Title

The competition dynamics of approach and avoidance motivations following interpersonal transgression.

Permalink

https://escholarship.org/uc/item/79j741ck

Journal

Proceedings of the National Academy of Sciences of the United States of America, 120(40)

Authors

Shen, Bo
Chen, Yang
He, Zhewen
et al.

Publication Date

2023-10-03

DOI

10.1073/pnas.2302484120

Peer reviewed

PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

The competition dynamics of approach and avoidance motivations following interpersonal transgression

Bo Shen[a,b], Yang Chen[c], Zhewen He[d], Weijian Li[a], Hongbo Yu[e,1], and Xiaolin Zhou[f,1]

Edited by Nils Kolling, University of Oxford, Oxford, UK; received February 27, 2023; accepted July 26, 2023 by Editorial Board Member Anna C. (Kia) Nobre

Two behavioral motivations coexist in transgressors following an interpersonal transgression—approaching and compensating the victim and avoiding the victim. Little is known about how these motivations arise, compete, and drive transgressors' decisions. The present study adopted a social interaction task to manipulate participants' (i.e., the transgressor) responsibility for another's (i.e., the victim) monetary loss and measure the participants' tradeoff between compensating the victim and avoiding face-to-face interactions with the victim. Following each transgression, participants used a computer mouse to choose between two options differing in the amount of compensation to the victim and the probability of face-to-face contact with the victim. Results showed that as participants' responsibility increased, 1) the decision weights on contact avoidance relative to compensation increased, and 2) the onset of the contact-avoidance attribute was expedited and that of the compensation attribute was delayed. These results demonstrate how competing social motivations following transgression evolve and determine social decision-making and shed light on how social-affective state modulates the dynamics of decision-making in general.

social transgression | social-affective state | mouse-tracking | decision dynamics | multiattribute decision

Interpersonal transgression, such as breaches of social norms, moral values, or mutual respect, are ubiquitous and inevitable in social life (1, 2). In its basic form, interpersonal transgression takes place in dyads consisting of a transgressor and a victim, in which the transgressor voluntarily or accidentally commits an action (or omission) that constitutes the transgression, and the victim is the recipient of such action (or omission). Interpersonal transgressions have costly consequences to both the transgressor and the victim, in the form of both material/physical and mental/psychological losses (2). On the victim's side, in addition to material/physical losses, a transgression poses a threat to their relationship with the transgressor and more broadly their status as a valued group member (3, 4). On the transgressor's side, committing a transgression may pose a threat to their moral self (5), damage their relationship with the victim (6, 7), and expose them to blame and even punishment from the victim or society (8).

Given these costly consequences, it is then critical for the physical and mental well-being of both sides, their relationship, and the cohesion of the group and the functioning of society at large that the transgressor seeks reconciliation with the victim. Past research has identified two ways in which the transgressors react to the transgression and the victim (9–12). On the one hand, a transgressor could adopt an approach tendency, such as genuinely expressing guilt and remorse, offering apologies to and seeking forgiveness from the victim, and providing material and/or psychological compensation (13–16). Extant research in social psychology and affective science has shown that being responsible for an interpersonal transgression not only elicits guilt in the transgressor but also leads the transgressor to adopt the approach tendency (7, 16–20).

On the other hand, transgressors may adopt an avoidance tendency following a transgression because approach tendencies may come at some costs on the transgressors' side. For example, acknowledging one's own fault threatens one's belief in a moral self, which is associated with aversive feelings. Exposing oneself to the devaluations and negative reactive attitudes of the victim is embarrassing and uncomfortable (21–24). Social contact with the victim also run the risk of inviting retribution from the victim in future interactions (25). For example, a study that manipulates the responsibility of the participants (i.e., the transgressors) in causing physical harm to another (i.e., the victim) shows that the participants would avoid direct eye contact with the victim during a virtual face-to-face meeting with the victim when the participants are fully responsible for the victim's harm; making eye contact with the victim elicits stronger emotional arousal in the transgressor (26). The costs of approaching the victim may lead the transgressors to avoid the victim, such as escaping and hiding from the victim, denying the transgression, denying the full

Significance

Following an interpersonal transgression, the transgressor may approach the victim to offer compensation and seek reconciliation or shun the victim to avoid confrontation and embarrassment. How do the underlying approach and avoidance motivations compete in the transgressor's mind and determine their socially adaptive or maladaptive choice? We addressed this question by triangulating an interpersonal transgression task, mouse-tracking, and computational modeling. We showed that the transgressor's responsibility in causing the victim's harm modulated the strength and onset of the approach and avoidance motivations. These findings suggest that the transgressor's social-affective state coordinates the dynamics of the competition between opposing social motivations and point to the potential cognitive mechanisms underlying psychological barriers to reconciliation.

Author contributions: B.S., Y.C., H.Y., and X.Z. designed research; B.S. and Y.C. performed research; B.S. contributed new reagents/analytic tools; B.S. analyzed data; and B.S., Z.H., W.L., H.Y., and X.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. N.K. is a guest editor invited by the Editorial Board.

[1] To whom correspondence may be addressed. Email: hongbo.yu@psych.ucsb.edu or xz104@psy.ecnu.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2302484120/-/DCSupplemental.

Published September 28, 2023.

humanness of the victim, and even shifting blame to the victim (27–30).

While past research has investigated the approach and avoidance tendencies of transgressors following transgression separately, these two behavioral tendencies or motivations need not be mutually exclusive. Indeed, some research has demonstrated that they can coexist as two competing motivations that jointly determine how a transgressor reacts to the transgression and the victim. For example, Amodio et al. (9) observed opposite approach-avoidance motivations from the transgressors' electroencephalography activities in the early and late stages of social transgression (9), suggesting that transgression-induced approach and avoidance motivations may dynamically change over time. The temporal dynamics of approach and avoidance motivations reveal the underlying cognitive processes following social transgression. For instance, the approach motivation may emerge rapidly, reflecting an intrinsic tendency to make amend and benefit the victim. Conversely, the avoidance motivation may emerge later after the transgressor is aware of the potential negative consequences of interacting with the victim, such as embarrassment and risks of retribution. This delayed emergence could be modulated by, for example, transgressor's responsibility in causing the interpersonal harm. The goal of the present study was to investigate whether the temporal prioritization between the two tendencies adjusts in accordance with the individual's moral and social-affective states, such as the responsibility in causing interpersonal harm and the social emotions associated with it.

To address these questions, we utilize a decision-making framework to examine how the participants prioritize the processing of specific decision attributes that are, respectively, linked to the motivations of compensation and contact avoidance. We combined an established interpersonal transgression task (16, 31) with mouse tracking technique and computational modeling (32). In the task, we varied the levels of the transgressor's responsibility in causing the victim's loss to create different social-affective states in the transgressor. To probe the transgressor's approach and avoidance motivations following a transgression, we asked the participant to make two-attribute binary choices with each option containing the amount of money allocated to the victim as compensation (targeting approach motivation) and the probability of having a face-to-face contact with the victim (targeting avoidance motivation). Since mouse-tracking is a sensitive tool that provides a temporally precise measure of the dynamics of multiple decision components or attributes (for a review, see ref. 33), we tracked the full dynamics of the decision process when the participants moved their mouse from the bottom center of the screen to choose one of the options on the upper-left and upper-right corners of the screen.

Computational modeling was applied to disentangle the impact of each decision attribute on the momentary moving speed of the mouse cursor, which enabled us to pinpoint the temporal dynamics of approach and avoidance motivations independently. We expected a temporal prioritization of the processing of approach/compensation motivation relative to the processing of avoidance/distancing motivation due to the prosocial nature of guilt (34, 35). Therefore, the attribute of monetary amount would have an earlier impact than that of the attribute of contact probability on the participants' mouse movements. We also predicted that the conflict between the approach and avoidance motivations would occur in the later stage of the decision when both motivations have accumulated strength in the process. We further predicted that such a temporal prioritization would be modulated by the transgressor's social-affective states. Specifically, we predicted that the transgressor's concern about social contact increases when their responsibility in causing the victim's loss increased. As a result, the temporal processing of contact avoidance would be expedited and the processing of monetary compensation would be delayed when the concern of avoidance motivation increased relative to that of compensatory motivation.

Results

The Interpersonal Paradigm Used to Simulate Social Transgression. The task involves an interpersonal paradigm to induce different levels of guilt and embarrassment in the transgressor following an unintentional social transgression (Fig. 1*A*). On each trial, each Player A was paired with a Player B to collaborate on a visual search task. Either one failing the task would cost Player B of each pair, but not Player A, ten monetary units (M.U.). Player A's responsibility in causing Player B's monetary loss gradually increases from the condition when only Player B failed the task (partner-failed), to the condition when both players failed the task (both-failed), and then in the condition when only Player A failed the task (self-failed). The number of trials across the conditions was balanced by a control algorithm (Fig. 1 *B*-*C*), without the participants' awareness (*SI Appendix*, Fig. S2).

On the trials where Player B lost money, Player A was subsequently asked to choose between two options, each characterized by two decision attributes: the amount of monetary compensation to Player B (i.e., attribute m) and the probability of experiencing face-to-face interaction with Player B through a video camera later on that trial (i.e., attribute p) (Fig. 1*D*). The two options were marked as Option 1(with m_1 and p_1) and Option 2 (with m_2 and p_2) for subsequent data analysis; however, the participants were not aware of these labels as the left or right positions of the two options were randomized across trials. The differences in the offer values between the two options were defined as $\Delta m \equiv m_1 - m_2$ and $\Delta p \equiv p_1 - p_2$ and controlled by a self-adaptive algorithm to optimize the task efficiency in revealing the participant's preferences within limited number of trials (Fig. 1*B* and *SI Appendix*, Methods). The social context was carefully designed to target Player A's approach motivation (e.g., compensating Player B) and social avoidance motivation (e.g., avoiding being seen by Player B) with the two decision attributes, respectively (*Methods*). It is possible that Player A's monetary preference and emotional reactions to face-to-face contact with Player B differ across individuals. For example, altruistic compensation may not be desirable for all participants, as some may prefer their partner to be worse off financially (36, 37); similarly, video contact with the victim may not be aversive for all of the participants. The possible combinations of the valence of m or p [i.e., appetitive (+) or aversive (−)] result in four types of preference (i.e., $m+p-$, $m+p+$, $m-p+$, and $m-p-$), which was well accommodated in the self-adaptive algorithm (Fig. 1*C* and *SI Appendix*, *Methods*).

To make a choice, Player A needed to move the mouse cursor from the bottom center of the screen to one of the options at the upper left and right corners of the screen (Fig. 1*D*; see details in *SI Appendix*, *Methods*). We tracked Player A's mouse movement as they made their choices. Previous studies show that time pressure is important for motivating an individual to consider the importance and the prioritization of different decision attributes (38, 39). The Player A participants were asked to make their choices within varying time limits (see the distribution of time limits in the inset of Fig. 1*A*; see details in *SI Appendix*, *Methods*); otherwise, their choice would be determined by a computer algorithm. The participants were asked to rate their emotions at the end of each trial (guilt and embarrassment for Player A; anger and embarrassment for Player B, which were not reported here).

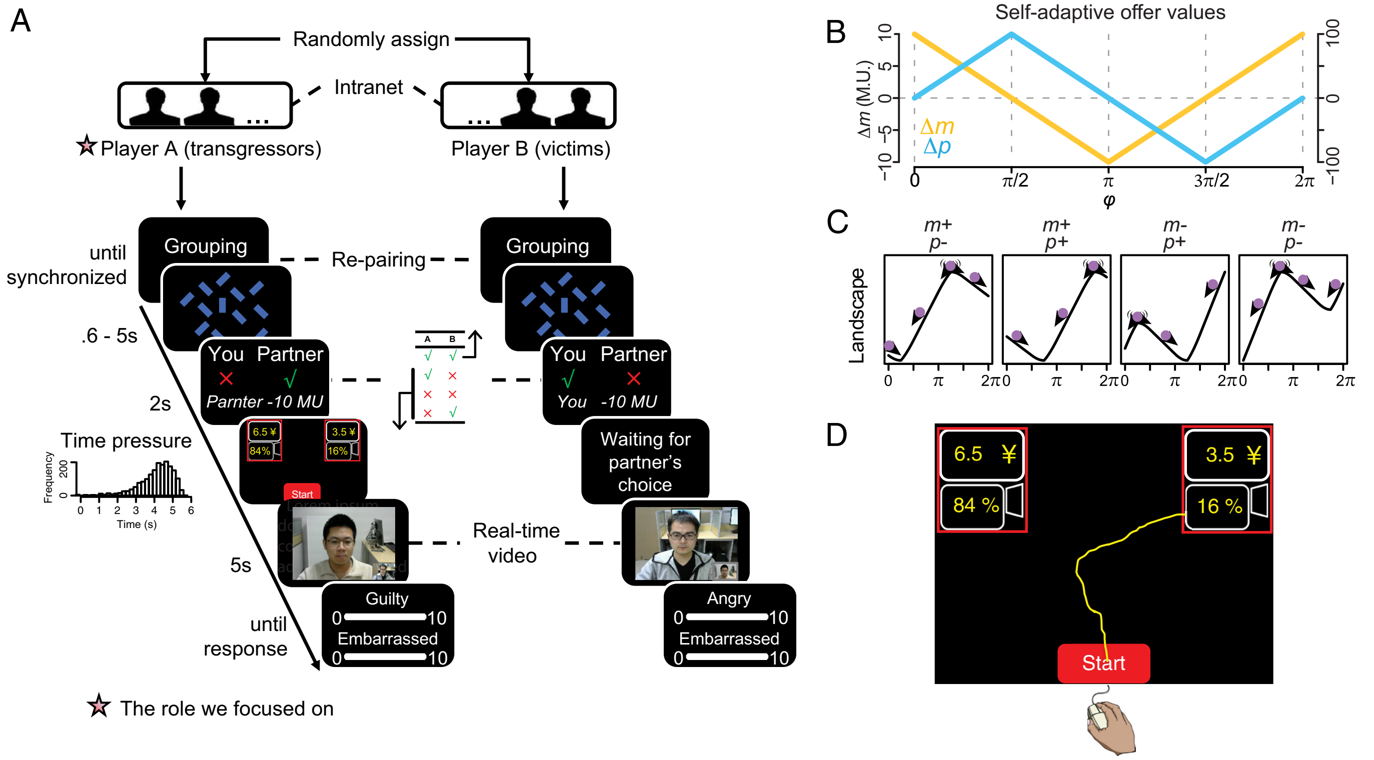

Fig. 1. Procedure and tasks. (*A*) After the random assignment of the roles in the beginning of the experiment (illustrated on the top), Player A and Player B were re-paired on each trial to go through the interactive task. The *Inset* in the middle shows the four outcomes from the cooperative visual search task, and the *Inset* on the left shows the empirical distribution of time pressure on Player A's binary choices. (*B*) Illustration of the self-adaptive algorithm to generate the values of the compensatory amount (m) and contact probabilities (p) in Option 1 (m_1, p_1) and Option 2 (m_2, p_2). The value differences between the two options ($\Delta m \equiv m_1 - m_2$ and $\Delta p \equiv p_1 - p_2$) consist of two out-of-phase triangular waves as a function of φ. (*C*) The landscape of adaptation. The self-adaptive algorithm in B led to different landscapes of adaptation for different types of preferences. The balls and arrows illustrate the offer values (corresponding to the φ value indicated on the x axis) and adaptation directions for each type of preference. The type of preferences is characterized by the perceived valence of m and p (i.e., appetitive: +; aversive: -) and is indicated on the top of each sub-panel. Each type of preference converged to a φ value that corresponds to an offer set that is optimal for revealing the information about the preference. (*D*) Example screen of Player A's choice.

Self-Reported Emotions and Final Decisions. As a manipulation check, we first examined whether the social interaction paradigm successfully induced the social emotions as we expected. To this end, we carried out repeated-measures ANOVA separately for Player A's guilt and embarrassment ratings. We found that the ratings of both guilt and embarrassment significantly differed across conditions (guilt: $F_{2,72} = 72.61$, $P < 0.001$; embarrassment: $F_{2,72} = 26.23$, $P < 0.001$). Post hoc paired t tests showed that both guilt and embarrassment ratings were significantly higher when the participants were more responsible in causing the partner's loss (guilt: self-failed > both-failed: $t_{36} = 7.94$, $P < 0.001$; both-failed > partner-failed: $t_{36} = 6.09$, $P < 0.001$; embarrassment: self-failed > both-failed: $t_{36} = 4.66$, $P < 0.001$; both-failed > partner-failed: $t_{36} = 3.65$, $P < 0.001$) (Fig. 2 *A* and *B*). These results confirmed the effectiveness of our paradigm in inducing the target emotions, suggesting that the social interaction contexts were engaging.

Further, because the emotion ratings were made after the choices about compensation and video contact, the self-reported emotions may reflect how these choices and social consequences influence participants' felt emotions. To pinpoint this potential influence, we regressed the emotion ratings against the amount of money compensated in the current trial and the occurrence of video contact (1 = video contact occurred, 0 = no video contact). In addition to the main effects of condition, as we reported above (guilt: $\beta = 2.60$, SE = 0.30, $t_{92} = 8.70$, $P < 0.001$; embarrassment: $\beta = 1.25$, SE = 0.25, $t_{142} = 5.03$, $P < 0.001$), the linear mixed regressions revealed a double dissociation with respect to the impacts of monetary compensation and video contact on guilt and embarrassment ratings of Player A (conditions were coded as ordinal variables in the regressions, with partner-failed = 1, both-failed = 2, and self-failed = 3). For guilt, the main effect of the monetary compensation was not significant ($\beta = 0.031$, SE = 0.073, $t_{662} = 0.42$, $P = 0.68$), whereas a significant interaction was found between the amount of money and the experimental conditions ($\beta = -0.065$, SE = 0.030, $t_{2076} = -2.13$, $P = 0.034$). Participants' guilt ratings significantly decreased as the amount of monetary compensation increased in self-failed and both-failed conditions but not so in the partner-failed condition (self-failed: $\beta = -0.17$, SE = 0.054, $t_{58} = -3.12$, $P = 0.003$; both-failed: $\beta = -0.12$, SE = 0.048, $t_{42} = -2.48$, $P = 0.017$; partner-failed: $\beta = -0.031$, SE = 0.031, $t_{82} = -1.00$, $P = 0.32$). Guilt ratings were not significantly modulated by the occurrence of video contact (main effect: $\beta = 0.72$, SE = 0.65, $t_{1841} = 1.12$, $P = 0.26$; interaction with conditions: $\beta = -0.056$, SE = 0.280, $t_{2070} = -0.20$, $P = 0.84$) (Fig. 2*A*). In contrast, ratings of embarrassment were significantly higher after video contact (main effect: $\beta = 1.99$, SE = 0.65, $t_{1246} = 3.07$, $P = 0.002$; interaction with conditions: $\beta = -0.32$, SE = 0.27, $t_{2093} = -1.19$, $P = 0.23$), but no such association was found between embarrassment ratings and the amount of monetary compensation (main effect: $\beta = 082$, SE = 0.07, $t_{684} = 1.17$, $P = 0.24$; interaction with conditions: $\beta = -0.041$, SE = 0.029, $t_{2084} = -1.39$, $P = 0.16$). These results demonstrated associations between the two emotions and the subsequent behaviors: the transgressor's feeling of guilt was alleviated after more money was compensated to the victim, while the feeling of embarrassment was intensified after the passive social contact. However, it is important to note that in our current design, self-reported emotions were measured only as psychological consequences of social

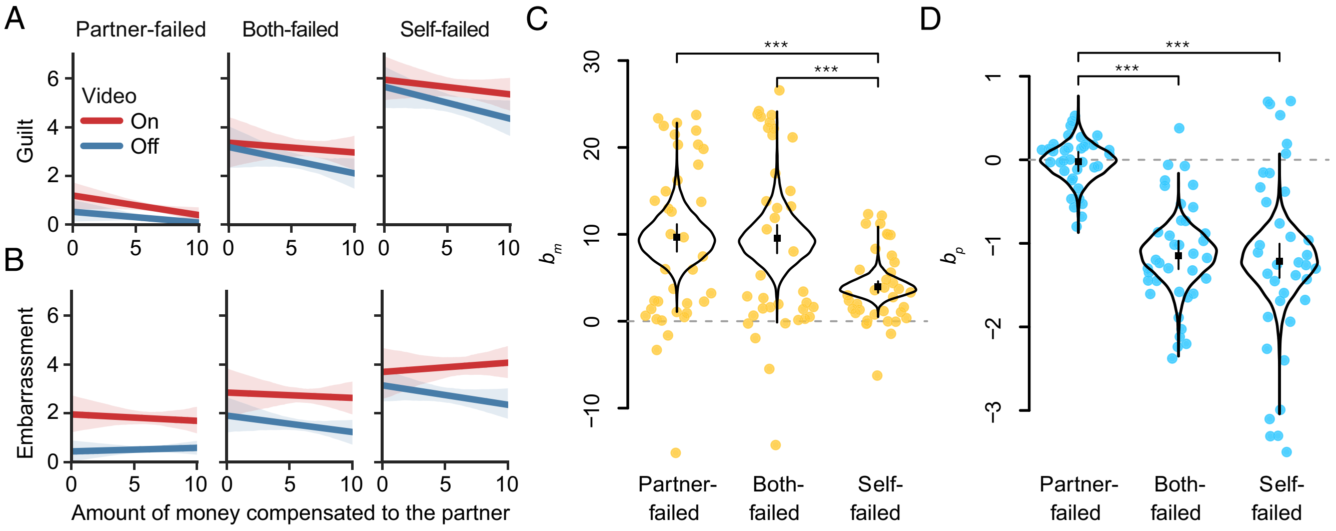

Fig. 2. Transgression-induced emotions and choice behavior. (*A* and *B*) Participants' (Players A's) self-reported guilt and embarrassment at the end of each trial. (*C* and *D*) The regression decision weights on monetary compensation (b_m, golden) and video contact probability (b_p, blue) estimated from the participants' binary choices. Colored dots indicate the random effects; black lines, squared dots, and whiskers indicate, respectively, the distributions, mean, and interquartile of the fixed effects.

events (e.g., transgression, compensation, and video contact). We acknowledge that in complex, life-like social interactions such as our current task, it is unlikely that behavioral tendencies (e.g., compensation and contact avoidance) are solely driven by any one discrete emotion. Instead, it is possible that multiple cognitive and affective processes jointly drive a behavioral tendency.

We hypothesized that as participants' perceived responsibility for interpersonal harm and their emotional reactions intensified, they would become more concerned about the embarrassing encounter during the video contact and/or a heightened probability of being frowned upon. This would consequently shift their decision-making process, resulting in a rebalanced trade-off between compensation and avoidance in their choices. To test this, we examined how the participants weighed the attributes of monetary compensation (i.e., Δm) and the probability of video contact (i.e., Δp) in their decisions. Positive or negative decision weights indicate that the corresponding attribute is appetitive or aversive to the participants, respectively. Hierarchical logistic regressions on participants' binary choices (Option 1 was coded as 1, Option 2 was coded as 0) showed that the fixed effects of participants' decision weights on monetary compensation (b_m) were significantly positive across all conditions, indicating a consistent preference for benefiting the victim (Fig. 2*C*, envelope lines surrounding the golden dots indicating the fixed effects of regressions; partner-failed: mean = 9.69, 95% CI = [5.49, 14.86]; both-failed: mean = 9.57, 95% CI = [5.04, 15.31]; self-failed: mean = 3.98, 95% CI = [2.20, 6.19]). On the contrary, the fixed effects of decision weights on video contact probability (b_p) were significantly negative in the both-failed and the self-failed conditions but not significantly different from 0 in the partner-failed condition. This pattern indicated that video contact with the victim was aversive to the participants when the participants were fully or partially responsible for the victim's loss and had no impact on the participants' choice when the participants were not responsible for the victim's loss (Fig. 2*D*, envelope lines surrounding the blue dots indicating the fixed effects of regressions; partner-failed: mean = −0.02, 95% CI = [−0.38, 0.33]; both-failed: mean = −1.15, 95% CI = [−1.71, −0.69]; self-failed: mean = −1.22, 95% CI = [−1.87, −0.64]). These results suggest that the transgressor shows both compensation-seeking and contact-avoidance tendencies after social transgression, confirming the hypothesis that the social contact avoidance motivation coexists with the compensation motivation in the social transgressor.

More interestingly, when we compared the difference in the decision weights across conditions (characterized by the estimated random effects; the golden and blue dots in Fig. 2 *C* and *D*), we found that b_m was significantly lower in the self-failed condition than in the partner-failed and the both-failed conditions (self-failed < partner-failed: t_{36} = −4.06, Cohen's D = −0.71, P < 0.001; self-failed < both-failed: t_{36} = −4.09, Cohen's D = −0.66, P < 0.001; no significant difference between partner-failed and both-failed: t_{36} = 0.01, Cohen's D = 0.00, P = 0.99). In contrast, b_p was more negative in the both-failed and self-failed conditions than in the partner-failed condition (both-failed < partner-failed: t_{36} = −11.31, Cohen's D = −2.21, P < 0.001; self-failed < partner-failed: t_{36} = −6.55, Cohen's D = −1.44, P < 0.001; no significant difference between self-failed and both-failed: t_{36} = −0.44, Cohen's D = −0.07, P = 0.66). This pattern suggests that the decision weight on benefiting the victim (i.e., monetary compensation) wanes as the concern about social contact increases (Fig. 2*C*). Since the collinearity between the estimated b_m and b_p was low (*SI Appendix*, Fig. S7), the trade-offs between the two decision weights reflect a genuine competition between the two motivations within the participants instead of an artifact introduced during parameter estimation. This reveals a unique aspect of the transgressors' social responses following a transgression. Previous studies of guilt and compensation have consistently shown that transgressors' compensation or willingness to compensate increases as their responsibility in causing the victim's harm increases (6, 16, 40). However, unlike the present study and everyday social interactions, none of these previous studies has a life-like social contact component in their experimental design. Therefore, they are not able to examine how the motives for benefiting the victim and avoiding social contact with the victim interact to determine the transgressor's response. To further investigate how the two motives evolve and interact over time, we examined the temporal dynamics of the processing of these two attributes based on the participants' mouse movement trajectories.

Asynchronous Processing of Compensation and Contact Avoidance. The choice trajectories provide detailed information about how the cognitive processes underlying decision-making unfold over time before the final decision is reached. The horizontal component of mouse movement velocity (v_h) reflects a participant's momentary preference for an option relative to the other (Fig. 3*A*). Our validation analysis confirmed that v_h toward a higher value option parametrically tracked the value difference between the two options (*SI Appendix*, Fig. S4). Thus, we regressed the attribute values (i.e., Δm and Δp) on v_h over 100 time bins equally distributed over the reaction time (RT) of each trial. The goal of this analysis is to reveal the dynamics of momentary decision weights on monetary compensation ($b_{m,t}$) and video contact probability ($b_{p,t}$) during the decision-making process. Hierarchical linear regression showed that $b_{m,t}$ ramped up to be significantly positive in the late stage of decision (around the 60th time bin) across all conditions (golden lines in Fig. 3*B*; asterisks above the shadow of 95% CI indicated significance after multicomparison correction described in *SI Appendix, Methods*), suggesting that the participants' preferences for monetary compensation gradually built up during the courses of decisions across all conditions. In contrast, $b_{p,t}$ drifted down to be significantly negative in the late stage of decision only in the both-failed and self-failed conditions but not in the partner-failed condition (blue lines in Fig. 3*B*), indicating that contact avoidance emerged to be a behaviorally relevant concern only in the both-failed and self-failed conditions. The general pattern of dynamics was consistent with the decision weights from binary choices.

To quantify the rising time of the dynamics, we calculated the earliest significant time point (ESTP) of each time series by bootstrapping (*SI Appendix, Methods*). ESTP indicates the earliest time point on a time series of the regression parameters that reaches statistical significance after multiple-comparison correction. Fig. 3*D* illustrates the distributions of ESTP for $b_{m,t}$ (golden) and $b_{p,t}$ (blue) across conditions. These distributions showed that the ESTP of $b_{m,t}$ was earlier than $b_{p,t}$ in the both-failed condition but show no significant difference from $b_{p,t}$ in the self-failed condition (both-failed condition: ESTP of $b_{m,t}$ = 69.02 ± 6.06, ESTP of $b_{p,t}$ = 89.27 ± 6.94, difference = −20.26 ± 9.20, 95% CI = [−30, −11]; self-failed condition: ESTP of $b_{m,t}$ = 73.17 ± 6.10, ESTP of $b_{p,t}$ = 75.96 ± 5.04, difference = −2.79 ± 7.98, 95% CI = [−15, 7]; partner-failed condition: ESTP of $b_{m,t}$ = 56.22 ± 14.38, ESTP of $b_{p,t}$ in the partner-failed condition was not compared because the $b_{p,t}$ dynamics did not survive multiple-comparison correction). These results suggest that the processing of monetary compensation arose earlier than that of contact avoidance in the both-failed condition but such a temporal advantage was diminished in the self-failed condition.

We further compared these dynamics across conditions and observed the shifting of dynamics across conditions for both attributes. The ESTP of $b_{m,t}$ was gradually delayed from the partner-failed condition to the both-failed condition then to the self-failed condition (Table 1). The delay from the partner-failed condition to the self-failed condition was significant (difference = 16.95 ± 15.84, 95% CI (one-tailed) = [3, 61]). In contrast, the

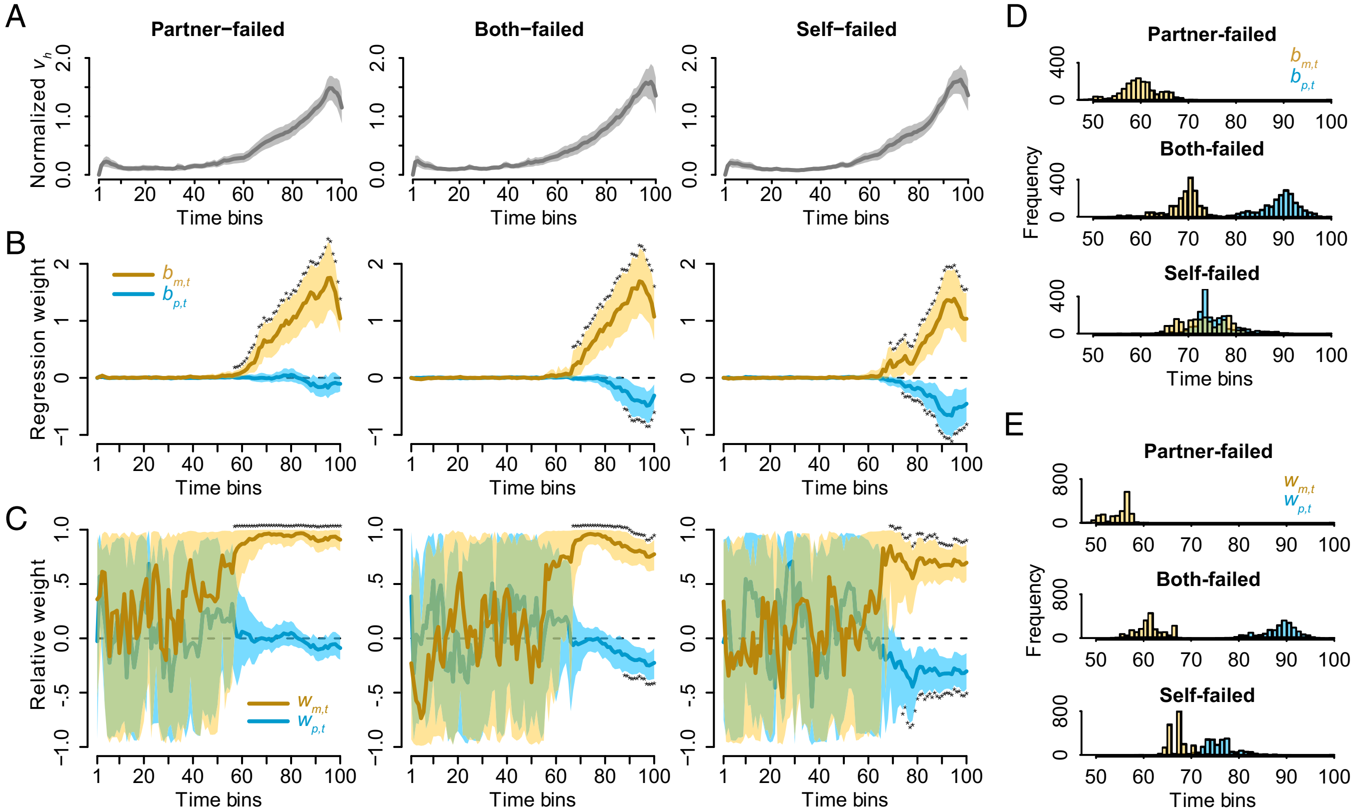

Fig. 3. The temporal dynamics of decision weights and the distributions of their earliest significant time point (ESTP). (*A*) The temporal dynamics of averaged velocity (v_h) over time bins. Shadows indicate 95% CI of v_h. (*B*) The temporal dynamics of regression weights on monetary compensation ($b_{m,t}$, golden) and contact probability ($b_{p,t}$, blue) over the three conditions. Lines and shadows indicate the mean value and 95% CI of the fixed effects, respectively. Asterisks indicate statistical significance after multicomparison correction. (*C*) The temporal dynamics of relative regression weights ($w_{m,t}$, golden and $w_{p,t}$, blue) after controlling for the magnitude of moving velocity. (*D*) The ESTP distributions of regression weights ($b_{m,t}$, golden and $b_{p,t}$, blue). (*E*) The ESTP distributions of $w_{m,t}$ and $w_{p,t}$ across three conditions reached the same conclusions as from *D*.

Table 1. Statistics on the ESTPs of dynamic decision weights

	Variable	Condition	Mean ± SD	95% CI
Time bins	$b_{m,t} - b_{p,t}$	Self	−2.79 ± 7.98	[−15, 7]
		Both	−20.26 ± 9.20	[−30, −11]
	$b_{m,t}$	Self - Both	4.16 ± 8.68	[−5, 16]
		Both - Partner	12.79 ± 15.73	[−1, 54]
		Self - Partner	16.95 ± 15.84	[3, 61]
	$b_{p,t}$	Self - Both	−13.31 ± 8.61	[−22, −2]
	$w_{1,t} - w_{2,t}$	Self	−9.08 ± 6.44	[−17, −3]
		Both	−27.08 ± 8.44	[−34, -19]
	$w_{1,t}$	Self - Both	5.79 ± 6.53	[−1, 12]
		Both - Partner	14.80 ± 19.01	[1, 60]
		Self - Partner	20.60 ± 18.75	[8, 66]
	$w_{2,t}$	Self - Both	−12.21 ± 8.43	[−20, −3]
Reaction times	$b_{m,t}$	Self - Both	0.09 ± 0.15 s	[−0.08, 0.30]
		Both - Partner	0.24 ± 0.28 s	[−0.01, 0.97]
		Self - Partner	0.33 ± 0.28 s	[0.08, 1.12]
	$b_{p,t}$	Self - Both	−0.22 ± 0.16 s	[−0.38, −0.02]
	$w_{1,t}$	Self - Both	0.12 ± 0.12 s	[−0.00, 0.23]
		Both - Partner	0.27 ± 0.34 s	[0.03, 1.08]
		Self - Partner	0.39 ± 0.33 s	[0.17, 1.20]
	$w_{2,t}$	Self - Both	−0.20 ± 0.15 s	[−0.34, −0.04]

ESTP of $b_{p,t}$ was significantly expedited in the self-failed condition than in the both-failed condition (difference = −13.31 ± 8.61, 95% CI = [−22, −2]). Because the time scales were normalized within each condition, comparing ESTPs across conditions might be affected by the differences in reaction time across conditions. To address this concern, we made corrections and rescaled the ESTPs according to the mean reaction time of each condition. The conclusions still held with the rescaled ESTPs (see Table 1 for a summary of statistics on the ESTPs and the corrected ESTPs across conditions).

Since $b_{m,t}$ and $b_{p,t}$ were regression weights to v_h, they carried not only the information about the participants' preferences but also the mechanical features of mouse movement. Both of the $b_{m,t}$ and $b_{p,t}$ dynamics ramped up to a peak value around the 95th time bin and were followed by a quick drop at the end of every time course (Fig. 3*B*), which were aligned with the dynamics of average v_h (Fig. 3*A*). This was due to the mechanical nature of the movement speed, which accelerated from the beginning of each trial to a peak and decelerated at the end of each trial. To rule out the potential bias that the movement speed may introduce to the ESTP distributions, we examined the dynamics of relative decision weights $w_{m,t}$ and $w_{p,t}$. Specifically, we obtain $w_{m,t}$ and $w_{p,t}$ by normalizing $b_{m,t}$ and $b_{p,t}$ with their summed magnitudes, i.e., $w_{m,t} = \dfrac{b_{m,t}}{|b_{m,t}| + |b_{p,t}|}$ and $w_{p,t} = \dfrac{b_{p,t}}{|b_{m,t}| + |b_{p,t}|}$ (Fig. 3*C*). $w_{m,t}$ and $w_{p,t}$ indicated the decision weights of one attribute relative to the other and, therefore, were free of the motor-related features on the time courses. The ESTPs bootstrapped from $w_{m,t}$ and $w_{p,t}$ resulted in the same conclusions as we obtained from $b_{m,t}$ and $b_{p,t}$. The ESTPs of $w_{m,t}$ were earlier than that of $w_{p,t}$ in the both-failed and self-failed conditions (both-failed condition: ESTP of $w_{m,t}$ = 61.42 ± 5.06, ESTP of $w_{p,t}$ = 88.50 ± 6.73, difference = −27.08 ± 8.44, 95% CI = [−34, −19]; self-failed condition: ESTP of $w_{m,t}$ = 67.22 ± 4.09, ESTP of $w_{p,t}$ = 76.30 ± 4.94, difference = −9.08 ± 6.44, 95% CI = [−17, −3]); the temporal advantage of $w_{m,t}$ relative to $w_{p,t}$ in the self-failed condition was significantly smaller than that in the both-failed condition (the 95% CIs above showed no overlapping) (Fig. 3*E* and Table 1). Comparing across conditions, the ESTPs of $w_{m,t}$ were delayed from partner-failed condition to both-failed condition and self-failed condition, and the ESTP of $w_{p,t}$ were expedited from both-failed condition to self-failed condition (ESTP of $w_{m,t}$, both-failed – partner-failed: difference = 14.80 ± 19.01, 95% CI = [1, 60]; self-failed – partner-failed: difference = 20.60 ± 18.75, 95% CI = [8, 66]; ESTP of $w_{p,t}$, self-failed – both-failed: difference = −12.21 ± 8.43, 95% CI = [−20, −3]). The conclusions were held after the corrections of the mean reaction time of each condition (results listed in Table 1).

Taken together, these results provide evidence that the processing of monetary compensation and contact avoidance is asynchronous. The participants' processing on monetary compensation was prioritized over contact avoidance across all conditions. However, as participants' motivation for contact avoidance relative to monetary compensation increased, the processing of contact avoidance was expedited and the processing of monetary compensation was delayed, suggesting that the processing adaptively prioritizes these attributes to match the participants' changing needs under different conditions.

Neural Circuit Model Predicts the Curvatures of the Competition Dynamics. An alternative interpretation of the above ESTPs results is that the temporal difference between *m* and *p* solely reflects the differences in the magnitude of decision weights instead of asynchronous processing of the attributes. In other words, the decision weight with a larger magnitude could be detected earlier in time, appearing as the time asynchrony we observed in the ESTP distributions. To address this concern, we applied a neural circuit model of multiattribute decision-making (41) to elucidate the impacts of magnitude and temporal asynchrony on the dynamics of the decision weights.

In this model, we assume that the two decision attributes, m and p, are integrated independently in the neural nodes of $\int M$ and $\int P$ (Fig. 4*A*). The inputs to the neural nodes are the difference in the values between Option 1 and Option 2 (i.e., Δm and Δp) and the outputs are the integrated decision signals (i.e., M_t and P_t). The two integrated signals then additively contribute to the final decision value (DV_t) and guide behavioral responses. The relative decision weights $w_{m,t}$ and $w_{p,t}$ are defined as the relative contributions of the integrated attribute values, i.e., M_t and P_t, to DV_t. At the end of the process, a positive DV_t leads to the choice of Option 1 and a negative DV_t leads to the choice of Option 2 (for details of the model, see *SI Appendix, Methods*). The temporal parameters t_m and t_p capture the time onsets when the attributes m and p start integrating, respectively. We tested the input magnitudes of Δm and Δp and the temporal asynchrony between t_m and t_p to the predicted dynamics of $w_{m,t}$ and $w_{p,t}$.

Fig. 4 *B–D* illustrate the temporal dynamics of the input values, Δm and Δp. To test the impact of magnitude, we varied the magnitude of Δp (indicated by the shade of blue) and fixed the magnitude of Δm. To test the impact of onset, we varied the onset of Δp (i.e., t_p) so that the process on the attribute of p has no delay (i.e., synchronous, Fig. 4*B*), a large delay (Fig. 4*C*), or medium delay (Fig. 4*D*) relative to the onset of Δm (i.e., t_m). Fig. 4 *E–G* show the integrated signals M_t and P_t from the corresponding situations. Both M_t and P_t increase after their onsets. Although the magnitude of Δp only affects the accumulation process of P_t, it affects the dynamics of both relative weights $w_{m,t}$ and $w_{p,t}$ since the relative weights are dependent on the ratio of M_t's and P_t's contributions to DV_t. Therefore, $w_{m,t}$ decreases when $w_{p,t}$ increases (Fig. 4 *H–J*). Importantly, time asynchrony impacts the fine-grained details of the dynamics. When the integration of the attribute p is delayed, the dynamics of $w_{m,t}$ shows nonmonotonic curvature in the late suppression when $w_{p,t}$ rises (Fig. 4 *I* and *J*); in contrast, when the two attributes are integrated synchronously, the dynamics of $w_{m,t}$ does not show such a late suppression even if the impact on its magnitude still exists (Fig. 4*H*). The curvatures of the late suppressions are different under different levels of time asynchrony. When p has a large delay relative to m, the dynamics of $w_{m,t}$ shows a quadratic curvature in the late suppression (Fig. 4*I*); when p has a medium delay relative to m, the dynamics of $w_{m,t}$ shows a kink of decrease right after $w_{p,t}$ rose (Fig. 4*J*).

To compare the prediction of the neural circuit model with the empirical dynamics, we fitted the neural circuit model to the dynamics of relative weights ($w_{m,t}$ and $w_{p,t}$) with two competing hypotheses, namely whether the processing of the two attributes is asynchronous or synchronous (*SI Appendix, Methods*). Specifically, we took the similar approach as described above, fixing the magnitude of Δm at 1 and fitting the magnitude of Δp to capture the participants' weight on contact avoidance relative to compensation. The results show that the asynchronous circuit with one additional parameter in each condition (i.e., t_m and t_p estimated indepe-ndently) outperformed the synchronous circuit (i.e., t_m and t_p estimated as an equal value) in all three conditions ($LL_{Asynchronous}$ = −989.8, $LL_{Synchronous}$ = −1,072.6; $AIC_{Asynchronous}$ = 2,001.7, $AIC_{Synchronous}$ = 2,161.1; $BIC_{Asynchronous}$ = 2,042.1, $BIC_{Synchronous}$ = 2,190.5). The model performance in each condition was visualized in Fig. 4*K* (asynchronous circuit) and Fig. 4*L* (synchronous circuit). Specifically, in the partner-failed condition, where the weight on contact avoidance was not significant, the two models perform similarly (Partner-failed: left panels in Fig. 4 *K* and *L*, $LL_{Asynchronous}$ = −316.4, $LL_{Synchronous}$ = −331.3). However, in the both-failed and self-failed conditions when the contact avoidance is a more serious concern, the asynchronous circuit tracks much closely on the empirical dynamics (Both-failed: middle panels in Fig. 4 *K* and *L*, $LL_{Asynchronous}$ = −313.8, $LL_{Synchronous}$ = −373.5, self-failed: right panels in Fig. 4 *K* and *L*, $LL_{Asynchronous}$ = −359.6, $LL_{Synchronous}$ = −367.7). Zooming in on these dynamics, we found that the asynchronous circuit captured the detailed pattern exceptionally well, particularly the nonmonotonic dynamics on $w_{m,t}$ (dashed lines tracking golden lines in Fig. 4*K*). In both the both-failed and self-failed conditions, the $w_{m,t}$ curves exhibited subsequent suppressions following the rise of $w_{p,t}$ (around time bin 77 in both-failed and time bin 62 in self-failed), due to the asynchronous/delayed time onsets of p relative to m. In contrast, the nonmonotonic dynamics was not observed in the synchronous circuit (dashed lines in Fig. 4*L*). In addition, the curvatures of the empirical dynamics were consistent with the model predictions. The empirical dynamics of $w_{m,t}$ showed quadratic curvature in the both-failed condition but demonstrated a sudden drop in the self-failed condition, since a larger delay of processing p relative to m in the both-failed condition but a smaller delay of processing p relative to m in the self-failed condition.

The time onsets estimated from the asynchronous circuit were consistent with the ESTPs results reported above, with the onsets of contact avoidance expedited and the onsets of compensation gradually delayed from partner-failed to self-failed condition (Partner-failed: $t_m = 44.2$, $t_p = 80.1$; both-failed: $t_m = 53.8$, $t_p = 77.3$; self-failed: $t_m = 58.8$, $t_p = 62.3$). Interestingly, the magnitudes of Δp estimated from the asynchronous circuit were comparable between the both-failed condition and the self-failed condition (Partner-failed: −0.20, both-failed: −0.50, and self-failed: −0.51), suggesting that the intensified final relative decision weight on contact avoidance in the self-failed condition might due to the time prioritization of this attribute. Other parameters estimated from the two models were reported in *SI Appendix, Methods*.

Taken together, the neural circuit model analysis teased apart the contributions of attribute magnitudes and temporal asynchrony to the dynamics of relative decision weights, suggesting that magnitude alone was not enough to result in the nonmonotonic curvatures observed in our empirical data. The modeling results suggest an asynchronous rather than synchronous processing on the two attributes and corroborate the findings based on the analysis of the ESTP distributions, namely, the dynamics of contact avoidance arise later than that of the monetary compensation. Moreover, the time onset for contact avoidance was expedited and the time onset for compensation was delayed from both-failed condition to self-failed condition.

Time-Varying DDM Reveals the Shifting of Asynchronous Processing over Conditions. To further dissociate the effects of condition on the onsets of decision attributes and the effects on the magnitude of the weights of the decision attributes, we applied a drift-diffusion model (DDM) to capture the onset of accumulation on each decision attribute while controlling for any difference induced by the drift rates of the attributes. To this end, we fit a time-varying drift diffusion model (tDDM), with a varying onset of accumulation on each decision attribute (Fig. 5*A*; cf. ref. 42). The model was applied to the reaction time and choice data but not to the mouse trajectories. The attributes of money (golden lines) and contact probability (blue lines) were accumulated with independent drift rates and summed into an integrated decision

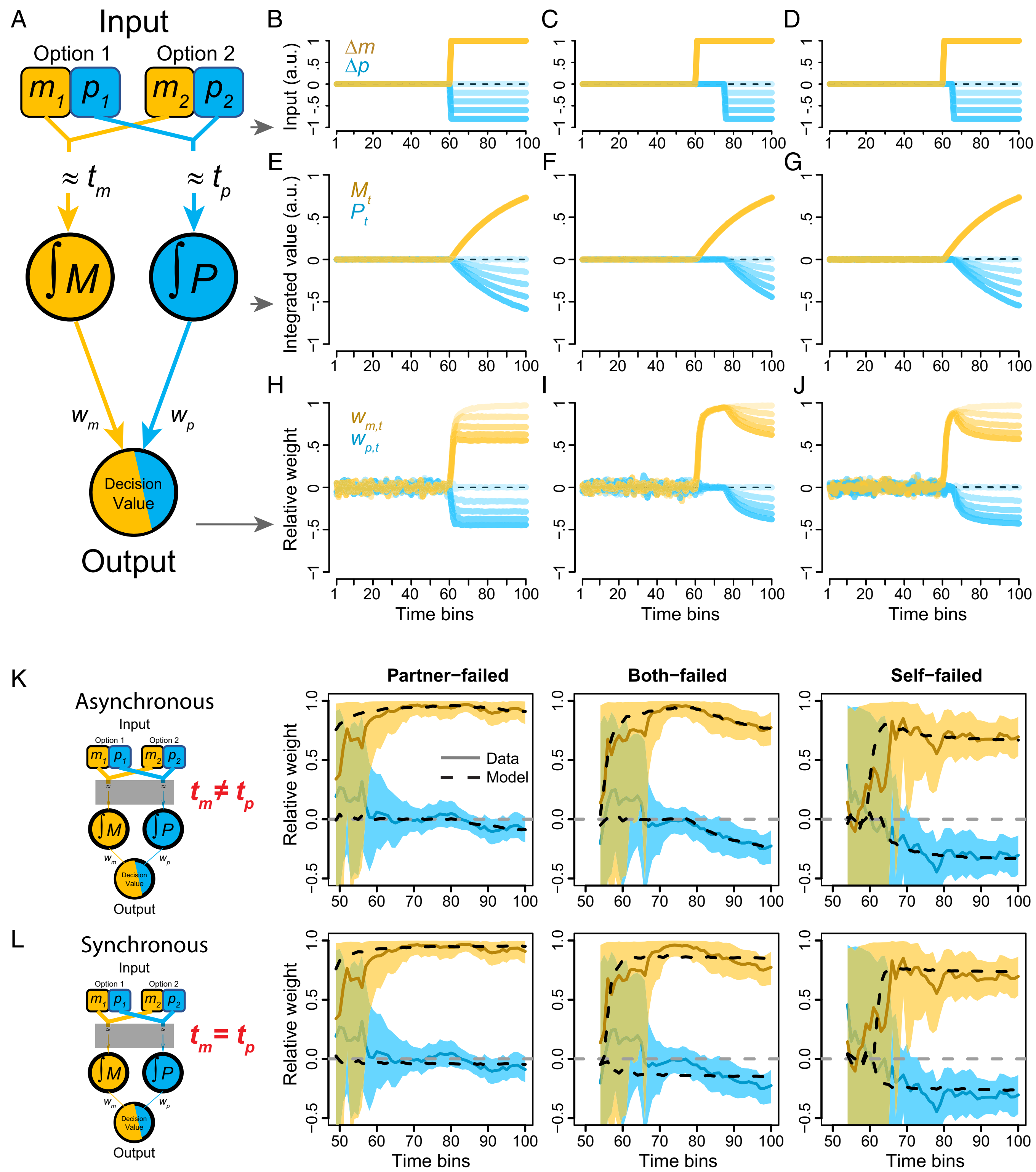

Fig. 4. The specification and results of the neural circuit model. (*A*) The neural circuit model gets input values of the two attributes *m* (golden) and *p* (blue) from Option 1 and Option 2. The values on different attributes are integrated into their corresponding neural nodes $\int M$ (golden) and $\int P$ (blue) with specific temporal delays t_m and t_p. The integrated values linearly contribute to the final decision node (decision value) to guide choice behavior, with their relative contribution notated as w_m and w_p. (*B*–*D*) The evolving of input value differences of each attribute $\Delta m \equiv m_1 - m_2$ (golden) and $\Delta p \equiv p_1 - p_2$ (blue) over time. Across the three panels, the magnitudes of Δp were varied and the magnitude of Δm was fixed; the onset of Δp varied from no delay (synchronous) (*B*), to having a large delay (*C*), and medium delay (*D*) relative to the onset of Δm. (*E*–*G*). The accumulated signals in *M* and *P* over time corresponding to the input settings in *B*–*D*. (*H*–*J*) The dynamics of relative decision weights $w_{m,t}$ (golden) and $w_{p,t}$ (blue) corresponding to the integrated signals in *E*–*G*. (*K*) By fitting the circuit with the asynchronous assumption, the circuit (dashed lines) captures well the rich and nonmonotonic dynamics in the empirical data (golden and blue lines indicating $w_{m,t}$ and $w_{p,t}$, respectively, with shadows indicating the 95% CI). (*L*) The circuit model with the synchronous assumption captures worse the empirical dynamics than the asynchronous circuit.

signal (black line). Different from the implementation of most sequential sampling models, which do not assume temporal priorities on different decision attributes (43–45), the tDDM assumes that the temporal onsets of the money attribute and the contact probability attribute can be asynchronous; it therefore includes a time asynchrony parameter (*tAsync*) to capture the relative onset time difference between these two attributes. Negative *tAsync* indicates the onset of money is later than that of contact probability (illustrated in the two left panels in Fig. 5*A*), and positive *tAsync* indicates the onset of money is earlier than that of contact probability (right panels in Fig. 5*A*). Model comparison between the tDDM and a standard DDM without time asynchrony (thus reducing one parameter) showed that tDDM fitted the data better than DDM ($LL_{tDDM} = -13{,}235.8$, $LL_{DDM} = -13{,}415.2$; $AIC_{tDDM} = 27{,}803.5$, $AIC_{DDM} = 27{,}940.4$) (see posterior predictive check on model fitting in *SI Appendix*, Fig. S8). Results from tDDM showed that *tAsync* significantly differed across conditions (ANOVA: $F_{2,72} = 3.32$, $P = 0.042$). Post hoc test showed that *tAsync* in the self-failed condition was more negative than in the both-failed and the partner-failed conditions (self-failed—both-failed: $t_{36} = -2.10$, $P = 0.043$, Cohen's D = −0.45; self-failed—partner-failed: $t_{36} = -2.19$, $P = 0.035$, Cohen's D = −0.38; both-failed—partner-failed: $t_{36} = 0.36$, $P = 0.72$, Cohen's D = 0.06) (Fig. 5*B*), indicating that the onset of processing on contact avoidance relative to monetary compensation was expedited in the self-failed condition than in the other two conditions. Other parameters, including the drift rates, showed no significant differences across conditions (decision threshold: $F_{2,72} = 0.09$, $P = 0.91$; bias: $F_{2,72} = 1.11$, $P = 0.33$; nondecision time: $F_{2,72} = 0.66$, $P = 0.52$; drift rate of money: $F_{2,72} = 0.24$, $P = 0.79$; drift rate of contact probability: $F_{2,72} = 1.45$, $P = 0.24$). This result further confirmed our findings that the processing of money and contact probability were asynchronous and the prioritization of the attributes varied as a function of the social interaction context. In addition, the fact that there was no difference in drift rate across conditions indicates that the difference in the final decision weights could be solely induced by the difference in the onset of attribute accumulation processes while the accumulation speed of each attribute was comparable.

Discussion

The present study investigated the temporal dynamics of approach-avoidance behavioral conflict in social transgressors after they unintentionally harm a victim. By adopting a social transgression task, we induced different levels of guilt-related social-affective states in the participants and examined how they made trade-offs between compensating the victim (i.e., the approach motivation) and avoiding face-to-face interaction with the victim (i.e., the avoidance motivation). Participants exhibited intensified competition between the motivations as guilt and embarrassment levels increased. By analyzing the participants' mouse movement trajectories during decision-making, we revealed that as participants' responsibility increased, the weight on contact avoidance increased and the onset of its dynamics was expediated; at the same time, the weight on monetary compensation decreased and the onset of its dynamics was delayed. Simulation from a neural circuit model and data fitting using time-varying DDM confirmed that the processing of the two attributes is asynchronous and temporally modulated by social-affective states.

The paradigm and the results of this study highlight the richness and complexity of the emotional and motivational states following interpersonal transgression. Traditional views in social psychology and affective science portrait a rather "sunny" or positive picture of how people react (or ought to react) to interpersonal transgression—as soon as they realize that the transgression has occurred, the transgressor would halt the damage, feel guilty, and make amends or apologize to the victim (6, 7). Part of the reason why most previous empirical evidence supports such a sunny view is that those studies predominantly examine what the participants would feel and how they may react in hypothetical scenarios, rather than what they actually feel and how they react in real-time social interactions. However, what one think they should do and what one actually do in social contexts differs in important ways, and they are associated with distinct neurocognitive processes (46). One noticeable exception to the traditional views of guilt and transgression is Amodio et al. (9) who demonstrated that both an avoidance tendency (i.e., refraining from damaging) and an approach tendency (i.e., making compensation) coexist in transgressors following an interpersonal transgression and that these tendencies have distinct temporal dynamics (9). However, even in this study, a critical aspect of real-world interpersonal transgressions was missing, namely face-to-face interaction between the transgressor and the victim. Both anecdotal and empirical evidence has shown that in a guilt state, the transgressor has a tendency to avoid or hide from confrontation with the victim (12, 26, 47). By adopting a measure of motivation with high temporal resolution

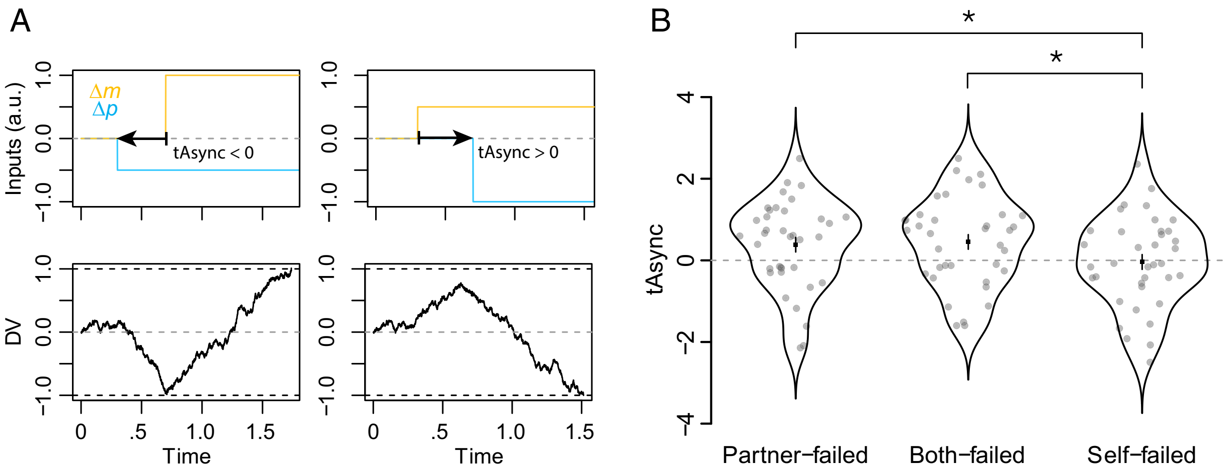

Fig. 5. The tDDM results. (*A*) Illustration of the tDDM hypothesis. The *Upper* panels illustrate the time onset of monetary input (Δ*m*, golden) and probability input (Δ*p*, blue). The *Bottom* panels illustrate the integrated decision signals (*DV*) over time. (*B*) The fitted *tAsync* in the three critical conditions. Gray dots and envelopes indicate the fitted values of individuals and the group-level distributions. Squared dots and whiskers indicate mean values and SEs in each condition.

(i.e., mouse-tracking), we are able to demonstrate that the two opposite tendencies do coexist in real-life interactions, and that the increase in the strength of one (i.e., avoiding confrontation) suppresses the manifestation of the other (i.e., compensating the victim). The current findings have implications for our understanding of the complex, and sometimes dialectic, nature of the emotional and motivational reactions in interpersonal transgression and other social interactive contexts, e.g., the coexistence of gratitude and indebtedness/guilt in helping contexts (48).

Our results have general implications about the role of motivation in orchestrating the temporal dynamics of decision-making. Recent theoretical work in decision-making emphasizes the role of metacognitive processes in the real-time control of decision-making process, such as guiding attention to optimize information seeking (38, 49, 50) and exerting cognitive efforts to achieve better task performance (51–53). Prior research has demonstrated the critical role of motivation in distributing cognitive efforts in the processing of single decision attributes (51, 53–55). However, empirical evidence regarding how motivation impacts the temporal dynamics of processing on multiple decision attributes has been lacking. The current study fills this gap by directly manipulating the participant's motivational states in a social transgression context and examining the temporal dynamics of the processing of two decision attributes related to the participant's internal needs (i.e., monetary compensation to the victim and avoiding face-to-face interaction with the victim). Tracking the participant's motor responses as the end process of decision execution may reflect an integrated effect of motivation on the latent cognitive processes underlying attribute processing and decision-making, including information sampling on these attributes, retrieving long-term memory about the values of these attributes, comparing the differences between options, and trade-offs between attributes. These processes are generally considered as (part of) cognitive control in previous studies (51–53). We show here that the fluctuation in motivational state modifies the relative decision weights on attributes, which are also reflected in the temporal prioritization of the attributes processing. Our findings extend the close relationship between motivation and cognitive control on single decision attributes to the coordination of multiple attributes processing and highlight the adaptive function of motivation in the exertion of cognitive control.

In conclusion, we demonstrate in a real-time social interaction paradigm that the compensatory and social-avoidance tendencies coexist and compete in a transgressor's social decision-making following an interpersonal transgression. We uncover the dynamics of the participants' decision-making using mouse tracking and computational modeling, demonstrating that the individuals process the attributes of compensation and social avoidance asynchronously over time and the temporal profiles are flexibly modulated by their social-affective states. These results shed light on the dynamics of complex emotional and motivational processing following social transgression and provide general implications for the mechanism of interaction between motivations and their related decision attributes' processing during decision-making.

Methods

Participants. Seventy-four right-handed college students participated in the experiment. Half of the participants ($N = 37$, age = 22.6 ± 2.1, 19 females) played the role of transgressors (the role we focused on), and the other half played the role of victims. The sample size was determined by power analysis (G*Power 3) (56) on a within-subject design one-sample t test, which showed that it requires at least 34 subjects to test the hypothesis on the temporal dynamic at α level 95% and power level 80% for a medium-level effect size (Cohen's d = 0.5) shown in a previous study using similar mouse-tracking methodologies (32). None of the participants reported a history of neurological or psychiatric illness. All the participants provided written informed consent before the experiment. Participants were paid based on their task performances (see details below). The experiment conformed to the Declaration of Helsinki and was approved by the Ethics Committee of the School of Psychological and Cognitive Sciences, Peking University.

Procedure.

Role assignment and preparation. Participants in the same session, consisting of 4, 6, or 8 individuals of the same sex, met in a computer room before the experiment. Half of them were assigned as Player A and the other half as Player B depending on the role concealed in a folded label each participant drew (Fig. 1*A*). Participants assigned to the same role were instructed to sit on the same side of the computer room in separate compartments, shielded from view of the participants in a different role on the other side of the testing room (for a photo of the testing room arrangement, see *SI Appendix*, Fig. S1*A*). Video cameras (Logitech WebCam C270) were placed in front of each participant so that they could see the paired partner through real-time video at designated time points during the experiment (see *SI Appendix*, Fig. S1*A* for the setup). Computers were connected via TCP-IP protocol using Matlab. Prior to the main task, all participants went through the same procedure for practicing the skills required in the main task (*SI Appendix*, Fig. S1 *B–D*).

The main task. All participants were instructed in the same room with the same information on the main task. Each player was endowed with 90 Chinese Yuan (about 14 US dollars) for the game. All players were informed that Player B, but not Player A, could lose up to 1/3 of their entire endowments (i.e., 30 RMB, or 840 M.U.), if the players failed their task in all trials (for the actual performance of our participants, Player B earned 80.79 ± 1.01 RMB, with an individual min of 78.87 and maximum of 82.51, less than the fixed amount of 90 RMB that Player A earned). Most of the participants experienced 84 trials, except one session (four pairs of participants) who experienced 68 to 69 trials due to a technical error. In the visual search task, both players saw 20 bars of the same color, either blue or orange, scattered in the central area (500 × 500 pixels) of the screen, and were asked to search for a target bar with an orientation that was either vertical or horizontal among the other distractors with random orientations. The color of the stimuli and the orientation of the target bar result in four possible associations with responses: blue & vertical, blue & horizontal, orange & vertical, and orange & horizontal. The participants' task was to identify which of the four possibilities was the case for each trial, by pressing one of four buttons, with two buttons on the keyboard ("Q" and "W") and the other two buttons on the mouse (left and right click). Pressing the wrong buttons or not responding within time constraints (see below) were counted as a failure.

When both were correct, the players saw "Pass" on the screen and proceeded to the next trial directly. When failure(s) occurred, Player A was asked to choose between two options presented at the top-left and top-right corners of the screen by moving the computer mouse from the bottom-center of the screen (Fig. 1*D*). Each option contains a certain amount of monetary compensation to Player B and a probability of engaging in video contact between the two players later on at the end of the trial. The offer values were controlled by a self-adaptive algorithm (*SI Appendix, Methods*).

At the end of the trial, the participants were asked to rate the intensities of two emotions (i.e., guilt and embarrassment for Player A, anger and embarrassment for Player B) they experienced on the current trial from 0 to 10.

Free money. Instead of having Player A compensate Player B using Player A's own money, here the money was from the experimenters. In this way, Player A's monetary self-interest would therefore have no impact on their choice. Player A's decision would be a result of the trade-off between two motivations we were interested in this study, namely, compensating (approaching) the victim and avoiding social contact with the victim. Admittedly, oftentimes in real-world social situations, compensation to the victims comes at the cost of the transgressors. Delineating how the transgressors' motivation for maximizing their self-interest is integrated with the motivations for benefiting and avoiding the victim is beyond the scope of the current study and calls for future research.

Confidential compensation and no communications during video contact. Video contact would be triggered when the chosen probability was greater than a randomly generated value (uniformly distributed between 0 and 100%), these two players would experience a 5-s long virtual face-to-face contact through real-time video; otherwise, a blank screen would be presented to both players for 5 s. If face-to-face contact with the victim was enacted, both Players were asked to keep a neutral facial expression and not to engage in communication using any speech, lip talk, facial expressions, or gestures during the video contact. Checking the video recordings during face-to-face contact periods confirmed that all participants complied well with the no-communication rule. Moreover, Player B was not informed about how much Player A compensated them on any trial. Instead, they would only know the cumulative amount they had received after the entire experiment was completed. The rationale of discouraging communication and keeping the amount of monetary compensation confidential was to prevent Player A from using face-to-face contact or the monetary compensation as a strategy to communicate with and seek forgiveness from Player B. In this manner, the binary choice was geared toward examining the trade-off between Player A's approach motivation (e.g., compensating Player B) and social avoidance motivation (e.g., avoiding being seen by Player B).

Hierarchical Logistic Regressions on Participants' Binary Choices. We estimated participants' decision weights b_m and b_p on monetary compensation and contact probability from their binary choices by using logistic regression (Eq. **1**),

$$\text{logit}\left(P_{\text{chosen Opt.1}}\right) \sim b_m \Delta m + b_p \Delta p, \qquad [\mathbf{1}]$$

where $\Delta m \equiv m_1 - m_2$ and $\Delta p \equiv p_1 - p_2$ indicate the differences in monetary amounts and probabilities between Option 1 and Option 2. Numeric values of Δm and Δp were scaled to the same range of $[-1, 1]$ in the regression to achieve better estimation reliability.

Parameters were estimated using Bayesian statistical inference method in RStan (57). Parameters of all participants in each estimation (separated by condition) were estimated by a single hierarchical model. Each line of estimation was sampled from eight chains, with each chain having 1,000 iterations of warm-up and another 1,000 iterations of sampling. Inspecting the sampling traces of each estimation confirmed that the number of warm-up iterations was sufficient for reaching convergence (see *SI Appendix, Results* for the model convergence report).

Estimation of the Decision Weight Dynamics. We estimated the impact of each attribute on the participants' momentary preference by running multiple regressions of the attributes' values on the participants' horizontal velocity (v_h); the result would indicate the transient tendency of toward or away from the options,

$$v_{h,t} \sim b_{m,t} \Delta m + b_{p,t} \Delta p. \qquad [\mathbf{2}]$$

The regressions were performed repetitively over 100 time bins, with t as the indicator of the time bin. Because $b_{m,t}$ and $b_{p,t}$ estimated here carried not only information about preference but also the mechanical features of mouse movement velocity, e.g., acceleration in the beginning of a trial and deceleration at the end of a trial. To remove the motor effect and keep only decision-related information, we normalized the $b_{m,t}$ and $b_{p,t}$ by their summed magnitude and obtained the relative decision weights, $w_{m,t}$ and $w_{p,t}$, where $w_{m,t} = \dfrac{b_{m,t}}{|b_{m,t}| + |b_{p,t}|}$ and $w_{p,t} = \dfrac{b_{p,t}}{|b_{m,t}| + |b_{p,t}|}$; the obtained values indicate the decision weights on one attribute relative to the other.

Parameters were estimated using the Bayesian statistical inference method mentioned above. Data from each condition and each time bin were estimated by a single hierarchical model, with all participants' data combined. The numbers of chains and iterations were set the same as above. The model convergence for the estimated parameters was checked and reported in *SI Appendix*, Fig. S6.

Data, Materials, and Software Availability. The data and code presented in this work will be openly available upon publication in Mendeley Data at http://doi.org/10.17632/jcng3638wx.1 (58).

ACKNOWLEDGMENTS. We thank Sensen Song for assistance in data collection, Dr. Lusha Zhu and Dr. Kunlin Wei for their suggestions on data analysis. This work was supported by National Natural Science Foundation of China (71942001, 31630034). B.S. was supported by China Postdoctoral Science Foundation (2018M642489).

Author affiliations: [a]School of Psychology, Zhejiang Normal University, Jinhua 321004, China; [b]Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016; [c]Graduate Program in Translational Biology, Medicine, and Health, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061; [d]Division of Biosciences, University College London, London WC1E 6BT, United Kingdom; [e]Department of Psychological and Brain Sciences, University of California Santa Barbara, Santa Barbara, CA 93106; and [f]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

1. W. Hofmann, D. C. Wisneski, M. J. Brandt, L. J. Skitka, Morality in everyday life. *Science* **345**, 1340–1343 (2014).
2. L. Woodyatt, M. Wenzel, T. G. Okimoto, M. Thai, Interpersonal transgressions and psychological loss: Understanding moral repair as dyadic, reciprocal, and interactionist. *Curr. Opin. Psychol.* **44**, 7–11 (2022).
3. W. A. Afifi, W. L. Falato, J. L. Weiner, Identity concerns following a severe relational transgression: The role of discovery method for the relational outcomes of infidelity. *J. Soc. Pers. Relatsh.* **18**, 291–308 (2001).
4. T. G. Okimoto, M. Wenzel, Bridging diverging perspectives and repairing damaged relationships in the aftermath of workplace transgressions. *Bus. Ethics Q.* **24**, 443–473 (2014).
5. N. Strohminger, S. Nichols, The essential moral self. *Cognition* **131**, 159–171 (2014).
6. R. F. Baumeister, A. M. Stillwell, T. F. Heatherton, Guilt: An interpersonal approach. *Psychol. Bull.* **115**, 243 (1994).
7. J. P. Tangney, J. Stuewig, D. J. Mashek, Moral emotions and moral behavior. *Annu. Rev. Psychol.* **58**, 345–372 (2007).
8. B. Parkinson, S. Illingworth, Guilt in response to blame from others. *Cogn. Emot.* **23**, 1589–1614 (2009).
9. D. M. Amodio, P. G. Devine, E. Harmon-Jones, A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychol. Sci.* **18**, 524–530 (2007).
10. T. J. Ferguson, H. Stegge, I. Damhuis, Children's understanding of guilt and shame. *Child Dev.* **62**, 827 (1991).
11. J. L. Freedman, S. A. Wallington, E. Bless, Compliance without pressure: The effect of guilt. *J. Pers. Soc. Psychol.* **7**, 117–124 (1967).
12. T. Schmader, B. Lickel, The approach and avoidance function of guilt and shame emotions: Comparing reactions to self-caused and other-caused wrongdoing. *Motiv. Emot.* **30**, 42–55 (2006).
13. U. Beyens, H. Yu, T. Han, L. Zhang, X. Zhou, The strength of a remorseful heart: Psychological and neural basis of how apology emolliates reactive aggression and promotes forgiveness. *Front. Psychol.* **6**, 1611 (2015).
14. X. Gao *et al.*, The mutuality of social emotions: How the victim's reactive attitude influences the transgressor's emotional responses. *NeuroImage* **244**, 118631 (2021).
15. M. E. McCullough, E. J. Pedersen, B. A. Tabak, E. C. Carter, Conciliatory gestures promote forgiveness and reduce anger in humans. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11211–11216 (2014).
16. H. Yu, J. Hu, L. Hu, X. Zhou, The voice of conscience: Neural bases of interpersonal guilt and compensation. *Soc. Cogn. Affect. Neurosci.* **9**, 1150–1158 (2014).
17. S. Čehajić-Clancy, D. A. Effron, E. Halperin, V. Liberman, L. D. Ross, Affirmation, acknowledgment of in-group responsibility, group-based guilt, and support for reparative measures. *J. Pers. Soc. Psychol.* **101**, 256–270 (2011).
18. K. Ohbuchi, M. Kameda, N. Agarie, Apology as aggression control: Its role in mediating appraisal of and response to harm. *J. Pers. Soc. Psychol.* **56**, 219–227 (1989).
19. E. Watanabe, Y. Ohtsubo, Costly apology and self-punishment after an unintentional transgression. *J. Evol. Psychol.* **10**, 87–105 (2012).
20. N. Weiss-Klayman, B. Hameiri, E. Halperin, Group-based guilt and shame in the context of intergroup conflict: The role of beliefs and meta-beliefs about group malleability. *J. Appl. Soc. Psychol.* **50**, 213–227 (2020).
21. S. Ceylan-Batur, A. K. Uskul, Preferred responses when honour is at stake: The role of cultural background, presence of others, and causality orientation. *Asian J. Soc. Psychol.* **25**, 336–347 (2022).
22. Y. Lin, N. Caluori, E. B. Öztürk, M. J. Gelfand, From virility to virtue: the psychology of apology in honor cultures. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2210324119 (2022).
23. K. Schumann, The psychology of offering an apology: understanding the barriers to apologizing and how to overcome them. *Curr. Dir. Psychol. Sci.* **27**, 74–78 (2018).
24. S. Shafa, F. Harinck, N. Ellemers, Sorry seems to be the hardest word: Cultural differences in apologizing effectively. *J. Appl. Soc. Psychol.* **47**, 553–567 (2017).
25. C. Gerlsma, V. Lugtmeyer, M. Van Denderen, J. De Keijser, Revenge and forgiveness after victimization: Psychometric evaluation of a Dutch version of the TRIM intended for victims and offenders. *Curr. Psychol.* **41**, 7142–7154 (2022).
26. H. Yu, Y. Duan, X. Zhou, Guilt in the eyes: Eye movement and physiological evidence for guilt-induced social avoidance. *J. Exp. Soc. Psychol.* **71**, 128–137 (2017).
27. S. Čehajić, R. Brown, R. González, What do I care? Perceived ingroup responsibility and dehumanization as predictors of empathy felt for the victim group *Group Process. Intergr. Relat.* **12**, 715–729 (2009).
28. N. Haslam, Dehumanization: An integrative review. *Personal. Soc. Psychol. Rev.* **10**, 252–264 (2006).
29. H. Xu, L. Bègue, R. Shankland, Guilt and guiltlessness: An integrative review. *Soc. Personal. Psychol. Compass* **5**, 440–457 (2011).
30. L. Niemi, L. Young, When and why we see victims as responsible: The impact of ideology on attitudes toward victims. *Pers. Soc. Psychol. Bull.* **42** (9), 1227–1242 (2016).
31. L. Koban, C. Corradi-Dell, P. Vuilleumier, Integration of error agency and representation of others pain in the anterior insula. *J. Cogn. Neurosci.* **25**, 258–272 (2013).
32. N. Sullivan, C. Hutcherson, A. Harris, A. Rangel, Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychol. Sci.* **26**, 122–134 (2015).
33. P. E. Stillman, X. Shen, M. J. Ferguson, How mouse-tracking can advance social cognitive theory. *Trends Cogn. Sci.* **22**, 531–543 (2018).
34. D. McGee, R. Giner-Sorolla, "How Guilt Serves Social Functions from Within" in *The Moral Psychology of Guilt*, B. Cokelet, C. J. Maley, Eds. (Rowman & Littlefield International Ltd., 2019), pp. 149-170.

35. I. E. De Hooge, "Improving Our Understanding of Guilt by Focusing on Its (Inter) Personal Consequences", in *The Moral Psychology of Guilt*, B. Cokelet, C. J. Maley, Eds. (Rowman & Littlefield International Ltd., 2019), pp. 131-148.
36. W. B. G. Liebrand, C. G. McClintock, The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *Eur. J. Personal.* **2**, 217–230 (1988).
37. R. O. Murphy, K. A. Ackermann, Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personal. Soc. Psychol. Rev.* **18**, 13–41 (2014).
38. J. Gottlieb, P.-Y. Oudeyer, Towards a neuroscience of active sampling and curiosity. *Nat. Rev. Neurosci.* **19**, 758–770 (2018).
39. R. Kurzban, A. Duckworth, J. W. Kable, J. Myers, An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* **36**, 661–679 (2013).
40. X. Gao *et al.*, Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7680–E7689 (2018).
41. W. W. Pettine, K. Louie, J. D. Murray, X.-J. Wang, Excitatory-inhibitory tone shapes decision strategies in a hierarchical neural network model of multi-attribute choice. *PLoS Comput. Biol.* **17**, e1008791 (2021).
42. S. U. Maier, A. Raja Beharelle, R. Polanía, C. C. Ruff, T. A. Hare, Dissociable mechanisms govern when and how strongly reward attributes affect decisions. *Nat. Hum. Behav.* **4**, 949–963 (2020).
43. R. M. Roe, J. R. Busemeyer, J. T. Townsend, Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychol. Rev.* **108**, 370–392 (2001).
44. J. S. Trueblood, S. D. Brown, A. Heathcote, The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychol. Rev.* **121**, 179–205 (2014).
45. M. Usher, J. L. McClelland, Loss aversion and inhibition in dynamical models of multialternative choice. *Psychol. Rev.* **111**, 757–769 (2004).
46. O. FeldmanHall *et al.*, What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition* **123**, 434–441 (2012).
47. S. L. Gable, Approach and avoidance social motives and goals. *J. Pers.* **74**, 175–222 (2006).
48. X. Gao *et al.*, The hidden cost of receiving favors: A theory of indebtedness. bioRxiv [Preprint], https://doi.org/10.1101/2020.02.03.926295 (Accessed 4 July 2023).
49. K. Friston *et al.*, Active inference and epistemic value. *Cogn. Neurosci.* **6**, 187–214 (2015).
50. L. Schulz, S. M. Fleming, P. Dayan, Metacognitive computations for information search: Confidence in control. *Psychol. Rev.* **130**, 604–639 (2023).
51. M. Botvinick, T. Braver, Motivation and cognitive control: From behavior to neural mechanism. *Annu. Rev. Psychol.* **66**, 83–113 (2015).
52. A. Shenhav, M. M. Botvinick, J. D. Cohen, The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
53. D. M. Yee, T. S. Braver, Interactions of motivation and cognitive control. *Curr. Opin. Behav. Sci.* **19**, 83–90 (2018).
54. C. Parro, M. L. Dixon, K. Christoff, The neural basis of motivational influences on cognitive control. *Hum. Brain Mapp.* **39**, 5097–5111 (2018).
55. J. D. Salamone, M. Correa, The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470–485 (2012).
56. F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
57. Stan Development Team, RStan: the R interface to Stan. (R package version 2.21.2., 2023). http://mc-stan.org/. Accessed 9 September 2023.
58. B. Shen, The Competition Dynamics of Approach and Avoidance Motivations Following Interpersonal Transgression. Mendeley Data, V1. https://doi.org/10.17632/jcng3638wx.1. Accessed 3 July 2023.