Personalities converge during enjoyable conversations

Christopher Welker[1], Thalia Wheatley[1,2],
Grace Cason[1], Catherine Gorman[1], Meghan Meyer[3]

[1] Dartmouth College
[2] Santa Fe Institute
[3] Columbia University

Keywords: conversation, personality, social cognition, self-perception, social interaction

Abstract

Based on current research, it's evident that the way people see themselves is shaped by their conversation partners. Historically, this literature has emphasized a one-way dynamic, specifically addressing how an individual's expectations can shape their partner's self-image. Given the reciprocal nature of conversation, we wondered whether conversation partners' self-views may mutually evolve. Using four-person round-robin conversation networks, we found that participants tended to have more similar self-views post-conversation than pre-conversation, an effect we term "inter-self alignment." Further, the more two partners' self-views aligned, the more they enjoyed their conversation and were inclined to interact again. This effect depended on both conversation partners becoming aligned. These findings suggest that the way we see ourselves is co-authored in the act of dialogue and as shared identities develop, the desire to continue the conversation increases.

Introduction

The people we interact with shape who we are. Although some aspects of our self-views are stable over time (Costa & McCrae, 1986; Diehl et al., 2006; Swann, 2012; Swann & Hill, 1982), self-views can shift in meaningful social interactions and different self-dimensions are activated in different social contexts (Burr, 1995; Gergen, 2011; Hampson, 1988; Markus & Wurf, 1987). To date, prior work has focused on unidirectional effects: e.g., how one person's expectations shape their partner's self-views (Fazio et al., 1981; Snyder, 1984). Yet conversation– the primary way humans interact to maintain social bonds (R. Dunbar, 1996; R. I. M. Dunbar, 2004; Mastroianni & Gilbert, 2021)– is characterized by a dynamic exchange between people. Given the reciprocal nature of conversation, do people's perceptions of themselves co-evolve? Here we test whether people's self-views converge during conversation.

Conversations are known to align many behavioral and psychological processes from how people speak (Brennan & Clark, 1996; Hawkins et al., 2020; Stolk et al., 2016; Wynn et al., 2023) to how they understand the world (Clark & Brennan, 1991; Coman et al., 2016; Pickering & Garrod, 2006; Sievers et al., n.d.). The possibility that self-views also converge during conversations is consistent with research on personality change during friendship. Although people tend to befriend similar others (Kossinets & Watts, 2009; Laursen, 2017; McPherson et al., 2001; Parkinson et al., 2018; Verbrugge, 1977), it is also true that friends influence each other to behave similarly (Aral et al., 2009; Ma et al., 2015) and even come to view themselves as more similar over time (van Zalk et al., 2020). For example, friends' self-reported extraversion tends to become more similar over months and years of friendship (van Zalk et al., 2020). While prior work has demonstrated that established relationships can induce similarity, we investigate whether this can happen during a single conversation.

According to shared reality theory, people are motivated to perceive that their own thoughts and inner experiences are aligned with others (Echterhoff et al., 2009; Echterhoff & Higgins, 2018; Rossignac-Milon et al., 2020). This motivation is consistent with findings that shared experience activates the brain's reward system supporting positive affect (Wagner et al., 2015). Given that people tend to feel happier after conversations than before (Reece et al., 2023), we wondered not only whether conversation aligns self-views, but whether this alignment predicts conversational enjoyment.

Of course, conversation is not monolithic (Cooney & Wheatley, n.d.). Conversations vary on many dimensions such as the goals of the conversation partners, the number of people involved, and the culture in which they take place (Bassetti & Liberman, 2021; Cooney et al., 2020; Yeomans et al., 2022). In particular, recent work has shown that conversations on "deep" topics that involve self-disclosure (e.g., sharing an embarrassing memory) lead to increases in conversation partners' perceived closeness relative to "shallow" conversations that are more impersonal (Mehl et al., 2010; Sprecher, 2021; Sun et al., 2020). Recent work has also shown

that self-views are most likely to change when participants are actively aware that they are reflecting on themselves (Schneider et al., 2022). We posit that self-disclosure and self-reflection may be more likely to take place during deep conversations. We manipulated topic depth with the prediction that deeper conversations would involve greater convergence of self-views and would be more enjoyable.

In summary, we propose that the degree to which conversation partners' self-views become more similar over the course of a conversation—a process we call "inter-self alignment"—should predict how enjoyable they find their conversation. To test this hypothesis, participants completed a series of trait ratings before and after having 10 minute conversations about either shallow or deep topics. Participants also rated how enjoyable they found the conversations. We then quantified the distance between partners' trait ratings before and after each conversation and determined whether those distances increased (inter-self divergence) or decreased (inter-self alignment) over the course of the conversation. We predicted that conversations that were more enjoyable would be marked by greater inter-self alignment. Further, we predicted that this inter-self alignment would be greater in deep, compared to shallow, conversations. Our data support our primary hypothesis: conversations with greater inter-self alignment were rated as more enjoyable. Our secondary hypothesis was not supported as this effect was equivalently robust in deep and shallow conversations. Further analyses revealed that the relationship between alignment and enjoyment was dyadic rather than individual (enjoyment was predicted by dyadic, but not individual-level alignment), was mediated by partner prediction accuracy (partners who aligned more were more accurate), that social aspects of the self were more readily aligned in conversation than other aspects, and that the relationship between alignment and enjoyment was relatively short-lived.

Methods

Participants

We recruited 104 participants (60.4% White/Caucasian, 4.2% Black/African American, 22.9% Asian, and 12.5% Other/Mixed Race and 67.7% female with a mean age of 19.22 years) via an online participant recruitment system at Dartmouth. All participants were at least 18 years old and were compensated with course credit. Data collection and experimental procedures were approved by the Committee for the Protection of Human Subjects at Dartmouth in Hanover, New Hampshire (CPHS ID: 00030517) and all participants provided informed consent before taking part in the study.

Sample size was determined based on results from a pilot study and is recorded in our pre-registration (AsPredicted #103006). Specifically, a separate set of 40 participants completed the conversation prompts and self-report scales of conversation enjoyment, closeness, and the trait ratings described below (see **Supplementary Materials** for a full description of the pilot study). Using the linear mixed model described in our results section below and reported on in our pre-registration, we observed a medium-sized relationship between inter-self alignment and conversation enjoyment in the pilot sample ($\beta = 0.30$, 95% CI = [0.00,0.60], $p = 0.053$). We

therefore ran power analyses using the R package simr to establish an appropriate sample size for Study 1 based on effect sizes from the pilot study (Green & MacLeod, 2016). We found the sample size needed to achieve 80% power to detect a difference between our main model with and without inter-self alignment as a predictor, testing whether including inter-self alignment significantly improved our ability to predict enjoyment. We then added one more pair of groups (one for each condition) to account for the fact that significant effect sizes from pilot studies are often overestimates of the true effect size. Based on our power analyses, we determined that we needed a sample size of 24 groups (144 conversations, 96 participants) to detect robust results.

Participants were recruited in groups of four and had conversations in a round-robin structure (i.e. each member of the group had a one-on-one conversation with every other group member [Figure 1]). If a group either did not finish all their conversations or was missing at least two survey responses, it was excluded from further analyses. Two groups failed to meet the criteria for inclusion, leaving us with a final sample size of 96 participants (24 groups; 144 one-on-one conversations).

The majority of our participants had not interacted with each other prior to their conversations. When participants rated how well they knew their partner before taking part in the study on a scale from 0 (Not at all) to 100 (Extremely) with a midpoint at 50 (Somewhat), the average response was an 8.29 (sd = 21.35) and over half of the participants (55.21%) responded with 0. These numbers are on par with other work studying conversations between strangers (Templeton et al., 2022). However, we control for the strength of any prior relationships between conversation partners by including the responses to this question as a covariate in our regression models.

Procedure

Each participant was assigned to a group with three other participants and, over the course of the study, had a 10 minute conversation with every other member of their group. We employed a between-subjects design; four-person groups were randomly assigned to talk about either deep or shallow topics during their conversations. The conversation prompts were selected from those used in a previous study of deep and shallow conversations (Kardas et al., 2022). Other than providing these prompts, we allowed participants to converse freely.

The study consisted of three phases (Figure 1, Right). In the first, we sent participants a baseline survey with a 60-item trait scale (Meyer et al., 2019) as well as the following personality trait scales: the self-monitoring scale (Snyder, 1974), the MacArthur Scale of Subjective Social Status (Adler et al., 2000), the Revised UCLA Loneliness Scale (Russell et al., 1980), and a self-report of social variety seeking (Elfenbein et al., 2023). We do not report results here from the personality trait scales as they were addressing a separate theoretical question. Participants also answered basic demographic questions (age, race, gender, education). Participants were sent the survey at least 24 hours before their first conversation, and finished the survey prior to that conversation.

In the second phase, participants had three conversations over Zoom – one with each other member of their group – with each conversation occurring on separate days. Although the conversations in our pilot study took place in person, we collected the round-robin data during the COVID-19 pandemic which prompted us to switch to a virtual format. Before each conversation, participants were instructed to hide "self view" and expand the Zoom window to fill the whole screen so that the experience would better resemble a face-to-face conversation. That is, they saw their conversation partner, but not themselves, during the conversation. The experimenter then briefly introduced the task, sent the participants the conversation prompts via the text chat, and left the Zoom room. After 10 minutes, the experimenter rejoined the room and sent the participants a link to a survey in which they filled out the 60-item trait scale another time along with questions about how many of the prompts they discussed during their conversation, the depth of their conversation, how well they knew their partner before the study, how much they enjoyed the conversation, how close they felt to their partner, and how similar they felt to their partner. In addition, participants rated their partner using the same 60-item trait scale. For additional information about the exact wording and range of the survey questions, please see our OSF repository (https://osf.io/7wtz2/?view_only=8ee86d31444649d9a9a9ed218b932883).

In the third phase, the 60-item trait scale was sent to participants 7 days after their final conversation. Importantly, the display order of the items in the 60-item trait scale was randomized every time it was presented to counter any effects of taking the survey multiple times.

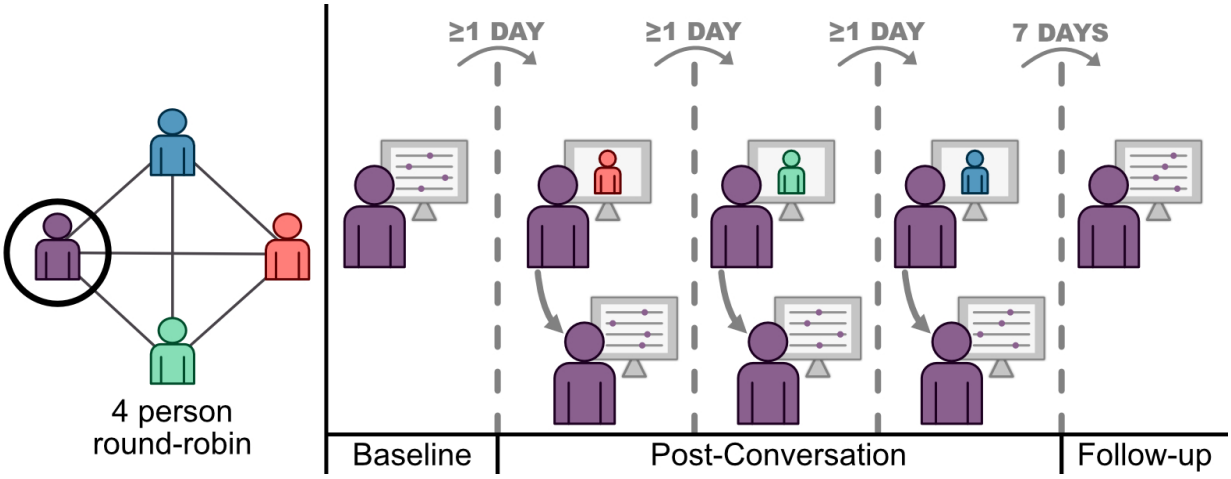

Figure 1: (Left) Participants were recruited in groups of four and had conversations in a round-robin structure. Each participant had a one-on-one conversation with every other member of their group. (Right) Depiction of an example participant's experience in the experiment. Participants were sent the baseline survey by email. At least one day later, they participated in a dyadic conversation with one of their group members, after which they filled out the post-conversation survey. They repeated this step two more times, once with each remaining group member. All conversations took place on separate days. Finally, one week after their final conversation, participants were sent the follow-up survey.

Variable creation

All variables were created in R (version 4.2.2).

Inter-self alignment

We defined inter-self alignment as the degree to which conversation partners' self trait ratings became more similar after the conversation compared to at baseline. Following steps detailed in our pre-registration, we operationalized inter-self alignment by comparing the Manhattan distance (Figure 2) between the partners' self-reported trait ratings in the baseline surveys to the distance between their self-reported trait ratings in the post-conversation surveys. Our use of Manhattan distance was inspired by prior work using similar scales (Meyer et al., 2019; Tan et al., 2015), however our results are not dependent on the use of this specific distance metric (see **Supplementary Materials** for additional analyses). We subtracted the distance between pairs' self ratings *after* their conversation from the distance between their self ratings *before* the conversation. This subtraction resulted in a value which, when positive, indicates that the pairs' trait ratings were more similar after the conversation than before and, when negative, indicates that they were more similar before the conversation than after the conversation. We mean-centered and z-scored participants' trait survey responses before calculating alignment to account for participant-level biases in scale use.

.

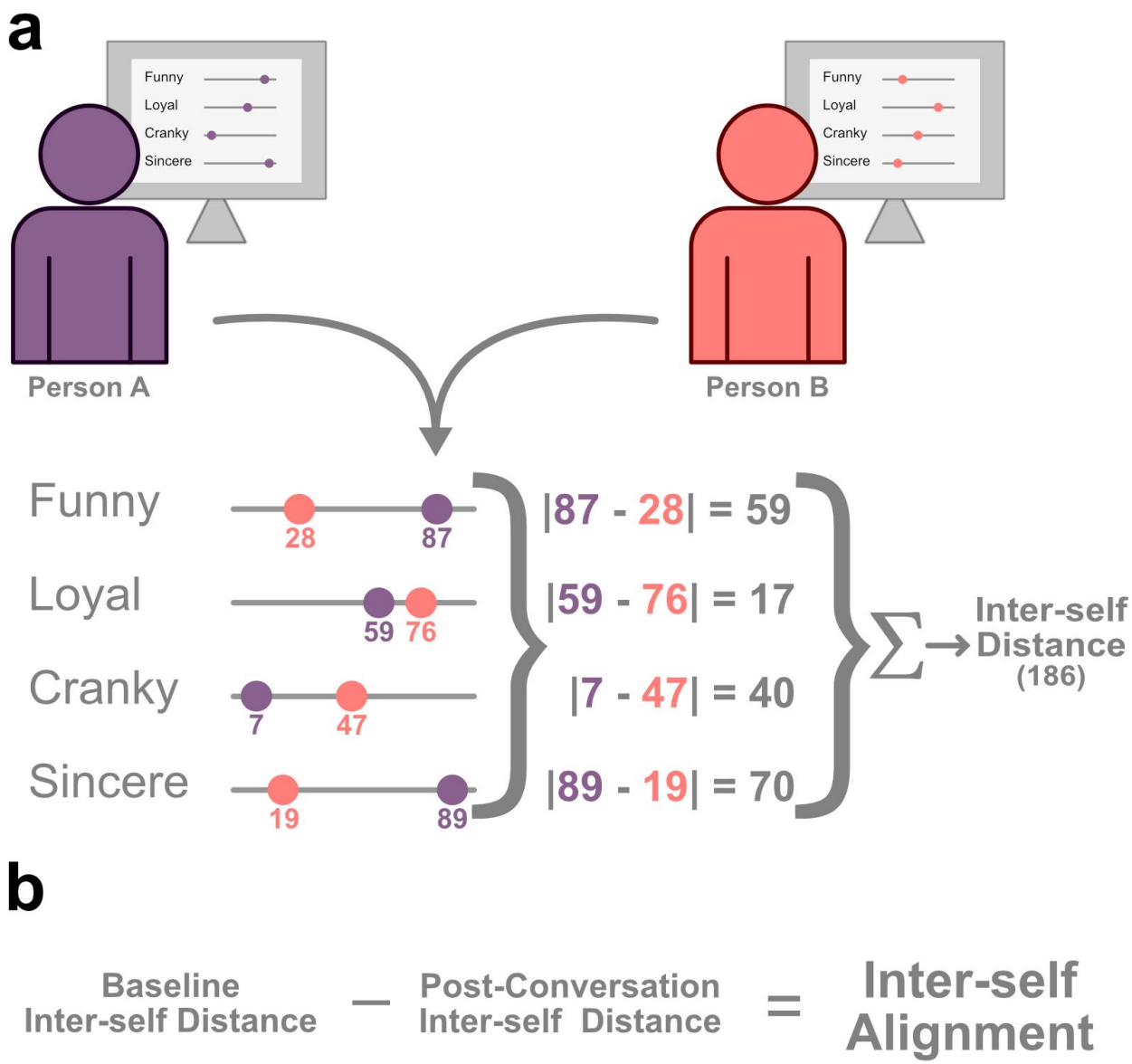

Figure 2: Schematic of how inter-self distance and alignment were calculated. a) Participants filled out 60-item trait scales from Meyer et al., (2019); in this illustrated example, we use four traits taken from the scale. To compute the distance between pairs' trait ratings, we calculated the sum of the absolute value of the differences between the two survey responses (i.e. the Manhattan distance). Responses were mean-centered and z-scored for each participant prior to distance calculation to account for any idiosyncratic differences in scale usage. b) Inter-self alignment was calculated by subtracting the distance between partners' trait ratings after the conversation from the distance between their trait ratings at baseline. Positive values indicate that participants' selves became more similar after the conversation and negative values indicate that participants' selves became less similar after the conversation.

Conversation enjoyment

Two questions from the post-conversation survey were related to conversation enjoyment: “How much did you enjoy the conversation that you just took part in?” and “If you were to see the person you just interacted with somewhere on campus, how likely would you be to stop and have a conversation with them?” Participants responded to both questions using a sliding scale from 0 (Not at all) to 100 (Extremely) with a midpoint at 50 (Moderately). We first examined the relationship between these two variables in a linear mixed effects model with random intercepts for the participant, conversation partner, dyad, and group. We found a significant, positive relationship between the two variables ($\beta = 0.66$, 95% CI = [0.57 ,0.75], $p < 0.001$) and so took their average to be an index of conversation enjoyment. These steps were detailed in our pre registration.

Self-reported closeness

Three questions from the post-conversation survey measured participants' perceived closeness with their conversation partner: “How emotionally close do you feel to the person you just talked with?”, “How similar, in terms of personality, temperament, major likes and dislikes, beliefs, and values, do you feel to the person you just talked with?”, and the inclusion of the other in the self scale (Aron et al., 1991). Participants responded to the first two questions with a sliding scale from 0 (Not at all) to 100 (Extremely) with a midpoint at 50 (Moderately) and followed standard protocol for the inclusion of the other in the self scale.

After confirming that the data were adequate for dimensionality reduction (Bartlett’s test of sphericity $\chi^2(3) = 263.78$, $p < .001$; Kaiser-Meyer-Olkin test KMO = 0.695), we mean-centered and z-scored the data and ran a PCA on the three questions related to perceived closeness. The elbow of the scree plot determined that one component was an appropriate approximation of the three variables, capturing a large amount of the shared variance between them (70.8%). We therefore used the first principal component of the PCA as a composite self-reported closeness score.

Partner prediction accuracy

To assess the accuracy of participants’ predictions of their conversation partners’ trait ratings, we found the Manhattan distance between participants’ ratings of their partners’ traits and their partners’ ratings of their own traits after their conversation. We then reversed the sign of the distances so that larger values would indicate greater partner prediction accuracy. Both sets of ratings were mean-centered and z-scored before distances were calculated.

Results

All analyses were carried out in R (version 4.2.2). Linear mixed effects models were implemented using the *lme4* package and were fit with maximum likelihood estimation; degrees of freedom and p-values were approximated using Satterthwaite’s method via the *lmerTest* package (Bates et al., 2015; Kuznetsova et al., 2017). We report standardized regression coefficients to increase interpretability.

Manipulation checks and scale validation

Conversation prompt effectiveness

Overall, most participants reported discussing all of their three assigned conversation prompts (mean = 2.81, median = 3, sd = 0.61). We used a linear mixed effects model to assess if the assigned prompts (deep or shallow) influenced the reported depth of participants' conversations. Random intercepts were included for each participant, conversation partner, dyad, and group to account for repeated measurements resulting from the round-robin design. The analysis revealed a significant, positive relationship between assigned topic and reported conversation depth ($\beta = 0.47$, 95% CI = [0.31,0.62], $p < 0.001$), indicating that participants in the deep topics condition reported having deeper conversations than those in the shallow topics condition.

Trait scale validation

The traits in the scale used here have been used in multiple studies investigating self- and other-knowledge (Dumas et al., 2002; Meyer et al., 2019; Rubin-McGregor et al., 2022). We ran analyses to confirm that the scale reliably captured our participants' unique views of themselves. For example, if a participant considers themself to be funny, then they should consistently rate themselves as funny every time they fill out the same scale. The extent to which they see themselves as funny may fluctuate with each conversation, but we expect that people have relatively stable self-views. To verify that subjects' responses to our scale were indeed consistent within each subject, we compared the distances between subjects' five scale responses (baseline, after each conversation, follow-up) to the distances between responses from different subjects, expecting distances to be smaller within subjects than between subjects.

To this end, we created a 460 x 460 model matrix with 5 x 5 squares along the diagonal set equal to 1 and the remainder of the matrix set to 2 (See Supplementary Figure 2, Left). Each of the 5 x 5 squares (n = 92) represented the distances between the five measurements of a single subject (at baseline, after each conversation, and at follow-up). Since four participants did not complete the follow-up survey, they were excluded from this analysis. We compared this model matrix to an equivalent matrix of Manhattan distances between subjects' trait scale responses (Supplementary Figure 2, Center) with a Mantel test. The test revealed a significant fit between our model and the observed data ($r = 0.09$, $p < 0.001$), suggesting that within subject distances were significantly smaller than between subject distances (Supplementary Figure 2, Right). This result implies that the trait scale reliably measures self-views.

General self-view alignment

Before testing our primary hypothesis – that greater inter-self alignment predicts greater conversation enjoyment – we first wanted to assess, across all subjects, whether selves converge during conversation.

To determine if conversations correspond with participants' self-views becoming more similar to their partners' self-views, we compared the distribution of observed alignment values

to 0 using a one-sample t-test. The distribution of alignment values fell significantly above 0 ($M_{alignment}$ = 6.77, 95% CI = [4.94, 8.59], t(143) = 7.34, $p < 0.001$; Figure 3a), suggesting that most pairs' personalities became more similar after the conversation compared to before the conversation. The distribution of inter-self alignment scores remained significantly greater than zero (Intercept = 6.85, 95% CI = [4.47, 9.23], $p < 0.001$) when controlling for baseline inter-self distance, assigned depth of conversation topics, and prior familiarity with their conversation partner in a linear mixed effects model which also included random intercepts for each group. The dependent variable, inter-self alignment, was not mean-centered or z-scored in this model in order to estimate its mean with the intercept term. Additionally, greater baseline inter-self distance predicted greater inter-self alignment (β = 4.06, 95% CI = [2.13, 5.99], $p < 0.001$). This relationship was expected given that, if two people are farther apart in trait space to begin with, they are able to move a greater distance towards each other than two people who started next to each other. Overall, we see that the distribution of alignment values falls significantly above zero.

As a test of the robustness of this effect, we ran a permutation-based analysis in which we compared the observed median value of inter-self alignment ($Med_{alignment}$ = 7.46) to a distribution of 5000 median alignment values obtained by randomly shuffling the order of traits in the inter-self alignment calculation. The observed median fell well above the distribution of medians from permuted values (maximum permuted median = 3.75; Figure 3b), once again suggesting that conversations aligned all pairs' self-views.

Next, we asked whether conversation-induced alignment was specific to dyadic sets. If a unique, co-created self-view emerges during each conversation, then inter-self alignment effects should be stronger within- versus between-dyadic pairs. To assess this, we compared the median inter-self alignment observed in our real pairs of dyads to a distribution of 5000 median inter-self alignment values generated from randomly assigned pseudo pairs. Inter-self alignment was calculated for the pseudo pairs exactly as it was for real pairs, except for each iteration of pseudo pairs, dyads consisted of randomly assigned participants. The observed median value of inter-self alignment fell in the middle of the permuted distribution with 30% of the permuted distribution being greater than the observed value (maximum permuted median = 10.55; Figure 3b).

Overall, significant results from the comparison to 0 and shuffled traits suggest that conversation tends to align personalities. However, results from the comparison to pseudo conversation pairs suggest that alignment in a conversation occurs in a general way across conversations, rather than in a dyad specific way.

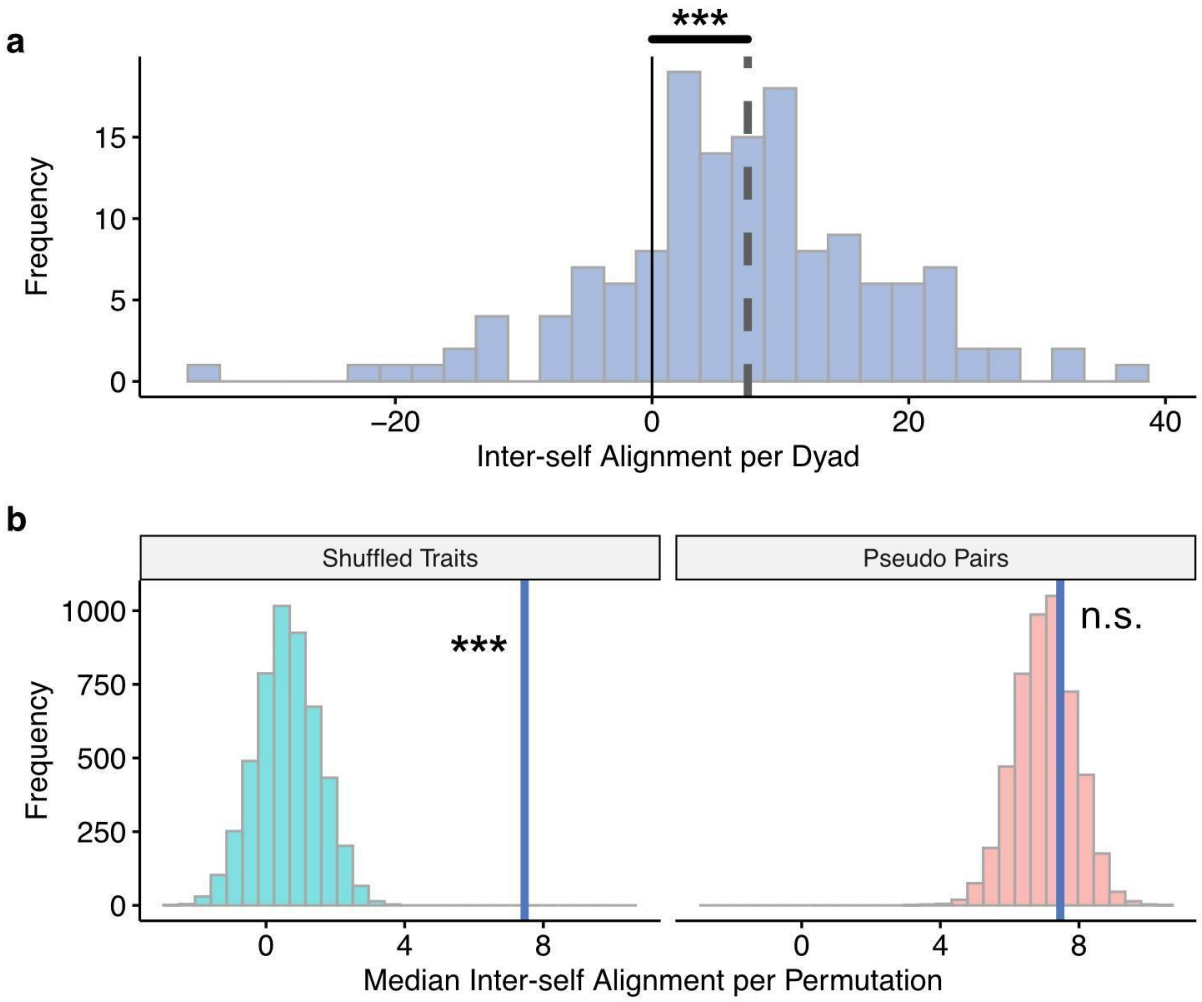

Figure 3: a) Histogram depicting the distribution of inter-self alignment scores per dyad. The dashed gray line represents the mean of the distribution, the solid black line marks 0, and *** denotes $p < 0.001$. b) Histograms showing null distributions of median inter-self alignment values generated from shuffling the order of traits before calculating alignment (blue/left) or from distributions of alignment values generated from pseudo pairs (red/right). Each null distribution is made of 5000 permuted values. The observed median (solid purple line) falls above the former, but not the latter null distribution.

Inter-self alignment predicts greater conversation enjoyment

Next, we returned to our primary hypothesis – that greater inter-self alignment should predict greater conversation enjoyment. We built a linear mixed effects model with one observation for each participant in each conversation (n = 288) to test the relationship between inter-self alignment and conversation enjoyment. The index of conversation enjoyment was included as the outcome variable, and inter-self alignment was our predictor of interest. We included the conversation topic (deep or shallow), the interaction between topic and inter-self alignment, and self-reports of how well the participants knew each other before the study as

covariates. Since the similarity of a dyad's personalities at baseline limits the extent to which they can align, we also controlled for participants' baseline inter-self distance. All continuous predictors were mean-centered and z-scored. To account for the repeated measurements inherent in a round-robin design, we included random intercepts for each participant, partner, dyad, conversation number (an ID variable coding for each participant's first, second, or third conversation), and group.

As predicted in our pre-registration, we found a significant, positive effect of inter-self alignment on conversation enjoyment (β = 0.18, 95% CI = [0.05, 0.30], p = 0.007; Figure 4a). In addition, the distance between partners' baseline personalities (β = -0.17, 95% CI = [-0.31, -0.03], p = 0.02) and how well the participants knew each other before the study (β = 0.28, 95% CI = [0.18, 0.38], p < 0.001) were also significantly related to conversation enjoyment.

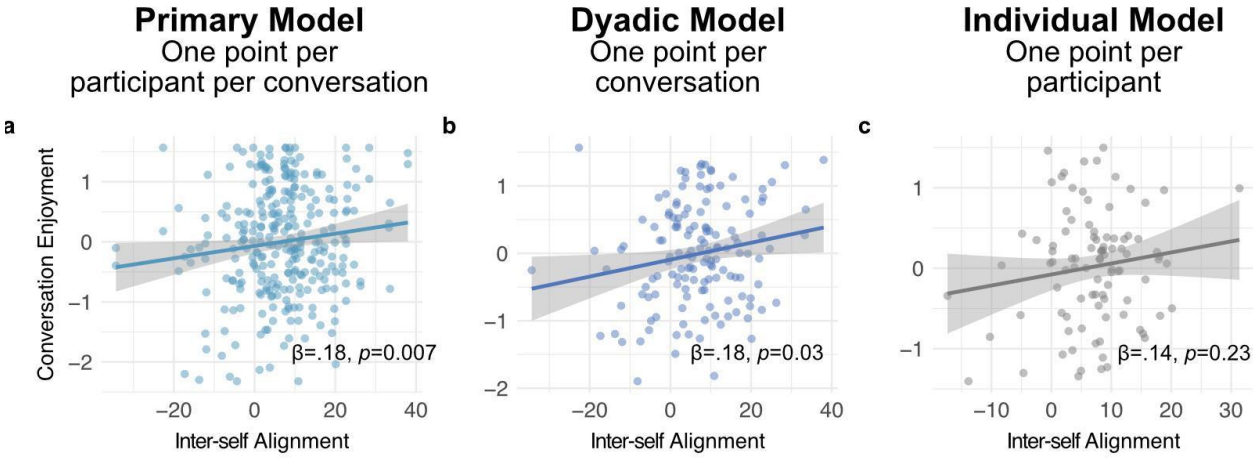

Figure 4: Scatter plots with regression lines showing the relationship between conversation enjoyment and inter-self alignment with each member of each conversation (a), each dyad (b), and each participant (c) treated as a single observation.

The relationship between inter-self alignment and conversation enjoyment is ephemeral

To examine the duration of the association between inter-self alignment and enjoyment, we ran the same model again, except inter-self alignment was calculated using inter-self distances from the baseline survey and the follow-up survey that took place one week after participation. The linear mixed effects model had identical random effects and covariates as the original model and the only significant predictor was the self-report of how well the participants knew each other before the study (β = 0.25, 95% CI = [0.12, 0.37], p < 0.001). The extent to which participants' personalities aligned from baseline to a week after the study was not related to how much they enjoyed their conversation (β = -0.01, 95% CI = [-0.15, 0.14], p = 0.907). Thus, inter-self alignment is meaningfully related to how much people enjoy talking to each other, but this effect seems to be relatively short-lived.

The relationship between inter-self alignment and conversation enjoyment manifests on a dyadic level

So far, results suggest that the more conversation partners align their self-views during a conversation, the more they enjoy talking with one another. Next, we asked whether this effect is driven by something specific individuals bring to all their conversations versus something attributable to specific conversation partners (i.e., dyads). In other words, do some people tend to align with all their conversation partners and does that trait-like tendency predict enjoyment? Or, is there something special about two particular people coming together to jointly align self-views that predicts more enjoyment for both of them?

Leveraging the round-robin structure of our data, we were able to determine whether inter-self alignment was related to conversation enjoyment at the level of the dyad and/or the individual. To do so, we averaged all relevant variables (i.e. conversation enjoyment, inter-self alignment, baseline inter-self distance, self-reports of how well participants knew each other prior to the study, assigned topic, and identification variables) within conversation pairs for the dyadic analysis and within each participant for the individual analysis. We then built linear mixed effects models with identical fixed effects structures as the original model. The random effects differed from the original model; both the dyadic and individual models only included random intercepts for each round robin group since the unit of analysis became coarser in each model.

In the dyadic model, inter-self alignment was a significant positive predictor of conversation enjoyment ($\beta = 0.18$, 95% CI = [0.02, 0.35], $p = 0.03$; Figure 4b). Baseline inter-self distance ($\beta = -0.23$, 95% CI = [-0.42, -0.05], $p = 0.01$) and how well the participants reported knowing each other before the study ($\beta = 0.39$, 95% CI = [0.24, 0.53], $p < 0.001$) were the only other significant predictors. In contrast, inter-self alignment was not a significant predictor of conversation enjoyment in the individual model ($\beta = 0.14$, 95% CI = [-0.08, 0.36], $p = 0.23$; Figure 4c), yet baseline inter-self distance ($\beta = -0.25$, 95% CI = [-0.52, -0.01], $p = 0.04$) and how well the participants reported knowing each other before the study ($\beta = 0.32$, 95% CI = [0.12, 0.53], $p = 0.001$) remained significant. The relationship between inter-self alignment and conversation enjoyment exists at the dyadic, but not the individual level of analysis. There seems to be something special about two particular people coming together to jointly align self-views that predicts more enjoyment for both of them.

Baseline inter-self similarity predicts greater perceived closeness

We next assessed whether inter-self alignment predicts perceived closeness to the conversation partner. We built a linear mixed effects model to identify the predictors of our secondary dependent variable, perceived closeness. This model was almost identical to the model used to identify the relationship between inter-self alignment and conversation enjoyment, except this model included perceived closeness as a dependent variable instead of conversation enjoyment.

Inter-self alignment was not a significant predictor of perceived closeness ($\beta = 0.12$, 95% CI = [-0.01, 0.25], $p = 0.06$), however baseline inter-self distance ($\beta = -0.19$, 95% CI = [-0.33, -

0.05], p = 0.008) and prior familiarity with the conversation partner (β = 0.40, 95% CI = [0.30, 0.50], p < 0.001) were both significantly related to perceived closeness.

Inter-self alignment, enjoyment, and partner prediction accuracy

Results reported above demonstrate that when self-views converge during conversation, the conversation is more enjoyable than when self-views do not converge. However, these results alone do not reveal the underlying cognitive mechanism that may link aligned self-views and enjoyment. Based on prior research, we wondered whether an underlying mechanism may be enhanced prediction: specifically, whether aligning self-views affords more accurate predictions of one's partner which, in turn, make conversations more enjoyable. This would be consistent with three bodies of past work indicating 1) that, when we mentally simulate others and make predictions about them, we incorporate information about them into our own self-views (Meyer et al., 2019; Rubin-McGregor et al., 2022; Schneider et al., 2022), 2) that accurate predictions are rewarding (Rutledge et al., 2014), and 3) that making more accurate predictions about social partners has been linked to both behavioral adaptability (adapting one's behavior to fit a partner's preferences) and having higher-quality social relationships (Hall et al., 2009; Schmid Mast & Hall, 2018).

To test this hypothesis, we first investigated if conversation partners who demonstrate greater inter-self alignment are also more accurate in predicting each other's traits. We quantified a participant's accuracy by calculating the distance between 1) participants' ratings of their partner's self-views and 2) their partner's ratings of their own self-views after the conversation, here termed "other-accuracy" (Figure 5a). We reversed the sign of these distances so that larger values indicated greater accuracy when predicting a partner's trait ratings. Note that "other-accuracy" is not mathematically redundant with inter-self alignment. Someone could accurately perceive that their partner is grouchy—leading to high other-accuracy—but not become any grouchier themselves over the course of their conversation—limiting their inter-self alignment.

We constructed a linear mixed effects model with other-accuracy as the outcome variable, inter-self alignment as the predictor of interest, and baseline inter-self distance, assigned conversation topic, and the familiarity with the conversation partner as covariates. We included random intercepts for each participant, conversation partner, dyad, conversation number, and group to account for repeated measurements. All variables were mean-centered and z-scored. Greater inter-self alignment significantly predicted other-accuracy(β = 0.65, 95% CI = [0.56, 0.74], p < 0.001; Figure 5b). Greater inter-self distance at baseline also predicted less other-accuracy (β = -0.51, 95% CI = [-0.61, -0.41], p < 0.001).

Other-accuracy also predicted greater conversation enjoyment (β = 0.32, 95% CI = [0.20, 0.43], p < 0.001; Figure 5b) in a linear mixed effects model controlling for inter-self distance at baseline, assigned conversation topic, and prior familiarity with the conversation partner. All terms were mean centered and z scored. Random intercepts were included for each participant, partner, dyad, conversation number, and group. Prior familiarity with the

conversation partner also predicted greater conversation enjoyment (β = 0.27, 95% CI = [0.17, 0.36], p < 0.001).

Given that other-accuracy related to both inter-self alignment and conversation enjoyment, we ran a mediation analysis with inter-self alignment as our independent variable, other-accuracy as a mediator, and conversation enjoyment as our dependent variable, using the R package *mediation* (Tingley et al., 2014). All regressions included controls for baseline inter-self distance, conversation topic, and prior familiarity with the conversation partner as well as a random intercept for each participant[1]. All variables were mean-centered and z-scored and we used nonparametric bootstrap estimation with 5000 Monte Carlo draws to estimate confidence intervals and p values. Other-accuracy fully mediated the relationship between inter-self alignment and conversation enjoyment (ACME = 0.26, 95% CI = [0.15, 0.38], p < 0.001; ADE = -0.06, 95% CI = [-0.21, 0.11], p = 0.53; Figure 5c).

[1] The R package only allows for the inclusion of one random effect, so we elected to include the one that routinely explains the most variance in the other reported mixed effects models (participants).

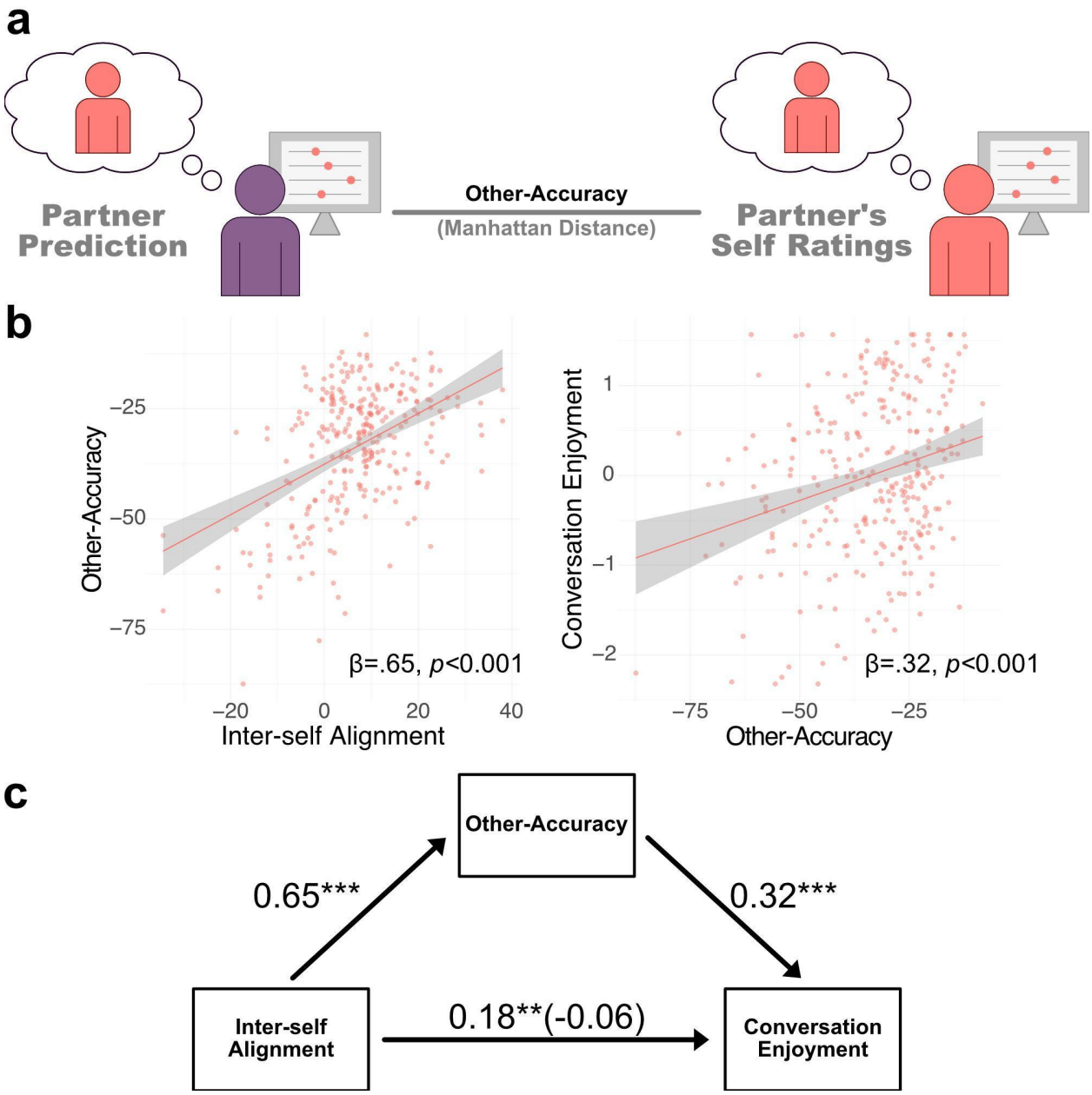

Figure 5: (a) Schematic of how partner prediction accuracy (or "other-accuracy") was calculated. We found the Manhattan distance between a participants' rating of their partners traits to their partners' ratings of their own traits. Distances were reversed scored so that larger values indicated greater accuracy in predicting their partner's trait ratings. Trait ratings were mean-centered and z scored for each participant prior to distance calculation to account for variations in scale usage across participants. (b) Scatter plots with regression lines showing the relationship between (left) inter-self alignment and other-accuracy and (right) other-accuracy and conversation enjoyment. (c) Depiction of how other-accuracy fully mediates the relationship between inter-self alignment and conversation enjoyment. Standardized regression coefficients are shown next to their corresponding arrows. The coefficient in parentheses reflects the relationship between inter-self alignment and conversation enjoyment when controlling for other-accuracy. All models control for baseline inter-self distance, reported prior familiarity with the

conversation partner, and depth of assigned topics. Random intercepts were included for each participant. * signifies $p < 0.05$, ** signifies $p < 0.01$, and *** signifies $p < 0.001$.

Exploring how selves align using the 3D Mind Model

Our results demonstrate three things: 1) peoples' views of themselves tend to align in conversations, 2) that alignment happens more during enjoyable conversations, and 3) the relationship between alignment and enjoyment is mediated by how well participants can predict their partners' self-views. We have not yet examined whether specific aspects of the self are more affected by conversation. Just as some mental states we experience, such as anger or sleepiness, are more or less likely to change from moment-to-moment, some aspects of the self might be more malleable in conversations (Honey et al., 2023; Kuppens et al., 2010; Sudhof et al., 2014; Thornton & Tamir, 2017). Similarly, aligning selves in certain ways may be more predictive of conversational enjoyment. To examine these questions, we used the 3D Mind Model—a well-established model of mental states—to reduce our trait space into three dimensions: valence, social impact, and rationality (Tamir et al., 2016; Tamir & Thornton, 2018; Thornton et al., 2022). Reducing the trait space into three dimensions allows for more tractable conclusions and results in less harsh thresholds from multiple comparisons corrections. The dimension of valence describes the extent to which mental states are either positive or negative, social impact describes the extent to which states are both intense and socially directed, and rationality describes the extent to which states are cognitive or emotional (Thornton & Tamir, 2020).

	Rank	Rationality	Social Impact	Valence
Top Traits	1	responsible	jealous	caring
	2	smart	friendly	sincere
	3	organized	passionate	enthusiastic
Bottom Traits	58	grouchy	wishy-washy	disorganized
	59	jealous	cold	neurotic
	60	cheerful	dull	unforgiving

Table 1: Table showing the traits which loaded the most and least onto each dimension of the 3D Mind Model.

The *affectr* R package allowed us to get weights for each dimension of the 3D Mind Model for all of our 60 traits (Thornton, 2018); see Supplementary Figure 4 for heatmap of entire weights matrix). Table 1 displays the traits which had the highest and lowest weights for each

dimension of the 3D Mind Model. We then used these weights to project participants' responses to the 60 traits into a three-dimensional space by multiplying a matrix of participants' trait scale responses by the 3D Mind Model weights matrix. This multiplication resulted in a new matrix with three 3D Mind Model scores for each participant which were used for a new set of exploratory analyses. As before, participants' trait ratings were mean-centered and z-scored prior to the multiplication to account for differences in scale usage across participants.

Selves align on social impact and rationality

To calculate alignment on each dimension of the 3D Mind Model, we found the absolute value of the difference between conversation partners' scores on each dimension in the baseline survey and in the post-conversation survey. These absolute differences describe the distance between dyads' scores on each dimension. We then subtracted the distance between dyads' scores after the conversation from the distance between their scores before the conversation to gauge how much each dyad aligned on each dimension of the 3D Mind Model. The resulting values are larger if conversation partners had more similar scores on a given dimension after the conversation compared to before the conversation.

Initially, to gauge whether pairs were more likely to align on certain dimensions of the 3D Mind Model, we compared the distribution of alignment values for each dimension to 0 with one-sample t-tests. Only rationality and social impact showed significant alignment across most participants ($M_{rationality} = 0.59$, $t_{rationality}(143) = 3.76$, 95% $CI_{rationality} = [0.28\ 0.91]$, $p_{rationality} < 0.001$; $M_{social\ impact} = 0.37$, $t_{social\ impact}(143) = 2.56$, 95% $CI_{social\ impact} = [0.08, 0.66]$, $p_{social\ impact} = 0.011$) and these tests remained significant after multiple comparisons correction via the Holm-Bonferroni method (Holm, 1979).

To gauge whether these effects were general across all participants or dyad-specific, we compared the observed medians of alignment on each dimension to null distributions of 5000 medians of alignment calculated from combinations of pseudo pairs. Alignment on all dimensions fell within the null distributions (13% of null results above median of rationality, 74% of null results above median of social impact, and 20% of null results above median of valence). These results suggest that all pairs tended to align on dimensions of the self-related to rationality and social impact, but that this was a general trend observed across all participants in the study.

Alignment on social impact and valence is predictive of enjoyment

Conversation partners who become more aligned in their self-views during a conversation also have more enjoyable conversations, but do all aspects of the self matter? If one dyad's conversation leads them to see themselves as funnier and another dyad's conversation leads them to feel more independent, will both dyads experience an equivalent boost in enjoyment? Or, are certain aspects of inter-self alignment more predictive of enjoyment than others? To answer these questions, we investigated the relationship between alignment on each dimension of 3D Mind Model space and conversation enjoyment.

To validate this dimensionally reduced space and check the robustness of our main findings, we re-ran our model relating inter-self alignment to conversation enjoyment using participants' 3D Mind Model scores in place of their trait ratings. We built linear mixed effects models with identical fixed and random effects structures as the original model except that inter-self alignment and baseline inter-self distance were calculated in 3D Mind Model space. Replicating our earlier results, participants whose personalities became more similar to their partners' tended to have more enjoyable conversations ($\beta = 0.22$, 95% CI = [0.08, 0.36], $p = 0.002$). Baseline distance between conversation partners' positions in 3D Mind Model space ($\beta = -0.15$, 95% CI = [-0.29, -0.01], $p = 0.03$) and self-reports of how well the conversation partners knew each other before the study ($\beta = 0.28$, 95% CI = [0.18, 0.38], $p < 0.001$) were also significant predictors of enjoyment.

To test whether alignment on any of the three 3D Mind Model dimensions was predictive of conversation enjoyment, we related alignment on each dimension to conversation enjoyment in three separate linear mixed effects models, controlling for baseline distance in 3D Mind Model space, assigned conversation topic, and self-reports of how well they knew their partner before the study. The models also included random intercepts for each participant, conversation partner, dyad, group, and conversation number. Alignment on social impact ($\beta = 0.15$, 95% CI = [0.02, 0.28], $p = 0.02$) and valence ($\beta = 0.17$, 95% CI = [0.04, 0.31], $p = 0.01$) were significant predictors of conversation enjoyment, and both remained significant after multiple comparison corrections using the Holm-Bonferroni method (Holm, 1979). Based on these results, we re-ran our mediation analysis two more times with alignment and other-accuracy generated from valence and social impact scores. Both models returned comparable results to the original model ($ACME_{social\ impact} = 0.08$, 95% CI = [0.03, 0.15], $p < 0.001$; $ADE_{social\ impact} = 0.06$, 95% CI = [-0.07, 0.18], $p = 0.374$; $ACME_{valence} = 0.12$, 95% CI = [0.03, 0.21], $p = 0.004$; $ADE_{valence} = 0.13$, 95% CI = [-0.02, 0.29], $p = 0.111$).

Discussion

It is well-established that talking with other people can change the way we think about the world (E. C. Baek & Falk, 2018; Brennan & Clark, 1996; Clark & Brennan, 1991; Coman et al., 2016; Hawkins et al., 2020; Pickering & Garrod, 2006; Sievers et al., n.d.; Stolk et al., 2016). Here, we demonstrate that conversation even changes the way we think about ourselves. Specifically, we show that conversation partners tend to think of themselves more similarly after a conversation than before, an effect we term "inter-self alignment." Further, the more partners' self-views aligned, the more they enjoyed their conversation. This effect manifested on a dyadic level, but not an individual level. An individual who tended to align with each of their conversation partners did not necessarily enjoy each of their conversations. Instead, enjoyable conversations were those in which *both* conversation partners became aligned with each other.

Contrary to our predictions, the relationship between inter-self alignment and enjoyment was not specific to deep conversations but manifested equally across both conversation conditions. This suggests that inter-self alignment does not require conscious reflection on the self but may be occurring in the background, consistent with spontaneous activity in self-

relevant neural regions known to occur during mind-wandering and at rest (Andrews-Hanna et al., 2010; Meyer & Lieberman, 2018; Ruby et al., 2013; Song & Wang, 2012).

The relationship between inter-self alignment and enjoyment was further mediated by how well participants were able to estimate their partners' own self-views. This suggests that developing more similar self-views over the course of a conversation is associated with having a better understanding of one's partner, leading to more enjoyable conversations. Exploratory analyses demonstrated that some aspects of the self converged more than others. Using the 3D Mind Model, we found that self-views tended to align on the dimensions of social impact and rationality, while aligning on the dimensions of social impact and valence predicted enjoyment. Social impact—capturing the degree to which trait ratings are socially directed (e.g. high ratings of adjectives like "jealous", "friendly", or "passionate")—was the only dimension that significantly predicted both inter-self alignment and conversation enjoyment. This suggests that conversations are most likely to facilitate social connection when people become a similarly social self, echoing the classic William James quote that "a man has as many social selves as there are individuals who recognize him" (James, 1890).

The inter-self alignment we observed was ephemeral, with enjoyment no longer being related to alignment a week after the conversation. This is not surprising given that these dyads were strangers who had a single, short conversation. Indeed, it would be disconcerting if every one-off conversation with a stranger radically altered our sense of self. However, repeated interactions could have a more enduring impact on self-views, consistent with research on friends becoming more similar over time (van Zalk et al., 2020). If so, the convergence of self-views through repeated conversation may be one mechanism underlying perceptions of self-other overlap (Aron et al., 1991; Galinsky et al., 2005; Tan et al., 2015; Waugh & Fredrickson, 2006) and may contribute to personality-based homophily in social networks (Deutsch et al., 1991; Laakasuo et al., 2020; McPherson et al., 2001). The prosocial strength of homophily is apparent in our findings as well. We show that trait-based homophily (i.e. baseline inter-self distance) is predictive of both conversation enjoyment and perceived social closeness between conversation partners. Further investigation is needed to determine the extent to which these effects compound at longer timescales.

It is important to note, however, that even temporary changes in self-views may have considerable benefits due to the fact that they predict enjoyable social interactions. Short periods of enjoyable activities, such as leisure or physical activity, promote overall wellbeing (Lathia et al., 2017; Zawadzki et al., 2015). Similarly, we find that enjoyable social interactions are those in which partners want to interact again. In turn, repeated interactions have been found to induce feelings of connectedness and wellbeing (Bernstein et al., 2018; Sun et al., 2020). Even if each individual conversation evokes an ephemeral change in affect, having a high quantity of enjoyable interactions could profoundly affect feelings of connectedness and wellbeing in the aggregate.

The present research also compliments and extends the literature on the effect of mental simulation on self-knowledge. It has been demonstrated that thinking about other people can

make participants think of themselves as being more similar to the simulated other (Meyer et al., 2019; Rubin-McGregor et al., 2022; Schneider et al., 2022). Thus, it is possible that the conversation partners in our studies who achieved better predictions of each other did so via mental simulation. More accurate simulation in conversations with greater inter-self alignment would also be consistent with work demonstrating that mental simulation can bolster feelings of social connection (E. Baek et al., 2019). Future research is needed to help clarify the link between inter-self alignment, mental simulation, and enjoyment.

Some limitations constrain the generalizability and scope of the conclusions which can be drawn from the reported findings. First, the relationship we observed between inter-self alignment and conversation enjoyment was not causal. We cannot know, therefore, whether inter-self alignment leads to greater conversational enjoyment, whether the inverse is true, or whether a third variable predicts both. Second, the participants in our study were recruited from an undergraduate population which limits the generalizability of our findings. Undergraduate students are younger than the general population and may have a less firmly established sense of self, potentially making them more willing to shift their self-views after enjoyable conversations. Third, although we varied context in terms of conversational depth and whether conversations took place in person vs. online, there are many other contexts that would be important to test in future studies (e.g., between close others, conversations of different lengths, or repeated conversations). Our particular sample consisted of students in a relatively novel social network who may have been more eager to connect with their peers compared to others in more established social networks.

Overall, our findings demonstrate that when conversation partners' self-views converge, they enjoy their conversations more. This effect was robust to whether the conversation was shallow or deep. The relationship between inter-self alignment and enjoyment was also short-lived, suggesting that conversation partners *momentarily* co-create perceptions of themselves. Future research is needed to understand whether and how these effects may compound with repeated interactions—i.e., to what extent we become the company we keep. We are excited by recent methodological and analytical advances that make these and other questions about social interaction newly tractable (Dingemanse et al., 2023; Redcay & Schilbach, 2019; Wheatley et al., n.d., 2019).

Our social understanding of the self parallels the social understanding of language. Just as Wittgenstein noted that "the meaning of a word is its use in the language" (Wittgenstein, 1953), we find that our view of the self is not static but dynamically forged in social interaction. Through the back-and-forth of dialogue, our self-views mutually evolve. The most enjoyable conversations are those in which we come to see ourselves in the same way.

References

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *19*(6), 586–592.

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., & Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *Journal of Neurophysiology*, *104*(1), 322–335.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21544–21549.

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60*(2), 241–253.

Baek, E. C., & Falk, E. B. (2018). Persuasion and influence: what makes a successful persuader? *Current Opinion in Psychology*, *24*, 53–57.

Baek, E., Tamir, D., & Falk, E. (2019). Considering others' mental states causally increases feelings of social bonding and information sharing. *Psyarxiv.com*. https://psyarxiv.com/nw43x/download?format=pdf

Bassetti, C., & Liberman, K. (2021). Making talk together: Simultaneity and rhythm in mundane Italian conversation. *Language & Communication*, *80*, 95–113.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models

Using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bernstein, M. J., Zawadzki, M. J., Juth, V., Benfield, J. A., & Smyth, J. M. (2018). Social interactions in daily life: Within-person associations between momentary social experiences and psychological and physical health indicators. *Journal of Social and Personal Relationships*, *35*(3), 372–394.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Burr, V. (1995). *An introduction to social constructionism*. Routledge.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* (pp. 127–149). American Psychological Association.

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. *Proceedings of the National Academy of Sciences*, *113*(29), 8171–8176.

Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Brooks, A. W. (2020). The many minds problem: disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, *31*, 22–27.

Cooney, G., & Wheatley, T. (in press). On Conversation. In D. T. Gilbert, S. Fiske, E. Finkel, & W. Mendes (Eds.), *Handbook of Social Psychology*.

Costa, P. T., & McCrae, R. R. (1986). Personality stability and its implications for clinical psychology. *Clinical Psychology Review*, *6*(5), 407–423.

Deutsch, F. M., Sullivan, L., Sage, C., & Basile, N. (1991). The Relations among Talking, Liking, and Similarity between Friends. *Personality & Social Psychology Bulletin*, *17*(4), 406–411.

Diehl, M., Jacobs, L. M., & Hastings, C. T. (2006). Temporal Stability and Authenticity of Self-Representations in Adulthood. *Journal of Adult Development*, *13*(1), 10–22.

Dingemanse, M., Liesenfeld, A., Rasenberg, M., Albert, S., Ameka, F. K., Birhane, A., Bolis, D.,

Cassell, J., Clift, R., Cuffari, E., De Jaegher, H., Novaes, C. D., Enfield, N. J., Fusaroli, R., Gregoromichelaki, E., Hutchins, E., Konvalinka, I., Milton, D., Rączaszek-Leonardi, J., … Wiltschko, M. (2023). Beyond Single-Mindedness: A Figure-Ground Reversal for the Cognitive Sciences. *Cognitive Science*, *47*(1), e13230.

Dumas, J. E., Johnson, M., & Lynch, A. M. (2002). Likableness, familiarity, and frequency of 844 person-descriptive words. *Personality and Individual Differences*, *32*(3), 523–531.

Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Dunbar, R. I. M. (2004). Gossip in Evolutionary Perspective. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, *8*(2), 100–110.

Echterhoff, G., & Higgins, E. T. (2018). Shared reality: Construct and mechanisms. *Current Opinion in Psychology*, *23*, iv – vii.

Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared Reality: Experiencing Commonality With Others' Inner States About the World. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *4*(5), 496–521.

Elfenbein, H. A., Jang, D., Kleinbaum, A. M., & Welker, C. (2023). *The choice that chooses you back: Introducing social variety-seeking as a trait predicting greater network brokerage through building connections with previously disconnected others*.

Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology*, *41*(2), 232–242.

Galinsky, A. D., Ku, G., & Wang, C. S. (2005). Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations: GPIR*, *8*(2), 109–124.

Gergen, K. J. (2011). *The Social Construction of Self*. Oxford University Press.

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E. (2009). Psychosocial Correlates of Interpersonal Sensitivity: A Meta-Analysis. *Journal of Nonverbal Behavior*, *33*(3), 149–180.

Hampson, S. E. (1988). *The Construction of Personality : An Introduction: Vol. 2nd ed*. Taylor &

Francis Routledge.

Hawkins, R. D., Frank, M. C., & Goodman, N. D. (2020). Characterizing the Dynamics of Learning in Repeated Reference Games. *Cognitive Science*, *44*(6), e12845.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, Theory and Applications*, *6*(2), 65–70.

Honey, C. J., Mahabal, A., & Bellana, B. (2023). Psychological Momentum. *Current Directions in Psychological Science*, *32*(4), 284–292.

James, W. (1890). *The principles of psychology, Vol I* (Vol. 697). Henry Holt and Co.

Kardas, M., Kumar, A., & Epley, N. (2022). Overly shallow?: Miscalibrated expectations create a barrier to deeper conversation. *Journal of Personality and Social Psychology*, *122*(3), 367–398.

Kossinets, G., & Watts, D. J. (2009). Origins of Homophily in an Evolving Social Network. *The American Journal of Sociology*, *115*(2), 405–450.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984–991.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26.

Laakasuo, M., Rotkirch, A., van Duijn, M., Berg, V., Jokela, M., David-Barrett, T., Miettinen, A., Pearce, E., & Dunbar, R. (2020). Homophily in Personality Enhances Group Success Among Real-Life Friends. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00710

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier People Live More Active Lives: Using Smartphones to Link Happiness and Physical Activity. *PloS One*, *12*(1), e0160589.

Laursen, B. (2017). Making and keeping friends: The importance of being similar. *Child Development Perspectives*, *11*(4), 282–289.

Ma, L., Krishnan, R., & Montgomery, A. L. (2015). Latent Homophily or Social Influence? An Empirical Analysis of Purchase Within a Social Network. *Management Science*, *61*(2), 454–473.

Markus, H., & Wurf, E. (1987). The Dynamic Self-Concept: A Social Psychological Perspective. *Annual Review of Psychology*, *38*(1), 299–337.

Mastroianni, A. M., & Gilbert, D. T. (2021). Do conversations end when people want them to? *Proceedings of the*. https://www.pnas.org/doi/abs/10.1073/pnas.2011809118

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, *27*(1), 415–444.

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: well-being is related to having less small talk and more substantive conversations. *Psychological Science*, *21*(4), 539–541.

Meyer, M. L., & Lieberman, M. D. (2018). Why people are always thinking about themselves: Medial prefrontal cortex activity during rest primes self-referential processing. *Journal of Cognitive Neuroscience*, *30*(5), 714–721.

Meyer, M. L., Zhao, Z., & Tamir, D. I. (2019). Simulating other people changes the self. *Journal of Experimental Psychology. General*, *148*(11), 1898–1913.

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332.

Pickering, M. J., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, *4*(2), 203–228.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews. Neuroscience*, *20*(8), 495–505.

Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., Glazer, T., Knox, D., Liebscher, A., & Marin, S. (2023). The CANDOR corpus: Insights from a large multimodal dataset of naturalistic conversation. *Science Advances*, *9*(13), eadf3197.

878 Rossignac-Milon, M., Bolger, N., Zee, K. S., Boothby, E. J., & Higgins, E. T. (2020). Merged
879 minds: Generalized shared reality in dyadic relationships. *Journal of Personality and Social*
880 *Psychology*. https://doi.org/10.1037/pspi0000266
881 Rubin-McGregor, J., Zhao, Z., & Tamir, D. I. (2022). Simulation induces durable, extensive
882 changes to self-knowledge. *Journal of Experimental Social Psychology*.
883 https://www.sciencedirect.com/science/article/pii/S0022103121001323
884 Ruby, F. J. M., Smallwood, J., Engen, H., & Singer, T. (2013). How self-generated thought
885 shapes mood--the relation between mind-wandering and mood depends on the socio-
886 temporal content of thoughts. *PloS One*, *8*(10), e77554.
887 Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The revised UCLA Loneliness Scale:
888 concurrent and discriminant validity evidence. *Journal of Personality and Social*
889 *Psychology*, *39*(3), 472–480.
890 Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural
891 model of momentary subjective well-being. *Proceedings of the National Academy of*
892 *Sciences of the United States of America*, *111*(33), 12252–12257.
893 Schmid Mast, M., & Hall, J. A. (2018). The Impact of Interpersonal Accuracy on Behavioral
894 Outcomes. *Current Directions in Psychological Science*, *27*(5), 309–314.
895 Schneider, M. J., Rubin-McGregor, J., & Tamir, D. (2022). *Simulation requires activation of self-*
896 *knowledge to change self-concept*. https://psyarxiv.com/92mru/download
897 Sievers, B., Welker, C., Hasson, U., Kleinbaum, A. M., & Wheatley, T. (in press). *How*
898 *consensus-building conversation changes our minds and aligns our brains*. Nature
899 Communications. https://psyarxiv.com/562z7/
900 Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social*
901 *Psychology*, *30*(4), 526.
902 Snyder, M. (1984). When Belief Creates Reality. In L. Berkowitz (Ed.), *Advances in*
903 *Experimental Social Psychology* (Vol. 18, pp. 247–305). Academic Press.

Song, X., & Wang, X. (2012). Mind wandering in Chinese daily lives--an experience sampling study. *PloS One*, *7*(9), e44423.

Sprecher, S. (2021). Closeness and other affiliative outcomes generated from the Fast Friends procedure: A comparison with a small-talk task and unstructured self-disclosure and the moderating role of mode of communication. *Journal of Social and Personal Relationships*, *38*(5), 1452–1471.

Stolk, A., Verhagen, L., & Toni, I. (2016). Conceptual Alignment: How Brains Achieve Mutual Understanding. *Trends in Cognitive Sciences*, *20*(3), 180–191.

Sudhof, M., Goméz Emilsson, A., Maas, A. L., & Potts, C. (2014). Sentiment expression conditioned by affective transitions and social forces. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1136–1145.

Sun, J., Harris, K., & Vazire, S. (2020). Is well-being associated with the quantity and quality of social interactions? *Journal of Personality and Social Psychology*, *119*(6), 1478–1496.

Swann, W. B., & Hill, C. A. (1982). When our identities are mistaken: Reaffirming self-conceptions through social interaction. *Journal of Personality and Social Psychology*, *43*(1), 59–66.

Swann, W. B., Jr. (2012). Self-verification theory. In E. Higgins (Ed.), *Handbook of theories of social psychology, Vol* (Vol. 2, pp. 23–42).

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212.

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 194–199.

Tan, Q., Zhan, Y., Gao, S., Fan, W., Chen, J., & Zhong, Y. (2015). Closer the relatives are, more intimate and similar we are: Kinship effects on self-other overlap. *Personality and*

Individual Differences, *73*, 7–11.

Templeton, E. M., Chang, L. J., Reynolds, E. A., LeBeaumont, M. D. C., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, *119*(4), e2116915119.

Thornton, M. A. (2018, August 13). *Constructing a 3-dimensional affect dictionary with word vector embeddings*. http://markallenthornton.com/blog/3daffect/

Thornton, M. A., & Tamir, D. I. (2017). Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(23), 5982–5987.

Thornton, M. A., & Tamir, D. I. (2020). People represent mental states in terms of rationality, social impact, and valence: Validating the 3d Mind Model. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *125*, 44–59.

Thornton, M. A., Wolf, S., Reilly, B. J., Slingerland, E. G., & Tamir, D. I. (2022). The 3d Mind Model characterizes how people understand mental states across modern and historical cultures. *Affective Science*, *3*(1), 93–104.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). *mediation: R package for causal mediation analysis*. https://dspace.mit.edu/handle/1721.1/91154?show=full

van Zalk, M. H. W., Nestler, S., Geukes, K., Hutteman, R., & Back, M. D. (2020). The codevelopment of extraversion and friendships: Bonding and behavioral interaction mechanisms in friendship networks. *Journal of Personality and Social Psychology*, *118*(6), 1269–1290.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social Forces; a Scientific Medium of Social Study and Interpretation*, *56*(2), 576–597.

Wagner, U., Galli, L., Schott, B. H., Wold, A., van der Schalk, J., Manstead, A. S. R., Scherer, K., & Walter, H. (2015). Beautiful friendship: Social sharing of emotions improves subjective feelings and activates the neural reward circuitry. *Social Cognitive and Affective*

Neuroscience, *10*(6), 801–808.

Waugh, C. E., & Fredrickson, B. L. (2006). Nice to know you: Positive emotions, self-other overlap, and complex understanding in the formation of a new relationship. *The Journal of Positive Psychology*, *1*(2), 93–106.

Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the Isolated Brain: The Promise and Challenge of Interacting Minds. *Neuron*, *103*(2), 186–188.

Wheatley, T., Thornton, M., Stolk, A., & Chang, L. J. (n.d.). The emerging science of interacting minds. *Perspectives in Psychological Science*.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York, NY, USA: Wiley-Blackwell.

Wynn, C. J., Barrett, T. S., Berisha, V., Liss, J. M., & Borrie, S. A. (2023). Speech Entrainment in Adolescent Conversations: A Developmental Perspective. *Journal of Speech, Language, and Hearing Research: JSLHR*, *66*(8S), 3132–3150.

Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2022). The Conversational Circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, *44*, 293–302.

Zawadzki, M. J., Smyth, J. M., & Costigan, H. J. (2015). Real-Time Associations Between Engaging in Leisure and Daily Health and Well-Being. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *49*(4), 605–615.