HHS Public Access

Author manuscript

J Exp Soc Psychol. Author manuscript; available in PMC 2021 March 01.

Published in final edited form as:

J Exp Soc Psychol. 2020 March ; 87: . doi:10.1016/j.jesp.2019.103916.

An association between biased impression updating and relationship facilitation: A behavioral and fMRI investigation☆

BoKyung Park[*], **Liane Young**

Boston College, United States of America

Abstract

Is ingroup bias associated with any benefit for maintaining close relationships? We examined the link between biased impression updating for ingroup members (i.e., friends) and relationship maintenance, as measured by the number of friends participants reported having (Studies 1 and 2). We also investigated the underlying neural basis of this possible effect, focusing on activity in the right temporo-parietal junction (RTPJ), a region of the social brain involved in moral updating (Study 2). Specifically, we tested whether selectively discounting negative information about close others, manifested in reduced impression updating, and indexed by reduced RTPJ activity, is related to maintaining close relationships. In Study 1, after imagining a friend and a stranger performing different positive and negative behaviors, participants who were reluctant to update how close they felt to their friend (friend-closeness) reported having more friends in real life. In Study 2, participants were led to believe that a friend and a stranger gave money to them or took money away from them, while they were in the scanner. Participants who engaged in less negative updating of friends versus strangers reported having more friends. Participants who engaged in less friend-closeness updating also showed reduced RTPJ activity when their friend took money from them, and this neural pattern was associated with reports of having more friends. Together, these findings suggest that selectively discounting close others' negative behavior is linked to maintaining close relationships, indicating a potential social benefit of ingroup bias.

Keywords

Impression updating; Ingroup bias; Relationship maintenance

1. Introduction

Ingroup bias, the tendency to judge ingroup members favorably compared to outgroup members (Tajfel, 1982), has been associated with a wide range of negative social outcomes: distorted perception about outgroups (Xiao, Coppin, & Van Bavel, 2016), reduced intergroup cooperation (Balliet, Wu, & Dreu, 2014; Sherif, 1966), and prejudice and discrimination (Brewer, 1999), to name a few.

☆This paper has been recommended for acceptance by Rachael Jack.

*Corresponding author. parkanj@bc.edu (B. Park).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2019.103916.

A key manifestation of ingroup bias is inaccurate impression formation and updating across group boundaries. Although accurately evaluating others is essential in many ways for successful social interactions (Shin, Kim, & Han, 2014; Zaki & Ochsner, 2011), people are nevertheless motivated to see ingroup members, including close others, in a more positive light (Brewer, 1999; Taylor & Brown, 1988). Ingroup bias can lead people to forget more negative information about ingroup members (Howard & Rothbart, 1980), discount ingroup members' harmful intentions in the service of blame mitigation (Monroe & Malle, 2019), and preferentially forgive ingroup members (Baumgartner, Götte, Gügler, & Fehr, 2012; Schiller, Baumgartner, & Knoch, 2014; Wohl & Branscombe, 2009). These findings reveal suboptimal processing of social information, specifically, information about others' negative behavior.

Relatively few studies have examined the potential association between bias in impression formation and updating and the maintenance or protection of close relationships, for example. Here we describe two mechanisms that might support such an association. First, extant literature on close relationships has revealed the relatively powerful impact of negative versus positive events on the quality of a relationship (for a review, see Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), given that negative behaviors are recognized better (Acitelli, Douvan, & Veroff, 1993; Alves et al., 2015; see also Kellermann, 1984) and reciprocated more (Levenson & Gottman, 1985; Rusbult, Johnson, & Morrow, 1986). This body of work suggests the possibility that selectively discounting close others' negative behavior may facilitate relationships. Alternatively, individuals' motivation to protect their social resources may be scaled according to the size of their social network. Thus, people with more social relationships may engage in selective discounting to a greater extent. Importantly, both accounts predict the same positive association that the present work establishes between the selective discounting of close others' negative behavior and relationship maintenance.

Previous research on brain regions for thinking about agents' mental states, also known as mentalizing, mindreading, or theory of mind (ToM), provides some clues to the neural mechanisms underlying impression updating. The mentalizing network has been associated with detecting social prediction error, i.e., the gap between expectation and observation (Koster-Hale & Saxe, 2013), and ultimately updating social impressions given new information (Baron, Gobbini, Engell, & Todorov, 2011; Boorman, O'Doherty, Adolphs, & Rangel, 2013; Mende-Siedlecki, Baron, & Todorov, 2013; Mende-Siedlecki, Cai, & Todorov, 2013; Thornton & Mitchell, 2018; see also Mende-Siedlecki & Todorov, 2016). For example, in a previous study, participants exhibited reduced activity in RTPJ, a key node of the mentalizing network, when a previously fair versus unfair target brought about a negative outcome (Kliemann, Young, Scholz, & Saxe, 2008). These fair targets were also judged as less blameworthy and less intentional. The behavioral and neural patterns indicate that failure to accurately detect or evaluate previously fair targets' negative behaviors may be related to reduced mentalizing, as indexed by activity in RTPJ.

Other work has examined this effect more directly in the context of ingroup bias. In one study, participants were presented with targets from their own group or an outgroup (Hughes, Zaki, & Ambady, 2017). Brain activity was measured when ingroup and outgroup

targets behaved in ways that were morally unexpected given their prior behaviors. Failure to recruit mentalizing regions in response to ingroup members' unexpected negative behaviors was associated with reduced impression updating. These findings again suggest that the failure to recruit mentalizing regions may underlie the failure to engage in negative impression updating of ingroup members.

The present work seeks to build on and extend the prior literature in three ways. First, we aim to replicate prior research showing that people engage in biased updating (by discounting ingroup members' bad behavior), and that this effect is accounted for by reduced recruitment of mentalizing regions such as RTPJ. Second, and critically, we also investigate the novel possibility that biased updating is associated with positive social outcomes. Our primary hypothesis is that overlooking close others' negative behavior, indexed by disengagement of mentalizing regions such as RTPJ, is associated with maintaining more close relationships, measured by the number of friends that people report having. Third, while research in social psychology often points to a *motivational* account for why people discount close others' negative behavior in the first place, a different theoretical account has also been proposed. Specifically, discounting or ignoring new information about a person that is inconsistent with one's strong prior knowledge or impression of the person (e.g., a close friend's unexpected negative behavior) can adhere to Bayesian principles and therefore reflect *Bayesian-rational* rather than biased or motivated updating (Gershman, 2019; Hahn & Harris, 2014). For example, previous research has reported that people are less influenced by their social partners' behavior when the behavior is inconsistent with their prior impressions of their partners (Fareri, Chang, & Delgado, 2012). People also experience greater social reward in response to their friend's positive behavior compared to a stranger's positive behavior (Fareri, Chang, & Delgado, 2015). However, little is known about whether these behavioral patterns are driven by Bayesian-rational or motivational processes. Our work represents a preliminary exploration of this question as well.

In Study 1, we examined the association between biased impression updating and self-reported number of close relationships. Specifically, we measured the degree to which participants updated their evaluations about a close friend versus a stranger when they learned new negative information about these agents, and we tested the association between update magnitude and reported number of friends. In Study 2, we investigated the underlying neural basis of this effect by presenting participants with positive and negative behaviors performed by a close friend and a stranger in the scanner and then examining the association between RTPJ activity and reported number of friends, as well as the degree to which participants updated their impressions of their friend.

2. Study 1

In Study 1, we explored whether participants' impression updates are linked to social outcomes. In particular, we predicted that participants showing resistance to impression updating for friends in the experimental paradigm also report having more friends in real life. For this study and Study 2, we report all measures, manipulations, and exclusions.

2.1. Method

2.1.1. Participants—One hundred and twenty-five Mturkers (37.6% female; age M = 35.34, S.D. = 10.65), divided into two groups, one week apart from each other,[2] were recruited for this study. Fourteen responses that did not fit our inclusion criteria were removed,[3] leaving 111 final participants. We found similar patterns when including all participants (See Supplementary Section 1). A sensitivity power analysis using a significance level $\alpha = 0.05$ and power $\beta = 0.80$ showed that this sample size is sufficient to detect an effect size of $f^2 = 0.07$.[4] Sample size was based on the expectation of a small to moderate effect size. The adjusted alpha boundary for this sequential sampling was $p < .0336$ (Lakens, 2014).

2.1.2. Material—We used eight different moral and immoral behaviors of a target person from prior work (Kim, Mende-Siedlecki, Anzellotti, & Young, in prep; Mende-Siedlecki, Baron, & Todorov, 2013). The moral and immoral behaviors were matched in terms of moral relevance, arousal, valence, and perceived frequency ($ps > 0.185$; see Supplementary Section 2).

2.1.3. Procedure—After providing informed consent, participants were asked to enter their gender-matched best friend's first name. Participants then completed the Inclusion of Other in Self scale (IOS; Aron, Aron, & Smollan, 1992), consisting of pairs of circles representing themselves and their friend, with increasing overlap between circles indicating increasing closeness. Those who chose the 4th pair or more on this 7-point scale (M = 5.32, S.D. = 0.90) were recruited. After the IOS question, participants answered the question regarding the number of hours per week they spent with or communicate with their friend (1 = Less than 30 min, 2 = 30 min–2 h, 3 = 2 h–4 h, 4 = 4 h–6 h, 5 = 6 h–8 h, 6 = 8 h–10 h, 7 = 10 h–12 h, 8 = 12 h–14 h, 9 = 14 h–16 h, 10 = 16 h–18 h, 11 = 18 h–20 h, 12 = over 20 h; M = 4.56, S.D. = 2.64), as a measure of the amount of prior experience they had with their friend.[5] After that, participants delivered two different ratings, trustworthiness (“To what extent is [Name] trustworthy?”) and closeness (“How close are you and [Name]?”), on 8-point scales, about their friend and a gender-matched stranger (“Jacob” for male participants and “Emily” for female participants) (“pre-exposure ratings”; see Supplementary Section 3b for correlations).

Next, participants were presented with 8 different behaviors and asked to imagine that their friend or the stranger (“target”) performed these behaviors in the past. Their friend's name and the stranger's name were piped into the behaviors. Participants first read two positive behaviors that one of the targets performed (e.g., “[Name] helped an elderly woman get up from her wheelchair”), followed by two negative behaviors performed by the same target (e.g., “[Name] spread rumors about a coworker's work ethic”). Afterwards, participants read

[2]After conducting a preliminary data analysis, we decided to double the sample size to capture a small to moderate effect size.
[3]We excluded responses from repeated IP addresses ($N = 2$), nonsensical responses to open-ended questions, such as “good” in a response to a prompt for participants' most unpleasant memory with their gender-matched best friend ($N = 8$), and responses where participates rated a stranger as closer to them than their friend ($N = 4$).
[4]Although our main statistical analyses used ordinal regressions, we based our power analyses on multiple regression models given that G*Power does not provide an ordinal regression option.
[5]Additionally, participants rated how much they liked their friend. Participants' IOS ratings, liking ratings, and the number of hours they spent with their friend were not associated with how many friends they reported having (Supplementary Section 3a).

two positive behaviors performed by the other target and two negative behaviors performed by that same target, consecutively. We counterbalanced the following conditions: (1) presentation order of two behaviors within each positive and negative block, (2) which target was presented first, and (3) which behaviors were presented as performed by friend or stranger. Because our primary interest is in how people update their impressions based on new negative behaviors, positive behaviors were always presented first, followed by negative behaviors. After imagining each behavior, participants made ratings of trustworthiness ("How trustworthy is [Name]?") and closeness ("How close are you and [Name]?") for the target agent of the behavior on 8-point scales, based on all the information available up to that point.[6]

Participants were then asked to evaluate both trustworthiness and closeness in general (as described above for the "pre-exposure ratings") for the target ("post-exposure ratings"), followed by questions about their total number of friends (1 = 0–2 friends; 2 = 3–5 friends; 3 = 6–9 friends; 4 = 10–19 friends; 5 = 20–49 friends; 6 = 50–99 friends; and 7 = More than 100 friends) and how often they make new friends (1 = Almost every week; 2 = A few weeks; 3 = Every month; 4 = A few months; 5 = Every year; 6 = A few years; 7 = About every 10 years; and 8 = More than every 10 years). To additionally measure the amount of prior experience participants had with their friend, participants answered an open-ended question about how long they have known their friend (years; M = 16.77, S.D. = 10.30), along with other items not explored here (See Supplementary Section 4).[7] Finally, participants answered demographic questions and were thanked and compensated. All procedures were approved by the Institutional Review Board at Boston College.

2.2. Analyses and results

To measure the extent to which participants updated their impressions of their friends, we subtracted participants' ratings for the last negative behavior of the target from their ratings for the last positive behavior of the target, generating four updating scores (updates in friend-closeness, friend-trustworthiness, stranger-closeness, and stranger-trustworthiness ratings) (for participants' impression updating after viewing only the positive behaviors, see Supplementary Section 6a). Approximately half of participants did not update their closeness ratings (51.4% for friend, 48.6% for stranger), so we recoded participants into three categories; those who never updated their closeness ratings ("No update" = 0; friend-closeness $N = 57$; stranger-closeness $N = 54$), those who decreased their closeness ratings ("Negative update" = −1; friend-closeness $N = 53$; stranger-closeness $N = 55$), and those who unexpectedly increased their closeness ratings ("Positive update" = +1; friend-closeness $N = 1$; stranger-closeness $N = 1$) after imagining negative behaviors. The findings were similar when excluding people who unexpectedly engaged in positive updating (See Supplementary Section 6b for changes in evaluations between the last positive behavior and the last negative behavior).

[6]How long participants took to make these ratings (reaction time) was not associated with their updating scores (See Supplementary Section 3c).

[7]We did run exploratory analyses with variables that might influence the number of friends participants reported having, including participants' extraversion and participants' socioeconomic status. Controlling for these variables did not change the results (See Supplementary Section 5).

We examined whether participants who were more resistant to updating their friend-closeness ratings reported having more friends. Ordinal regression on the number of friends that participants reported having revealed that those who never updated, and those who positively updated, reported having more friends, significant under the adjusted alpha boundary (Lakens, 2014), Estimate = 0.73, S.E. = 0.34, Wald = 4.61, Odds ratio = 2.07, p = .032, 95% CI for Estimate = [0.06, 1.39], compared to those who negatively updated their friend-closeness ratings (Fig. 1[8]). This effect remained consistent after we controlled for how often participants made new friends, how many hours per week they spent in direct contact with or in remote communication with their friend, and how long they have known their friend (See Supplementary Section 5), as well as updates in friend-trustworthiness, stranger-closeness, and stranger-trustworthiness. Importantly, when entered in the same model, none of the updates in other impressions was significantly associated with the number of friends participants reported having, $ps > 0.48$, Odds ratio = 0.88–1.07.[9]

As a validity check, we subtracted participants' closeness ratings in their pre-exposure evaluation (i.e., before they imagined any behaviors) from those in their post-exposure evaluation (i.e., after they imagined all behaviors), generating post minus pre closeness updates. Using this new metric, we found again that participants who less negatively updated their friend-closeness ratings reported having more friends, Estimate = 0.61, S.E. = 0.24, Wald = 6.40, Odds ratio = 1.85, p = .011, 95% CI for Estimate = [0.14, 1.09]. Again, these effects did not change after we controlled for how often participants made new friends and updates in friend-trustworthiness, stranger-closeness, and stranger-trustworthiness[10] (See Supplementary Section 3d for correlations between covariates in regression models; Supplementary Section 7 for analyses matched with Study 2).

2.3. Study 1 discussion

In Study 1, we found that people who were more resistant to updating how close they felt to their friend after learning about their friend's negative behavior reported having more friends. Importantly, controlling for the number of hours per week participants spent with their friend in person or in remote communication, and the number of years participants have known their friend did not change the results. The results suggest that people may discount negative information about their friends because they are motivated to maintain their positive impressions of their friends. In other words, biased rather than rational impression updating might account for close relationship maintenance. Also, although we did not expect different updating effects for friend-closeness and friend-trustworthiness, participants' updates in friend-trustworthiness, as well as stranger-closeness and stranger-trustworthiness, were not associated with reports of number of friends.

[8]We did not plot “Positive update” group in the figure as this group was an N of 1.

[9]We used raw update values for trustworthiness ratings given that more participants varied in these ratings, compared to closeness ratings. However, even when we categorized participants into “Never update,” “Negative update,” and “Positive update” groups as with the closeness ratings, we found no significant effects.

[10]A majority of the participants did not update their ratings between before and after imagining the behaviors especially for friends (77.3% for friend-closeness, 68.2% for friend-trustworthiness, 67.3% for stranger-closeness, and 32.7% for stranger-trustworthiness). However, we found similar effects after replacing the continuous post minus pre closeness updates with categorized closeness updates (No update, Negative update, and Positive update).

In this study, participants were asked to *imagine* that their friend and the stranger committed the hypothetical behaviors described in experimental stimuli. In the next study, we asked participants to bring their own friends in person to the scanning session, where participants and their friends engaged in ostensibly real-time interactions while participants' brain activity was measured. Thus, Study 2 aimed to further explore the mechanisms that contribute to biased impression updating and relationship maintenance.

3. Study 2

Study 2 extended Study 1 in two ways. First, we had participants bring their close friends with them to the scan session, where participants then observed positive and negative behaviors performed by their friend. Moreover, these behaviors directly affected the participants rather than a third party. Thus, unlike Study 1, in which participants were asked to merely imagine that their friend engaged in certain behaviors, in Study 2, participants were led to believe that their friends were actually behaving positively and negatively toward the participants in real time.

Second, we examined the underlying neural mechanisms by which people discount their friend's negative behavior and maintain close relationships. Specifically, we predicted that disengagement of mentalizing in response to friends' negative behavior (indexed by reduced RTPJ activity) would be correlated with (1) reduced updates in friend-closeness, suggesting that decreased RTPJ activity accounts for biased impression updating for close others, and (2) a greater number of friends, supporting the argument that overlooking or ignoring close others' negative behavior may be linked to maintaining more close relationships.

3.1. Method

3.1.1. Participants—We recruited thirty right-handed and neurologically and psychologically intact participants to bring a close gender-matched friend with them to the scanner. Participants who selected pairs of circles equal to or greater than 4-points on the same IOS scale as in Study 1 were recruited. Six participants were excluded from further analyses,[11] leaving a total of 24 participants in the final sample. Since this dataset was initially collected for a different study (Park, Fareri, Delgado, & Young, 2019), we did not determine the sample size in advance. All procedures were approved by the Institutional Review Boards at Boston College and the Massachusetts Institute of Technology.

3.1.2. Social judgment task—Participants were informed that they would play a game (“Social Judgment Task”) with two different targets, their friend and a gender-matched stranger whom they met at the scanner. Participants were instructed as follows: In the game, there are two different roles, “Player 1” and “Player 2.” On each trial, each player receives $20 for use in that trial. Player 1 decides how much to give to or take from Player 2 in $5 increments; Player 2 passively views Player 1's decision. After that, participants were told

[11]Because of excessive head movement (three participants), a structural abnormality (one participant), a deviant expectation about their friend (one participant), or completing fewer than half of the trials (one participant). Participants who completed trials equal to or more than 50% of all trials were included in the final sample.

that they were randomly assigned to play as Player 2, and they would see their friend and the stranger taking turns as Player 1. In reality, all Player 1 decisions were pre-programmed.

As Player 2, participants were asked to make either closeness or trustworthiness ratings about Player 1 in each trial after seeing Player 1's decision. During the game, participants were first presented with the identity (friend or stranger) of Player 1 on each trial (2 s), followed by rating type (2 s), closeness or trustworthiness. After a jittered fixation (2–6 s), participants observed how much Player 1 gave or took (2 s; $5, $10, $15, or $20); and then after another jittered fixation (2–6 s), participants were able to make their ratings on an 8-point scale (4 s). Each trial was divided by a jittered fixation (2–6 s) (Fig. 2). Participants were told that one of the trials would be randomly selected, and that they and Player 1 would receive the amount of money received in that trial in addition to their base compensation.

3.1.3. Procedure—Prior to the scan session, participants completed the IOS scale and answered the question regarding the number of hours per week they spent with their friend (1 = Less than 30 min, 2 = 30 min–2 h, 3 = 2 h–4 h, 4 = 4 h–6 h, 5 = 6 h–8 h, 6 = 8 h–10 h, 7 = 10 h–12 h, 8 = 12 h–14 h, 9 = 14 h–16 h, 10 = 16 h–18 h, 11 = 18 h–20 h, 12 = over 20 h; M = 9.79, S.D. = 3.39) in a prescreening survey. After that, participants arrived at the scan session with their friend, and met a gender-matched confederate (stranger). Pre-scan impressions were measured by questions asking the participants and their friend to rate how trustworthy they felt their friend and the stranger to be, and how close they felt to their friend and to the stranger [“Pre-scan evaluation”]. All three people were then instructed together about the game structure; real participants were escorted to a separate scanning area. There, participants were told that they were assigned to play as Player 2 in the game and that they would make trustworthiness and closeness ratings about Player 1 at the end of each trial. Participants then completed 8 practice trials of the game and entered the scanner. Once inside the scanner, participants completed 192 trials of the game (16 trials in each of 12 runs, total time = 74 min 24 s), while functional scans were acquired.

After completing the Social Judgment Task, participants were presented with two runs of a theory-of-mind localizer task (ToM; 10 trials in each run; total time = 9 min 4 s) (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011), composed of conditions in which participants had to infer another person's mental states (“belief” condition) or physical representations of an object (“photo” condition). Afterwards, participants exited the scanner, made the general trustworthiness and closeness ratings again for their friend and the stranger [“Post-scan evaluation”],[12] and were debriefed and compensated.

To assess the number of participants' close relationships, we contacted all participants 2–7 months after the scanning session for a follow-up survey. Sixteen out of 24 total participants completed this survey. Although this small sample size requires conservative interpretation

[12]The post-scan survey included exploratory items: how negative/positive, aroused, and surprised participants felt when their friend or the stranger gave or took a certain amount; how participants would explain their friend's and the stranger's positive and negative behaviors; how much participants trusted other people in their daily life; how participants would plot varying relationships (ranging from stranger to best friend) on a scale; and demographic questions. We did not analyze these exploratory items with the exceptions of how participants would explain their friend's and the stranger's positive and negative behaviors (Supplementary Section 8), how much they trusted other people in their daily life (Supplementary Section 5), and how participants would plot varying relationships (ranging from stranger to best friend), to rule out potential confounding effects (Supplementary Section 8).

for the following findings (see Yarkoni, 2009 for relevant discussion), we proceed in analyzing the data as initial evidence regarding the association between mentalizing activity and maintenance of close relationships. Participants who responded to our follow-up survey answered questions about the number of their friends (1 = 0–2 friends; 2 = 3–5 friends; 3 = 6–9 friends; 4 = 10–19 friends; 5 = 20–49 friends; 6 = 50–99 friends; and 7 = More than 100 friends) and how often they make new friends (1 = Almost every week; 2 = A few weeks; 3 = Every month; 4 = A few months; 5 = Every year; 6 = A few years; 7 = About every 10 years; and 8 = More than every 10 years) along with other items.[13] A sensitivity power analysis using a significance level $\alpha = 0.05$ and power $\beta = 0.80$ showed that this sample size was sufficient to detect correlations ≥0.64 ($N = 16$) or ≥ 0.54 ($N = 24$).[14]

3.1.4. FMRI acquisition and preprocessing—We used a 3T Siemens Magnetom Prisma-FIT scanner outfitted with a 32-channel head coil at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at the Massachusetts Institute of Technology. Functional scans were acquired while participants were playing the Social Judgment Task and the ToM Task. Thirty-two 3 × 3 × 3 mm slices of gradient echo T2* weighted echo-planar images (EPI) provided whole brain coverage (TR = 2 s, TE = 30 ms, flip angle = 90°). Before the functional scans, high-resolution T1-weighted anatomical scans were acquired (TR = 2.53 s, TE = 1.69 ms) while participants were looking at a blank screen.

Brain data were analyzed using Analysis of Functional Neural Images (AFNI; AFNI_16.2.06 version) software (Cox, 1996). The first six functional scans before the task of each run were removed to compensate for magnet stabilization. All other images were de-obliqued, concatenated across runs, slice timing corrected (using the first slice as a reference), motion corrected (using the third volume as a reference and Fourier interpolation), spatially smoothed (using a 3D isotropic Gaussian kernel of an 8 mm full width at half maximum), normalized by the average activity over the entire task to generate percent signal change, and high-pass filtered omitting frequencies < .01 Hz (process as described in Wu, Samanez-Larkin, Katovich, & Knutson, 2014).

A spherical VOI (radius = 8 mm) centered on the RTPJ coordinates [57, −58, 19] derived from the whole-brain ToM t-test map contrasting belief condition versus photo condition was constructed (See Supplementary Section 10 for ToM findings).[15] Percent signal change (PSC) data within this VOI were extracted for each condition within the 2 (Agent: Friend, Stranger) X 2 (Valence: Taking, Giving) X 2 (Task: Closeness, Trustworthiness) X 2 (Amount: Low [$5, $10], High [$15, $20]) design. Since we found similar patterns across Task and Amount levels, PSC data were averaged again across Task and Amount levels.

[13]Other items measured in Study 2 include how much participants liked their friend, participants' most pleasant and unpleasant memory with their friend, and how likely they would be to play an economic game with their friend versus a random person. We analyzed two of these items for exploratory purposes: how much participants liked their friend (See Supplementary Section 3) and participants' pleasant and unpleasant memory recall (See Supplementary Section 9).

[14]Since eight participants did not respond to our request to complete the follow-up survey, this was the maximum level of power we could achieve. Post-hoc power analyses revealed that our power ranged from 0.45 to 0.83. Since G*Power does not provide an option for ordinal regressions, we used correlation models to calculate power.

[15]Other brain regions in the ToM network (dorsomedial prefrontal cortex, left TPJ, and precuneus) did not show the same patterns as RTPJ.

Thus, final RTPJ activity values were averaged within friend-taking, friend-giving, stranger-taking, and stranger-giving conditions. Sampling was delayed by 4 s to account for the hemodynamic lag to peak (Knutson, Rick, Wimmer, Prelec, & Loewenstein, 2007).[16]

3.2. Results

To test whether changes in participants' evaluations about friend versus stranger were associated with the reported number of friendships, we created two indices. First, we subtracted participants' pre-scan evaluation from their post-scan evaluation, creating closeness and trustworthiness updating scores for friend and stranger (See Supplementary Section 6c for the pre-scan versus post-scan evaluations comparisons). Then we subtracted updating scores for stranger from those for friend to control for an overall tendency to update before and after the game, respectively for closeness and trustworthiness, generating "post minus pre closeness updates" and "post minus pre trustworthiness updates". An ordinal regression on the number of friends participants reported having with post minus pre closeness updates [friend-stranger], controlling for how often participants made new friends,[17] revealed that the less negatively participants updated friend-closeness ratings compared to stranger-closeness ratings, the more friends they reported having, Estimate = 0.88, S.E. = 0.37, Wald = 5.54, Odds ratio = 2.41, $p = .019$, 95% CI for Estimate = [0.15, 1.61] (Fig. 3A; see Supplementary Section 12 for the effect of how often participants made new friends). These patterns remain significant even after we controlled for the number of hours per week participants spent with their friend (Supplementary Section 5). We ran the same analysis after substituting post minus pre closeness updates with post minus pre trustworthiness updates [friend-stranger], and found that biased post minus pre trustworthiness updates in favor of one's friend were also associated with reports of having more friends, Estimate = 1.36, S.E. = 0.54, Wald = 6.33, Odds ratio = 3.88, $p = .012$, 95% CI for Estimate = [0.30, 2.41].

Next, we tested whether participants' RTPJ activity during the Social Judgment Task was associated with the degree to which they updated their impressions about friend versus stranger, as well as how many friends they reported having. RTPJ activity from the friend-taking, friend-giving, stranger-taking, and stranger-giving conditions were entered in the model. The regression analysis on participants' post minus pre closeness updates [friend-stranger] revealed that only activity from the friend-taking condition significantly explained updates in closeness ratings, B = −9.63, S.E. = 3.57, $\beta = -0.62$, $t = -2.70$, $p = .014$ (Fig. 3B). RTPJ activity from the other conditions was not significantly associated with post minus pre closeness impression updates, $|\beta s| < 0.30$, $ps > 0.23$. Controlling for the number of hours per week participants spent with their friend did not change the results (Supplementary Section 5). We ran the same model after substituting post minus pre closeness updates with post minus pre trustworthiness updates [friend-stranger], but RTPJ activity from the friend-taking condition did not show a significant effect, $\beta = -0.15$, $p = .492$ (See Supplementary Section 13).

[16] We found similar patterns of findings with the beta values extracted from a general linear model (Supplementary Section 11).
[17] In additional analyses we found similar patterns without controlling for how often participants made new friends.

Finally, to test for associations between participants' RTPJ activity during the Social Judgment Task and maintenance of close relationships,[18] we conducted an ordinal regression on the number of friends participants reported having with RTPJ activity from the friend-taking condition as the predictor,[19] controlling for how often participants made new friends.[20] We found that greater RTPJ activity in response to the friend's taking behavior was associated with reports of having fewer friends, Estimate = −0.10, S.E. = 0.05, Wald = 4.44, Odds ratio = 0.91, $p = .035$, 95% CI for Estimate = [−0.19, −0.01] (Fig. 3C; see Supplementary Section 12 for the effect of how often participants made new friends). These effects also emerged in analyses controlling for the number of hours per week participants spent with their friend, and RTPJ activity in the stranger-taking, friend-giving, and stranger-giving conditions (See Supplementary Section 3e for correlations between variables in regression models; see Supplementary Section 3f for correlations between measures of participants' prior experiences and their updating for friends).

3.3. Study 2 discussion

Study 2 provided a conceptual replication and extension of Study 1. Participants who engaged in less negative updating of friend-closeness compared to stranger-closeness reported having a greater number of friends. Moreover, these participants also showed reduced RTPJ activity when their friend took money from them, which was in turn associated with having more friends. These findings indicate that participants who disengaged in mentalizing in response to their friend's negative behavior during the Social Judgment Task reported maintaining more social relationships in real life. These findings also suggest that selectively discounting negative information about close others, manifested in reduced impression updating, and indexed by decreased RTPJ activity, can be linked to maintaining close relationships. Moreover, the same patterns emerged in analyses controlling for the number of hours per week participants spent with their friend, suggesting that the relationship between update resistance and relationship maintenance was not simply due to the strength of participants' priors about their friend.

4. General discussion

Is ingroup bias in the form of asymmetric updating for friends versus strangers accompanied by social success in the form of more friendships? In two studies, we found that selective processing of information about friends within an experimental task was associated with reporting a greater number of friendships in real life. In Study 1, participants who were reluctant to update how close they felt to their friend after imagining their friend performing negative behaviors reported having more friends. Study 2 explored a potential mechanism for this effect in examining activity in RTPJ, a region that has been implicated in encoding and integrating mental state information for moral judgment (Decety & Cacioppo, 2012; Young, Cushman, Hauser, & Saxe, 2007). We found that reduced RTPJ activity in response to a friend's negative behavior was associated with reporting a greater number of friends in

[18]Again, we would like to emphasize that this finding should be interpreted conservatively. See Yarkoni, 2009 for relevant discussion.
[19]To facilitate the interpretation of the odds ratio, we multiplied RTPJ PSC by 100 (converted from ranging between [−0.26 ~ 0.19] to ranging between [−26 ~ 19]).
[20]In additional analyses we found similar patterns without controlling for how often participants made new friends.

real life. Moreover, reduced RTPJ activity in response to a friend's negative behavior also accounted for less negative updating in closeness ratings for one's friend versus a stranger. Together, the findings of Studies 1 and 2 suggest that neglecting the negative behavior of friends may be associated with maintaining close relationships.

We offer two possible accounts of the association between the selective discounting of close others' negative behavior and relationship maintenance. First, selectively discounting negative information about close others may facilitate relationship maintenance and perhaps the broadening of one's social network. Previous literature has demonstrated that the quality of relationships is more powerfully impacted by negative versus positive events (Baumeister et al., 2001), possibly because negative behaviors are more easily recognized and reciprocated (Acitelli et al., 1993; Alves et al., 2015; Levenson & Gottman, 1985; Rusbult et al., 1986; see also Kellermann, 1984). Other work has shown that people's greater motivation to bond with their partners leads them to discount their partners' unresponsive behavior; instead people blame their own failure to disclose their needs to their partners, a pattern that in turn predicts greater trust in partners (Lemay & Clark, 2015; Lemay & Melville, 2014). Thus, selectively discounting close others' negative behavior may help people to build and maintain relationships. Alternatively, in a reversal of this relationship, it may be the case that people who manage many social relationships are under greater cognitive load, which can lead to greater ingroup bias (Dreu, Dussel, & Velden, 2015). Future work is needed to explore these different accounts.

In both studies, we used two different impression dimensions: closeness and trustworthiness. Closeness concerns the relationship and may be influenced by a number of subjective factors such as shared experiences and similarity. Trustworthiness, meanwhile, concerns the character of the target and may be influenced more by the nature of the target's behaviors. Although closeness and trustworthiness are highly correlated, we considered the initial possibility that people are more resistant to updating closeness ratings compared to trustworthiness ratings especially for friends. This was indeed the case in Study 1, though similar patterns emerged for closeness and trustworthiness in Study 2. We note that, in Study 1, participants were asked to *imagine* that their friend committed some bad behavior that also did not directly impact the participants themselves. By contrast, in Study 2, participants believed they were interacting with their friend (whom they brought with them in person to the scan session) in real time, and also participants were the direct target of their friend's behaviors. Given these differences, we think it is possible that the friend versus stranger contrast and the negative behaviors were more salient in Study 2, leading to more robust effects across impression dimensions.

Although general believability was enhanced in Study 2 as described above, a separate question is whether participants perceived new information about their friend versus the stranger as similarly useful, reliable, or informative. Put plainly, participants already have access to abundant information about their friend, and any new data, especially data that appear inconsistent with their strong prior impressions, might be perceived as less useful or reliable, compared to new data about the stranger. Thus, in neglecting to update their impressions of their friend, participants might not be showing *bias* per se but rather a form of Bayesian-rational updating, that is, discounting new evidence inconsistent with strong priors

(Hahn & Harris, 2014). Importantly, however, controlling for the amount of participants' prior experience with their friend (measured by the number of years they have known their friend [Study 1] and the number of hours per week they spent with their friend [Studies 1 and 2]) did not change the results. Moreover, participants' prior experience with their friend was not associated with the extent of biased updating either. Thus, above and beyond the strength of participants' priors, participants' motivation to maintain their impressions of their friend appears to be a key factor. Furthermore, discounting new evidence that is not consistent with one's prior knowledge might also rely on being able to generate an alternative explanation to account for the unexpected behavior (e.g., my friend took money from me because she will share the spoils after the experiment is over) (Gershman, 2019); elsewhere, we have proposed that generating this account would require increased mentalizing (Kim, Park, & Young, 2019; Park, Kim, & Young, in press). Given our finding of *reduced* RTPJ activity associated with biased impression updating, we think it is more likely that participants in the present research were motivated to preserve their impressions of their friends and to protect their relationships, in a biased fashion.

We recognize several limitations of the current work. First, because of the correlational nature of the data, additional research is required to investigate the causal nature of the observed effect. Directly manipulating participants' motivation to update (or not) may be one approach to take in future work. Second, while Study 2 represents an initial attempt to explore the neural basis of biased updating for close others, its sample size is limited. Replicating the findings with a larger sample is an important next step. Finally, this research focuses on social network size. An open question is whether centrality within one's social network (Weaverdyck & Parkinson, 2018) may have a similar or different effect. Research exploring centrality and other network features can clarify how social processing affects (or is affected by) social status and success.

The present work raises a number of other interesting questions for future work as well. For example, what kind of evidence and how much of it would be required for people to update their impressions of their friends and reconsider existing relationships? One possibility is that sufficient evidence of a friend's immorality may lead to a “turning point” in a relationship, when one confronts the choice to leave the relationship. Indeed, previous research on partner choice has suggested the primacy of moral signals: people care about whether potential partners are cooperative (Hardy & Van Vugt, 2006; Jordan, Hoffman, Nowak, & Rand, 2016; Pleasant & Barclay, 2018; Sylwester & Roberts, 2010), and people are also more likely to choose fair versus rich partners (Raihani & Barclay, 2016). Other work on essentialism reveals that people perceive moral traits to be especially essential to identity (Strohminger & Nichols, 2015) and base their impressions of others more on morality than any other characteristics, such as warmth or competence (Brambilla, Carraro, Castelli, & Sacchi, 2019; Goodwin, 2015). Future work can consider whether people are less likely to show biased impression updating for non-moral traits.

Accurately forming and revising impressions about others is critical for effectively navigating the social world. However, resisting impression updates may also be linked to maintaining strong relationships. The current research advances our understanding of the potential benefits of ingroup bias, at a mechanistic level, such as social bonding within one's

network. We are optimistic that future work may uncover ways to apply these findings to enhancing intergroup cooperation and negotiation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded by a grant from the John Templeton Foundation [5107321] awarded to Liane Young and NIH Shared instrumentation grant [S10OD021569] awarded to McGovern Institute for Brain Research at MIT. The authors would like to thank Mauricio Delgado, Dominic Fareri, Peter Mende-Siedlecki, Minjae Kim, Emma Alai, and Mariel Kronitz for their contributions.

References

Acitelli LK, Douvan E, & Veroff J (1993). Perceptions of conflict in the first year of marriage: How important are similarity and understanding? Journal of Social and Personal Relationships, 10, 5–19.

Alves H, Unkelbach C, Burghardt J, Koch AS, Krüger T, & Becker VD (2015). A density explanation of valence asymmetries in recognition memory. Memory and Cognition, 43, 896–909. [PubMed: 25772462]

Aron A, Aron EN, & Smollan D (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. Journal of Personality and Social Psychology, 63(4), 596–612.

Balliet D, Wu J, & Dreu CKD (2014). Ingroup favoritism in cooperation: A metaanalysis. Psychological Bulletin, 140(6), 1556–1581. [PubMed: 25222635]

Baron SG, Gobbini MI, Engell AD, & 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. [PubMed: 21030482]

Baumeister RF, Bratslavsky E, Finkenauer C, & Vohs KD (2001). Bad is stronger than good. Review of General Psychology, 5(4), 323–370.

Baumgartner T, Götte L, Gügler R, & Fehr E (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. Human Brain Mapping 33, 1452–1469. [PubMed: 21574212]

Boorman ED, O'Doherty JP, Adolphs R, & Rangel A (2013). The behavioral and neural mechanisms underlying the tracking of expertise. Neuron, 80, 1558–1571. [PubMed: 24360551]

Brambilla M, Carraro L, Castelli L, & Sacchi S (2019). Changing impressions: Moral character dominates impression updating. Journal of Experimental Social Psychology, 82, 64–73.

Brewer MB (1999). The psychology of prejudice: Ingroup love or outgroup hate? Journal of Social Issues, 55(3), 429–444.

Cox RW (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research, 29(3), 162–173. [PubMed: 8812068]

Decety J, & Cacioppo S (2012). The speed of morality: A high-density electrical neuroimaging study. Journal of Neurophysiology, 108, 3068–3072. [PubMed: 22956794]

Dodell-Feder D, Koster-Hale J, Bedny M, & Saxe R (2011). FMRI item analysis in a theory of mind task. Neuroimage, 55, 705–712. [PubMed: 21182967]

Dreu CKWD, Dussel DB, & Velden FST (2015). In intergroup conflict, self-sacrifice is stronger among pro-social individuals, and parochial altruism emerges especially among cognitively taxed individuals. Frontiers in Psychology, 6, 572. [PubMed: 25999888]

Fareri DS, Chang LJ, & Delgado MR (2012). Effects of direct social experience on trust decisions and neural reward circuitry. Frontiers in Neuroscience, 6, 148. [PubMed: 23087604]

Fareri DS, Chang LJ, & Delgado MR (2015). Computational substrates of social value in interpersonal collaboration. The Journal of Neuroscience, 35(21), 8170–8180. [PubMed: 26019333]

Gershman SJ (2019). How to never be wrong. Psychonomic Bulletin and Review, 10.3758/s13423-018-1488-8.

Goodwin GP (2015). Moral character in person perception. Current Directions in Psychological Science, 24(1), 38–44.

Hahn U, & Harris AJL (2014). What does it mean to be biased: Motivated reasoning and rationality In Brian HR (Ed.). Psychology of learning and motivation. San Diego, CA: Academic Press.

Hardy CL, & Van Vugt M (2006). Nice guys finish first: The competitive altruism hypothesis. Personality and Social Psychology Bulletin, 32(10), 1402–1413. 10.1177/0146167206291006. [PubMed: 16963610]

Howard JW, & Rothbart M (1980). Social categorization and memory for in-group and out-group behavior. Journal of Personality and Social Psychology, 38(2), 301–310.

Hughes BL, Zaki J, & Ambady N (2017). Motivation alters impression formation and related neural systems. Social Cognitive and Affective Neuroscience, 12(1), 49–60. [PubMed: 27798250]

Jordan JJ, Hoffman M, Nowak MA, & Rand DG (2016). Uncalculating cooperation is used to signal trustworthiness. Proceedings of the National Academy of Sciences, 113(31), 8658–8663. 10.1017/CBO9781107415324.004.

Kellermann K (1984). The negativity effect and its implications for initial interaction. Communication Monographs, 51, 37–55.

Kim M, Mende-Siedlecki P, Anzellotti S, & Young L The neural signatures of updating strong and weak impressions. Manuscript in preparation.

Kim M, Park B, & Young L (2019). The psychology of rational versus non-rationalbelief updating. Manuscript submitted for publication

Kliemann D, Young L, Scholz J, & Saxe R (2008). The influence of prior record on moral judgment. Neuropsychologia, 46, 2949–2957. [PubMed: 18606175]

Knutson B, Rick S, Wimmer GE, Prelec D, & Loewenstein G (2007). Neural predictors of purchases. Neuron, 53(1), 147–156. [PubMed: 17196537]

Koster-Hale J, & Saxe R (2013). Theory of mind: A neural prediction problem. Neuron, 79, 836–848. [PubMed: 24012000]

Lakens D (2014). Performing high-powered studies efficiently with sequential analyses: Sequential analyses. European Journal of Social Psychology, 44, 701–710.

Lemay EP, & Clark MS (2015). Motivated cognition in relationships. Current Opinion in Psychology, 1, 72–75.

Lemay EP, & Melville MC (2014). Diminishing self-disclosure to maintain security in partner's care. Journal of Personality and Social Psychology, 106(1), 37–57. [PubMed: 23937643]

Levenson RW, & Gottman JM (1985). Physiological and affective predictors of change in relationship satisfaction. Journal of Personality and Social Psychology, 49, 85–94. [PubMed: 4020618]

Mende-Siedlecki P, Baron SG, & Todorov A (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. The Journal of Neuroscience, 33(50), 19406–19415. [PubMed: 24336707]

Mende-Siedlecki P, Cai Y, & Todorov A (2013). The neural dynamics of updating person impressions. Social Cognitive and Affective Neuroscience, 8(6), 623–631. [PubMed: 22490923]

Mende-Siedlecki, & Todorov A (2016). Neural dissociations between meaningful and mere inconsistency in impression updating. Social Cognitive and Affective Neuroscience, 11(9), 1489–1500. [PubMed: 27217118]

Monroe AE, & Malle BF (2019). People systematically update moral judgments of blame. Journal of Personality and Social Psychology, 10.1037/pspa0000137.

Park B, Fareri D, Delgado M, & Young L (2019). How theory-of-mind brain regions process prediction error across relationship contexts. Manuscript submitted for publication

Park B, Kim M, & Young L (in press). An examination of accurate versus “biased” mentalizing in moral and economic decision-making In Gilead M & Ochsner KN (Eds). The Neural Basis of Mentalizing. New York: Springer.

Pleasant A, & Barclay P (2018). Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. Psychological Science, 29(6), 868–876. [PubMed: 29708860]

Raihani NJ, & Barclay P (2016). Exploring the trade-off between quality and fairness in human partner choice. Royal Society Open Science, 3(11), 160510 10.1098/rsos.160510. [PubMed: 28018638]

Rusbult CE, Johnson DJ, & Morrow GD (1986). Impact of couple patterns of problem solving on distress and nondistress in dating relationships. Journal of Personality and Social Psychology, 50, 744–753.

Schiller B, Baumgartner T, & Knoch D (2014). Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. Evolution and Human Behavior, 35, 169–175.

Sherif M (1966). In common predicament Boston: Houghton Mifflin.

Shin YS, Kim HY, & Han S (2014). Neural correlates of social perception on response bias. Brain and Cognition, 88, 55–64. [PubMed: 24859091]

Strohminger N, & Nichols S (2015). Neurodegeneration and identity. Psychological Science, 26(9), 1469–1479. [PubMed: 26270072]

Sylwester K, & Roberts G (2010). Cooperators benefit through reputation-based partner choice in economic games. Biology Letters, 6(5), 659–662. 10.1098/rsbl.2010.0209. [PubMed: 20410026]

Tajfel H (1982). Social psychology of intergroup relations. Annual Review of Psychology, 33, 1–39.

Taylor SE, & Brown JD (1988). Illusion and well-being: A social psychological perspective on mental health. Psychological Bulletin, 103(2), 193–210. [PubMed: 3283814]

Thornton MA, & Mitchell JP (2018). Theories of person perception predict patterns of neural activity during mentalizing. Cerebral Cortex, 28, 3505–3520. [PubMed: 28968854]

Weaverdyck ME, & Parkinson C (2018). The neural representation of social networks. Current Opinion in Psychology, 24, 58–66. [PubMed: 29886253]

Wohl MJA, & Branscombe NR (2009). Group threat, collective angst, and ingroup forgiveness for the war in Iraq. Political Psychology, 30(2), 193–217.

Wu CC, Samanez-Larkin GR, Katovich K, & Knutson B (2014). Affective traits link to reliable neural markers of incentive anticipation. NeuroImage, 84, 279–289. [PubMed: 24001457]

Xiao YJ, Coppin G, & Van Bavel J (2016). Clarifying the role of perception in intergroup relations: Origins of bias, components of perception, and practical implications. Psychological Inquiry, 27(4), 358–366. 10.1080/1047840X.2016.1237822.

Yarkoni T (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power–commentary on Vul et al. (2009). Perspectives on Psychological Science, 4(3), 294–298. [PubMed: 26158966]

Young L, Cushman F, Hauser M, & Saxe R (2007). The neural basis of the interaction between theory of mind and moral judgment. Proceedings of the National Academy of Sciences of United States of America, 104(20), 8235–8240.

Zaki J, & Ochsner K (2011). Reintegrating the study of accuracy into social cognition research. Psychological Inquiry, 22(3), 159–182.

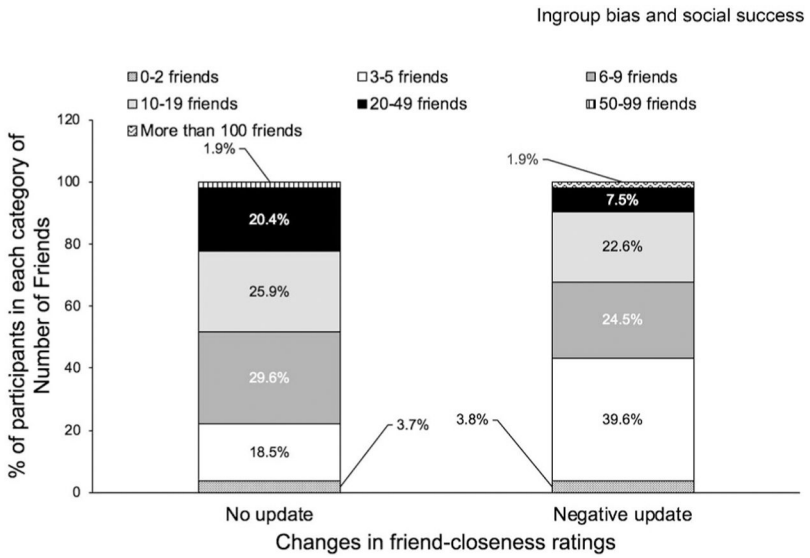

Fig. 1. Distribution of participants' reported number of friends by updates in friend-closeness ratings.

Participants who did not update friend-closeness ratings reported having a greater number of friends. For visualization, we depicted the % of participants who chose each category.

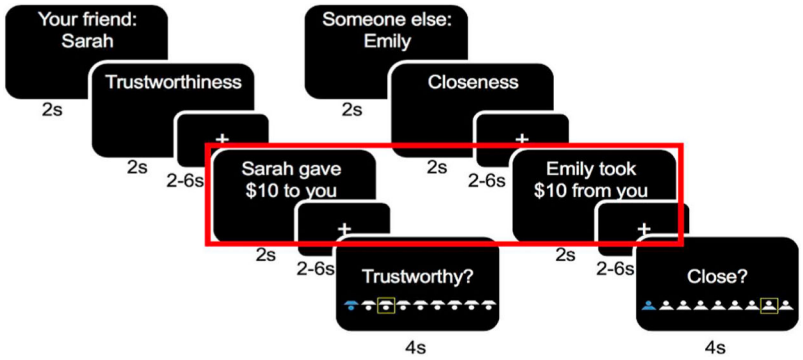

Fig. 2. Representative trials of Social Judgment Task.

Participants viewed the name of Player 1 (2 s); the rating that they would make in the trial (2 s); a jittered fixation cross (2–6 s); the decision of the Player 1 (2 s); a jittered fixation cross (2–6 s); and made their ratings (4 s). Each trial was divided by another jittered fixation cross (2–6 s). For further analyses, we focused on the phase when participants viewed Player 1's decision, marked in the red box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

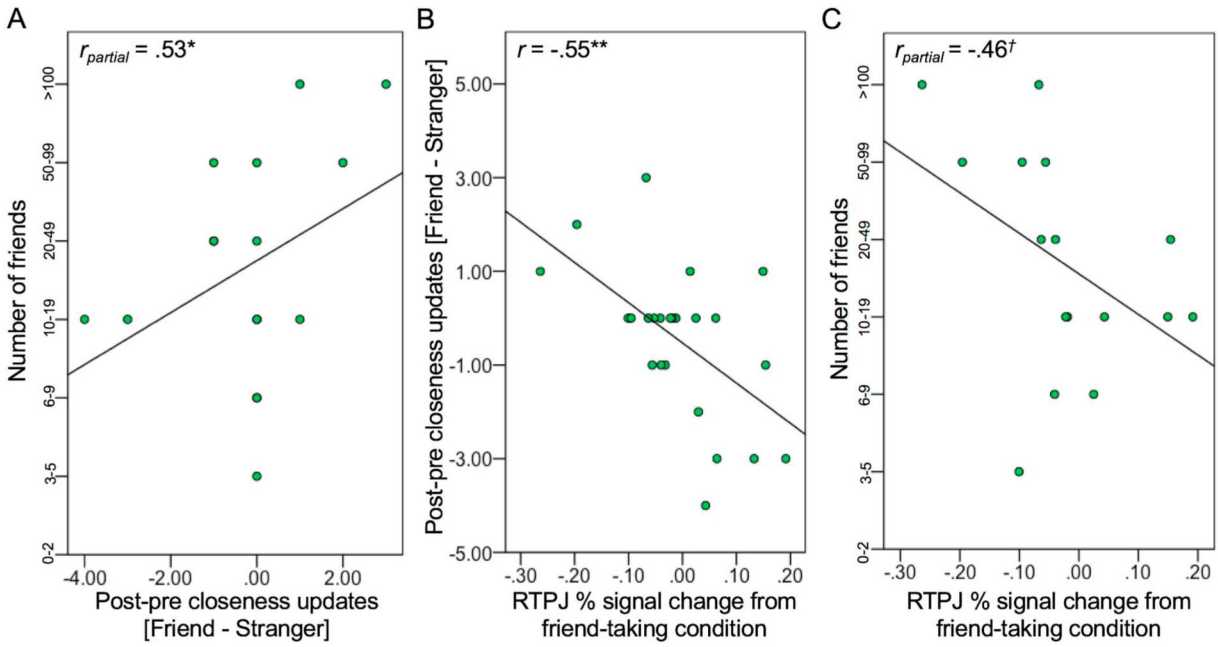

Fig. 3. Associations between post minus pre closeness updates, RTPJ activity, and the number of friends in Study 2.

(A) Participants who updated friend-closeness ratings less negatively compared to stranger-closeness ratings reported having a greater number of friends. (B) Participants who showed decreased RTPJ activity when their friend took money from them engaged in less negative updating for friend-closeness ratings compared to stranger-closeness ratings. (C) Participants who showed decreased RTPJ activity when their friend took money from them reported having a greater number of friends. For the visualization purpose, the zero-order correlations are depicted. $r_{partial}$ = Correlation coefficients after controlling for how often participants make new friends. $\dagger p < .10$, $*p < .05$, $**p < .01$.