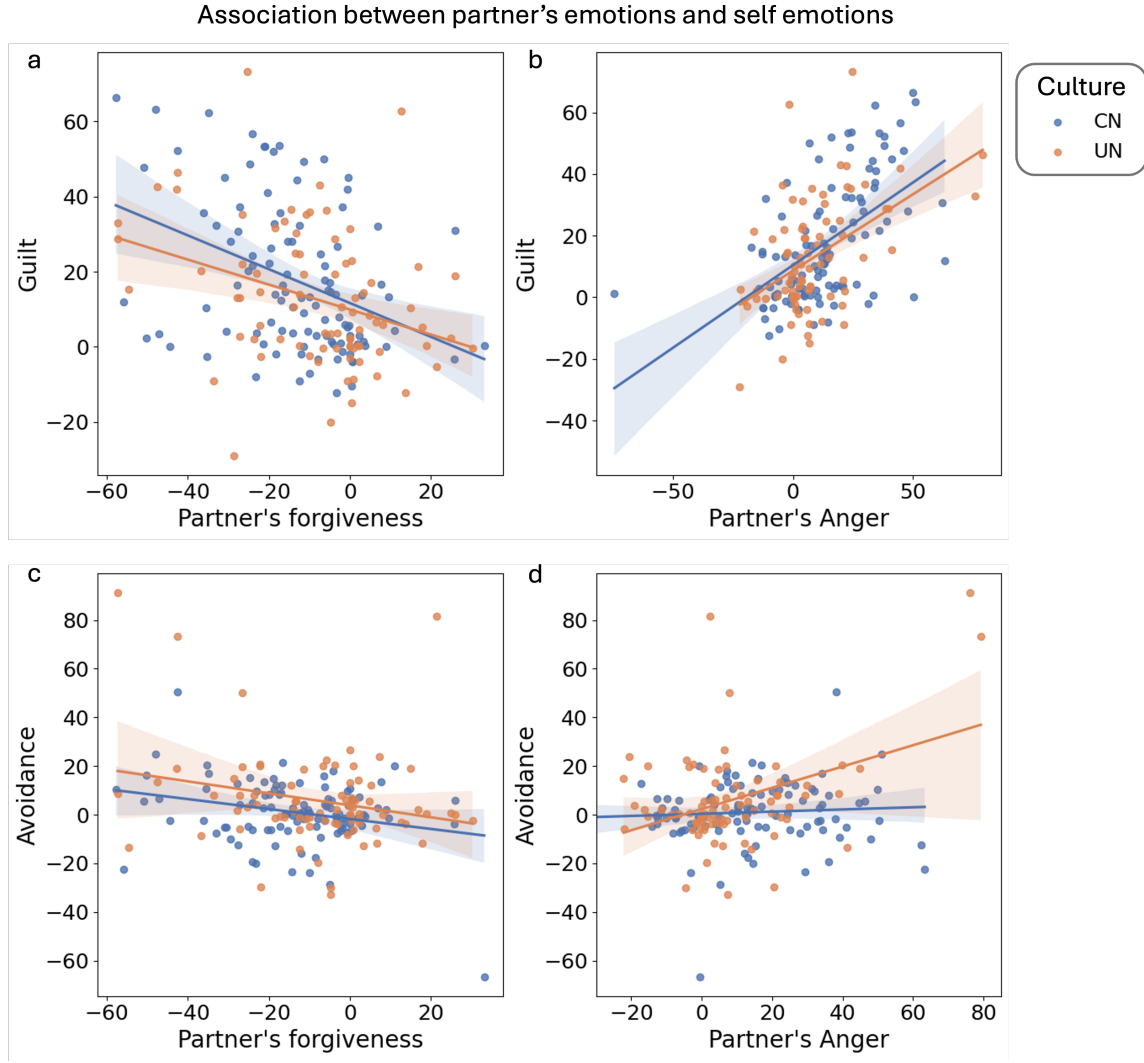


Figure 4: Cultural similarities and differences in the associations between partner emotions and self-emotions. Panels (a) and (b) display the association between an individual's guilt and their expectation of partners' forgiveness and anger, respectively. Panels (c) and (d) illustrate the association between an individual's avoidance tendency and their expectation of partners' forgiveness and anger. A cultural difference is evident in avoidance tendencies: U.S. participants showed increased avoidance with higher expected anger, whereas Chinese participants tended to face partner's anger, indicated by lower avoidance tendency. CN: Chinese sample; UN: U.S. sample.

We then test the association on the individual level. The IS-RSA analysis on the Chinese sample revealed a significant positive association between social trait judgment style and expected partner emotion (permutation test with 1,000 iterations, corrected $r = .05$, $p < .001$, Figure 5(b)). This result indicates that similarity in social trait judgment style, including all traits mentioned above is closely linked to a corresponding pattern of inferring partner's emotion in the experimental context. More specifically, if two participants have similar social judgment pattern, for example, perceiving others as warmer or more trustworthy, they also appear to estimate partner's emotion in a closer way. Together, these findings support an association between trait perception and emotion expectation at the individual level.

However, although we observed association between social trait judgment style and emotion expectation in Study 1, we failed to replicate the result on the U.S. sample in Study (permutation 1000 times, $r = 0.0001$, corrected $p = 0.526$, Figure 5(a)). Potential limitations will be discussed.

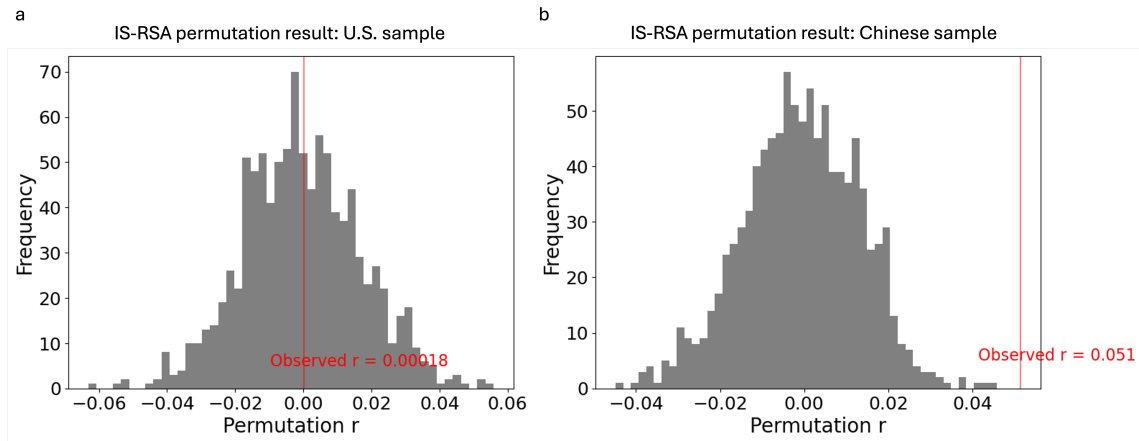


Figure 5: Results from IS-RSA. The gray histogram showed the null distribution of 1,000 times permutation, while the red line showed the observed correlation coefficient. (a) Results from the U.S. sample (b) Results from the Chinese sample.

3.3 Discussion

Study 2 not only confirmed most of the findings from Study 1 but also provided insights into motivational tendencies, as well as cultural uniformity and differences. In both American

and Chinese samples, we observed a strong link between expected partners' emotions and individuals' emotional responses. This supports the adaptationist theory of self-conscious emotion, which posits that emotions serve adaptive functions to optimize social interaction outcomes. As such, the functionality is supposed to be uniform and can be found in every culture, despite the ideological difference. Specifically, our results indicate that guilt-related emotional and behavioral processes—manifested as compensatory actions—closely track the potential undervaluation of a social partner, as indicated by the partner's negative emotional reactions. These adaptive responses, such as active apologies and remedial actions, help repair relationships.

Despite these cultural differences, we observed a general monotonic effect of guilt and expectations of partner emotions on motivational tendencies: increased guilt intensified both approach and avoidance tendencies; a forgiving partner was associated with reduced guilt and correspondingly lower approach and avoidance, whereas an angry partner elicited greater guilt alongside heightened approach and avoidance. Intuitively, approach and avoidance appear to be opposing tendencies that should not coexist. However, in the current study, both emerged simultaneously. This may reflect the distinction between behavioral reactions and motivational tendencies. Participants were placed in a hypothetical setting and likely experienced internal conflicts following wrongdoing. Consequently, we observed that both approach and avoidance tendencies increased in response to heightened guilt and anticipated partner anger.

However, cultural differences also emerged. While American participants' avoidance tendency increased when they bore greater responsibility and was associated with expected partner anger, Chinese participants showed no such inclination to avoid when more responsible and maintained consistently low avoidance regardless of anticipated partner anger. These findings align with previous research on cultural differences in relationship restoration. Individuals from individualistic cultures, such as the U.S., tend to avoid direct confrontation and expressions of anger, whereas those from collectivistic cultures are more likely to offer apologies explicitly and frequently (Barnlund and Yoshioka, 1990; Wagatsuma and Rosett, 1986),

reflecting differences in self-representation. Some researchers suggest that North Americans typically exhibit a strong need for self-enhancement (Heine and Lehman, 1999; Heine et al., 1999), which motivates them to avoid publicly admitting individual mistakes. In contrast, East Asians often maintain a self-critical and self-improving orientation (Heine and Lehman, 1999; Kim, 2003). In these cultures, directly addressing potentially face-threatening situations is seen as a demonstration of modesty and a necessary step for self-improvement (Kim, 2003). In the current study, when bearing greater responsibility for their partner's suffering, Chinese participants continued to show relatively low avoidance, indicating a preference for addressing the situation directly. Moreover, their avoidance tendency was not modulated by partner anger, suggesting a distinct pathway between emotion and behavioral motivation. These correlational findings lay a foundation for future exploration of cross-cultural differences.

Moving to the link between partner emotion and social trait judgments, we observed slight differences compared to Study 1. The Chinese sample displayed a positive inter-subject correlation similar to that observed among U.S. participants in Study 1, where similarities in social trait judgments were associated with similarities in emotion expectation. In contrast, the U.S. student sample did not exhibit this association. Considering that the U.S. sample was relatively small and unrepresentative, which is composed solely of university students and has a biased gender ratio, further research is needed to disentangle these differences and clarify the incongruence.

4 General Discussion

In conclusion, this study introduced an interactive social scenario that enables researchers to examine self-conscious emotions in real-time interpersonal exchanges. We observed a robust circulation of emotion between the interactive dyad: the individual's emotions closely track their expectation of the partner's emotional state, a pattern that persisted across different samples and cultures. Moreover, we identified an inter-subject association between

social trait judgments and the expectation of partners' emotions, linking the general social judgment style to situational social contexts. This finding has profound implications, deepening the understanding of impaired daily social interactions observed in certain clinical groups, which may stem from altered trait perception.

Emotions are involved in every stage of social interactions and permeate every aspect of our lives, serving both as a motivator for actions and as a consequence of our decisions. Self-conscious emotions, in particular, are grounded in social evaluation and emerge within interactive settings. Our findings underscore the necessity of studying emotions in their natural interpersonal contexts, where both the emotion generator and the social partner are considered. This study demonstrates that expectations regarding partners' emotional and behavioral reactions can significantly influence the individual's own emotional experience. Not only does our work provide empirical support for the adaptive perspective of self-conscious emotion (Sznycer, 2019), but it also takes an important first step toward disentangling the dynamics of emotion generation in an interpersonal context. Future research could investigate the temporal progression of partners' emotion expectation and the subsequent emergence of self-emotion, as well as how these processes evolve in response to various real-time information.

Extending our discussion further, we explored the relationship between social trait judgments and context-specific emotion expectation—a process closely linked with mentalizing. An inter-subject correlation was identified in both U.S. and Chinese datasets, despite the trait judgment task being entirely independent of the interactive transgression task. This result provides initial evidence for the interrelationship between static social trait perception and situation-specific emotion expectation.

Future studies could take several directions. First, although our study did not reveal a robust, direct, linear relationship between social perception style and emotion expectation, it does demonstrate their connection in a more complex, inter-subject manner. Moreover, it may be possible to further elucidate this connection and build predictive models of emotion expectation using social perception style. Previous literature supports this possibility: individuals with

autism spectrum disorder have been shown to exhibit both altered social trait perception (Cao et al., 2023; Yu et al., 2022; Zhao et al., 2024) and impaired mentalizing ability (see Chung et al. (2014) for a meta-analysis). Their consistently skewed social perception suggests a potential link between specific social trait representational patterns and hindered mentalizing processes. A more comprehensive model that accounts for the nonlinearity of this relationship could yield meaningful insights.

Second, it is important to note that, due to the limitations of our design, we were only able to examine differences between subjects rather than variations within subjects across different contexts. Our current paradigm and the independence of our two experiments do not allow us to capture the nuances of one's emotional reactions and expectations across varying situations, such as interacting with a harsh partner compared to a kind partner. Future research should consider modifying the paradigm to incorporate greater flexibility, thereby enabling researchers to track emotional changes in response to diverse situational and interpersonal factors.

Additionally, all interactions in the current study were one-shot events. While this design allowed us to isolate the basic effects of emotion generation without added complexity, self-conscious emotions are inherently embedded in a broader social climate. The processes underlying emotion generation—such as the need for relational repair and the emphasis on others' valuation—may only be fully realized in multiple rounds of interaction or within a genuine social circle. Consequently, some essential components of self-conscious emotion may be missing from our current findings.

To sum up, this study not only elucidates the intricate interplay between expectation of partner's emotion and the generation of self-conscious emotion but also highlights the importance of embedding emotion research within authentic social contexts. Our findings reveal that an individual's expectation about their partner's reactions—which are closely tied to their social trait judgment styles—profoundly shape their own emotional experiences. This work establishes a critical foundation for future research aimed at unraveling the complex and multifaceted dynamics of emotion generation in real-world social interactions.

References

- Ahler, D. J., Citrin, J., Dougal, M. C., and Lenz, G. S. (2017). Face value? Experimental evidence that candidate appearance influences electoral choice. *Political Behavior*, 39(1):77–102.
- Barnlund, D. C. and Yoshioka, M. (1990). Apologies: Japanese and American styles. *International Journal of Intercultural Relations*, 14(2):193–206.
- Baron, S. G., Gobbini, M. I., Engell, A. D., and Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, 6(5):572–581.
- Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin*, 115(2):243–267.
- Blair, I. V., Judd, C. M., and Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science*, 15(10):674–679.
- Cao, R., Zhang, N., Yu, H., Webster, P. J., Paul, L. K., Li, X., Lin, C., and Wang, S. (2023). Comprehensive social trait judgments from faces in autism spectrum disorder. *Psychological Science*, 34(10):1121–1145.
- Chelnokova, O., Laeng, B., Eikemo, M., Riegels, J., Løseth, G., Maurud, H., Willoch, F., and Leknes, S. (2014). Rewards of beauty: The opioid system mediates social motivation in humans. *Molecular Psychiatry*, 19(7):746–747.
- Chung, Y. S., Barch, D., and Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bulletin*, 40(3):602–616.
- de Hooge, I. E., Breugelmans, S. M., and Zeelenberg, M. (2008). Not so ugly after all:

- When shame acts as a commitment device. *Journal of Personality and Social Psychology*, 95(4):933–943.
- De Hooze, I. E., Zeelenberg, M., and Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition and Emotion*, 21(5):1025–1042.
- DeBruine, L. and Jones, B. (2017). Face Research Lab London Set. https://figshare.com/articles/Face_Research_Lab_London_Set/5047666.
- Finn, E. S., Glerean, E., Khojandi, A. Y., Nielson, D., Molfese, P. J., Handwerker, D. A., and Bandettini, P. A. (2020). Idiosyncrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage*, 215:116828.
- Fiske, S. T., Cuddy, A. J. C., and Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2):77–83.
- Frijda, N. H., Kuipers, P., and Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, 57(2):212–228.
- Gao, X., Yu, H., Peng, L., Gong, X., Goh, J. Y. Y., Jiang, C., and Zhou, X. (2021). The mutuality of social emotions: How the victim’s reactive attitude influences the transgressor’s emotional responses. *NeuroImage*, 244:118631.
- Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., and Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences*, 115(33):E7680–E7689.
- Gilbert, P. (1997). The evolution of social attractiveness and its role in shame, humiliation, guilt and therapy. *The British Journal of Medical Psychology*, 70(2):113–147.
- Gilbert, P. (2000). The relationship of shame, social anxiety and depression: The role of the evaluation of social rank. *Clinical Psychology & Psychotherapy*, 7(3):174–189.

- Heine, S. J. and Lehman, D. R. (1999). Culture, self-discrepancies, and self-satisfaction. *Personality and Social Psychology Bulletin*, 25(8):915–925.
- Heine, S. J., Lehman, D. R., Markus, H. R., and Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106(4):766–794.
- Hofstede, G. H. (1978). Value systems in forty countries: Interpretation, validation and consequence for theory. Technical report, European Institute for Advanced Studies in Management.
- Kim, S.-H. (2003). Korean cultural codes and communication. *International Area Review*, 6(1):93–114.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. Technical report, NIMH Center for the Study of Emotion and Attention.
- Leith, K. P. and Baumeister, R. F. (1998). Empathy, shame, guilt, and narratives of interpersonal conflicts: Guilt-prone people are better at perspective taking. *Journal of Personality*, 66(1):1–37.
- Lenz, G. S. and Lawson, C. (2011). Looking the part: Television leads less informed citizens to vote based on candidates’ appearance. *American Journal of Political Science*, 55(3):574–589.
- Li, Z., Yu, H., Zhou, Y., Kalenscher, T., and Zhou, X. (2020). Guilty by association: How group-based (collective) guilt arises in the brain. *NeuroImage*, 209:116488.
- Lin, C., Adolphs, R., and Alvarez, R. M. (2017). Cultural effects on the association between election outcomes and face-based trait inferences. *PLOS ONE*, 12(7):e0180837.
- Lin, C., Adolphs, R., and Alvarez, R. M. (2018). Inferring whether officials are corruptible from looking at their faces. *Psychological Science*, 29(11):1807–1823.
- Lin, C., Keles, U., and Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. *Nature Communications*, 12(1):5168.

- Ma, D. S., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47:1122–1135.
- Martin, D. S. (1978). Person perception and real-life electoral behaviour. *Australian Journal of Psychology*, 30(3):255–262.
- Oosterhof, N. N. and Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32):11087–11092.
- Roseman, I. J., Wiest, C., and Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology*, 67(2):206–221.
- Scherer, K. R. and Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2):310–328.
- Schmader, T. and Lickel, B. (2006). The approach and avoidance function of guilt and shame emotions: Comparing reactions to self-caused and other-caused wrongdoing. *Motivation and Emotion*, 30(1):42–55.
- Su, C. and Hynie, M. (2019). A cross-cultural study on the experience of shame and guilt. In Zangeneh, M. and Al-Krenawi, A., editors, *Culture, Diversity and Mental Health—Enhancing Clinical Practice*, pages 243–262. Springer International Publishing.
- Sznycer, D. (2019). Forms and functions of the self-conscious emotions. *Trends in Cognitive Sciences*, 23(2):143–157.
- Sznycer, D. (2022). Value computation in humans. *Evolution and Human Behavior*, 43(5):367–380.
- Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., and Halperin, E. (2016). Shame

- closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences*, 113(10):2625–2630.
- Sznycer, D., Xygalatas, D., Agey, E., Alami, S., An, X.-F., Ananyeva, K. I., Atkinson, Q. D., Broitman, B. R., Conte, T. J., Flores, C., Fukushima, S., Hitokoto, H., Kharitonov, A. N., Onyishi, C. N., Onyishi, I. E., Romero, P. P., Schrock, J. M., Snodgrass, J. J., Sugiyama, L. S., Takemura, K., Townsend, C., Zhuang, J.-Q., Aktipis, C. A., Cronk, L., Cosmides, L., and Tooby, J. (2018). Cross-cultural invariances in the architecture of shame. *Proceedings of the National Academy of Sciences*, 115(39):9702–9707.
- Tangney, J. P. (1990). Assessing individual differences in proneness to shame and guilt: Development of the self-conscious affect and attribution inventory. *Journal of Personality and Social Psychology*, 59(1):102–111.
- Tangney, J. P. and Fischer, K. W. (1995). *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*. Guilford Press.
- Tangney, J. P., Wagner, P., and Gramzow, R. (1992). Proneness to shame, proneness to guilt, and psychopathology. *Journal of Abnormal Psychology*, 101(3):469–478.
- Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728):1623–1626.
- Tooby, J. and Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In *Handbook of emotions*, pages 114–137. Guilford Press, 3rd edition.
- Tracy, J. L. and Robins, R. W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry*, 15(2):103–125.
- Tracy, J. L., Robins, R. W., and Tangney, J. P. (2013). *The Self-Conscious Emotions: Theory and Research*. Guilford Publications.

- van 't Wout, M. and Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108(3):796–803.
- Wagatsuma, H. and Rosett, A. (1986). The implications of apology: Law and culture in Japan and the United States. *Law & Society Review*, 20(4):461–498.
- Wicker, F. W., Payne, G. C., and Morgan, R. D. (1983). Participant descriptions of guilt and shame. *Motivation and Emotion*, 7:25–39.
- Willis, J. and Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7):592–598.
- Wiltermuth, S. S. and Cohen, T. R. (2014). "I'd only let you down": Guilt proneness and the avoidance of harmful interdependence. *Journal of Personality and Social Psychology*, 107(5):925–942.
- Wong, Y. and Tsai, J. (2007). Cultural models of shame and guilt. In *The Self-Conscious Emotions: Theory and Research*, pages 209–223. Guilford Press.
- Yu, H., Cao, R., Lin, C., and Wang, S. (2022). Distinct neurocognitive bases for social trait judgments of faces in autism spectrum disorder. *Translational Psychiatry*, 12(1):104.
- Yu, H., Duan, Y., and Zhou, X. (2017). Guilt in the eyes: Eye movement and physiological evidence for guilt-induced social avoidance. *Journal of Experimental Social Psychology*, 71:128–137.
- Yu, H., Gao, X., Shen, B., Hu, Y., and Zhou, X. (2024). A levels-of-analysis framework for studying social emotions. *Nature Reviews Psychology*, 3(3):198–213.
- Yu, H., Hu, J., Hu, L., and Zhou, X. (2014). The voice of conscience: Neural bases of interpersonal guilt and compensation. *Social Cognitive and Affective Neuroscience*, 9(8):1150–1158.

Zhao, S., Cao, R., Lin, C., Wang, S., and Yu, H. (2024). Differences in the link between social trait judgment and socio-emotional experience in neurotypical and autistic individuals. *Scientific Reports*, 14(1):5400.