Finding Agreement: fMRI-hyperscanning reveals that dyads explore in mental state space to align opinions.

*Sebastian P.H. Speer[1], Haran Sened[2], Laetitia Mwilambwe-Tshilobo[3,4],Emily B. Falk [4,5,6,7], Lily Tsoi[8], Shannon M. Burns[#,9,10], Diana I. Tamir[#,1,3]

[1]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, USA.
[2]Department of Counseling and Human Development, University of Haifa, Haifa, Israel.
[3]Department of Psychology, Princeton University, Princeton, New Jersey, USA.
[4]Annenberg School for Communication, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[5]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[6]Wharton Marketing Department, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[7]Operations, Information, and Decisions Department, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[8]Department of Psychology, Caldwell University, Caldwell, New Jersey, USA.
[9]Department of Psychological Science, Pomona College, Claremont, California, USA.
[10]Department of Neuroscience, Pomona College, Claremont, California, USA.

*Corresponding author: Sebastian Speer. Email: sspeer@princeton.edu

# These authors jointly supervised this work: Shannon Burns, Diana Tamir

Abstract

Many prize synchrony as the ingredient that turns a conversation from a debate into a delightful duet. Yet, learning from other people's diverging opinions can also foster understanding and agreement, satisfy curiosity, and spur imagination. Using fMRI hyperscanning and natural language processing we tested how two debaters navigate conflictual conversations to find agreement. Dyads (N=60) discussed pressing societal problems while being instructed to either *persuade* their partner or *compromise* with each other. Our analysis uncovered three key insights: First, individuals instructed to compromise rather than persuade tended to agree more at the end of the session. Second, hyperscanning and linguistic analyses revealed that encouraging compromise resulted in increased exploration during conversations; dyads given compromise instructions traversed more diverse mental states and topics. Third, heightened exploration was linked to greater eventual agreement. Notably, the effect of the compromise instruction on agreement was entirely mediated by the degree of exploration. Together, these results suggest that trying to find agreement may be spurred by exploration, something that happens when people are motivated to compromise but not persuade.

Introduction

People do not always agree with each other. They disagree with their friends, colleagues, or strangers on basic questions, such as what to eat for dinner, and on big societal questions, such as whether and how to address climate change or how to reduce the cost of education. Such disputes impede progress and can have serious consequences for human well-being and social and environmental stability. How can people approach and engage in decision-making conversations to maximize their ultimate consensus? In this study, we use fMRI hyperscanning and natural language processing to pioneer an investigation of the neural and linguistic dynamics that support mutual agreement in a real-time conversation.

Some conversations naturally foster agreement and harmony, while others deepen conflict and amplify discord. A rich foundation of work on negotiation conversations points to at least two factors that determine which course a conversation will take: motives and strategies (J. M. Brett, 2000; J. Brett & Thompson, 2016). First, the outcome of decision-making conversations often depends on partners' motives for the conversation—such as their underlying opinions, goals, and needs. The motives, in turn, influence the second factor–communication strategies–that people use to bridge disconnects and pursue agreement. To reach agreement, people might use strategies for aligning their thoughts and feelings by *finding common ground.* They might also use a strategy for *exploring new ground*, offering fresh, unexpected ideas or diverging perspectives. Together, motives and strategies interact to determine the effectiveness of the negotiation process and the final agreements reached.

Social motives shape the negotiation process

Partners' motives for the conversations powerfully shape the dyad's potential for agreement (De Dreu et al., 2000; Olekalns & Smith, 1999, 2003b, 2003a). Social motives can be thought of as the goals in conversations (J. Brett & Thompson, 2016). Dyads can bring at least two categories of motives to decision-making conversations: Competitive and cooperative.

Competitive motives are characterized by maximizing one's own gains, or maximizing the difference between the self and other (Beersma & De Dreu, 2002; De Dreu et al., 2001; Messick & McClintock, 1968; Pruitt & Rubin, 1986). When pursuing competitive motives, people often have a specific end goal in mind, leading them to direct the conversation toward a particular outcome. This focus on achieving a desired result can narrow a person's thinking and reduce cognitive flexibility, making it harder to consider alternative viewpoints (Brynjarsdottir et al., 2012; Hommel, 2015; Zhang et al., 2020). In addition, people are often motivated to resist their competitor's persuasion, to retain their sense of freedom and control (Knowles & Linn, 2004; Wheeler et al., 2007). Thus, competitive negotiators may hesitate to support and build on each others' ideas (Decety et al., 2004; Lu et al., 2019). This typically results in a failure to identify tradeoffs that could have created more value and higher joint gains (Kong et al., 2014). Together, this work suggests that competitive motives will lead to more rigid, constrained, and focused conversations, and result in lower levels of agreement.

By contrast, cooperative motives are often characterized by prosocial intentions focused on maximizing outcomes for the self and others. Dyads in a cooperative mindset pursue a shared goal, and actively process, learn, and build on each others' ideas (Hargadon & Bechky, 2006; Vera & Crossan, 2005)**.** As a result, the cooperative mindset has been linked to more creativity and cognitive flexibility (Bittner et al., 2016; Carmeli et al., 2015; Gelfo, 2019; Hon et al., 2014; Laureiro-Martínez & Brusoni, 2018) and seeing truth as more subjective (M. Fisher et al., 2017). Cooperative motives are linked to integrative negotiation behaviors (De Dreu et al., 2000;

Deutsch, 1973)**,** which predominantly consist of sharing information about interests and opinions and then forging tradeoffs to produce high joint gains (Weingart et al., 1990). As a result, cooperative negotiators consistently achieve higher joint gains and create more value than competitive behaviors (De Dreu et al., 2000; Kong et al., 2014; Olekalns & Smith, 1999, 2003a, 2003b)**.** Collectively, these findings suggest that cooperative motives lead to more creative, flexible, and synergistic conversations, and result in higher levels of agreement.

Mapping motives onto the interaction dynamics

If cooperative social motives are associated with more flexibility and agreement, while competitive motives are linked to a narrowed focus and limited mutual benefit, how do these different motives manifest in interaction dynamics?

Two important dynamics that unfold in dyads during conversations are alignment and exploration (Burns et al., 2024): Psychologists and neuroscientists have prized *alignment* of mental states as an indicator of shared understanding, or a path to conflict resolution (Tunçgenç & Cohen, 2018; Valdesolo & DeSteno, 2011; Wiltermuth & Heath, 2009). Agreement is achieved when two people's mental states converge: people agree more when they show greater linguistic (Manson et al., 2013), physiological (Behrens et al., 2020), and neural alignment (Liu et al., 2017; Wang et al., 2022; H. Zhang et al., 2023). That is, alignment may facilitate agreement because it enhances joint attention, mutual understanding, and empathy, all critical elements for cooperative behavior (Liu et al., 2017; Lu et al., 2019; Wang et al., 2022)**.** Indeed, alignment has been associated with positive social outcomes in related contexts, such as emotional support (Doré & Morris, 2018), interpersonal liking (Ireland et al., 2011; Putman & Street, 1984; Street et al., 1983), social influence (Ludwig et al., 2013), compliance (Jacob et al., 2011; Ramseyer & Tschacher, 2011) and greater understanding between managers and employees (Balconi, 2010). However, it is also possible that alignment merely reflects shared understanding, but does not create it.

Exploring new ground, a process by which people communicate diverse opinions, can foster understanding and collaboration (Perc & Szolnoki, 2008). While exploration of alternative views may initially appear as a divergence in mental states, dyads who more thoroughly explore may be better poised to reach eventual agreement. When individuals engage with diverse viewpoints and thoughtfully process the information, the resulting attitude change is more durable and meaningful than when individuals simply start out in agreement, which may reflect superficial consensus without lasting impact (Petty & Cacioppo, 1986). Entertaining diverse ideas within one's own mind can also mitigate polarized attitudes (Sassenberg & Winter, 2024). A similar process may be at play when two minds entertain diverse ideas. Indeed, exploration across minds has been linked to beneficial social outcomes, such as more secure attachment (Nguyen et al., 2024), improved emotion regulation (Lee & Fujita, 2023; Powers & LaBar, 2019), and more effective problem solving (Goldstone et al., 2024).

Given its association with divergent and flexible thinking, a cooperative mindset may encourage the *exploration of new ground* during a conversation*.* The cooperative mindset fosters more cognitive flexibility and creativity and is focused on information sharing (Decety et al., 2004; Gilson & Shalley, 2004; Hargadon & Bechky, 2006; Van Knippenberg et al., 2004; Vera & Crossan, 2005). This may explain why cooperation goals promote more eventual agreement. In contrast, a competitive mindset might lead to less exploration through the pursuit of winning or persuading one's partner of a particular point*.* A competitive negotiator with a clear end state in mind might narrow the focus of a conversation, inhibiting the mutual sharing and processing of broad information that is useful for meaningful agreement.

The current study: How do we find agreement?

This study aimed to uncover the emergent neural and linguistic dynamics that help people reach mutual agreement, a question we addressed in four parts:

First, to replicate previous findings (De Dreu et al., 2000; Kong et al., 2014; Olekalns & Smith, 1999, 2003a, 2003b), we investigated how social motives influence eventual agreement. Dyads engaged in real-time conversations on how to allocate money to address two contentious issues, air pollution and the cost of tuition. They provided their opinions on how to allocate money before and after the conversation, which allowed us to measure the extent to which the conversation helped them find agreement. We experimentally induced a competitive mindset by instructing half of the dyads to persuade each other (both partners received the same instruction); we induced a cooperative mindset in the other half of dyads by instructing them to compromise with their conversation partner. We then tested if dyads instructed to compromise reached higher levels of agreement than dyads instructed to persuade.

Next, we addressed an open question in the literature by examining how these social motives shaped dyads' interaction dynamics. Specifically, we examined how the cooperation and persuasion motives related to two conversational strategies – exploration vs.alignment. To this end, we measured exploration in two spaces: dyads' mental state as measured through their neural and linguistic dynamics, and topic trajectories over the course of the conversations.

Mental states offer us a window into an individual's thoughts and feelings throughout the conversation. Two people's coordinated motion within mental state space thus reflects the extent to which they feel free to venture into unconstrained territory or stay constrained within a safe mental space. Prior work has established that three dimensions, namely social impact, rationality, and valence, named the 3D mind model, accurately represent people's mental states (Tamir et al., 2016; Tamir & Thornton, 2018; Thornton & Tamir, 2020). Dyads might evidence exploration by drifting farther away from each other over time in mental state space, or by exploring many novel, diverse mental states over the conversation. We measure both types of exploration in neural and linguistic data here, as divergence (increasing distance over time) and novelty (Ravreby et al., 2022), respectively. The topics of conversation unveils a perspective into if dyads explore diverse content. We used topic modeling to assess what dyads talked about throughout their conversations. Dyads might evidence exploration by covering more topics, more topic switches, and more distant topic jumps. We measure each of these types of topic exploration here.

We used both neuroimaging and linguistic data to track participants' exploration through mental state space and topic space. fMRI-hyperscanning provides us with unprecedented insight into how the interaction partners reciprocally shape each other's emotions, thoughts, and mental states in real time, without interrupting the flow of the conversation. Our neuroimaging approach allows us to pinpoint linguistic markers associated with successful conversations (Speer et al., 2024). These insights can then be applied to analyze much larger datasets in the future, thanks to the efficiency and scalability of natural language processing (NLP).To track individuals' location in mental state space, we used neural decoding models that translated whole-brain neural activity patterns into a location in 3D mental state space (Kragel et al., 2018), and NLP that translated a speaker's language into a location in 3D mental state space. To track dyads' location in topic space, we used BERTopic to extract topic representations for each sentence in every conversation across all dyads. With these measures in hand, we tested if cooperative dyads engaged in more exploration than dyads instructed to persuade.

Third, we tested whether the interaction dynamics also predict agreement. To this end, we investigated if dyads who engaged in more mental state and topic exploration agreed more by the end of the conversation.

Finally, we tested whether the interaction dynamics mediate the effect of motives on agreement. To accomplish this, we conducted a mediation analysis to test whether the instructions to compromise lead to greater agreement by encouraging greater exploration. We implemented this multi-modal approach in combination with naturalistic free-flowing conversations to capture the bidirectional and emergent properties of social interactions.

Results

The instruction to compromise amplifies the positive effect of conversation on agreement

As a first step, we tested how social motives (*compromise vs. persuade*) influence eventual agreement. Prior to each conversation, participants were exposed to the discussion problem and five different solutions to address the problem. Participants independently allocated 100 million dollars to the five different solutions both before and after the conversation, which allowed us to measure how similar dyads' opinions were before and after the conversation. Participants agreed significantly more after the conversations than before (Figure 1: $t(216) = 6.96$, $p < 0.001$). This increased agreement was observed in both conditions (compromise: $t(108) = 6.36$, $p < 0.001$; persuade: $t(106) = 3.51$, $p < 0.001$). Next, we tested how different social motives influence this relationship. We found a significant interaction effect between time and condition (Figure 1: $t(216) = 2.13$, $p = 0.03$) indicating that instruction to compromise amplified this effect: participants in the compromise condition agreed significantly more than participants in the persuade condition after the conversation (Figure 1: $t(54) = 3.28$, $p = 0.002$) but not before the conversation (Figure 1: $t(54) = 0.84$, $p = 0.40$). So, talking to each other and exchanging perspectives appears to increase agreement and being in a compromising mindset enhances this effect.

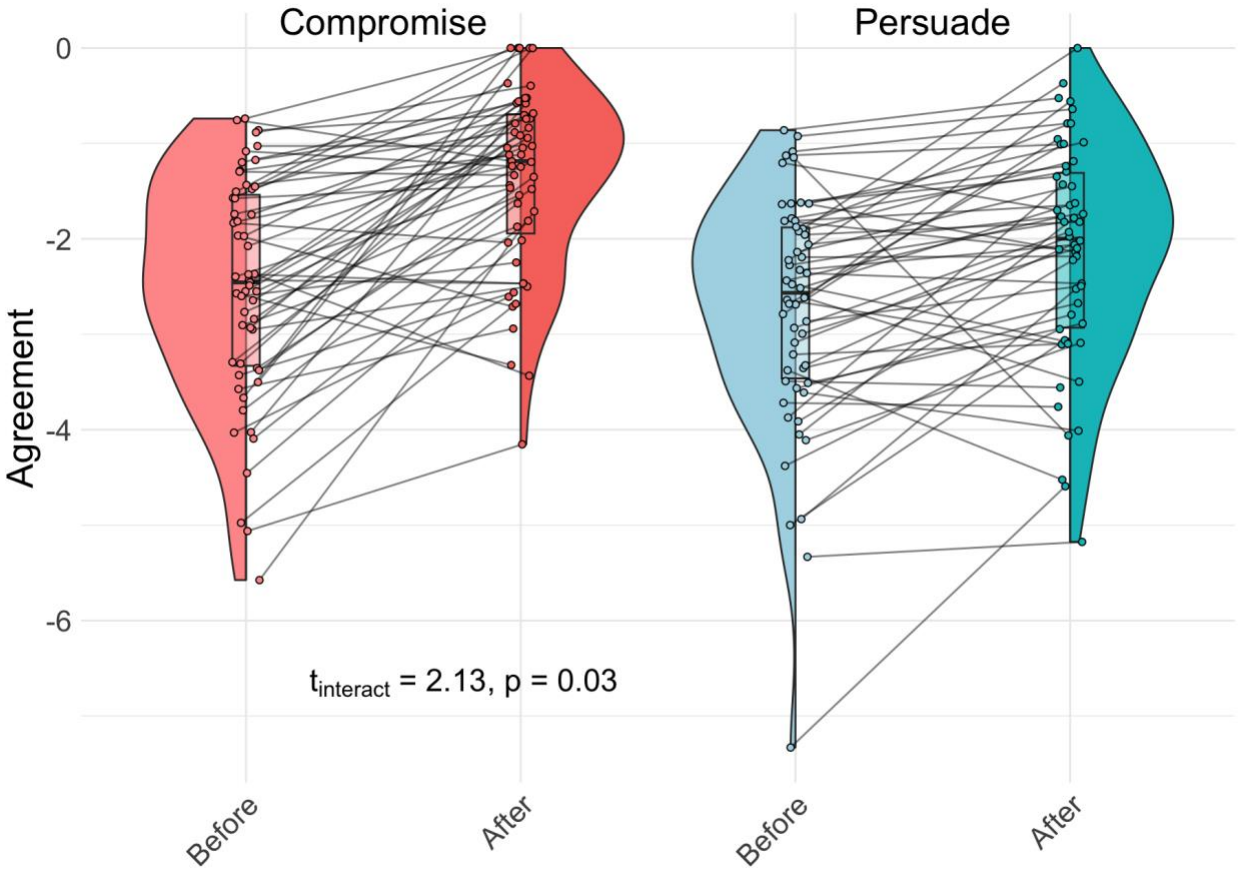


Figure 1. Violin plots illustrating that participants agree significantly more after the conversation as compared to before the conversation and that this effect is amplified in the compromise

condition. Agreement represents the inverse distance in opinion space (agreement = -1* distance; a smaller distance represents higher agreement).

Instruction to compromise increases exploration

Social motives matter: being instructed to compromise increased the chance of finding agreement. Next, we investigated how instructions to compromise vs. persuade impacted interaction dynamics. We quantified dynamics by tracking how dyads moved through mental state space and topic space. First, to measure how dyads moved through mental state space, we located each partner on 3 mental state dimensions throughout the conversation. To locate dyads in neural mental state space, we applied a neural decoding model to both partners whole-brain neural activity at each moment of the conversation; to locate dyads in linguistic mental state space, we used natural language procession on the language each speaker produced over the course of the conversation (Speer et al., 2024; Supplementary Material 1). We measure mental state exploration as the change in distance between participants over time (divergence) and the absolute amount of exploration (novelty). Second, we measured how dyads moved through topic space. We applied unsupervised machine learning to extract the topic of each turn and calculate the semantic distance between adjacent topics (Supplementary Material 2).

Mental state exploration. Using the decoded neural coordinates, we calculated the distance between the two conversation partners in mental state space at each time point. A smaller distance represented greater alignment of mental states between the dyad. Increasing distance over time represented more exploration. We expected the instruction to compromise to induce exploration and the instruction to persuade to reduce exploration. We employed multilevel models to examine if dyads diverged over the course of a conversation. Analyses revealed that dyads instructed to compromise diverged in neural mental state space, whereas dyads instructed to persuade converged (two-way interaction between time and social motive: β = -0.0005, 95% CI [-0.0007, -0.0003], *SE* = 0.00008, $p < 0.0001$; Figure 2A). Dyads motivated to compromise aligned more initially than dyads who were assigned to persuade and then diverged in mental state space until their distance was larger than the persuaders. The linguistic analysis replicated this interaction effect (β = -0.05, 95% CI [-0.08, -0.02], *SE* = 0.016, p = 0.001; Figure 2B), such that dyads in the compromise condition diverged in mental state space whereas participants in the persuade condition converged.

To measure novelty, we calculated the difference in distribution of mental state locations from turn to turn. We expected the compromise condition to be associated with more novelty. The neural novelty analysis revealed a trend suggesting that dyads in the compromise condition exhibited significantly more novel mental states than participants in the persuade condition ($t(53)$ = 1.95, p = 0.056; Figure 3A). The linguistic analyses revealed a significant effect in the same direction, such that dyads in the compromise condition exhibited significantly more novel mental states than dyads in the persuade condition ($t(53)$ = 3.19, p = 0.002; Figure 3B).

Together, these findings suggest that the instruction to compromise increases mental state exploration both in terms of relative distance (divergence) and absolute novelty.

Topic exploration. To measure topic exploration, we applied unsupervised machine learning to extract the topic of each turn and calculate the semantic distance between them (Supplementary Material 2). We expected dyads in the compromise condition to explore more topic space, which we measured in several ways. To parallel the mental state analyses, we first looked at alignment vs.divergence in topic space. This analyses revealed a two-way interaction

effect between turns and social motives on topic distance: Dyads instructed to compromise explored increasingly more distinct topics than dyads instructed to persuade (β = -0.003, 95% CI [-0.006, -0.0005], *SE* = 0.001, *p* = 0.02; Figure 2C). Thus, similar to the mental state findings, dyads instructed to compromise diverge in topic space over time, whereas dyads instructed persuade align in topic space.

Mirroring the mental state novelty analyses, we found that participants in the compromise condition exhibited more topic novelty, as evidenced by longer distances between topics (*t(53)* = 3.56, *p* = 0.0008; Figure 3C). In addition, participants instructed to compromise generated significantly more topics (*t(53)* = 2.76, *p* = 0.007) and showed a trend towards switching topics more frequently (*t(53)* = 1.99, *p* = 0.051).

Collectively, these findings demonstrate that dyads who are instructed to compromise explore a more expansive topic space by switching between more distant topics more rapidly, whereas conversationalists motivated to persuade stick with similar topics for longer.

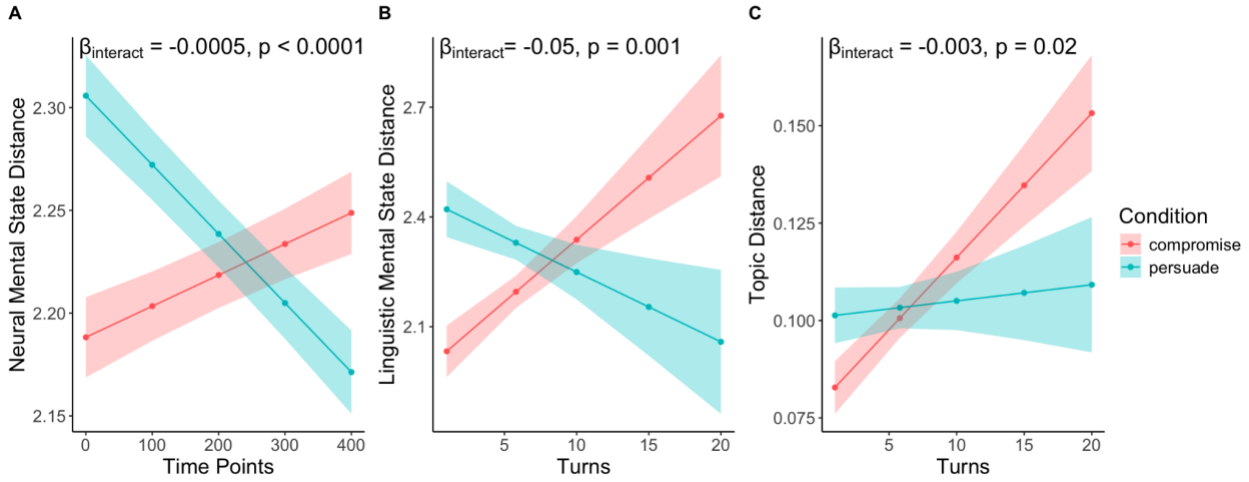


Figure 2. The instruction to compromise increases exploration. Compromisers diverge while persuaders align in neural (A) and linguistic mental state space (B) and in topic space (C).

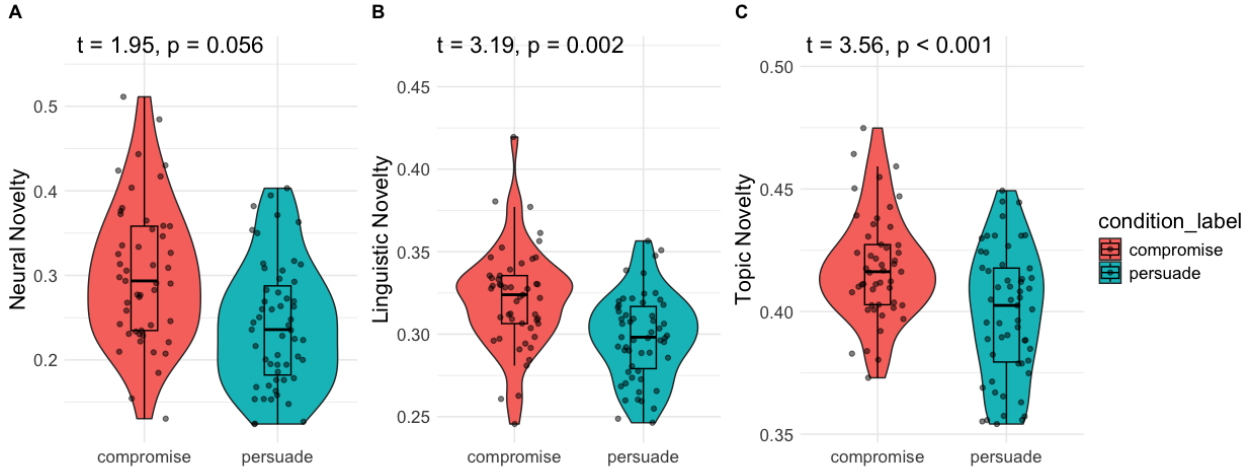


Figure 3. Dyads instructed to compromise exhibit more novelty in their mental state and topic trajectories. (A-B) Dyads in the compromise condition have more diverse mental state trajectories than dyads in the persuade condition and (C) jump longer distances on average between topics.

Exploration is associated with more agreement

After seeing that cooperative motives increased both agreement and exploration, we next tested if exploration enhances agreement. We expected that exploration through both mental state and topic space would result in more agreement.

Mental state exploration. We fit two multilevel regression models (neural and linguistic) to investigate how agreement predicts increases in the distance between partners in mental state space over time. Each model predicted distance from agreement, and time points (within-dyads, for neural) or turns (within-dyads, for linguistic) as predictors, and controlled for initial difference in opinion (distance in opinion space before the conversation). The analyses revealed significant interactions between time points and agreement, such that dyads who agreed more diverged more in neural mental state space ($\beta = 0.0001$, SE = 0.00004, 95% CI [0.00004, 0.0002], $p = 0.003$; Figure 4A). Dyads who agreed more started with higher mental state alignment than dyads who disagreed more and then diverged in neural mental state space until their distance was larger than the dyads who disagreed more. No significant effects were found for the effect of agreement on linguistic mental state exploration.

We also found that more neural mental state novelty was associated with significantly more agreement ($\beta = 4.08$, 95% CI [0.62, 7.55], SE = 0.1.72, $p = 0.02$; Figure 5A). Similarly, linguistic novelty was linked to more agreement ($\beta = 10.86$, 95% CI [4.39, 17.33], SE = 3.21, $p = 0.002$; Figure 5B). Together, this suggests that more mental state exploration leads to more agreement.

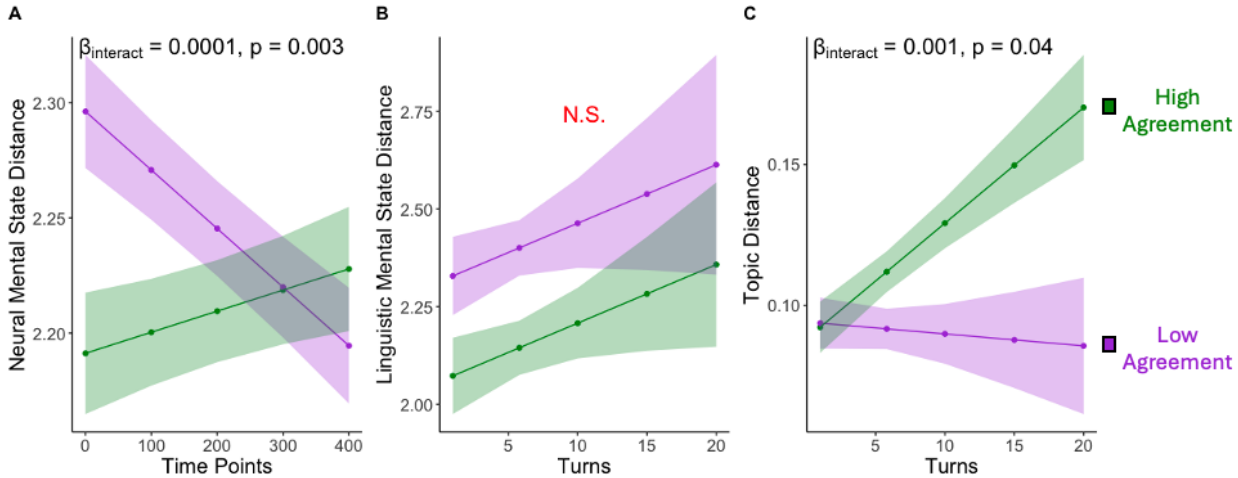


Figure 4. Dyads who explored more mental state and topic space found more agreement. These plots show the fitted values of the two-way interaction effects for agreement (at +1.5 SD and -1.5 SD for each outcome) and time points predicting divergence in neural and linguistic mental state space and topic space derived from the multilevel models described above.

Topic exploration. Paralleling the mental state analyses, we fit a multilevel regression model to investigate how agreement predicts changes in distance in topic space over time. We found that dyads who diverged more in topic space over time found more agreement, whereas dyads who

did not diverge ended up disagreeing more (β = 0.001,95% CI [0.00005, 0.003], SE = 0.0006, p = 0.042; Figure 4C).

Similar to the mental state novelty findings, we found a trend towards greater topic novelty predicting more agreement (β = 5.96, 95% CI [-0.74, 12.66], SE = 3.33, p = 0.08; Figure 5C). In addition, dyads who generated more topics (β = 0.03, 95% CI [0.002, 0.059], SE = 0.01, p = 0.039) found more agreement. No significant effects were found for topic switches. In sum, these findings imply that more topic exploration leads to more agreement.

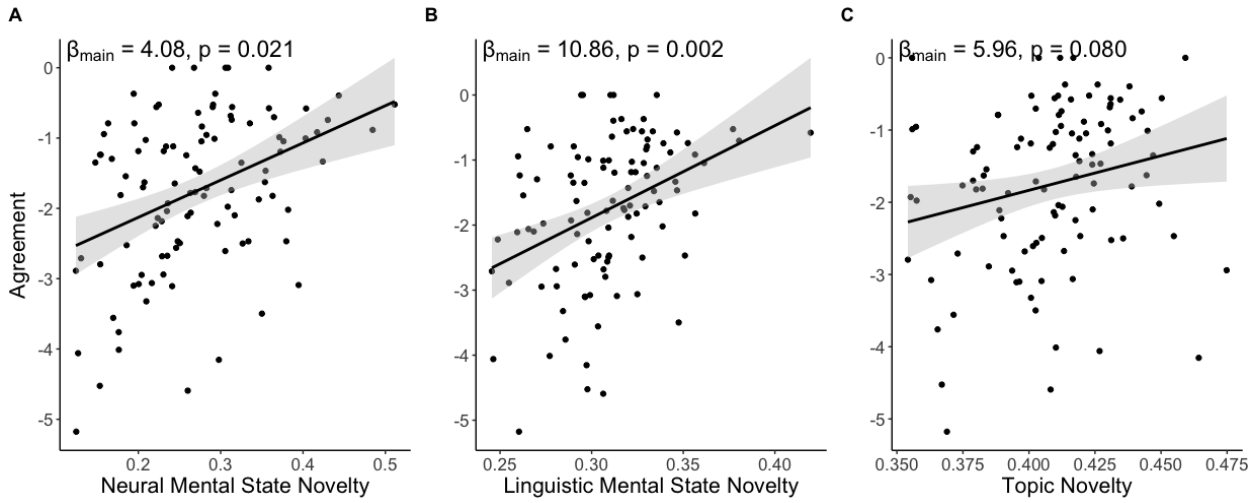


Figure 5. Dyads who exhibit more novelty in their topic and mental state trajectories agree more. The grey ribbons represent the 95% confidence intervals around the fitted values.

Exploration mediates the effect of social motive on agreement

So far we have shown that being motivated to compromise increased both agreement and exploration. Exploration, in turn, increased agreement. Here, we tested whether the effect of social motives (compromise vs.persuade) on agreement was mediated by exploration. To this end, we created a composite measure of exploration by computing the mean over the neural mental state, linguistic mental state, and topic novelty. We used the novelty measure instead of the divergence measure because it captures the absolute rather than the relative amount of exploration and thus renders this measure more interpretable. We used a composite across the three measures because analyses revealed that all three modalities (neural, linguistic and topic) tap into the same underlying construct: exploration (see Supplementary Material 3). We performed a mediation analysis that revealed that social motive was a significant predictor of exploration (β = 0.025, 95% CI [0.009, 0.04], *SE* = 0.008, *p* = 0.003; Figure 6), and exploration significantly predicted agreement (β = 7.77, 95% CI [1.88, 13.68], *SE* = 2.92, *p* = 0.011; Figure 6). After controlling for exploration, the direct effect of social motive on agreement was no longer significant (β = 0.32, 95% CI [-0.08, 0.73], *p* = 0.12; Figure 6) indicating full mediation. The indirect effect of social motive through exploration was significant (β = 0.19, 95% CI [0.04, 0.41], *p* = 0.008; Figure 6) based on 5000 bootstrap resamples. The total effect of social motive on agreement was significant (β = 0.51, 95% CI [0.11, 0.92], *p* = 0.012; Figure 6). These findings show that the effect of motivation to compromise on agreement is fully mediated by exploration. Though the composite measure best predicts agreement (see Supplementary Material 4), mediation analyses with each individual measure of exploration replicated this same pattern of results (Supplementary Material 5).

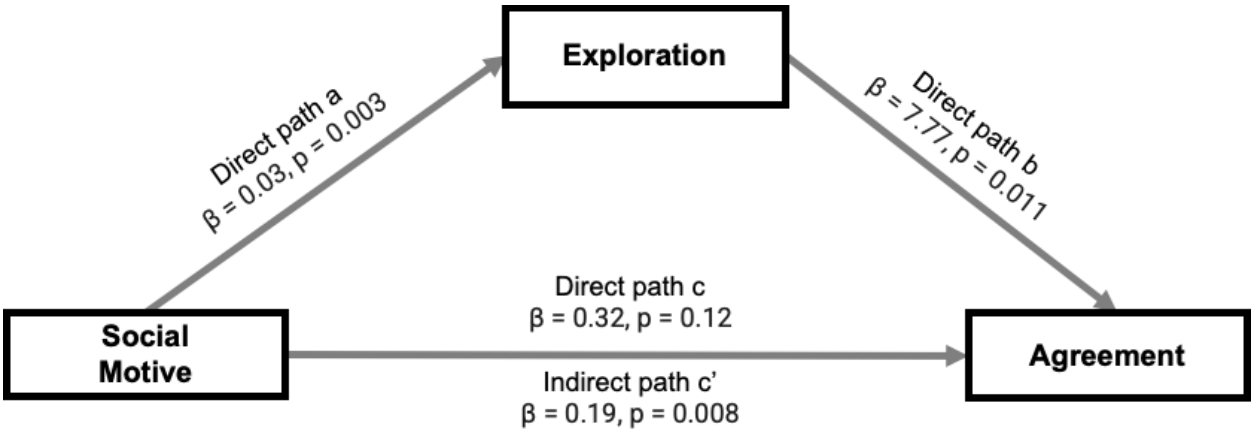


Figure 6. Composite Exploration mediates the positive effect of the instruction to compromise on agreement.

Discussion

Reaching agreement can facilitate a wide range of human goals and achievements, serving as a cornerstone for cooperation, trust-building, and successful conflict resolution. When and how do people navigate conversation to find agreement with others? In this study, we used fMRI-hyperscanning to investigate how people align or explore as they discuss pressing societal problems. Our analyses yielded three key insights that collectively guide us toward an answer. First, people instructed to compromise as opposed to persuade ultimately agreed more at the end of the conversation. Second, compromise increased exploration. Third, dyads who engaged in more exploratory conversations when discussing contentious topics were more likely to eventually find agreement. Finally, the effect of compromise on agreement was fully mediated by the extent of this exploration. Thus, compromising may facilitate agreement by promoting greater exploration during conversations.

This study highlights the key role of exploration in conversation, using advanced tools to identify exploratory dynamics. We measured exploration using several convergent measures. First, we tracked exploration through two key grounds through which dyads need to interact in conversation: mental stats and topics. The mind's capacity to represent others' mental states enables individuals to understand and predict others' thoughts and feelings, which is crucial in negotiation. Unlike traditional neuroimaging metrics that fail to capture mental states encoded in patterns across the whole brain, our mental state decoding approach provides a tractable neurobiological signature of this process and assesses the dynamic interactions between two minds during negotiation. Combining this approach with parallel linguistic mental state tracking helps identify linguistic markers of negotiation success that can then be applied to much larger datasets due to NLP's scalability. Topic modeling provides insights into how many diverse aspects of a discussion problem dyads consider, revealing the creativity and flexibility of their problem-solving approach. It also helps us understand how the dynamics of the topics dyads discuss shape and interact with their mental state dynamics. Second, we measured exploration using two types of data: neural and linguistic. Finally, we operationalized exploration in two ways: divergence and novelty. Both measures revealed that dyads in a cooperative mindset explored more neural and linguistic mental state space. We reproduced these findings in topic space: Dyads who agreed more generated more topics and jumped between more distant topics than dyads who failed to find consensus. Thus, across analyses, we show that a cooperative mindset is linked to more exploration. We also directly tested convergent validity across all of our metrics and modalities and found significant positive correlations (see Supplementary Material 3). Convergent evidence across these vastly different modalities (neural and linguistic) and spaces (mental state and topic) enhances the validity of the conclusion that exploration facilitates agreement. This multi-modal corroboration also strengthens the robustness of the findings by demonstrating the reliability and accuracy of our measures. Our findings, therefore, show promise for generalizability.

We observed that dyads in the compromise condition are more aligned at the beginning of the conversation, whereas dyads in the persuade condition start off more distant in mental state space (see Figure 2). This may reflect the differing motivational orientations, the cooperative vs.competitive mindset, with which the dyads enter the conversation due to their different instructions. Because individuals with a cooperative mindset focus on mutual interests and shared goals—fostering open communication and a willingness to consider different perspectives— their minds may be more aligned at the negotiation's outset. In contrast, those with a competitive mindset prioritize their own positions and aim to convince the other person, leading to less alignment in their initial thinking. Initial alignment may be a necessary condition for later exploration that we found to be essential for agreement. In fact, in related work it has

been found that strangers in affiliative conversation who are most successful in forging social bonds initially align their mental states, and once they reach a critical threshold, start diverging (Sened et al., 2024). It is important to note that this initial gap in mental state alignment cannot be due to differences in participants' starting opinions, which were similar across conditions (see Figure 1).

Our results highlighted that dyadic exploration in particular, as opposed to exploring individually, was associated with increased agreement. This finding is compatible with previous studies (for meta analysis see: Kong et al., 2014). Our findings also extend past work by emphasizing the role of emergent properties of dyadic interaction rather than being driven by only one exploratory conversant who pulls their partner along. We demonstrate across modalities that dyadic exploration consistently outperformed individual exploration in predicting agreement (see supplementary material 3). This suggests that it is not enough for one person to just rove individually through topic space, but that finding consensus requires mutual coordination in exploration. Reciprocity may be a key mechanism contributing to the synergistic effects of dyadic interaction (Gouldner, 1960). It has been found that the amount of reciprocal information sharing predicted individual gains (Liu & Wilson, 2011). So, agreement may emerge from mutually sharing information and building on each other's ideas. This collaborative exploration may render dyadic interaction greater and more complex than simply adding individual actions. Another benefit of dyadic exploration could be that one conversation partner can assist the other to overcome a conversational impasse and break free from stagnation (Druckman, 1986; Putnam & Jones, 1982; Weingart et al., 1990). Previous negotiation research suggests that impasses can be disrupted by: (1) one partner making a comment about the negotiation process itself (Brett et al., 1998); or (2) recognizing that progress has stalled and that there is nothing to lose by sharing information more openly (Druckman, 1986). Together, this work highlights that to understand the determinants of agreement, it is essential to study how partners dynamically shape each other's opinions in naturalistic conversation rather than focusing on isolated individuals.

Lastly, we found that the effect of compromise on agreement is fully mediated by exploration. This suggests that the strong association of a cooperative mindset and agreement can be fully attributed to more exploratory interaction dynamics. Therefore, while social motives set the stage, it is the nature of the interaction dynamics - represented by neural, linguistic and topic trajectories - that ultimately determines the outcome. This highlights the pivotal role of interaction dynamics as the mediating mechanism between motives and outcomes, emphasizing that fostering exploratory communication processes can lead to more successful negotiations. Thus, even when parties share compatible motives, ineffective negotiation dynamics can provoke misunderstandings and obstruct agreements. Conversely, effective dynamics can bridge divides even when initial motives are at odds.Therefore, the focus in negotiation research and practice may shift toward enhancing interaction dynamics to improve outcomes.

Our findings align well with a foundational model of negotiation outcomes (J. M. Brett, 2000; J. Brett & Thompson, 2016). The model states that the negotiators' social motives directly influence the outcome potential and indirectly influence the actual outcome by affecting the interaction dynamics used. We find that a cooperative mindset, induced by instructions to compromise, created more agreement than a competitive mindset induced by instructions to persuade. This is consistent with previous research repeatedly showing that cooperative motives are associated with more agreement (De Dreu et al., 2000; Olekalns & Smith, 1999, 2003b, 2003a; Weingart et al., 2007)**.** According to the model, a cooperative mindset should map onto the use of specific behaviors and dynamics. The cooperative motive is manifested in

explicitly sharing information where the conversation partners exchange questions and answers to learn about each others' perspectives and generate shared insight (L. Thompson & Hastie, 1990; L. L. Thompson, 1991)**.** This allows the two conversation partners to learn about their negotiation interests, motives, and priorities (R. Fisher et al., 2011; Walton & McKersie, 1965). However, this work left open the question of *how* social motives (cooperative vs.competitive) map onto the interaction dynamics that eventually determine the outcome of a negotiation. We contribute to the negotiation literature by showing that cooperative motives map onto exploratory interaction dynamics: dyads who were in a cooperative mindset explored more topics and mental state space, which was rewarded with more agreement. In contrast, participants in a competitive mindset slowly converged in topic and mental state space, resulting in less consensus.

In today's increasingly polarized world, where people often find themselves trapped in echo chambers that intensify disagreement, developing interventions for healthier communication is essential. As workplaces and global organizations become more diverse, bringing together individuals from various cultures, the need for inclusive dialogue becomes even more pressing. At the same time, addressing global challenges like climate change, economic inequality, and public health crises requires cooperative efforts across divides. Our findings open promising avenues for developing interventions targeted at reducing polarization, miscommunication and conflict. Simply instructing participants to find compromise can lead to significantly higher levels of agreement than instructing participants to persuade. Thus, encouraging compromise in conflictual situations may be a promising first step. Beyond that, training employees, politicians and teachers in integrative negotiation and communication skills can have powerful effects on bridging divides (Lewicki et al., 2024; Thompson & DeHarpport, 1994).

While this study focused on what neural and conversational dynamics lead to agreement, a promising avenue for future research would be to explore what differentiates successful persuaders from unsuccessful ones. Are there reliable strategies that grant the persuader the power to pull the other towards themselves in opinion space? In addition, while this study focused on investigating conversations in which both partners had the same social motives (both competitive or both cooperative), the real world is full of situations where these motives diverge. Future studies may benefit from exploring how the interaction dynamics change when the motives do not align. Lastly, it may be fruitful to explore what happens when conflict between conversation partners increases. For instance, future studies may draw from a more diverse sample of participants with more polarized opinions. Further, future studies may design experiments in which participants' financial incentives are pitted against each other so that there are direct financial consequences for the decisions the conversation partners make. Higher tension in conflictual conversations can shift the dynamics, which can lead to more rigid positions and reactive strategies, and, in turn, may hinder collaboration and resolution.

Exploration also helps forge social bonds in the contexts of affiliative conversations (Sened et al., 2024; Speer et al., 2024). Dyads who explored more topics and more mental state space ended up enjoying their conversation more and felt closer and more similar to their conversation partners. This prompts the question of when exploration proves advantageous and in which situations it may be less so. In educational environments, such as classroom discussions and tutoring, encouraging students to explore different topics may enhance critical thinking and deeper understanding. In addition, exploratory conversation may assist in identifying gaps in knowledge and tailor the learning experience (De Felice et al., 2023). Similarly, in counseling and therapy, empathic conversations that encourage clients to explore their thoughts and emotions may lead to more personal insights and growth (Kussman et al., 2024). In addition, an exploratory approach, characterized by open questions and active listening, builds trust and

creates a safe space (Brooks et al., 2015; Jones, 2011; Yeomans et al., 2019). In contrast, in emergency and crisis situations, when time is of the essence and immediate action is required, concise and direct communication is crucial. Exploration may cause delays and increase ambiguity, uncertainty and confusion which may exacerbate the situation. Lastly, in legal contexts, exact language and high precision are required to avoid misunderstandings. Exploring topics outside the agreed scope can lead to unintended obligations, which may reveal strategic information or create vulnerabilities. By thoughtfully applying exploration where it is most beneficial and recognizing when it's not appropriate, people can enhance the efficacy of their communication across various settings.

This study addresses a key question in conflict resolution: What leads to mutual agreement in conversations? The findings show that, overall, conversations increase agreement. However, not all conversations are equally conducive to the alignment of opinions: Particularly conversations that are aimed at finding compromise as opposed to persuading lead to more agreement. Notably, this effect can fully be explained by the more exploratory nature of conversation with the motivation to compromise. These compromise-oriented conversations were characterized by more novel mental state trajectories and more diverse and distant topics. These insights inform broader debates on how to reach agreement and provide useful strategies for navigating disagreement constructively.

Methods

The primary aim of the study was to investigate how people navigate conversations to achieve mutual agreement.To this end, we used fMRI hyper-scanning: 60 dyads engaged in real-time conversations on discussion topics aimed at stimulating diverging opinions. Half of the recruited dyads were instructed to compromise, whereas the other half were instructed to persuade. We first tested how social motives (compromise vs.persuade) relate to agreement. Secondly, we examined how the social motives influence the interaction dynamics (*alignment* vs. *exploration*). Third we assessed how the interaction dynamics influence agreement. Lastly, we assessed whether the interaction dynamics mediate the effect of motives on agreement.

Participants

A total of 60 dyads (120 participants) engaged in a real-time conversation while the two dyad members were simultaneously scanned using fMRI hyperscanning. Due to technical issues, the data for 4 dyads remained incomplete. So the final dataset consisted of 56 dyads, which were randomly assigned to compromise (29 dyads; n = 58 participants; age 17 - 36, Mage= 19.94, SDage = 3.05; 31 women, 27 men, 0 non-binary, African American = 1, Asian = 12, Caucasian = 28, Other = 17), or to persuade (27 dyads; n = 54 participants; age 18-23, Mage = 19.66, SD= 1.27; 38 women, 15 men, 1 non-binary, African American = 6, Asian = 17, Caucasian = 22, Other = 9). All dyads were randomly paired and were unacquainted before the study. All participants had to be at least 18 years old to be eligible for the study and had to be cleared for undergoing an MRI scan (e.g. no metal implants, heart valves etc.).

Protocol

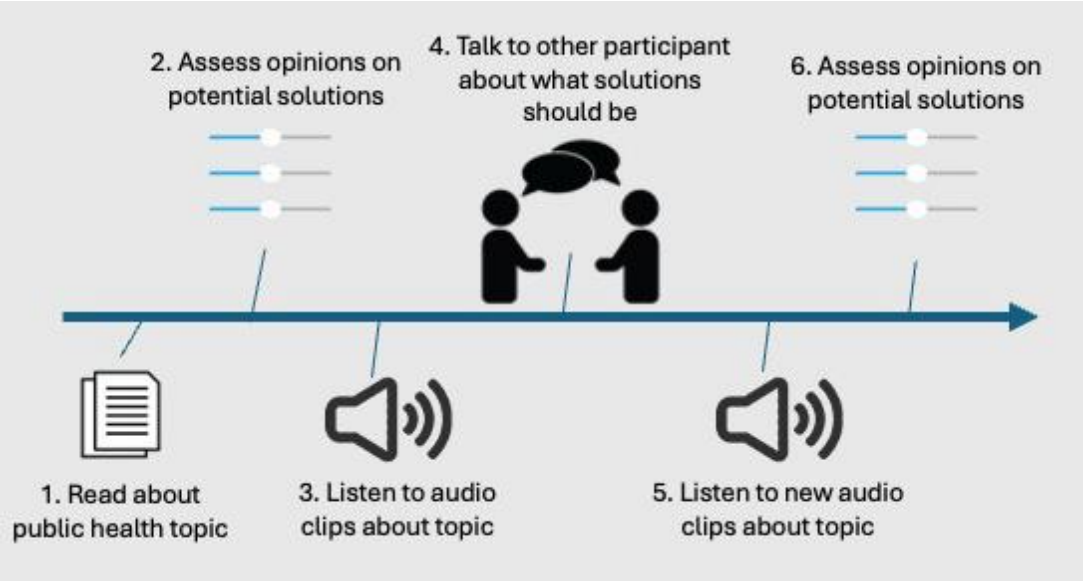


Figure 7. The order of the tasks. Participants first read about the public health topic then provide their opinions about potential solutions. Afterwards, they listen to an audio clip about the topic. Next, they discuss the topic with their conversation partner, before providing their opinion about the solutions a second time. Lastly, they evaluate the conversations.

Participants were simultaneously scanned in two MRI scanners while they learned about and discussed two topics (air pollution, cost of tuition). These topics were chosen based on behavioral pretesting on what current event issues inspire the most diverse opinions regarding

solutions. To introduce them to the discussion topics they were asked to read about the topics (see Figure 7). Next, participants were presented with five different potential solutions to address the problem they just learned about (the order of the topics, air pollution, and cost of tuition, was counterbalanced). Participants were then asked how $100 million should be allocated across the five options to address the discussion problem. They could take as much time as they wanted to answer the questions. Afterwards they would listen to 3 news clips about the discussion topics for ~7 minutes.The news articles were chosen for their descriptive reporting of people experiencing consequences of the the public health problem, rather than promoting any particular solution to the problem, in order to minimize influence on participant opinions. Next, just before starting the conversation dyads were assigned to one of two between-dyad conditions: (i) Compromise: In this condition, dyads were asked to find a compromise on their opinions on the discussion topics presented. (ii) Persuade: In this condition, dyads were instructed to persuade their conversation partner of their opinion. Both participants received the same instruction in each condition (i.e. both were instructed to compromise or both were instructed to persuade). We chose a between-subject design because it would be hard for dyads to switch goals (persuasion vs.compromise) from one trial to the next. Participants then conversed with the other member of their dyad for 10 minutes. Trials began with the conversation prompt (3 sec), followed by 10 min of turn-taking (see Figure 8). Each discussion topic was collected in a separate run. The prompt for the conversation was about deciding how to allocate financial support to each of the five potential solution options. The topic they discussed first was counterbalanced across dyads. The conversations were freeform, which meant that participants could say whatever they wanted.Participants indicated the end of a turn by pressing a button to transfer the speaker role to the conversation partner. After participants finished the conversation they again watched ~ 7 minutes of 3 news clips (different from the first viewing) and again answered how they would allocate $ 100 million to five different solutions addressing the problem, before moving on to the next discussion problem.

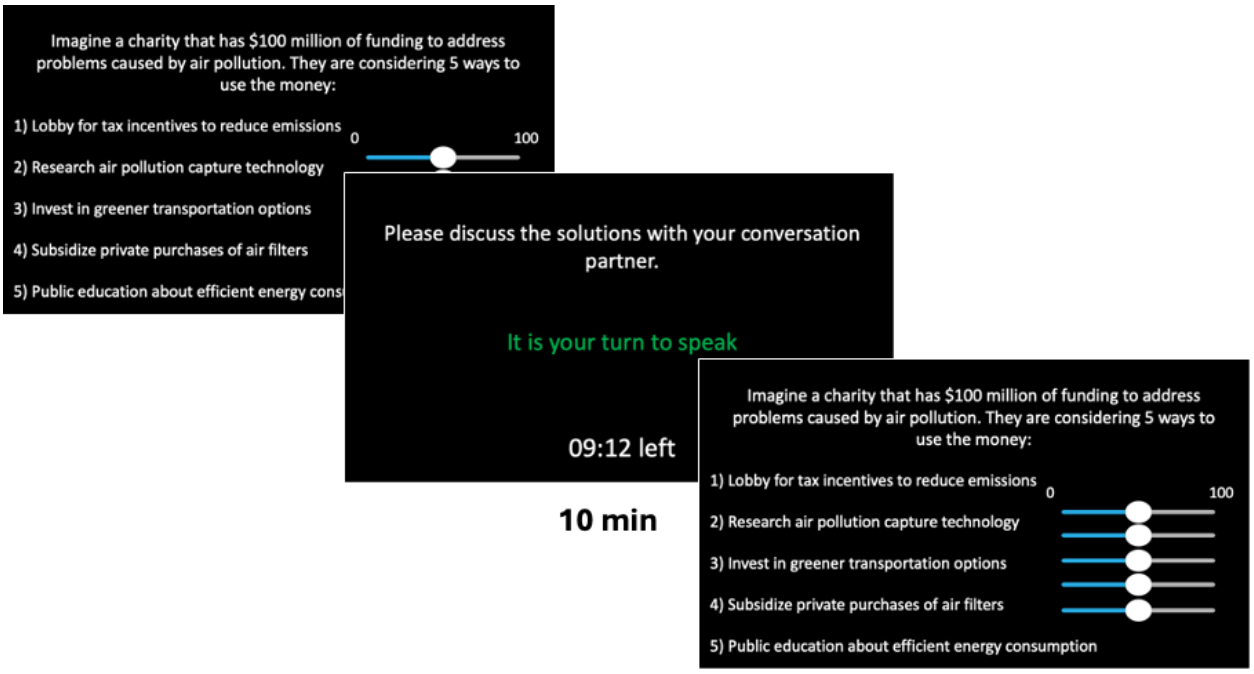


Figure 8. An example trial as seen from a participant's perspective. Dyads viewed the discussion prompt and allocated $100,000,000 across the five proposed solutions for

addressing the problem. Subsequently, they were assigned speaker and listener roles and the speaker started speaking. The speaker could then press a button to ‘pass the mic’ to the listener. In addition, participants saw how much time there was left in the current trial at the bottom of the screen. After they finished the conversation they were again asked to allocate money across the 5 different solutions.

Speech recording

We acquired recordings of the conversations’ content during the fMRI scan using a specialized MR-compatible recording system (FOMRI II; Optoacoustics Ltd). This system employs two optical microphones positioned orthogonally. One microphone captures background noise (reference), while the other captures both background noise and the speaker's speech (signal). A dual-adaptive filter subtracts the reference input from the signal channel using a least mean square method. To ensure effective subtraction, the reference signal undergoes adaptive filtering, with the filter gains continuously adjusted based on the residual signal and reference input. To prevent filter divergence during speech, a voice activity detector is incorporated into the algorithm. Additionally, a speech enhancement spectral filtering algorithm further preprocesses the speech output for real-time enhancement.

fMRI acquisition

The fMRI scans for the pairs were conducted using both a 3T Siemens Skyra MRI system and a 3T Siemens Prisma MRI system, with identical scanning parameters employed for both systems. Functional scans were obtained covering the entire brain in an interleaved manner, with a slice thickness of 3.0 mm and an in-plane resolution of 3.0 × 3.0 mm, and a flip angle of 80°. The echo time (TE) was set at 28 ms, and the repetition time (TR) was 1500 ms. Additionally, a T1-weighted image was obtained to serve as an anatomical reference, with a resolution of 1.0 × 1.0 × 1.0 mm, 176 sagittal slices, a flip angle of 9°, TE of 2.98 ms, and TR of 2300 ms. To minimize head movement during scanning, subjects' heads were secured using foam padding.

Preprocessing

The fMRI data whereas preprocessed using fMRIPrep version 20.2.0, a Nipype-based tool chosen for its robustness and reproducibility in preprocessing diverse datasets. fMRIPrep automatically adapts workflows based on best-in-class algorithms, enabling high-quality preprocessing with minimal manual intervention. Detailed methods for preprocessing can be found in (Speer et al., 2024).

In brief, each T1-weighted (T1w) volume was corrected for intensity nonuniformity and skull stripped. Spatial normalization was performed using the International Consortium for Brain Mapping (ICBM) 152 Nonlinear Asymmetrical template version 2009c through nonlinear registration. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter was carried out on the brain-extracted T1w. Field map distortion correction involved coregistering the functional image to the same-subject T1w image, followed by boundary-based registration to the T1w with 9 degrees of freedom. Motion correction, field distortion correction, and spatial normalization were applied in a single step using Lanczos interpolation. Physiological noise regressors were extracted using CompCor, with six components each for temporal (tCompCor) and anatomical (aCompCor) variants.

Because speaking induces motion artifacts and potential distortions in the magnetic field, an additional denoising step was necessary to clean the data after preprocessing. Consequently,

we conducted further confound regression on the preprocessed data. This regression involved incorporating various regressors, including six head motion parameters, their cosine, first derivatives, squares, signals from white matter and cerebrospinal fluid, along with their derivatives and squared derivatives, as regressors of no interest. The residuals obtained from this regression, which represent the cleaned data, were then utilized for subsequent analysis. These regressors were selected based on a systematic assessment comparing different confound models (Speer et al., 2024). Since we were interested in changes in mental state that do not operate on a second to second scale, we applied temporal gaussian smoothing of signal to increase the signal to noise ratio. Specifically, we applied a gaussian smoothing kernel with a width of 6 TRs (9s) to the data. The cleaned data were segmented into individual prompts to examine the trial effects over time. To prevent confounding our analysis with the BOLD response to the onset and offset of the discussion prompt and task switching, we truncated each trial by removing the first and last 7 TRs (10.5 s) as recommended by previous research(Nastase et al., 2019).

Measures

Measuring agreement

We measured participants' quantitative allocation of money across five solutions to the discussion problems as our primary behavioral measure, reflecting how much they valued each option. These allocations formed opinion profiles for each participant. By calculating the Mahalanobis distance between the profiles of dyad members before and after their conversations, we assessed their levels of agreement to determine if discussions increased or decreased consensus.

EMeasuring mental state locations

Individual locations. For our main analyses, we wanted to test how the instruction to compromise vs.persuade shape alignment (convergence vs. divergence) in mental state space and how this relates to agreement. To do this, we needed to locate and track each participant's mental state at each moment in the conversation. To accomplish this in the neural data, we used decoding models trained on independent data to extract mental state locations from whole-brain activity patterns (Speer et al., 2024). These decoding models were applied to each volume of the preprocessed and denoised conversation data for each participant to decode to what extent these three dimensions are expressed, and thus which mental states are represented in each participant's mind at each moment in the conversation. All the fMRI analyses reported to this point were conducted using custom Python scripts (https://osf.io/qsnyj/). We also measured dyads' mental state trajectories by analyzing their language. To this end, we employed a natural language processing (NLP) algorithm, affectr (https://github.com/markallenthornton/affectr), to extract mental state location in 3D space based on the words participants used (Supplementary Material 1).

Two types of exploration. In this study we use two ways of operationalizing conversation strategies (*finding common ground* vs. *exploring new ground*). First, we look at the distance between conversation partners over time, which allows us to examine to what extent they *converge* or *diverge.* Second, we investigate how much *novelty* the conversations contain. This provides us with a more holistic perspective on the two conversation strategies.

Dyadic divergence. For the neural data, we used the values for the three dimensions for each subject and time point to compute the Mahalanobis distance between the two conversation partners at each time point (TR). This gave us a measure of how distant the two partners are to

each other in mental state space, where a smaller distance represents more similar states. The Mahalanobis distance was chosen due to its insensitivity to the scale of the variables entered and because it removes redundant information from correlated variables. As a consequence, it has become the standard distance measure employed for multivariate fMRI analyses (Allefeld & Haynes, 2014; Bobadilla-Suarez et al., 2020; Nili et al., 2014). For the linguistic data, we used the same approach with the only difference that we computed the distance in mental state space between each turn for each conversation. We exclusively used trials with more than three turns (2 trials were removed), because the number of samples must exceed the number of dimensions (N = 3: rationality, social impact & valence) to calculate the covariance matrix needed to estimate the Mahalanobis distance. Participants signaled the end of their turn by pressing a button, facilitating the demarcation of turns. Trials with an excessive number of turns were excluded to ensure meaningful analysis with NLP tools. The 1.5 * interquartile range rule identified trials with over 21 turns as outliers, resulting in the removal of these trials (4 trials, < 4%) from the NLP analysis. This resulted in the exclusion of one participant. Additionally, the number of words per turn was included as an additional predictor in regression models (detailed below) to account for its effects.

Dyadic novelty. To test how much dyads explore mental state space as they are engaged in a conversation we adapted an additional approach from a previous study that investigated movement synchrony (Ravreby et al., 2022). This method tests how different (not-repetitive) each conversation turn is to the preceding turns, and accordingly, assess a dyads mental state *novelty*. For the neural data, for each dyad, we used the Kolmogoriv-Smirnov distance (K-S distance; Massey, 1951) to compute the distance between the empirical distribution of the Z-scored mental state locations of both partners of a given conversation turn to each of the previous conversation turns. So, for each conversation turn the distribution of mental state locations of both conversation partners was compared to all the distributions of previous turns. All turns with less than 3 timepoints were removed from analysis (6% of all turns). Each test involves comparing two turns at a time. The two-sample K-S test, which checks if two samples originate from the same distribution, is particularly appropriate for our situation, because it is sensitive to differences in both the location and shape of the empirical cumulative distribution functions (CDFs) of each pair of turns. The distance between the empirical CDFs is small when the distributions are similar (it is zero when they are identical) and approaches one when they are very different. Dor each dyad, we first compared the first and the second turn using the K-S distance test. Since there are no turns before the first one, we also assigned it as the novelty score of the first segment. Subsequently, we compared the third turn to each of the previous turns: we compared the third turn and the second turn, and the third turn and the first turn. The average statistic of the K-S distance tests of the two comparisons served as the dyadic novelty score of the third segment. This procedure was repeated for each dyad up to the last turn N. Thus, each turn was assigned a dyadic novelty score, which allowed us to track the novelty changes along the time domain. In order to have a single dyadic novelty score for each participant, we averaged the exploration scores of all the turns. To measure novelty in linguistic mental state space, the same approach as above was used with the difference that we were only able to extract mental state locations of the person speaking. So, here we compared distributions of words between sentences instead of turns. Each sentence had an average of 21 words (SD = 14) with a minimum of 3 words, which made it suitable for our analysis. So, for each trial we extracted the mental state location of each word the speaker said. We then compared the distribution of mental state location of each new sentence to all the previous sentences as outlined above.

Measuring topic exploration

Individual locations. The previous analyses extracted the conversation partners' locations in mental state space but did not delve into the topics discussed. To tackle this question, we employed topic modeling to investigate the relationship between neural and linguistic mental dynamics and conversation content. Using BERTopic, we extracted topic representations for each sentence in every conversation across all dyads (Supplementary Material 2).

Dyadic divergence. Here, we calculated the distance in topic space from one turn to the next using the cosine distance. We adopted an approach used in a previous hyperscanning study focusing on affiliative conversations (Speer et al., 2024): The choice of cosine distance was motivated by the high-dimensional embedding space (>300 dimensions) in which topics are encoded, surpassing the number of samples (in this case, participants), rendering the computation of Mahalanobis distance unfeasible. Cosine distance was chosen over the Euclidean distance because the Euclidean distance encounters problems in high dimensional space (Aggarwal et al., 2001). In addition, the cosine similarity has been established as the standard distance measure for NLP. It is also the default metric in BERTopic because it does not take into account the magnitude of vectors, which is helpful when working with text data where word counts are influenced by the length of the documents, which often vary considerably in NLP analyses.

Dyadic Novelty. Third, we measured dyadic topic novelty by computing the pairwise cosine distance between each topic within a dyad and then averaged these distances. This allowed us to gauge the extent of content space exploration by each dyad.

Number of Topics. First, we considered the variance in the number of topics generated.

Topic Switches. Second, we examined the frequency of transitions between topics, including shifts back to previously discussed topics within a conversation segment.

Analyses

Investigating how social motives relates to agreement

As a first step, we tested whether conversations in general increase agreement. We entered the dyads distance in opinion before and after the conversation into a multilevel (topics nested in dyads) two-sample t-test to test whether dyads agreed more after the conversation. Next, we added the social motive as a predictor to the model to investigate whether there are difference in how much dyads align their opinions contingent on whether they were instructed to compromise or persuade.

Exploring how the social motives influences interaction dynamics

Dyadic Divergence. To test how the instruction to compromise vs.persuade shaped conversation strategy (convergence vs. divergence) in mental state and topic space we fit regression models to our data. Given the nested structure of our data (time points within trials within participants), we conducted multilevel analyses to test the effect of social motive (instruction to *persuade* vs. *compromise*) and time (time points and trials) on mental state and topic alignment. Since each discussion prompt constituted an independent conversation, we were particularly interested in time points (for neural data) and turns (for linguistic and topic data) within each prompt. Thus, in combination with conversation goal, time points within trials were considered the primary independent variable. The dependent variable for the neural and

the linguistic data was the continuous Mahalanobis distance with a Gaussian link representing the distance in mental state space between the conversation partners. For the topic data the dependent variable was the continuous cosine distance representing the distance in topic space with a Gaussian link. The social motive condition served as a dyad-level predictor, whereas time will serve as a trial-level (discussion topic) and a within-trial-level (time points for neural data and turns for topic and linguistic data) predictor. The models allowed for random intercepts within participants. The multilevel models were implemented in R(R Core Team, 2022) using the *nlme* package (Pinhero & Bates, 2023). This modeling approach allowed us to test how the instruction to compromise vs.persuade shapes exploration of content and mental state space in conversation over time and how the different measures relate to each other.

Dyadic Novelty. We applied a multilevel (topics nested in dyads) two-sample t-test to test whether our measures of individual and dyadic mental state and topic novelty differ between dyads instructed to compromise and persuade.

Investigating how interaction dynamics relate to agreement

Next, we examined how convergence vs. divergence and novelty in mental state space and topic space predicted agreement. To examine how convergence vs. divergence affects agreement, we fit three multilevel regression models to predict each of the three distance metrics, respectively: Mahalanobis distance in neural and linguistic mental state space and cosine distance in topic space. Each model includes agreement, trial number, and time points (within-dyads) as predictors and also controlled for initial difference in opinion (distance in opinion space before the conversation).

Testing whether interaction dynamics mediate the effect of social motives on agreement

In addition, we wanted to see how the effect of social motives (compromise vs.persuade) on agreement is mediated by interaction dynamics. To address this, we created a composite measure of exploration by calculating the mean over the neural and linguistic mental state novelty and the topic exploration measure (mean cosine topic distance). This resulted in a composite exploration score for each trial for each dyad. To examine the potential mediating effect of exploration on the relationship between conversation goal and agreement, we conducted a mediation analysis using the *mediate* function from the *mediation* package in R (Tingley et al., 2014).

We used nonparametric bootstrapping with 5,000 simulations to estimate confidence intervals for the indirect effect. The average causal mediation effect (ACME) and the average direct effect (ADE) were calculated, along with the total effect. Statistical significance of the mediation was assessed using 95% confidence intervals. All analyses were performed in R, version 4.2.1.

Supplementary Material

Supplementary Material 1: AffectR algorithm.

We implemented the affectR package to decode how dyads traverse linguistic mental state space based on the words they chose to say in the conversation. affectR allows us to measure the sentiment of words along the three mental state dimensions (rationality, social impact, and valence). The affectR package was created using ratings of mental states words on 16 psychological dimensions(Tamir et al., 2016). The ratings were collected from 1205 participants on Amazon Mechanical Turk, with aggregate values freely available online (https://osf.io/3qn47/). Principal component analysis (PCA) revealed that these 16 dimensions could be reduced to the three mental state dimensions, which proved able to explain patterns of brain activity associated with thinking about others' mental states (Tamir et al., 2016). To expand the volume of the dictionary from 166 words to two million words, the fastText library (https://fasttext.cc/) was employed. A support vector regression (SVR) model with a radial basis function was used to predict the ratings of 166 mental state words along three principal component dimensions. This approach was feasible because these 166 words are part of the fastText corpus. The SVR model was then utilized to estimate 3D dimension scores for all 2 million words in the fastText corpus, resulting in the creation of a comprehensive weighted dictionary. This approach has been implemented in a previous neuroimaging study (Thornton & Tamir, 2020) to show that the 3d mind model does not only capture how mental states are represented in the brain but also in language. Further, in more recent studies, the same method was applied to study whether the 3d mind model characterizes how people understand mental states across modern and historical cultures (Thornton et al., 2022) and to explore what mental state trajectories lead to stronger social connection in affiliative conversations (Speer et al., 2024).

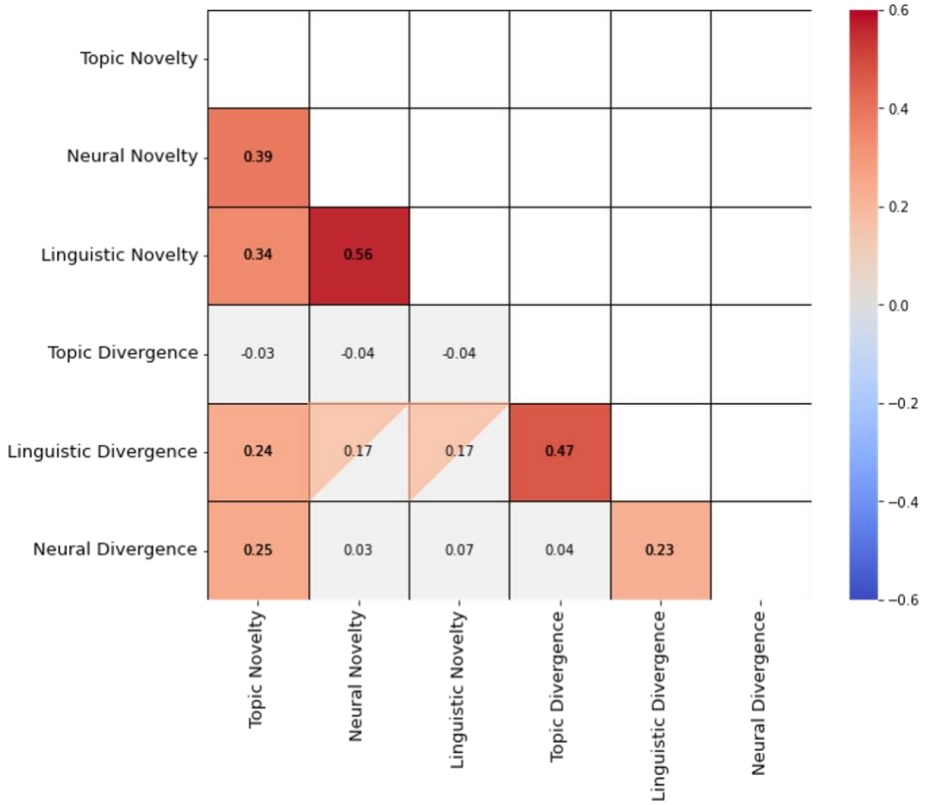
Supplementary Material 2: Topic modeling using BERTopic.

We applied topic modeling to explore how the neural and linguistic dynamics relate to the content of the conversations. Topic modeling is gaining popularity in psychology and social neuroscience and has previously been successfully used in the domain of marketing on social media conversations to drive brand strategy (Resnik et al., 2015) and has been implemented in clinical psychology to differentiate between the language of depressed and non-depressed individuals (Putnick & Bornstein, 2016) and to detect self-injurious thoughts and behavior on online platforms (Franz et al., 2020). We employed BERTopic (https://maartengr.github.io/BERTopic/index.html), which is a topic modeling technique that leverages hugging-face (https://huggingface.co/) transformers and class-based term frequency-inverse document frequency (cTF-DF) to create dense clusters providing easily interpretable topics while keeping important words in the topic descriptions. The benefit of BERTopic, as opposed to more traditional algorithms, is that it selects the number of topics automatically and does not require the specification of number of topics beforehand, which enables a purely data-driven exploration of topics. In addition, BERTopic does not only rely on word-level but also uses sentence-level embeddings, which has the advantage of considering the semantic relationship between words in a corpus, leading to more informative topics. Another advantage

is that due to the nature of the semantic embeddings, textual preprocessing (stemming, lemmization, stopword removal, etc.) is not needed.

Supplementary Material 3: Correlation between exploration across modalities

To test whether there is internal consistency between the neural, linguistic and topic exploration measures (novelty and divergence), we computed the pearson correlation between all six measures. To obtain measures for neural, linguistic and topic divergence we extracted the random slopes for each participant from the three multilevel models depicted in Figure 3. That is, the multilevel model fit on the neural data, the linguistic data and the topic modeling data. The correlation analysis revealed that the six measures of exploration are overall positively correlated which suggests that our measures of exploration are internally consistent (see Supplementary Figure 1). All three measures of novelty are strongly significantly positively correlated. The divergence measures also show positive correlation: The observation that the neural and linguistic divergence are correlated was expected because these two measures exist in the same mental state space. The observation that linguistic and topic divergence are correlated was expected because they are both based on linguistic data. As a result, the fact that the correlation between topic and neural divergence was not significant was expected to reveal the weakest correlation, because the conceptual and methodological distance was largest. We also observed significant correlations between topic novelty and linguistic and neural divergence, which was expected because topic novelty was also based on distances of individual data points whereas neural and linguistic novelty was based on comparisons between distributions. We also observed marginally significant correlation between linguistic divergence and neural and linguistic novelty. Together, these results suggest that our measures of exploration are internally consistent.



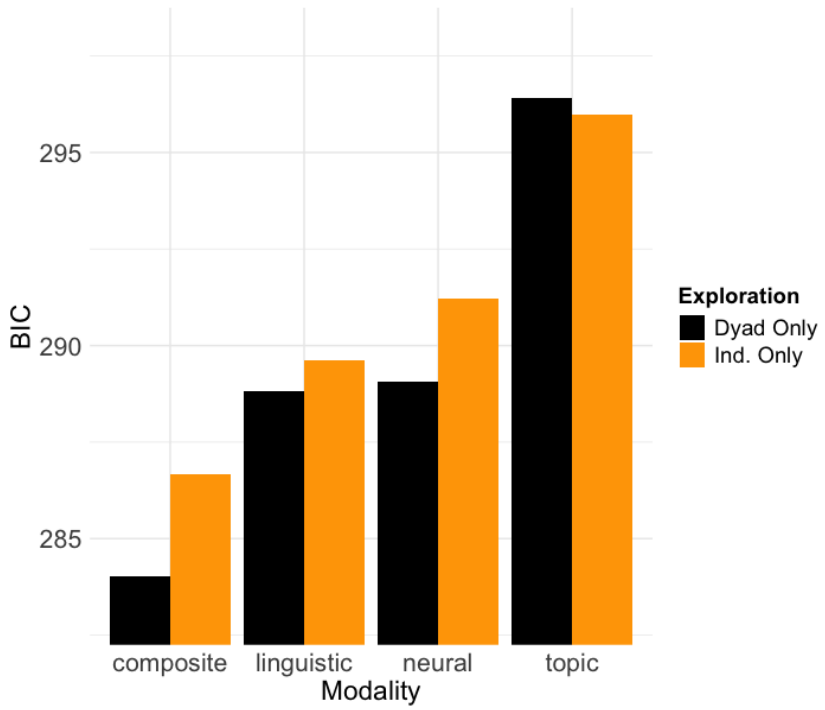
Supplementary Figure 1. Correlation between neural, linguistic and topic novelty and divergence. Colored squared are significant at $p < 0.05$. Squared that are filled with a triangle are marginally significant ($p < 0.10$). Grey squares are not significant.

Supplementary Material 4: Individual vs.Dyadic Exploration across modalities

We examined whether exploration is mostly driven by one of the partners being exploratory or whether there are synergistic effects of both partners exploring. To this end, we performed a model comparison where we fit a model with only individual exploration and a model with only dyadic exploration predicting agreement (while also controlling for the number of turns and the initial distance in opinions). Individual neural and linguistic exploration was computed the same as the dyadic neural exploration with the only difference that here we only focus on the turns of one of the participants at a time. For individual topic exploration, the mean distance between topics introduced by one participant at a time was computed. To compare the models we used the Bayesian Information Criterion (BIC; lower values represent higher fit). The model comparison revealed that for all measures of exploration, except for topic exploration, dyadic exploration was a better predictor than the two individual partners' exploration. This suggests that the combined dyadic exploration is greater than the sum of its parts and there are synergistic effects of dyadic exploration.

This analysis also reveals that the dyadic neural and linguistic exploration models predict agreement equally well, whereas the topic exploration measure does a poorer job at predicting agreement. The analysis also shows that these models explain different portions of the variance

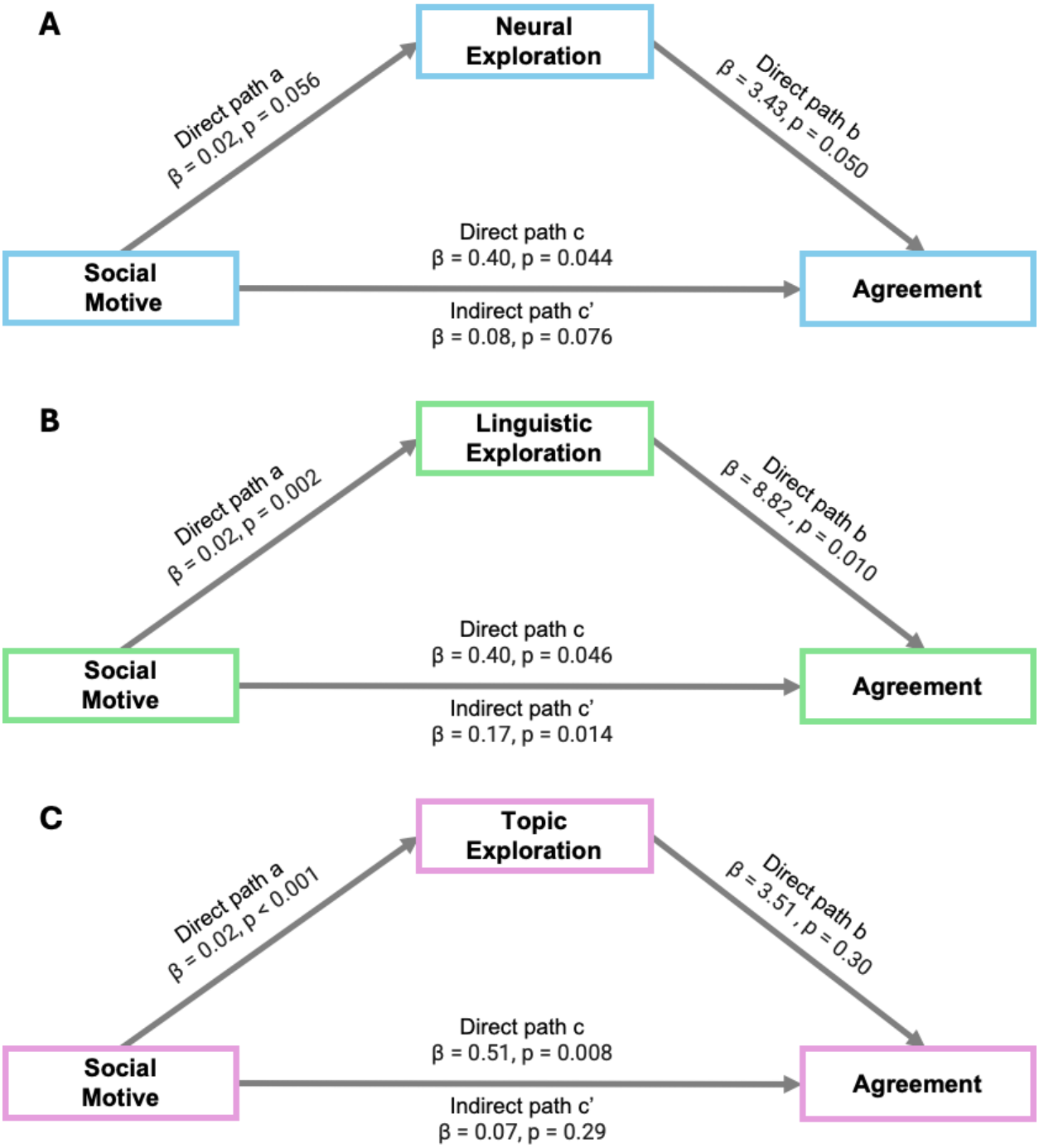
in agreement across dyads as the composite measure of all three of these measures does substantially better in predicting agreement than each individual modality.



Supplementary Figure 2. Dyadic exploration is a better predictor than individual exploration. This graph shows the model fit (BIC; lower is better) for a model with both participants' individual exploration or their dyadic exploration as predictor of agreement across the 3 modalities and their composite.

Supplementary Material 5: Mediation analyses robustness check

To test the robustness of our mediation analysis, we also conducted mediation analysis with each individual measure of exploration as the mediation. The analysis revealed a similar pattern as the findings in Supplementary Material 3: The mediation analysis revealed significant and marginally significant mediation effects for the linguistic and neural mental state exploration measures respectively. No significant mediation was found for the topic exploration measure, where only the effect of social motive on exploration was significant (see Supplementary Figure 2). In addition, these analyses show that these measures appear to capture different portions of variance in dyads exploration behavior as the composite measure performs best in predicting agreement and is the only one that fully mediates the effect of social motive on agreement.



Supplementary Figure 3. Exploration mediates the positive effect of compromise on agreement. Mediation analyses for the neural A), linguistic B) and topic C) exploration measures.

References

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. In J. Van den Bussche & V. Vianu (Eds.), *Database Theory—ICDT 2001* (Vol. 1973, pp. 420–434). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44503-X_27

Allefeld, C., & Haynes, J.-D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *NeuroImage*, *89*, 345–357. https://doi.org/10.1016/j.neuroimage.2013.11.043

Balconi, M. (2010). The Sense of Agency in Psychology and Neuropsychology. In B. Michela (Ed.), *Neuropsychology of the Sense of Agency: From Consciousness to Action* (pp. 3–22). Springer Milan. https://doi.org/10.1007/978-88-470-1587-6_1

Beersma, B., & De Dreu, C. K. W. (2002). Integrative and Distributive Negotiation in Small Groups: Effects of Task Structure, Decision Rule, and Social Motive. *Organizational Behavior and Human Decision Processes*, *87*(2), 227–252. https://doi.org/10.1006/obhd.2001.2964

Behrens, F., Snijdewint, J. A., Moulder, R. G., Prochazkova, E., Sjak-Shie, E. E., Boker, S. M., & Kret, M. E. (2020). Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports*, *10*(1), 19609. https://doi.org/10.1038/s41598-020-76539-8

Bittner, J. V., Bruena, M., & Rietzschel, E. F. (2016). Cooperation goals, regulatory focus, and their combined effects on creativity. *Thinking Skills and Creativity*, *19*, 260–268. https://doi.org/10.1016/j.tsc.2015.12.002

Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., & Love, B. C. (2020). Measures of Neural Similarity. *Computational Brain & Behavior*, *3*(4), 369–383. https://doi.org/10.1007/s42113-019-00068-5

Brett, J. M. (2000). Culture and Negotiation. *International Journal of Psychology*, *35*(2), 97–104. https://doi.org/10.1080/002075900399385

Brett, J. M., Shapiro, D. L., & Lytle, A. L. (1998). BREAKING THE BONDS OF RECIPROCITY IN NEGOTIATIONS. *Academy of Management Journal*, *41*(4), 410–424. https://doi.org/10.2307/257081

Brett, J., & Thompson, L. (2016). Negotiation. *Organizational Behavior and Human Decision Processes*, *136*, 68–79. https://doi.org/10.1016/j.obhdp.2016.06.003

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart People Ask for (My) Advice: Seeking Advice Boosts Perceptions of Competence. *Management Science*, *61*(6), 1421–1435. https://doi.org/10.1287/mnsc.2014.2054

Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C., & Sengers, P. (2012). Sustainably unpersuaded: How persuasion narrows our vision of sustainability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 947–956. https://doi.org/10.1145/2207676.2208539

Burns, S. M., Tsoi, L., Falk, E. B., Speer, S. P. H., & Tamir, D. I. (under review). Interdependent minds: Quantifying the dynamics of successful social interaction. *Current Directions in Psychological Science*.

Carmeli, A., Dutton, J. E., & Hardin, A. E. (2015). Respect as an engine for new ideas: Linking respectful engagement, relational information processing and creativity among employees and teams. *Human Relations*, *68*(6), 1021–1047. https://doi.org/10.1177/0018726714550256

De Dreu, C. K. W., Evers, A., Beersma, B., Kluwer, E. S., & Nauta, A. (2001). A theory-based measure of conflict management strategies in the workplace. *Journal of Organizational Behavior*, *22*(6), 645–668. https://doi.org/10.1002/job.107

De Dreu, C. K. W., Weingart, L. R., & Kwon, S. (2000). Influence of social motives on integrative negotiation: A meta-analytic review and test of two theories. *Journal of Personality and*

Social Psychology, *78*(5), 889–905. https://doi.org/10.1037/0022-3514.78.5.889

De Felice, S., Hamilton, A. F. D. C., Ponari, M., & Vigliocco, G. (2023). Learning *from* others is good, *with* others is better: The role of social interaction in human acquisition of new knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1870), 20210357. https://doi.org/10.1098/rstb.2021.0357

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage*, *23*(2), 744–751. https://doi.org/10.1016/j.neuroimage.2004.05.025

Deutsch, M. (1973). *The resolution of conflict*. Yale University Press.

Doré, B. P., & Morris, R. R. (2018). Linguistic Synchrony Predicts the Immediate and Lasting Impact of Text-Based Emotional Support. *Psychological Science*, *29*(10), 1716–1723. https://doi.org/10.1177/0956797618779971

Druckman, D. (1986). Stages, Turning Points, and Crises: Negotiating Military Base Rights, Spain and the United States. *Journal of Conflict Resolution*, *30*(2), 327–360. https://doi.org/10.1177/0022002786030002006

Fisher, M., Knobe, J., Strickland, B., & Keil, F. C. (2017). The Influence of Social Interaction on Intuitions of Objectivity and Subjectivity. *Cognitive Science*, *41*(4), 1119–1134. https://doi.org/10.1111/cogs.12380

Fisher, R., Ury, W., & Patton, B. (2011). *Getting to yes: Negotiating agreement without giving in.* Penguin.

Franz, P. J., Nook, E. C., Mair, P., & Nock, M. K. (2020). Using Topic Modeling to Detect and Describe Self-Injurious and Related Content on a Large-Scale Digital Platform. *Suicide and Life-Threatening Behavior*, *50*(1), 5–18. https://doi.org/10.1111/sltb.12569

Gelfo, F. (2019). Does Experience Enhance Cognitive Flexibility? An Overview of the Evidence Provided by the Environmental Enrichment Studies. *Frontiers in Behavioral Neuroscience*, *13*, 150. https://doi.org/10.3389/fnbeh.2019.00150

Gilson, L. L., & Shalley, C. E. (2004). A Little Creativity Goes a Long Way: An Examination of Teams' Engagement in Creative Processes. *Journal of Management*, *30*(4), 453–470. https://doi.org/10.1016/j.jm.2003.07.001

Goldstone, R. L., Andrade-Lotero, E. J., Hawkins, R. D., & Roberts, M. E. (2024). The Emergence of Specialized Roles Within Groups. *Topics in Cognitive Science*, *16*(2), 257–281. https://doi.org/10.1111/tops.12644

Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review*, *25*(2), 161. https://doi.org/10.2307/2092623

Hargadon, A. B., & Bechky, B. A. (2006). When Collections of Creatives Become Creative Collectives: A Field Study of Problem Solving at Work. *Organization Science*, *17*(4), 484–500. https://doi.org/10.1287/orsc.1060.0200

Hommel, B. (2015). Between Persistence and Flexibility. In *Advances in Motivation Science* (Vol. 2, pp. 33–67). Elsevier. https://doi.org/10.1016/bs.adms.2015.04.003

Hon, A. H. Y., Bloom, M., & Crant, J. M. (2014). Overcoming Resistance to Change and Enhancing Creative Performance. *Journal of Management*, *40*(3), 919–941. https://doi.org/10.1177/0149206311415418

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, *22*(1), 39–44. https://doi.org/10.1177/0956797610392928

Jacob, C., Guéguen, N., Martin, A., & Boulbry, G. (2011). Retail salespeople's mimicry of customers: Effects on consumer behavior. *Journal of Retailing and Consumer Services*, *18*(5), 381–388. https://doi.org/10.1016/j.jretconser.2010.11.006

Jones, S. M. (2011). Supportive Listening. *International Journal of Listening*, *25*(1–2), 85–103. https://doi.org/10.1080/10904018.2011.536475

Knowles, E. S., & Linn, J. A. (Eds.). (2004). *Resistance and persuasion*. Lawrence Erlbaum Associates.

Kong, D. T., Dirks, K. T., & Ferrin, D. L. (2014). Interpersonal Trust within Negotiations: Meta-Analytic Evidence, Critical Contingencies, and Directions for Future Research. *Academy of Management Journal*, *57*(5), 1235–1255. https://doi.org/10.5465/amj.2012.0461

Kragel, P. A., Koban, L., Barrett, L. F., & Wager, T. D. (2018). Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. *Neuron*, *99*(2), 257–273. https://doi.org/10.1016/j.neuron.2018.06.009

Kussman, M. A., Speer, S. P. H., Huelskamp, L., Gwei, S., Nguyen, T., Xue, S., & Tamir, D. I. (2024). Linguistic Signatures of Empathic Strategies Predict Social Outcomes. *Society for Personality and Social Psychology 2024 Conference, Denver, CO.*

Laureiro-Martínez, D., & Brusoni, S. (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal*, *39*(4), 1031–1058. https://doi.org/10.1002/smj.2774

Lee, D. S., & Fujita, K. (2023). From whom do people seek what type of support? A regulatory scope perspective. *Journal of Personality and Social Psychology*, *124*(4), 796–811. https://doi.org/10.1037/pspi0000405

Lewicki, R., Saunders, D., & Barry, B. (2024). *Negotiation* (9th ed.). McGraw-Hill.

Liu, T., Saito, G., Lin, C., & Saito, H. (2017). Inter-brain network underlying turn-based cooperation and competition: A hyperscanning study using near-infrared spectroscopy. *Scientific Reports*, *7*(1), 8684. https://doi.org/10.1038/s41598-017-09226-w

Lu, K., Xue, H., Nozawa, T., & Hao, N. (2019). Cooperation Makes a Group be More Creative. *Cerebral Cortex*, *29*(8), 3457–3470. https://doi.org/10.1093/cercor/bhy215

Ludwig, S., de Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, *77*(1), 87–103. https://doi.org/10.1509/jm.11.0560

Manson, J. H., Bryant, G. A., Gervais, M. M., & Kline, M. A. (2013). Convergence of speech rate

in conversation predicts cooperation. *Evolution and Human Behavior*, *34*(6), 419–426. https://doi.org/10.1016/j.evolhumbehav.2013.08.001

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78. https://doi.org/10.1080/01621459.1951.10500769

Meina Liu, & Wilson, S. R. (2011). The Effects of Interaction Goals on Negotiation Tactics and Outcomes: A Dyad-Level Analysis Across Two Cultures. *Communication Research*, *38*(2), 248–277. https://doi.org/10.1177/0093650210362680

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1–25. https://doi.org/10.1016/0022-1031(68)90046-2

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, nsz037. https://doi.org/10.1093/scan/nsz037

Nguyen, T., Kungl, M. T., Hoehl, S., White, L. O., & Vrtička, P. (2024). Visualizing the invisible tie: Linking parent–child neural synchrony to parents' and children's attachment representations. *Developmental Science*, e13504. https://doi.org/10.1111/desc.13504

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Olekalns, M., & Smith, P. L. (1999). Social Value Orientations and Strategy Choices in Competitive Negotiations. *Personality and Social Psychology Bulletin*, *25*(6), 657–668. https://doi.org/10.1177/0146167299025006002

Olekalns, M., & Smith, P. L. (2003a). SOCIAL MOTIVES IN NEGOTIATION: THE RELATIONSHIPS BETWEEN DYAD COMPOSITION, NEGOTIATION PROCESSES AND OUTCOMES. *International Journal of Conflict Management*, *14*(3/4), 233–254.

https://doi.org/10.1108/eb022900

Olekalns, M., & Smith, P. L. (2003b). Testing the relationships among negotiators' motivational orientations, strategy choices, and outcomes. *Journal of Experimental Social Psychology*, *39*(2), 101–117. https://doi.org/10.1016/S0022-1031(02)00520-6

Perc, M., & Szolnoki, A. (2008). Social diversity and promotion of cooperation in the spatial prisoner's dilemma game. *Physical Review E*, *77*(1), 011904. https://doi.org/10.1103/PhysRevE.77.011904

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* Springer Science & Business Media.

Pinhero, J., & Bates, D. (2023). *nlme: Linear and Nonlinear Mixed Effects Models.* R Core Team. https://CRAN.R-project.org/package=nlme.

Powers, J. P., & LaBar, K. S. (2019). Regulating emotion through distancing: A taxonomy, neurocognitive model, and supporting meta-analysis. *Neuroscience & Biobehavioral Reviews*, *96*, 155–173. https://doi.org/10.1016/j.neubiorev.2018.04.023

Pruitt, D., & Rubin, J. E. (1986). *Social conflict: Escalation, stalemate and settlement.* New York, NY: Random House.

Putman, W. B., & Street, R. L. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language*, *1984*(46). https://doi.org/10.1515/ijsl.1984.46.97

Putnam, L. L., & Jones, T. S. (1982). Reciprocity in negotiations: An analysis of bargaining interaction. *Communication Monographs*, *49*(3), 171–191. https://doi.org/10.1080/03637758209376080

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation

for Statistical Computing. https://www.R-project.org/.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, *79*(3), 284–295. https://doi.org/10.1037/a0023419

Ravreby, I., Shilat, Y., & Yeshurun, Y. (2022). Liking as a balance between synchronization, complexity and novelty. *Scientific Reports*, *12*(1), 3181. https://doi.org/10.1038/s41598-022-06610-z

Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V.-A., & Boyd-Graber, J. (2015). Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in Twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 99–107. https://doi.org/10.3115/v1/W15-1212

Sassenberg, K., & Winter, K. (2024). Intraindividual Conflicts Reduce the Polarization of Attitudes. *Current Directions in Psychological Science*.

Sened, H., Speer, S. P. H., Cooney, G., Reece, A. G., & Tamir, D. I. (2024). The Temporal Arc of Successful Conversations. *Society for Personality and Social Psychology 2024 Conference, San Diego, CA.*

Speer, S. P. H., Mwilambwe-Tshilobo, L., Tsoi, L., Burns, S. M., Falk, E. B., & Tamir, D. I. (2024). Hyperscanning shows friends explore and strangers converge in conversation. *Nature Communications*, *15*(1), 7781. https://doi.org/10.1038/s41467-024-51990-7

Street, R. L., Brady, R. M., & Putman, W. B. (1983). The Influence of Speech Rate Stereotypes and Rate Similarity or Listeners' Evaluations of Speakers. *Journal of Language and Social Psychology*, *2*(1), 37–56. https://doi.org/10.1177/0261927X8300200103

Su, Q., Li, A., Zhou, L., & Wang, L. (2016). Interactive diversity promotes the evolution of cooperation in structured populations. *New Journal of Physics*, *18*(10), 103007. https://doi.org/10.1088/1367-2630/18/10/103007

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

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*, *113*(1), 194–199. https://doi.org/10.1073/pnas.1511905112

Thompson, L., & DeHarpport, T. (1994). Social Judgment, Feedback, and Interpersonal Learning in Negotiation. *Organizational Behavior and Human Decision Processes*, *58*(3), 327–345. https://doi.org/10.1006/obhd.1994.1040

Thompson, L., & Hastie, R. (1990). Social perception in negotiation. *Organizational Behavior and Human Decision Processes*, *47*(1), 98–123. https://doi.org/10.1016/0749-5978(90)90048-E

Thompson, L. L. (1991). Information exchange in negotiation. *Journal of Experimental Social Psychology*, *27*(2), 161–179. https://doi.org/10.1016/0022-1031(91)90020-7

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*, *125*, 44–59. https://doi.org/10.1016/j.cortex.2019.12.012

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. https://doi.org/10.1007/s42761-021-00089-z

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). **mediation**: *R* Package for Causal Mediation Analysis. *Journal of Statistical Software*, *59*(5). https://doi.org/10.18637/jss.v059.i05

Tunçgenç, B., & Cohen, E. (2018). Interpersonal movement synchrony facilitates pro-social behavior in children's peer-play. *Developmental Science*, *21*(1), e12505.

https://doi.org/10.1111/desc.12505

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion*, *11*(2), 262–266. https://doi.org/10.1037/a0021302

Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work Group Diversity and Group Performance: An Integrative Model and Research Agenda. *Journal of Applied Psychology*, *89*(6), 1008–1022. https://doi.org/10.1037/0021-9010.89.6.1008

Vera, D., & Crossan, M. (2005). Improvisation and Innovative Performance in Teams. *Organization Science*, *16*(3), 203–224. https://doi.org/10.1287/orsc.1050.0126

Walton, R. E., & McKersie, R. B. (1965). *A behavioral theory of labor relations.* McGraw-Hill.

Wang, L.-S., Cheng, J.-T., Hsu, I.-J., Liou, S., Kung, C.-C., Chen, D.-Y., & Weng, M.-H. (2022). Distinct cerebral coherence in task-based fMRI hyperscanning: Cooperation versus competition. *Cerebral Cortex*, *33*(2), 421–433. https://doi.org/10.1093/cercor/bhac075

Weingart, L. R., Brett, J. M., Olekalns, M., & Smith, P. L. (2007). Conflicting social motives in negotiating groups. *Journal of Personality and Social Psychology*, *93*(6), 994–1010. https://doi.org/10.1037/0022-3514.93.6.994

Weingart, L. R., Thompson, L. L., Bazerman, M. H., & Carroll, J. S. (1990). TACTICAL BEHAVIOR AND NEGOTIATION OUTCOMES. *International Journal of Conflict Management*, *1*(1), 7–31. https://doi.org/10.1108/eb022670

Wheeler, S. C., Briñol, P., & Hermann, A. D. (2007). Resistance to persuasion as self-regulation: Ego-depletion and its effects on attitude change processes. *Journal of Experimental Social Psychology*, *43*(1), 150–156. https://doi.org/10.1016/j.jesp.2006.01.001

Williams, M., Belkin, L. Y., & Chen, C. C. (2020). Cognitive Flexibility Matters: The Role of Multilevel Positive Affect and Cognitive Flexibility in Shaping Victims' Cooperative and Uncooperative Behavioral Responses to Trust Violations. *Group & Organization Management*, *45*(2), 181–218. https://doi.org/10.1177/1059601120911224

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and Cooperation. *Psychological Science*, *20*(1), 1–5. https://doi.org/10.1111/j.1467-9280.2008.02253.x

Yeomans, M., Brooks, A. W., Huang, K., Minson, J., & Gino, F. (2019). It helps to ask: The cumulative benefits of asking follow-up questions. *Journal of Personality and Social Psychology*, *117*(6), 1139–1144. https://doi.org/10.1037/pspi0000220

Zhang, H., Yang, J., Ni, J., De Dreu, C. K. W., & Ma, Y. (2023). Leader–follower behavioural coordination and neural synchronization during intergroup conflict. *Nature Human Behaviour*, *7*(12), 2169–2181. https://doi.org/10.1038/s41562-023-01663-0

Zhang, W., Sjoerds, Z., & Hommel, B. (2020). Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking. *NeuroImage*, *210*, 116572. https://doi.org/10.1016/j.neuroimage.2020.116572