Supplementary Materials for

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

**Park, B. and Young, L.**

*Supplementary Section 1: Analyses with the full sample in Study 1*

We ran the analyses without excluding any participants (N = 125) and found similar patterns as reported in the main paper. First, ordinal regression with the categorized closeness updating for friend while imagining the behaviors of their friend or a stranger ("Negative update" = -1, "No update" = 0, "Positive update" = +1) on the number of friends that participants reported having revealed that those who never updated, and those who positively updated, reported having a directionally greater number of friends compared to those who negatively updated their friend-closeness ratings (Estimate = .38, S.E. = .28, Wald = 1.88, Odds ratio = 1.47, $p$ = .171, 95% CI for Estimate = [-.17, .93]). Similarly, when switching the categorized updates participants showed while they were imagining the behaviors with their changes in closeness ratings between post-exposure evaluation (i.e., after they imagined all behaviors) and pre-exposure evaluation (i.e., before they imagined any behaviors), we found that participants who less negatively updated their friend-closeness ratings reported having more friends (Estimate = .52, S.E. = .21, Wald = 5.80, Odds ratio = 1.67, $p$ = .016, 95% CI for Estimate = [.10, .93]).

*Supplementary Section 2: Stimuli pretesting in Study 1*

This pretesting was conducted for a separate fMRI project (Kim, Mende-Siedlecki, Anzellotti, & Young, in prep). In total, 1,700 vignettes describing moral or immoral behaviors were pre-tested (1,200 from Mende-Siedlecki, Baron, & Todorov, 2013; 500 new generated) in terms of the moral relevance ("How morally relevant is this behavior?" ranging from 0 [not at all relevant] to 8 [extremely relevant]), perceived frequency ("How many people, out of 100, have done this behavior?"), arousal ("How emotionally arousing is this behavior?" ranging from 0 [not at all] to 8 [extremely]), and valence ("How positive/negative is this behavior?" ranging from -4 [extremely negative] to 4 [extremely positive]), along with other items for the purpose of this project. We took absolute values of valence ratings to compare the magnitude of the valence between moral and immoral behaviors. Participants' pretested ratings of each story were averaged to a single value for each variable. Please see (Kim et al., in prep) for further information.

A series of independent t-tests on the eight behaviors selected for the present research revealed that the moral and immoral behaviors were not significantly different in moral relevance (moral M = 6.13, S.E. = .20; immoral M = 6.15, S.E. = .10; $p$ = .934), perceived frequency (moral M = 20.67, S.E. = 2.05; immoral M = 22.78, S.E. = 1.41; $p$ = .428), arousal (moral M = 4.52, S.E. = .35; immoral M = 3.89, S.E. = .23; $p$ = .186), or absolute valence (moral M = 2.32, S.E. = .29; immoral M = 2.28, S.E. = .27; $p$ = .931).

*Supplementary Section 3: Correlations between variables*

a. Correlation coefficients (Pearson's *r*) between the number of friends participants reported having on the one hand and Inclusion of Other in Self (IOS) ratings, how much participants liked their friend (liking), and how many hours spent per week with their friend (hours) on the other hand.

| | | Number of friends |
|---|---|---|
| Study 1 | IOS | .12 |
| | Liking | .11 |
| | Hours | -.03 |
| Study 2 | IOS | -.15 |
| | Liking | .35 |
| | Hours | -.46† |

Note. IOS = Inclusion of Other in Self scale. †$p < .10$.

b. Correlation coefficients (Pearson's *r*) between the pre-exposure ratings in Study 1

| Pre-exposure ratings | friend-closeness | friend-trustworthiness | stranger-closeness | stranger-trustworthiness |
|---|---|---|---|---|
| friend-closeness | - | | | |
| friend-trustworthiness | .59*** | - | | |
| stranger-closeness | -.21* | -.20* | - | |
| stranger-trustworthiness | .10 | .06 | .36*** | - |

Note. Correlations between pre-exposure ratings (Study 1). *$p < .05$, ***$p < .001$.

c. Correlation coefficients (Pearson's *r*) between participants' updating for their friend and log-transformed reaction time

Study 1

| | Reaction time when imagining negative behaviors [Friend-Stranger] |
|---|---|
| Closeness updating [while imagining the behaviors] | -.08 |
| Trustworthiness updating [while imagining the behaviors] | .12 |
| Closeness updating [Post minus pre] | .08 |
| Trustworthiness updating [Post minus pre] | -.01 |

Note. Reaction times were log-transformed and collapsed across closeness and trustworthiness ratings. Closeness updating while participants were imagining the behaviors is categorized (0 = Never update, -1 = Negative update, +1 = Positive update) as described in the main paper.

Study 2

In Study 2, we subtracted ratings on a given trial from ratings on the previous trial respectively for friend-closeness, friend-trustworthiness, stranger-closeness, and stranger-trustworthiness conditions, taking the absolute value of these scores as an index of trial-by-trial updating. We conducted a linear mixed-effects regression with the log-transformed trial-by-trial RT on trial-by-trial updating, including individual participants as random effects. The effect of RT on trial-by-trial updating was not significant, B = -.03, S.E. = .06, $t$ = -.49, $p$ = .626 (for further analyses, please see Park, Fareri, Delgado, & Young, 2019).

d. Correlation coefficients (Pearson's *r*) between covariates in regression models in Study 1.

*Updates while participants were imagining the behaviors.*

| | Updates in friend-closeness | Updates in friend-trustworthiness | Updates in stranger-closeness | Updates in stranger-trustworthiness | How often people make new friends |
|---|---|---|---|---|---|
| Updates in friend-closeness | - | | | | |
| Updates in friend-trustworthiness | -.55*** | - | | | |
| Updates in stranger-closeness | .19† | -.25* | - | | |
| Updates in stranger-trustworthiness | -.25* | .45*** | -.16† | - | |
| How often people make new friends | -.11 | .03 | .18† | .13 | - |

Note. Correlations between variables controlled in regression models (Study 1). Updates in friend-closeness and stranger-closeness are categorized (larger numbers indicate more positive updates; 0 = Never update, -1 = Negative update, +1 = Positive update). Other updating scores are computed by subtracting participants' ratings for the last negative story from their ratings for

the last positive story; larger numbers indicate more negative updates. *†p < .10, *p < .05, ***p < .001*.

*Updates before and after participants imagined all behaviors.*

| | Updates in friend-closeness | Updates in friend-trustworthiness | Updates in stranger-closeness | Updates in stranger-trustworthiness | How often people make new friends |
|---|---|---|---|---|---|
| Updates in friend-closeness | - | | | | |
| Updates in friend-trustworthiness | .53*** | - | | | |
| Updates in stranger-closeness | .01 | .04 | - | | |
| Updates in stranger-trustworthiness | .04 | .04 | .34*** | - | |
| How often people make new friends | -.04 | .004 | -.06 | -.17† | - |

Note. Correlations between variables controlled in regression models (Study 1). Updating scores are computed by subtracting participants' 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). *†p < .10, ***p < .001*.

e. Correlation coefficients (Pearson's *r*) between covariates in regression models in Study 2.

*Behavioral analyses*

| | Closeness updates [Friend – Stranger] | Trustworthiness updates [Friend – Stranger] | How often people make new friends |
|---|---|---|---|
| Closeness updates [Friend – Stranger] | - | | |
| Trustworthiness updates [Friend – Stranger] | .53** | - | |
| How often people make new friends | .16 | .50† | - |

Note. Correlations between behavioral covariates in Study 2. *†p < .10, **p < .01*

*Neural analyses*

| | RTPJ activity [friend-taking] | RTPJ activity [friend-giving] | RTPJ activity [stranger-taking] | RTPJ activity [stranger-giving] | How often people make new friends |
|---|---|---|---|---|---|
| RTPJ activity [friend-taking] | - | | | | |
| RTPJ activity [friend-giving] | .46* | - | | | |

| | | | | | |
|---|---|---|---|---|---|
| RTPJ activity [stranger-taking] | .36† | .42* | - | | |
| RTPJ activity [stranger-giving] | .26 | .06 | -.35† | - | |
| How often people make new friends | .08 | -.42 | .03 | -.12 | - |

Note. Correlations between neural covariates and how often people make new friends in Study 2. *†p < .10, *p < .05*

f. Correlation coefficients (Pearson's *r*) between participants' updating for their friend and measures of prior experience with the friend

| | | Closeness updating | Trustworthiness updating |
|---|---|---|---|
| Study 1 | How many years participants have known their friend | .09 | -.17 ($p$ = .081) |
| | How many hours per week participants spend with their friend | -.02 | -.01 |
| Study 2 | How many hours per week participants spend with their friend | -.03 | .01 |

Note. Study 1 closeness updating index is categorized (0 = Never update, -1 = Negative update, +1 = Positive update) as described in the main paper; Study 2 closeness and trustworthiness updating metrics are controlled for updating for strangers, matching the indices used in the main paper.

*Supplementary Section 4. Descriptions and findings with exploratory survey items in Study 1.*

For exploratory purposes, in Study 1, after participants read all of the stories about their friend and a stranger, they also answered a series of questionnaires intended to capture different aspects of their social interactions. These questionnaires included the positive relations with others subscale (e.g., "I enjoy personal and mutual conversations with family members or friends") from the Psychological Well-being Scale (Ryff, 1989; Ryff & Keyes, 1995), the individual loyalty subscale (e.g., "If I make a promise to a friend, I will keep it") from the Individual and Group Loyalty Scale (Beer & Watson, 2009), Relational-Interdependent Self-Construal Scale (e.g., "When I think of myself, I often think of my close friends or family also") (Cross, Bacon, & Morris, 2000), and the Unidimensional Relationship Closeness Scale with the friend whose name they submitted to the survey (e.g., "My relationship with [friend's name] is close.") (Dibble, Levine, & Park, 2011), along with other items[1].

In exploratory analyses, we found that the more negatively participants updated their trustworthiness ratings for a stranger after reading negative stories about the stranger, the greater relational satisfaction (Pearson's $r = .28$, $p = .004$), greater loyalty ($r = .45$, $p < .001$), greater relational self-construal ($r = .28$, $p = .003$), and closer relationships with their friend ($r = .29$, $p = .003$) they had. These effects remained similar after controlling for updates in friend-closeness, friend-trustworthiness, and stranger-closeness ratings. These findings suggest that participants who are more satisfied with their current relationships in real life might be more likely to rate strangers more negatively within the experimental paradigm, perhaps resulting in a favorable comparison with their friends. It is also possible that participants who are more

[1] Other items included participants' most pleasant and unpleasant memory with their friend, how long is their oldest friendship, loyalty subscale from Moral Foundations Questionnaire (Graham, Haidt, & Nosek, 2008), extraversion and emotional stability subscales from Ten Item Personality Measure (Gosling, Rentfrow, & Swann, 2003), how much participants liked their friend, and how many hours they spend or communicate with their friend. Our findings remained similar after we statistically controlled for these variables. Additionally, we also measured how many followers they have on their social media account(s), but 24.3% - 60.4% of participants did not have the account(s) that we asked about (facebook, instagram, and twitter). Thus, to secure enough statistical power, these variables were not included in the analyses any further.

sensitive to others' negative behavior might be more selective in choosing friends, enhancing the quality of their close relationships and increasing commitment to these relationships.

*Supplementary Section 5: Effect of biased updating on the number of friends after controlling for different variables*

a. Study 1: Effect of friend-closeness updating while participants were imagining their friend's behavior on the number of friends they reported having, controlling for the following variables

| | Odds ratio | Estimate from ordinal regression [95% CI for Estimate] |
|---|---|---|
| IOS | 2.07 | .73 [.06, 1.39] |
| Liking | 2.04 | .71 [.05, 1.38] |
| Extraversion | 1.66 | .51 [-.17, 1.18] ($p$ = .143) |
| SES | 2.04 | .71 [.04, 1.38] |
| How many hours per week participants spent/ communicated with their friend | 2.06 | .73 [.06, 1.39] |
| How long participants have known their friend | 2.07 | .73 [.06, 1.40] |

Note. Friend-closeness updating was categorized ("No update" = 0, "Negative update" = -1, and "Positive update" = +1); SES = Socioeconomic status (low income, lower middle income, middle income, upper middle income, and upper income); statistics acquired from ordinal regressions.

b. Study 1: Effect of post minus pre friend-closeness updating on the number of friends participants reported having, controlling for the following variables

| | Odds ratio | Estimate from ordinal regression [95% CI for Estimate] |
|---|---|---|
| IOS | 1.84 | .61 [.13, 1.09] |

| | | |
|---|---|---|
| Liking | 1.83 | .60 [.13 1.08] |
| Extraversion | 1.77 | .57 [.09, 1.05] |
| SES | 1.87 | .63 [.15, 1.10] |
| How many hours per week participants spent/ communicated with their friend | 1.85 | .61 [.14, 1.09] |
| How long participants have known their friend | 1.88 | .63 [.15, 1.11] |

Note. Friend-closeness updating was calculated by subtracting 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); SES = Socioeconomic status (low income, lower middle income, middle income, upper middle income, and upper income); statistics acquired from ordinal regressions.

c. Study 2: Effect of post minus pre closeness updating on the number of friends participants reported having, controlling for the following variables

| | Odds ratio | Estimate from ordinal regression [95% CI for Estimate] |
|---|---|---|
| IOS | 2.33 | .84 [.12, 1.57] |
| Liking | 2.42 | .88 [.15, 1.61] |
| Trusting other people in daily life | 2.42 | .88 [.16, 1.60] |
| SES | 2.28 | .82 [.07, 1.57] |

| | | |
|---|---|---|
| How many hours per week participants spent with their friend | 2.21 | .80 [.02, 1.57] |

Note. SES = Socioeconomic status (low income, lower middle income, middle income, upper middle income, and upper income); statistics acquired from ordinal regressions controlling for how often participants made new friends.

d. Study 2: Effect of RTPJ activity from friend-taking condition on the number of friends participants reported having, controlling for the following variables

| | Odds ratio | Estimate from ordinal regression [95% CI for Estimate] |
|---|---|---|
| IOS | .91 | -.09 [-.18, -.01] |
| Liking | .90 | -.11 [-.21 -.01] |
| Trusting other people in daily life | .90 | -.11 [-.21, -.01] |
| SES | .86 | -.16 [-.28, -.03] |
| How many hours per week participants spent with their friend | .89 | -.12 [-.21, -.02] |

Note. SES = Socioeconomic status (low income, lower middle income, middle income, upper middle income, and upper income); statistics acquired from ordinal regressions controlling for how often participants made new friends.

*Supplementary Section 6: Impression updating*

a. Impression updating after reading positive stories in Study 1.

To explore participants' impression updating after reading only positive stories, we subtracted participants' pre-exposure ratings (before they imagined any behaviors) from their ratings in the second positive story, and ran paired t-tests comparing participants' changes in ratings for friend versus stranger. We found that participants updated more positively for a stranger than for their friend in both closeness and trustworthiness dimensions (closeness: friend M = .02, S.E. = .07, stranger M = .88, S.E. = .12; trustworthiness: friend M = .14, S.E. = .06, stranger M = 1.75, S.E. = .15), *ps* < .001. However, we would like to note that this effect might be due to the fact that participants already gave very high ratings for their friend before they imagined any behaviors (pre-exposure ratings for friend: closeness M = 7.40, S.D. = .79; trustworthiness M = 7.38, S.D. = .82; pre-exposure ratings for stranger: closeness M = 1.76, S.D. = 1.58; trustworthiness M = 4.12, S.D. = 1.54), and then they were presented with two positive stories where they could not give any higher ratings.

b. Comparisons between evaluations at the last positive story and those at the last negative story in Study 1.

To examine how much people updated after considering target agents' (hypothetical) negative behaviors, we ran a 2 (Agent: Friend, Stranger) X 2 (Task: Closeness, Trustworthiness) repeated ANOVA on participants' updates (ratings for the last positive story – ratings for the last negative story). We found a significant main effect of Agent, $F(1,109) = 14.56$, $p < .001$, partial eta-squared = .12, indicating that participants updated their evaluation about a stranger (M = 2.48, S.E. = .13) more than their evaluation about their friend (M = 1.92, S.E. = .16). A significant main effect of Task, $F(1,109) = 253.54$, $p < .001$, partial eta-squared = .70, showed that participants updated less for closeness ratings (M = 1.09, S.E. = .12) than for trustworthiness ratings (M = 3.31, S.E. = .16). However, these effects were modified by a

significant Agent X Task interaction, $F(1,109) = 31.55$, $p < .001$, partial eta-squared = .22. While participants updated their closeness ratings for friend and stranger similarly (Friend M = 1.11, S.E. = .15; Stranger M = 1.07, S.E. = .14, $p = .827$, 95% CI = [-.29, .37]), they updated trustworthiness ratings for the stranger (M = 3.88, S.E. = .18) more than for their friend (M = 2.73, S.E. = .19), $p < .001$, 95% CI = [-1.54, -.77] (Figure S1). Thus, participants were particularly less reluctant to update their ratings for the stranger's trustworthiness.

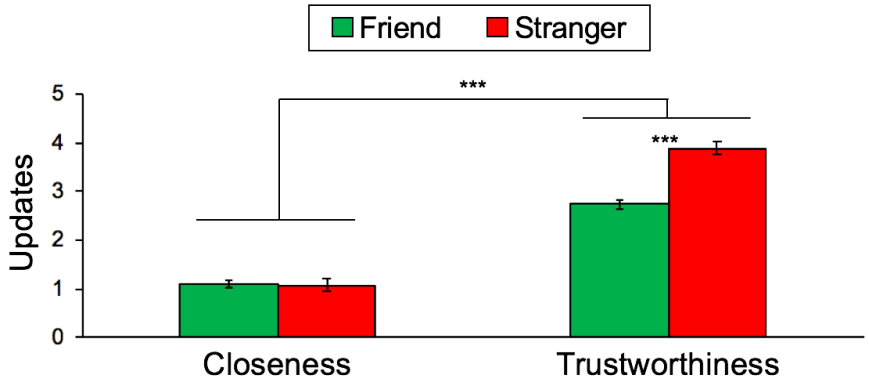



Figure S1. *Updates in closeness and trustworthiness ratings in Study 1.* Participants were more likely to update their trustworthiness ratings than their closeness ratings in general, and were more likely to update stranger-trustworthiness, after reading two consecutive negative versus positive behaviors. ***$p < .001$.

c. Pre-scan versus post-scan evaluations in Study 2.

To examine participants' general impression updates before and after the Social Judgment Task in Study 2, we compared their pre-scan and post-scan impression evaluations of their friend and of the stranger. We conducted a 2 (Agent: Friend, Stranger) X 2 (Task: Closeness, Trustworthiness) X 2 (Time: Pre-scan, Post-scan) repeated-measures ANOVA on participants' ratings outside the scanner.

There were significant main effects of Agent ($F[1, 23] = 351.29$, $p < .001$, partial eta-squared = .94) and Task ($F[1, 23] = 149.37$, $p < .001$, partial eta-squared = .87), indicating that

participants gave more positive evaluations to their friend (M = 7.03, S.E. = .14) than to the stranger (M = 2.90, S.E. = .15), and for the trustworthiness ratings (M = 5.69, S.E. = .14) than for the closeness ratings (M = 4.24, S.E. = .08)[2]. These effects were modified by a significant interaction between Agent X Task, $F(1, 23) = 93.11$, $p < .001$, partial eta-squared = .80, indicating that participants rated closeness for stranger (M = 1.44, S.E. = .11) lower than trustworthiness for stranger (M = 4.35, S.E. = .25), $p < .001$, while these ratings did not differ for friend (closeness M = 7.04, S.E. = .16; trustworthiness M = 7.02, S.E. = .15), $p = .873$.

More importantly, as predicted, these effects were modified by a significant 3-way interaction between Agent X Task X Time, $F(1, 23) = 4.53$, $p = .044$, partial eta-squared = .16 (Figure S2). Although participants perceived their friend as less trustworthy in the post-scan evaluation (Pre-scan M = 7.33, S.E. = .14; Post-scan M = 6.71, S.E. = .24), $p = .025$, 95% CI for difference = [.09, 1.17], partial eta-squared = .20, they did *not* significantly change their closeness ratings for their friend (Pre-scan M = 7.00, S.E. = .15; Post-scan M = 7.08, S.E. = .20), $p = .575$, 95% CI for difference = [-.39, .22], partial eta-squared = .01. Participants perceived the stranger as less trustworthy in the post-scan evaluation (Pre-scan M = 4.75, S.E. = .27; Post-scan M = 3.96, S.E. = .30), $p = .007$, 95% CI for difference = [.24, 1.35], partial eta-squared = .27. They rated the stranger as marginally closer in the post-scan evaluation (Pre-scan M = 1.17, S.E. = .13; Post-scan M = 1.71, S.E. = .20), $p = .050$, 95% CI for difference = [-1.08, .001], partial eta-squared = .16[3]. These findings suggest that participants were more

[2] The Task main effect was qualified by Time, Task X Time $F(1, 23) = 19.67$, $p < .001$, partial eta-squared = .46, showing that participants rated closeness higher in the post-scan evaluation (pre-scan M = 4.08, S.E. = .08; post-scan M = 4.40, S.E. = .12), $p = .025$, while they rated perceived trustworthiness lower in general in the post-scan evaluation (pre-scan M = 6.04, S.E. = .15; post-scan M = 5.33, S.E. = .21), $p = .006$. This effect seems to be driven by their increased closeness ratings in the post-scan evaluation for the stranger.

[3] We interpret this pattern in the context of the floor effect of participants' pre-scan closeness ratings for stranger. Participants first made closeness ratings upon barely meeting the confederate (stranger), and 91.7% of the participants rated the stranger a 1 on the closeness scale, which corresponded to "A total stranger." The game, involving ostensible interactions with the stranger, may have contributed to the slight increase in the closeness ratings.

protective toward their perception of closeness with their friend than their perception of friends' trustworthiness, and their evaluations for the stranger.

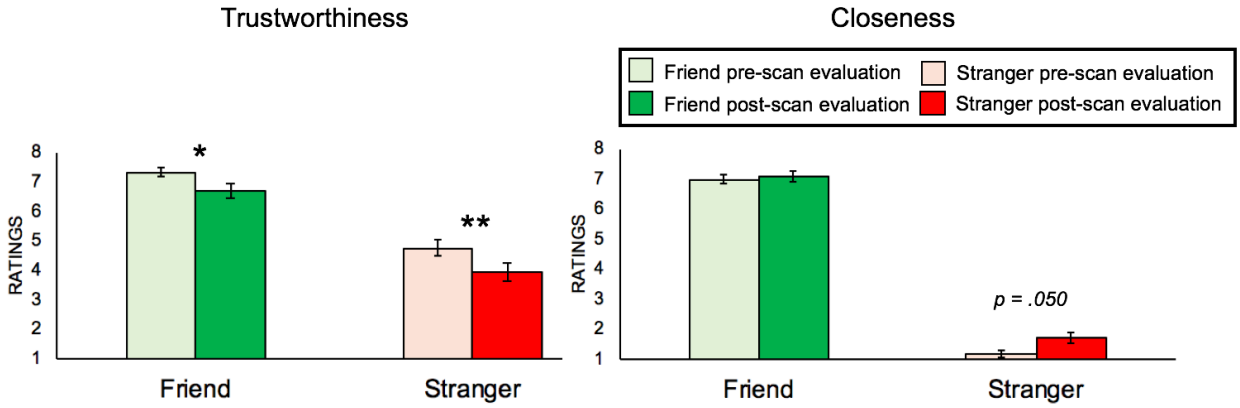



Figure S2. *General impression updates in Study 2:* Participants' trustworthiness ratings were decreased after the Social Judgment Task for both friend and stranger (Left). Participants rated the stranger marginally closer to them after the game. Closeness ratings for the friend remained the same (Right). Error bars represent standard errors (S.E.). *$p < .05$, **$p < .01$.

*Supplementary Section 7: Study 1 analysis matched with Study 2*

To apply the same analyses for Study 1 and Study 2, in Study 1, we subtracted participants' closeness ratings in their pre-exposure evaluation (i.e., before they read any stories) from those in their post-exposure evaluation (i.e., after they read all of the stories), generating pre-post closeness updates for friend and stranger, respectively. We then subtracted updating scores for the stranger from those for their friend to control for overall tendency to update before and after reading the stories, respectively for closeness and trustworthiness, as in Study 2. Using this new metric, controlling for how often participants made new friends as in Study 2, we found that participants who less negatively updated their friend-closeness ratings compared to their stranger-closeness ratings reported having more friends, Estimate = .34, S.E. = .15, Wald = 5.31, Odds ratio = 1.40, $p$ = .021, 95% CI for Estimate = [.05, .62]. Substituting post minus pre closeness updates [friend-stranger] with post minus pre trustworthiness updates [friend-stranger], we could find only marginal association between trustworthiness updating and the number of friends participants reported having, Estimate = .18, S.E. = .10, Wald = 3.18, Odds ratio = 1.19, $p$ = .074, 95% CI for Estimate = [-.02, .37]. This pattern suggests that participants who less negatively updated their friend-trustworthiness ratings compared to stranger-trustworthiness ratings reported having marginally more friends.

*Supplementary Section 8. Descriptions and findings with exploratory survey items in Study 2.*

a. Participants' explanation styles for their friend and a stranger

In Study 2, after participants exited the scanner, participants were asked to read four different hypothetical scenarios. The scenarios described situations in which their friend or the stranger put money in a meter for an unknown driver (positive) and laughed at a person who tripped on the curb (negative) (Fuhrman, Bodenhausen, & Lichtenstein, 1989). Participants used a 9-point scale, rating to what extent they think each incident reflects an aspect of the agent (1) or reflects an aspect of the situation (9) [agent versus situation], and the extent to which this behavior would be stable over time (1) or variable over time (9) [stable versus variable]. We averaged across the agent versus situation ratings and stable versus variable ratings for each scenario. After that, we submitted these scores to a 2 (Agent: Friend, Stranger) X 2 (Valence: Positive, Negative) repeated ANOVA model. There were no significant effects of Agent, Valence, or the interaction between Agent X Valence, *ps* > .38.

b. Participants' subjective distance from various relationships.

In Study 2, after participants exited the scanner, we also measured participants' subjective distance between themselves and different social targets (best friend, close friend, friend, casual friend, acquaintance, vague acquaintance, and stranger) after they exited the scanner as below:

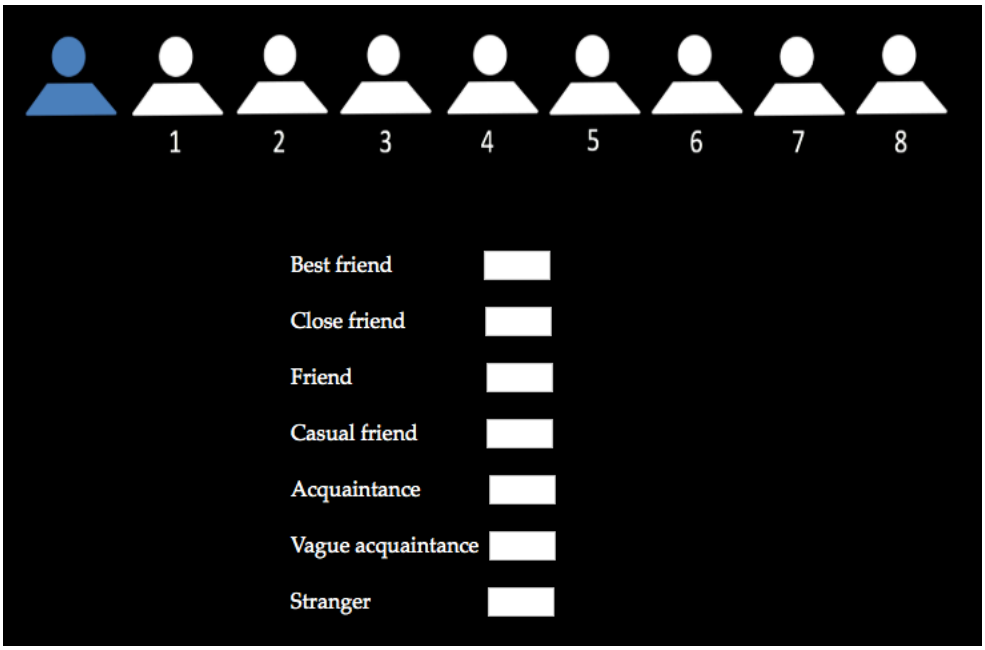
"Please assign the following words describing different relationships to the white human figures that you think correspond to the relationship. Each human figure has unique numbers. Please write down the number in the box next to the relationship word that corresponds to the figure. For example, if you think 'Stranger' corresponds to the human figure with number '8', please write down '8' next to 'Stranger.'

You do not have to assign numbers to all the relationship words. In other words, if you think some relationship words are not corresponding to the human figures in the scale, you can skip

using them. For example, if you think ‘Stranger’ does not correspond to any of the human figures on the scale, you can leave the box next to ‘Stranger’ empty.

Also, you do not have to use all numbers. In other words, if you think some numbers on the scale are not corresponding to the relationship words provided below, you can skip using them. For example, if you think the human figure with number ‘8’ does not correspond to any of the relationships provided below, you do not need to write down ‘8’ in any of the boxes.

You can assign the same number to more than one figure. For example, if ‘Vague acquaintance’ and ‘Stranger’ both mean ‘8’ on the scale, then you can put in ‘8’ for both of them.”




Participants were verbally instructed to think that the blue figure represented participants themselves.

Following the instruction that participants can assign the same number to more than one figure, 37.5% of participants assigned the same distance to more than one relationship category. This suggests that participants did not simply assign the numbers in their order but actually evaluated the distance for each relationship. As a manipulation check, participants assigned the closest human figure (distance 1) as corresponding to best friend (M = 1.04, S.D.

= .20), followed by close friend (M = 2.04, S.E. = .46), friend (M = 3.48, S.D. = .59), casual friend (M = 4.57, S.D. = .50), acquaintance (M = 5.52, S.D. = .82), vague acquaintance (M = 6.75, S.D. = .61), and stranger (M = 7.94, S.D. = .22). Importantly, these distances were not correlated with the number of friends participants reported having, $|rs| < .38$, $ps > .17$, suggesting that, regardless of the number of friends participants reporting having in the real world, subjective distance across various levels of social relationships was similar across our participants.

*Supplementary Section 9. Associations between participants' impression updating and their use of emotion words in memory recall*

In Studies 1 and 2, for exploratory purposes, participants were asked to describe their most pleasant and unpleasant moments with their friend in the time they have known each other. We used the Linguistic Inquiry and Word Count, 2007 program (LIWC; Pennebaker, Francis, & Booth, 2001; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007) to quantify word use. Specifically, after removing the clauses that the participants repeated from the question (e.g., "The most pleasant moment with my friend was…", "The most unpleasant moment with my friend was…"), we analyzed the percentage of mutually exclusive positive emotion (e.g., love, nice, sweet; Pennebaker et al., 2007) and negative emotion (e.g., hurt, ugly, nasty; Pennebaker et al., 2007) word use, to assess participants' quality of interaction with their friend.

Although the degrees to which participants updated their closeness and trustworthiness ratings about their friend and stranger were not associated with the use of emotional words in their memory recall in Study 1, participants' RTPJ activity during the Social Judgment Task was significantly associated with their use of emotional words in Study 2. Specifically, we found that participants who showed greater RTPJ activity in response to their friend's taking behavior during the Social Judgment Task used more negative emotion words when they recalled the most unpleasant memory with their friend, Pearson's $r$ = .59, $p$ = .034, 95% CI = [-.02, .88]. These findings suggest that participants who downplayed their friend's taking behavior less, thus those who might be more sensitive to their friend's negative behavior, experienced or recalled greater negative emotion when they recalled a social event. In contrast, participants' RTPJ response to their friend's giving behavior was not associated with their use of positive emotion words when they recalled the most pleasant memory with their friend, Pearson's $r$ = -.07, $p$ = .802. RTPJ activity in response to stranger's giving or taking was not significantly associated with emotional word use, $ps$ > .135.

*Supplementary Section 10. Activation in response to Theory of Mind task, belief versus photo in Study 2*

| Region | x | y | z | Peak Z | Voxels |
|---|---|---|---|---|---|
| LTPJ | -46 | -64 | 22 | 6.95 | 1434 |
| **R Superior temporal gyrus/ RTPJ** | 57 | -58 | 19 | 5.82 | 1199 |
| L Superior frontal gyrus/ L dmPFC | -7 | 47 | 37 | 5.64 | 918 |
| L Precuneus | -4 | -58 | 28 | 6.28 | 572 |
| L Middle frontal gyrus | -37 | 35 | 24 | -5.13 | 292 |
| R Middle frontal gyrus | 44 | 35 | 24 | -4.95 | 111 |
| L Inferior temporal gyrus | -56 | -50 | -18 | -4.84 | 87 |
| R Medial frontal gyrus | 5 | 52 | -2 | 4.38 | 67 |
| R Precuneus | 29 | -65 | 41 | -4.20 | 51 |
| L Precuneus | -22 | -68 | 35 | -4.23 | 41 |

Note. Uncorrected $p < .001$, cluster > 35 continuous voxels, corrected $p < .05$, regions of interest in bold. R; right, L; left, TPJ; temporo-parietal junction, dmPFC; dorsomedial prefrontal cortex.

*Supplementary Section 11. Findings with beta values*

To run a confirmatory analysis of the findings reported in the main paper, we conducted a whole-brain analysis on the functional scans acquired in Study 2. We constructed a general linear model (GLM, ordinary least-squares regression) including fourteen regressors of interest. Two regressors marked the time participants were presented with the other players' name, (1) when the name of friend was presented and (2) when the name of stranger was presented. Four regressors marked when participants were presented with the type of the task (closeness or trustworthiness), followed by the name of their friend or the stranger. Given that participants already knew whom they were about to evaluate in the given trial, we marked the timings of (3) friend-closeness, (4) friend-trustworthiness, (5) stranger-closeness, and (6) stranger-trustworthiness conditions. Four additional regressors marked when participants viewed the decision of their friend or the stranger, (7) friend-taking, (8) friend-giving, (9) stranger-taking, and (10) stranger-giving. Finally, four regressors marked when participants delivered their evaluations, in response to (11) friend-taking, (12) friend-giving, (13) stranger-taking, and (14) stranger-giving behaviors. To minimize the influence of physiological confounds, eight regressors of no interest were also included: six modeling head movement, one sampling white matter activity, and one sampling cerebrospinal fluid activity (Chang & Glover, 2009). Before they were submitted to the model, regressors of interest were convolved with a canonical gamma variate hemodynamic delay (Cohen, 1997).

We extracted the beta coefficients from the RTPJ ROI during the friend-taking, friend-giving, stranger-taking, and stranger-giving conditions. Because of the high collinearities between the beta values, ranging between $.48 < rs < .79$, $ps < .018$, we subtracted beta values of the stranger-taking condition from those of the friend-taking condition, and we subtracted beta values of the stranger-giving condition from those of the friend-giving condition, to account for the RTPJ response to strangers. Confirming our findings, controlling for the effect of beta values from the giving condition, lower beta values for friend-taking compared to stranger-taking were

significantly associated with more positive updates for friend, $r = -.48$, $p = .020$. Moreover, lower beta values for friend-taking compared to stranger-taking were marginally associated with reporting having more friends, $r = -.48$, $p = .062$, an association that remained marginal after controlling for beta values from giving conditions and how often participants made new friends, $r = -.48$, $p = .079$.

*Supplementary Section 12. The effect of how often participants made new friends on how many friends participants reported having in Study 2.*

In the ordinal regression on the number of friends participants reported having with general closeness updates, controlling for how often participants made new friends, we found that those who make new friends less often had a smaller number of friends, Estimate = -1.09, S.E. = .43, Wald = 6.50, Odds ratio = .34, $p$ = .011, 95% CI for Estimate = [-1.93, -.25]. However, we found a similar association between participants' general closeness updates and the number of friends they had without controlling for how often they made new friends.

Additionally, another ordinal regression on the number of friends participants reported having with their RTPJ activity from friend-taking condition, controlling for how often participants made new friends, also revealed that participants who make new friends less often had a significantly lower number of friends, Estimate = -.74, S.E. = .36, Wald = 4.19, Odds ratio = .48, $p$ = .041, 95% CI for Estimate = [-1.45, -.03]. But the association between participants' RTPJ activity and the number of friends remained similar without controlling for how often they make new friends.

*Supplementary Section 13. Association between participants' RTPJ activity and general trustworthiness updates in Study 2.*

For exploratory purposes, we ran regression analyses on participants' general trustworthiness updates [friend-stranger] with RTPJ activity averaged within friend versus stranger conditions, for giving versus taking trials separately. We found that post minus pre changes in trustworthiness were tracked by RTPJ activity in the friend-giving condition, B = -13.21, S.E. = 3.80, *ß* = -.69, *t* = -3.48, *p* = .003. The lower RTPJ activity participants showed in response to their friend's giving behavior, the more positively participants changed their trustworthiness ratings for their friend than for the stranger. Given that decreased efforts for mentalizing is often associated with more positive evaluations (Hughes, Zaki, & Ambady, 2017; Kliemann, Young, Scholz, & Saxe, 2008; Park, Blevins, Knutson, & Tsai, 2017), this pattern suggests that participants who did not experience the need for mentalizing when their friend gave money might evaluate their friend more positively than the stranger after the game. RTPJ activity from other conditions was not significantly related with changes in trustworthiness ratings, *ps* > .074.

References

Beer, A. & Watson, D. (2009). The individual and group loyalty scales (IGLS): Construction and preliminary validation. *Journal of Personality Assessment, 91*(3), 277-287, DOI: 10.1080/00223890902794341

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

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

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of Personality and Social Psychology, 78*(4), 791-808.

Dibble, J. L., Levine, T. R., & Park, H. S. (2011). The unidimensional relationship closeness scale (URCS): Reliability and validity evidence for a new measure of relationship closeness. *Psychological Assessment, 24*(3), 565-572.

Fuhrman, R. W., Bodenhausen, G. V., & Lichtenstein, M. (1989). On the trait implications of social behaviors: Kindness, intelligence, goodness, and normality ratings for 400 behavior statements. *Behavior Research Methods, Instruments, & Computers, 21*(6), 587-597.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504-528.

Graham, J., Haidt, J., & Nosek, B. A. (2008). *Moral Foundations Questionnaire.* Retrieved from http://www.moralfoundations.org/questionnaires

Hughes, B. L., Zaki, J., & Ambady, N. (2017). Motivation alters impression formation and related neural systems. *Social Cognitive and Affective Neuroscience, 12*(1), 49-60.

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

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia, 46*, 2949-2957.

Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *The Journal of Neuroscience, 33*(50), 19406-19415.

Park, B., Blevins, E., Knutson, K., & Tsai, J. L. (2017). Neurocultural evidence that ideal affect match promotes giving. *Social Cognitive and Affective Neuroscience, 12*(7), 1083-1096.

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.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001* (p. 71). Mahwah, NJ: Erlbaum.

Pennebaker, J. W., Chung, C., K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007.* www.LIWC.Net.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*, 1069–1081.

Ryff, C. D. & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*, 719–727.