Exposure to the Views of Opposing Others with Latent Cognitive Differences Results in Social Influence—But Only When Those Differences Remain Obscured

February 2023

Abstract

Cognitive differences can catalyze social learning through the process of one-to-one social influence. Yet the learning benefits of exposure to the ideas of cognitively dissimilar others often fail to materialize. Why do cognitive differences produce learning from interpersonal influence in some contexts but not in others? To answer this question, we distinguish between cognition that is *expressed*—one's public stance on an issue and the way in which supporting arguments are framed—and cognition that is *latent*—the semantic associations that underpin these expressions. We theorize that, when latent cognition is *obscured*, one is more likely to be influenced to change one's mind on an issue when exposed to the opposing ideas of cognitively dissimilar, rather than similar, others. When latent cognition is instead *observable*, a subtle similarity-attraction response tends to counteract the potency of cognitive differences—even when social identity cues and other categorical distinctions are inaccessible. To evaluate these ideas, we introduce a novel experimental paradigm in which participants: (a) respond to a polarizing scenario; (b) view an opposing argument by another whose latent cognition is either similar to or different from their own and is either observable or obscured; and (c) have an opportunity to respond again to the scenario. A pre-registered study (N=1,000) finds support for our theory. A supplemental study (N=200) suggests that the social influence of latent cognitive differences operates through the mechanism of argument novelty. We discuss implications of these findings for research on social influence, collective intelligence, and cognitive diversity in groups.

Cognitive differences are often assumed to be a catalyst for social learning. Such learning occurs in part through social influence: When exposed to the opposing ideas of cognitively dissimilar others, people are more likely to find those ideas novel and thus to update their prior beliefs and change their minds about a focal issue (Marsden and Friedkin 1993; Page 2019; Becker, Brackbill, and Centola 2017). Yet the learning benefits of exposure to the ideas of cognitively dissimilar others often fail to materialize (Baldassarri and Bearman 2007; Guilbeault, Becker, and Centola 2018; Bail, Argyle, Brown, Bumpus, Chen, Hunzaker, Lee, Mann, Merhout, and Volfovsky 2018). Why do cognitive differences result in learning from interpersonal influence in some contexts but not in others?

We propose that the answer partly lies in the distinction between expressed and latent cognition and in the observability of the latter to a potential influence target. By *expressed cognition*, we refer to one's public stance on a given issue and the arguments one makes to support this view. Our conceptualization of *latent cognition* draws on a growing body of work in cultural sociology that focuses on schemas, the semantic associations through which people understand and develop their point of view on a focal issue (Wood, Stoltz, Van Ness, and Taylor 2018; Hunzaker and Valentino 2019; Cerulo, Leschziner, and Shepherd 2021; Boutyline and Soter 2020).[1] Two organizational leaders who disagree on the strategic choice of growing the business organically or instead through a joint venture with an external partner can still vary in the extent to which they are cognitively similar or different at a deeper level—that is, the degree to which their schemas of such concepts as partnership or control, which in turn inform their stances and supporting arguments about the appropriate growth strategy, diverge from one another (Goldberg 2011). Because these schemas are not directly accessible to others, we refer to them as latent.

How does one's latent cognitive similarity or dissimilarity to opposing others relate to social influence—that is, the likelihood of changing one's mind about an issue after being exposed to the arguments of another individual who has a different stance? On one hand, one might find the ideas

[1] Our conceptual arguments and measurement strategy, described in greater detail below, focus on the semantic associations people have about a specific concept rather than the full set of associations that exist in their minds. Associations about a specific concept are sometimes referred to as construals rather than schemas. Given that sociologists frequently use "schema" as an omnibus term, we adopt this terminology to remain consistent with the prevailing nomenclature.

of opposing others whose latent cognition is similar to one's own to be more persuasive because such arguments may be more compatible with one's own way of thinking. On the other, the ideas of opposing others whose latent cognition is dissimilar to one's own might prove more convincing because they are more likely to be novel and challenge one's taken-for-granted assumptions. We propose that, when similarities and differences between one's own and another individual's latent cognition are obscured, the novelty advantage of ideas produced by a cognitively dissimilar other will outweigh the familiarity edge of ideas emanating from someone who is cognitively similar. In other words, people are more likely to be influenced by the arguments of opposing others whose latent cognition is dissimilar, rather than similar, to their own.

Next we consider the possibility that people may be less open to another person's opposing viewpoint if they are aware of how this person sees the world differently than they do. Although the semantic associations that exist in others' minds are not directly accessible, they can be inferred to varying degrees through indirect means—for example, based on their past actions or the symbols they choose to display. We theorize that when latent cognitive similarities and differences are observable, a subtle similarity-attraction response will be activated when one engages with others' views (Byrne 1997). As a consequence, one will prefer to engage with the less novel or challenging ideas of a person whose latent cognition is similar to one's own rather than the potentially more provocative ideas of an individual whose latent cognition is dissimilar. Whereas considerable prior work, spanning sociology and social psychology, has highlighted that people tend to be drawn to the ideas of others with whom they share a categorical identity and often eschew the ideas of outgroup members (Abrams and Hogg 1990; Macy, Kitts, Flache, and Benard 2003; Cialdini and Goldstein 2004; DellaPosta, Shi, and Macy 2015; Guilbeault, Becker, and Centola 2018), we depart from the existing literature in proposing that mere knowledge of the relationship between one's own and another's latent cognition can extinguish the potency of cognitive differences in exerting social influence—even in the absence of social identity cues about the individual (e.g., whether the person is of the same gender or race) or knowledge of other categorical distinctions that might exist (e.g., whether the person has the same personality type).

To evaluate these ideas, we introduce a novel experimental paradigm that departs from

prevailing approaches that compare the outcomes of one group of individuals who are exposed to a potential influence source and another group that is not exposed. Our technique instead exposes all participants to the ideas of an opposing other but varies: (a) whether the idea originates from someone whose latent cognition is similar to or different from their own; and (b) whether latent cognitive similarities or differences are obscured or observable. We report the results of a pre-registered study (N=1,000) that finds support for our theory. We also report the results of a supplemental study (N=200), which suggests that the social influence of latent cognitive differences operates through the mechanism of argument novelty. We discuss implications of this work for research on social influence, collective intelligence, and cognitive diversity in groups.

INTERACTION, COGNITIVE DIFFERENCES, AND INFLUENCE

Social influence can occur at the dyadic level (Hogg and Turner 1987; Friedkin and Johnsen 2011; Rivera, Soderstrom, and Uzzi 2010), as well as in the context of social groups and broader networks (Postmes, Haslam, and Swaab 2005; Baldassarri and Bearman 2007). Yet even learning in group contexts often arises in part through the process of one-to-one social influence. Consider, for example, a board of directors that is debating a shift in a firm's growth strategy. A director who was initially opposed to growth through a joint venture with an external partner may come around to support this strategy through a series of one-on-one consultations with the CEO and other management team members prior to the formal board vote on approving the joint venture. Our conceptual arguments pertain to this foundational process of one-to-one influence.

Whether in the context of dyads or larger social groups, the conditions under which exposure to differing viewpoints leads to social influence have been longstanding concerns for both sociologists and social psychologists (French 1956; DeGroot 1974; Marsden 1981). Sociological work in this vein has focused on network-analytic models in which individuals' views can change endogenously through interaction and exogenously based on the conditions through which their initial views formed (Marsden and Friedkin 1993; Baldassarri and Bearman 2007; Centola 2021). Meanwhile, social psychological research has instead emphasized individual motivations—for ex-

ample, for accurate information, self-consistency, and identity affirmation—as well as properties of the messages conveyed during interaction (Wood 2000; Cialdini and Goldstein 2004; Cialdini and Sagarin 2005).

A unifying theme across these literatures centers on how cognitive differences can, in some cases, produce social learning through interpersonal influence and, in other instances, fail to do so or even lead people to become more entrenched in their views. Exemplifying the former perspective, Balietti et al. (2021) report that dyads in which one member was exposed to the views of another who had the opposite stance on the question of whether government policy should further redistribute wealth subsequently became *less* polarized in their views. Moreover, when participants were also informed about incidental, nonpolitical similarities they had with the argument source, they reported feeling closer and more open to the source. In a similar fashion, Feinberg and Willer (2015) report that individuals exposed to opposing arguments on such controversial topics as universal health care and same-sex marriage were more apt to be influenced to change their minds—but only when the arguments were framed in language consistent with their moral values.

Although not always directly related to influence outcomes, supportive evidence for the learning benefits of cognitive diversity also comes from research on its role in team performance: When team members approach problems from different perspectives, they can jointly develop novel insights that no individual could have conceived of independently (Pelled, Eisenhardt, and Xin 1999; Amabile, Conti, Coon, Lazenby, and Herron 1996; Aggarwal and Woolley 2019; Lix, Goldberg, Srivastava, and Valentine 2022). For instance, de Vaan, Stark, and Vedres (2015) find that creative teams are most likely to learn from one another in ways that result in breakthrough innovations when they exhibit the network property of "structural folding"—members of one cohesive group have shared memberships with another cohesive group—and when the cognitive distance between the groups is high. Hong and Page (2004) meanwhile report results of simulation studies demonstrating that groups composed of randomly assembled problem solvers outperform groups that consist of high-ability problem solvers. This difference arises because randomly assembled groups are cognitively divergent, whereas groups whose members are selected on the basis of demonstrated ability also tend to be cognitively convergent.

In contrast to these findings, another set of studies demonstrates how various identity-based mechanisms can undermine the efficacy of cognitive differences in producing social influence. When social identities are salient, people are more likely to change their views on a topic when they are exposed to ideas from others who have similar social identities and more likely to entrench in their views when exposed to ideas from different-identity group members (Mark 2003; DellaPosta, Shi, and Macy 2015). For example, Bail and colleagues (2018) report that exposure to messages from those with opposing political ideologies led Republicans to become significantly more, not less, conservative in their subsequently reported attitudes and Democrats to become slightly more, not less, liberal in their later-reported views—although this latter shift was not statistically significant.

In the context of face-to-face deliberations about the rights of sexual minorities in Poland, Wojcieszak (2011) shows that those who initially held extreme views on the topic and perceived disagreement with their interlocutor exhibited greater opinion polarization following the discussion. In an online experiment, Guilbeault, Becker, and Centola (2018) find that exposure to opposing beliefs in bipartisan social networks improved the accuracy of judgments about climate change among both conservatives and liberals to the point of eliminating belief polarization. When people were primed to think about partisanship by being exposed to the logos of the two main political parties in the U.S., however, the social learning benefits of cognitive difference evaporated. Consistent with this finding, Liu and Srivastava (2015) find that increased interaction between United States senators with opposing political identities, which corresponds to cognitive difference, led to greater divergence, rather than convergence, in their voting behavior. Indeed, this tendency for "negative" social influence, fueled in part by identity-based mechanisms, has been documented in a wide range of simulation studies (Macy, Kitts, Flache, and Benard 2003; Flache and Macy 2011).

Together, these accounts paint a picture in which cognitive differences between people fuel social influence, while awareness of these differences dampens people's willingness to learn from one another. Previous work attributes this suppressive effect to social identity. We argue, in contrast, that people may fail to be influenced by opposing ideas if they are aware of their cognitive alignment with the argument source—even in the absence of any identity-related signals.

EXPRESSED VERSUS LATENT COGNITION

When two people come into contact with one another to discuss or debate an issue, some facets of their cognition are revealed through discourse (Lix, Goldberg, Srivastava, and Valentine 2022). Such *expressed* forms of cognition include the stance one takes on a given issue, as well as the ways in which one frames one's supporting arguments. Other facets of cognition, such as mental models (Carley and Palmquist 1992; Converse, Cannon-Bowers, and Salas 1993), are instead *latent* in that they are typically not directly accessible to others.

Latent Cognition in the Form of Schemas

Our conceptualization of latent cognition is grounded in the growing body of work in cultural sociology on schemas, which encompass the broad array of concepts through which people view the world and the semantic associations they make between these concepts (d'Andrade 1995; DiMaggio 1997; Srivastava and Banaji 2011; Hunzaker 2016; Gauchat and Andrews 2018; Miles, Charron-Chénier, and Schleifer 2019). A core insight from this work is that two individuals may take different stances, based on a wide range of arguments, on a focal issue but still vary in the extent to which they have a common understanding about concepts that are closely related to the focal issue (Goldberg 2011; Goldberg and Stein 2018). In other words, misalignment on stances does not always imply misalignment on the schemas that inform the stance and the logic supporting it. For example, two organizational leaders who are on opposite sides of a debate about whether to launch a new joint venture may nevertheless have similar schemas about such concepts as "control" or "risk." Is one more likely to be influenced by opposing arguments that emanate from another whose schemas are similar to or different from one's own?

Figures 1 and 2 illustrate our conceptualization of this question. In Figure 1, the thought bubble on the left, which represents latent cognition, depicts a person's schemas as a set of associations relative to a focal concept.[2] Returning to the illustrative example of a debate within a

[2]Our depiction builds on connectionist models of cognition (Tenenbaum, Kemp, Griffiths, and Goodman 2011). In this view, activating a focal concept leads to the subsequent activation of a set of related concepts through the process of "spreading activation" (Collins and Loftus 1975).

board of directors about whether to pursue growth through a joint venture or organically, certain concepts such as "cooperation," "control," and "risk" might be especially salient and might therefore shape how a given director understands the issue, thinks about the tradeoffs, and formulates a point of view. The thought bubble on the right, which represents expressed cognition, depicts the public stance a person ultimately takes on the choice, as well as the way in which she frames the arguments and supporting rationale for this view to her fellow directors. Although concepts such as "control" and "risk" may partly inform a person's views, in this example, the focal concept of "cooperation" is most relevant in shaping the individual's public stance and supporting arguments.

[Figure 1 about here.]

Figure 2 depicts the two scenarios of interest. In Panel A, the focal individual, 1, is exposed to the supporting arguments of another person, 2, who has an opposing stance on the joint venture decision but whose latent cognition is similar given the high degree of overlap in their schemas of the salient concept of "cooperation." In Panel B, the focal individual, 1, is exposed to the supporting arguments of another person, 3, who also has an opposing stance on the joint venture decision and whose latent cognition is dissimilar given the limited overlap in their schemas of "cooperation."

[Figure 2 about here.]

Social Influence When Latent Cognitive Differences are Obscured

Even in the absence of social identity cues, people discussing substantive topics can become aware of the latent cognitive similarities and differences that exist between them. For example, in the one-on-one deliberations between board members on whether to launch a new joint venture, a focal individual may draw inferences about the other party's schemas based on prior discussions of other issues involving principles of corporate control and risk management or based on their functional backgrounds. Yet such discussions can also occur in ways that keep latent cognitive differences mostly hidden from view—for example, if board members were to write out their arguments in favor of their views and those arguments were then anonymously circulated among the group.

We first consider the relationship between latent cognitive differences and social influence when the former are obscured. In such a context, people may, on one hand, be drawn to the opposing ideas and arguments of individuals with similar schemas because their ways of thinking are logically compatible. In our example, those in favor of initiating a joint venture who think of the concept of "cooperation" as being closely associated with "compromise" might, for instance, find opposing arguments that emphasize the adverse consequences of misaligned goals and interests of the two partnering organizations to make logical sense. Yet because "cooperation" and "compromise" are already strongly associated in their minds, it is also likely that they have already thought of and dismissed such arguments. In contrast, when such individuals are confronted with opposing arguments generated by others who construe "cooperation" as being more closely tied to "coordination" than to "compromise," they are more likely to be exposed to new ways of thinking about tradeoffs associated with the choice—for example, the level of effort and cost that might be involved to coordinate activities and decisions with joint venture leaders. This exposure might, in turn, lead them to update their beliefs and withdraw their support for the joint venture.

Extrapolating from this stylized example, we propose that, when latent cognitive differences are obscured, the novelty advantage of cognitive dissimilarity will outweigh the familiarity benefits of cognitive similarity in producing social influence. Thus, we hypothesize:[3]

Hypothesis 1: When latent cognition is obscured, one will be more likely to be socially influenced to change one's mind about a given topic when exposed to opposing arguments from another individual with dissimilar, rather than similar, latent cognition.

Social Influence When Latent Cognitive Differences are Observable

When latent cognitive differences are observable, we theorize that a subtle similarity-attraction mechanism (Byrne 1997) will be activated that will undermine the potency of cognitive differences in producing social influence—even when social identity cues and other forms of categorical distinction are absent. In particular, when people become aware that the opposing ideas they are encountering

[3]Hypotheses were pre-registered on Open Science Foundation. A blinded version of the preregistration can be found at: https://osf.io/7mqwv/?view$_o nly = 907df7b6d82046ccbe077539eba52061$

come from someone whose latent cognition is similar to their own, they will develop a preference for engaging with those ideas. Conversely, when they become aware that opposing ideas emanate from an individual whose latent cognition is different from their own, they will pay less attention to and even discount those ideas.

Support for this perspective comes from self-categorization theory in social psychology (Turner, Hogg, Oakes, Reicher, and Wetherell 1987; Turner and Reynolds 2011), which finds that people tend to be more persuaded by information that comes from an ingroup source than from an outgroup source (Turner, Wetherell, and Hogg 1989; Hogg, Turner, and Davidson 1990). Whereas self-categorization theory relies on the existence of social groups with which people might associate their selves and whose stereotypical tendencies shape their subsequent attitudes and behavior, we instead consider the consequences for social influence of merely knowing that an opposing other is cognitively similar or different.

We propose that the observability of latent cognition will induce people to differentially engage with the less novel or challenging ideas of cognitively similar others rather than the potentially more provocative ideas of cognitively dissimilar others. Returning to our motivating example of a managerial debate about launching a new joint venture, an individual who construes "cooperation" as being closely associated with "compromise" might find the opposing arguments of another person who instead perceives "cooperation" as being closely tied to "coordination" to be novel and thus persuasive; however, given that people tend to avoid engaging with and pay less attention to information from dissimilar information sources (Turner 1991), knowing that this colleague has a different schema might also lead the focal individual to discount the colleague's novel information. We therefore expect:

Hypothesis 2: The observability of latent cognitive similarities and differences will negatively moderate the effects of cognitive dissimilarity in producing social influence.

Scope Conditions and Relevant Contexts

We expect our theory to apply in contexts wherein people lack direct or indirect information about each other's latent schemas. In such settings, the novelty of opposing ideas generated by cognitively different others will garner attention and influence a focal individual's thinking. Under what conditions can we expect our theory to be operative? We propose two scope conditions: when interactions occur in the absence of identity-related cues and without the use of ideology-revealing language.

First, we expect latent schemas to be less easily inferred through interaction when identity-related cues—for example, related to gender, race, political affiliation, and class—are obscured. Thus, we would not expect our theory to apply to typical interactions on such platforms as Twitter and Facebook wherein partisan signaling is salient and widespread (Taylor, Muchnik, Kumar, and Aral 2022; Bail, Argyle, Brown, Bumpus, Chen, Hunzaker, Lee, Mann, Merhout, and Volfovsky 2018). Instead, our theory is better suited to explaining the dynamics of social influence on platforms in which people exchange opinions largely anonymously. Consider someone who arrives to the micro-blogging platform Reddit and reads the first comment they see on a given thread. This user may not seek out or acquire any identifying information about the anonymous Reddit user who posted the comment. Such interactions are not only common on Reddit, but they can also have a striking influence on people's opinions and behaviors (Moyer, Carson, Dye, Carson, and Goldbaum 2015; Mancini, Desiderio, Di Clemente, and Cimini 2022). Other relevant contexts include platforms like Gigster and Topcoder, in which software developers collaborate online to write code and can choose how much identity-related information to share (Lix, Goldberg, Srivastava, and Valentine 2022).

Second, we expect our theory to apply when the arguments people are exposed to are mostly substantive and do not use language that implicitly reveals their ideology. For example, in a debate about abortion rights, if one person's arguments invoked the term "unborn children" and the other's reasoning centered on the term "war on reproductive rights," both sides would be revealing their political ideology. In doing so, they would also "leak" their underlying schemas related to the

policy question at hand. Thus, we would not expect our theory to apply to highly politicized debates (e.g., gun control, government regulation) in which the language people use to construct their arguments is readily associated with a well-recognized ideology. Rather, our theory is better suited to explaining social influence on issues that are either too new or too nuanced to fit a well-recognized ideology, since the presence of such ideologies can lead people to associate each other's opinions with particular identities and thereby activate ingroup / outgroup biases. Examples of the latter might include whether and how to regulate new AI technology, whether and how to require remote workers to return to the office, and how to handle common leadership dilemmas in the workplace. We test our theory using an experimental paradigm that satisfies both of these scope conditions: interactions happen in ways that do not reveal identity-related information and on a topic (a leadership dilemma) about which ideological information is not automatically revealed through language.

METHOD

Setting Up the Experimental Paradigm

To test our hypotheses, we needed an experimental paradigm that would allow us to: (1) present participants with a decision task that was likely to elicit polarized views so we could connect them to the supporting arguments of others who could potentially persuade them to change their mind; (2) have control over whether participants were connected to peers whose latent cognition was dissimilar or similar to their own; and (3) have the ability to toggle whether or not the latent cognitive similarity or difference between a participant and the source of a given argument was obscured or instead observable. We accomplished these aims through four preliminary steps.

1. Identifying A Polarizing Scenario

We first identified from the organizational behavior literature a situational judgment test (hereafter referred to as an SJT), which represents a hypothetical leadership dilemma with a finite set of choices (Peus, Braun, and Frey 2013; McDaniel and Whetzel 2005). This scenario is similar to

our motivating example of a board debating different growth strategies in that one can formulate a set of logical arguments to support alternative points of view. It differs in that it focuses on the choice that an individual leader would make with respect to a given subordinate. Yet one can readily imagine how such a decision would be susceptible to social influence from other leaders. For example, in the context of a matrix organization, two leaders may need to confer with each other to reach agreement on how to respond to an underperforming subordinate who has a solid-line reporting relationship to one leader and a dotted-line relationship to the other. One advantage of choosing a hypothetical scenario rather than a currently raging policy debate is that our study participants had neither prior experience with the scenario nor prior exposure to the ideas of people who disagree with them about it. On the other hand, the use of a hypothetical scenario raises questions about external validity, which we address in the discussion section below.

We pretested a broader set of existing SJTs, modified them such that they involved binary choices rather than more than two choices, and ultimately selected one that was polarizing (i.e., an SJT that elicited a bimodal distribution of opinions with few respondents holding moderate views) (Fig. A1 in the Appendix confirms that the distribution of responses to this SJT is bimodal as intended).Appendix A provides further details on our procedure for selecting and configuring the SJT and provides a full description of the vignette.

2. Measuring Latent Cognitive Similarities and Differences

To measure latent cognitive similarities and differences between participants, we followed recent sociological work in using a semantic association task (Mohr 1998; Hunzaker and Valentino 2019; Srivastava and Banaji 2011). Such tasks involve asking participants to report the associations that come to mind when prompted with a particular concept of interest. Hunzaker and Valentino (2019), for example, use this approach to show that Republicans and Democrats have different conceptual associations related to poverty. Yet a limitation of their method is that it conflates interpretations of concepts with their valence. For instance, respondents are asked to associate multivocal concepts such as “immigrant” with terms such as “lazy” or “dishonest” that have strong positive or negative

connotations. It is therefore unclear whether the cognitive differences identified through their approach reflect differing interpretations of a concept or different feelings about it.

We therefore somewhat modified the approach used by Hunzaker and Valentino (2019) by implicitly controlling for valence when eliciting participants' conceptual associations about the focal concept related to our hypothetical scenario. To choose concepts that are potentially associated with "leader," we drew from social psychological research indicating that people commonly and instinctively perceive others through the dual lenses of warmth and competence (Fiske, Cuddy, and Glick 2007). Warmth-competence perceptions are also central to people's assessments of organizational leaders (Chemers 2014). We chose our associated concepts from 64 positive and negative words related to dimensions of warmth and competence (Rosenberg, Nelson, and Vivekananthan 1968; Fiske, Cuddy, and Glick 2007). We asked participants to select four words that were associated in their minds with "leader" from a menu of eight words in a given subset (repeated for 6 distinct subsets). We supplemented the warmth-competence terms with neutral terms. This design resulted in the selection of terms within a given set that all had a comparable valence. Overall, each participant encountered six blocks of eight terms, one of which had unambiguously positive valence, one of which had unambiguously negative valence, two of which had mixed (i.e., including both positive and negative terms) valence, and two of which had neutral valence. Table 1 lists the words we ultimately selected. Figure A2 in the appendix verifies that participants who provided similar word associations via this method also rated opposing arguments for the SJT in a similar manner, suggesting that our association task detects cognitively meaningful similarities and differences that are relevant to the SJT.

[Table 1 about here.]

3. Generating Opposing Arguments and Assessing Reference Group Schemas

Next we had to generate a set of opposing arguments for the SJT and assess the schemas of those who came up with each argument. We did so by conducting a preliminary study with a sample of

self-identified organizational leaders[4] (N=200) on the Prolific platform (https://www.prolific.co/). We refer to this group as our Reference Sample. Participants in the Reference Sample completed the SJT and the schema elicitation task. We asked them to not only tell us their stance on the SJT but also to write out their arguments in support of their view.

4. Evaluating and Selecting Opposing Arguments

Finally, we conducted a second preliminary study on Prolific with a different sample of self-identified organizational leaders (N=200), which we refer to as our Evaluation Sample. They were presented with the arguments generated by Reference Sample participants and asked to assess the arguments on such dimensions as novelty and informativeness. Based on these evaluations, we identified ten arguments of comparable length and quality that were in support of each side of the SJT dilemma. Given that the SJT involved a binary choice, this yielded a total of twenty arguments. Examples of these arguments are included in Appendix A.

Main Study: Sample and Procedure

Based on pre-registered power analyses, we recruited 1,000 self-identified organizational leaders who had not participated in any of the preliminary studies for our main study. These participants had the following profile: their mean age was 37.5 (σ = 10.91) years. 40% identified as Female. 10.5% identified as Black or African American; 4% as Multiracial; 6% as Asian; 5% as Hispanic/Latino; 72% as White; and 3.5% in other categories.[5]

[4] Across all studies, participants were classified as organizational leaders if they answered "Yes" to all of the following pre-screening questions provided by Prolific: (1.) "At work, do you have any supervisory responsibilities? In other words, do you have the authority to give instructions to subordinates?" (2.) "Do you have any experience being in a management position?" (3.) "Does your work require you to regularly interact with other employees (e.g. co-workers, colleagues, subordinates, assistants)?" and (4.) "At work, how many people do you have the authority to give instructions to? That is, how many subordinates do you have?" Responses to 4 that identified one or more subordinates were coded as a "Yes" response for our purposes.

[5] 10.4% listed their highest level of education as High School Diploma (or lower); 15.2% as some College; 40.3% as Bachelor's Degree; 29.1% as Master's Degree; and 5% as MD or PhD. The participants reported the following number of years at work: less than 1 year 10%; 1-5 years 37%; More than 5 years 53%. The main sectors of employment in which our participants work were as follows: Legal 1.4%; Arts 2.8%; Agriculture & Food 1.7%; Medicine 8.2%; Finance 6.7%; Government 8.1%; Architecture & Construction 3.7%; Education 13.1%; Hospitality & Tourism 3.8%; Business and Administration 4.7%; and Information Technology (IT) 9.2%; Manufacturing 4.6%; Retail 6.7%; Science, Technology, Engineering and Mathematics 6.7%; Social Sciences 2%; Transportation and Logistics 3.1%.

Participants first completed the schema elicitation task and were then presented with the SJT and asked to make a decision about it. We randomly counterbalanced which of the two binary choices appeared on the left side of the decision slider. Participants were then shown an argument from someone in the Reference Sample (described above) who had made the *opposite* choice on the SJT.[6] We followed a 2x2 design, with participants being assigned to conditions of latent cognitive *similarity* versus *dissimilarity* and *observable* versus *obscured* latent cognition. We developed an algorithm to choose arguments that were generated by a Reference Sample participant who was cognitively similar to or different from the focal study participant, and we programmed our survey to present this argumentn real time. Latent cognitive similarity was assessed based on the Jaccard Index of participants' schema elicitation task responses—that is, the proportion of associated concepts that overlapped between our focal participant and the participant from our Reference sample who generated a given opposing argument.[7] In the latent cognitive similarity condition, the algorithm selected an argument generated by a Reference Sample participant whose schema elicitation task responses overlapped significantly with those of the main study participant, whereas in the latent cognitive dissimilarity condition it selected an argument from one whose responses had little overlap.

Participants in the obscured condition only saw the opposing argument with no information provided about the source of the argument, while those in the observable condition were shown a number (i.e., the Jaccard Index) representing the degree of overlap between their responses to the schema elicitation task and those of the person whose opposing argument they were viewing. To provide some context for this number, they were told that the number was either above or below the mean level of overlap (i.e., the mean overlap between the focal participant and all participants

[6]Prior experimental studies find that in estimation and problem solving tasks similar to our paradigm, the base rate of people changing their opinion as a result of independent reflection (i.e., in the absence of social influence) is strikingly low (Centola 2022). For example, in two separate studies, less than 5% of participants changed their views as a result of reflection: one in the context of lay people evaluating climate data (Guilbeault, Becker, and Centola 2018) and another in the context of clinicians recommending treatment actions (Centola, Guilbeault, Sarkar, Khoong, and Zhang 2021). Given the extent to which this has been demonstrated in prior research, our current study exposes all participants to opposing arguments to focus on identifying different conditions under which this exposure can lead to social influence

[7]Specifically, the Jaccard Index measures the intersection of the schema elicitation task responses from participant i and j, divided by the union of the schema elicitation task responses from participant i and j. In this sense, the Jaccard Index measures the overlap of participants' semantic associations, while normalizing this overlap by controlling for the overall number of unique word selections made by both participants.

in the Reference Sample).

Figure 3 provides an overview of the conditions and procedure we used in this study.[8]

[Figure 3 about here.]

As Figure 4 indicates, our randomization procedure was successful: The cognitive distance between a main study participant and the Reference Sample participant whose argument they were exposed to is strongly predicted by experimental condition.

[Figure 4 about here.]

After viewing the opposing argument, participants were shown the SJT a second time and were given the opportunity to revisit their decision about it. Consistent with our pre-registration, we focused on two dependent variables. The first is whether a participant changed her mind about the binary choice associated with an SJT. We classified a change in mind as movement across the midpoint of the 100-point slider scale, which ranged from -50 (the strongest possible position in favor of Option 1) to 50 (the strongest possible position in favor of Option 2), between the first and second time a person was presented with the SJT. For example, a participant was coded as having changed her mind if her initial response was -12 (i.e., slightly in favor of Option 1) and her subsequent response was 9 (i.e., slightly in favor of Option 2). The second dependent variable is the degree to which a person's response shifted in the direction of the opposing view—whether or not they made a categorical shift in their opinion. By this measure, a person who shifted from -40 to -10 (i.e., moving from strong to lukewarm support for Option 1) would be assessed as being more strongly influenced than someone who shifted from -40 to -38 (i.e., moving only slightly toward the opposing view) or from -40 to -43 (i.e., moving further away from the opposing view).[9]

[8] The following link provides access to our survey instrument: bit.ly/social$_i nfluence_s tudy. To preserve anonymity, the page for elicitin$

[9] This study is an improved replication of three pre-registered studies. A document summarizing these initial studies and their results is provided here: https://osf.io/mzrdt. One of these studies replicates the latent cognition observable condition while revealing cognitive similarity by categorizing participants as either a "Type-X" or "Type-Y"manager, such that managers in the observable latent cognitive similarity condition were told that the argument presented was produced by a manager of the same type, and managers in the observable latent cognitive dissimilarity condition were told that the argument presented was produced by a manager of the different type. The results

Supplemental Mechanism Study

Our theory posits that the mechanism through which cognitive differences produce social influence is argument novelty. To test this idea and to establish that our findings are not an artifact of our choice of SJT, we conducted a pre-registered supplemental study using a different sample of self-identified organizational leaders from Prolific (N=200).[10] The study sample and procedure are reported in Appendix C.

RESULTS

Before reporting our main results, we note that there was no significant difference in the initial decisions made by participants—that is, before exposure to an opposing argument—across all conditions (Kruskal-Wallis H Test, $p = 0.89$, $\chi^2 = 0.63$; see Figure A1 in Appendix A for details). Moreover, as reported in Appendix B, each argument was equally likely to appear across all conditions, suggesting that the randomization algorithm worked in the way we intended it to.

In support of Hypothesis 1, we find the expected difference between the two conditions in the likelihood of opposing arguments changing participants' minds: 21% of managers in the obscured latent cognitive similarity condition changed their decision, whereas 31% of managers in the obscured latent cognitive dissimilarity condition did so. This difference is statistically significant ($p = 0.01$, Two-sample Proportion Test). The initial choice made by managers had no significant effect on their likelihood of changing decision: In the latent cognitive similarity / obscured condition, 22% of participants who initially provided option 1 changed their decision and 21% of participants who initially provided option 2 changed their decision ($p = 1$, Two-sample Proportion Test); likewise, in the latent cognitive dissimilarity / obscured condition, 35.5% of participants who initially provided option 1 changed their decision and 27% of participants who initially provided option 2 changed their decision ($p = 0.19$, Two-sample Proportion Test).

from this study are consistent with our theory and show that signaling cognitive similarity through arbitrary group identities has a stronger effect on disrupting the influence edge of cognitive difference as compared to the numerical manipulation.

[10] A blinded version of the preregistration can be found at: https://osf.io/8mc59/?view$_o nly$ = 9376fe02456$a4bc$39b1605$aaa5b4f$653.

Table 2 reports results of a logistic regression based on data from the two conditions in which latent cognition was obscured. To account for potential differences in the strength or quality of arguments to which participants were exposed, the model includes argument fixed effects. This model estimates that participants in the obscured latent cognitive dissimilarity condition were 2.52 times ($p < 0.001$) more likely to change their recommended decision relative to those in the latent cognitive similarity condition.

[Table 2 about here.]

Table 3 tests Hypothesis 1 using our second dependent variable: the magnitude of opinion change in the direction of the opposing argument—whether or not a person made a categorical change in their decision. Table 3 reports the results of an OLS regression with argument fixed effects, which again focuses on participants in the two conditions where latent cognition was obscured. In further support of Hypothesis 1, the results indicate that participants in these conditions who were exposed to cognitively dissimilar arguments made larger adjustments to their initial views in the direction of the opposing argument relative to those who were exposed to cognitively similar arguments. This difference was significant ($\beta = 6.72$, $p = 0.01$).

[Table 3 about here.]

Turning to Hypothesis 2, compared to participants in the condition in which latent cognition was obscured, those in the condition in which latent cognition was observable were not more likely to change their decision when exposed to an opposing argument from a dissimilar rather than a similar other. In fact, we observe the *opposite* trend: 32% of participants in the observable latent cognitive similarity condition changed their decision, whereas 24% of managers in the observable latent cognitive dissimilarity condition did so—a difference that is marginally significant ($p < 0.07$, Two-sample Proportion Test) and in the opposite direction as when latent cognition is obscured.

Table 4 reports results of a logistic regression that more formally tests Hypothesis 2 using the full sample from the main study. The model includes indicator variables indicating whether latent

cognition was dissimilar (rather than similar) and whether it was observable (rather than obscured). Hypothesis 2 is tested using the interaction of the two indicator variables. Again including argument fixed effects, we find strong evidence in support of Hypothesis 2: When cognitive dissimilarity is observable, opposing arguments are significantly less likely to change participants' decisions, as compared to when cognitive dissimilarity is obscured ($p < 0.01$, OR=0.37, β=-0.98, SE=0.29).

[Table 4 about here.]

Table 5 tests Hypothesis 2 using our second dependent variable of the magnitude of opinion change in the direction of the opposing argument. It reports the results of two OLS regressions that include argument fixed effects. The negative and significant interaction term in Model 1 demonstrates that Hypothesis 2 holds even when we consider the continuous measure of social influence rather than the binary one. Model 2 shows that this result is robust to controlling for the strength of a participant's initial stance on the issue (i.e., the first numerical response to the SJT). The mean magnitudes of opinion change were 16.7 for those in the obscured latent cognitive similarity condition, 19.5 for those in the obscured latent cognitive dissimilarity condition, 20.4 for those in the observable latent cognitive similarity condition, and 15.6 for those in the observable latent cognitive dissimilarity condition.

[Table 5 about here.]

In Appendix C, we describe the results of the supplemental study described above (N=200) that lends support for the theorized mechanism underpinning the relationship between cognitive dissimilarity and social influence: argument novelty. As reported in Figure C2 and Table C1 in Appendix C, we find that perceived argument novelty mediates the effect of cognitive dissimilarity on social influence. Not only were arguments generated by cognitively dissimilar individuals perceived as more novel, but they were also positively associated with participants' self-reported willingness to change their mind. This result holds across two distinct SJTs.

DISCUSSION

The goals of this article have been to: (1) examine how the degree of cognitive alignment or misalignment between individuals relates to their ability to influence one another; (2) uncover the contexts in which such influence is more or less likely to occur; and (3) investigate the mechanisms through which cognitive dissimilarity can sometimes yield social influence. We recruited 1,200 self-identified organizational leaders from an online platform to participate in two pre-registered studies. The results of our main experimental study (N=1,000) demonstrate that obscured cognitive differences promote social influence: Absent cues about social identity, other forms of categorical membership, and latent cognition, participants exposed to the ideas of opposing others with dissimilar latent cognition were more likely to change their minds about a decision than were participants exposed to the ideas of opposing others with similar latent cognition. Our main study also shows how the observability of latent cognition can nullify the influence edge of cognitive differences: Mere knowledge of the fact that a given argument was generated by someone who is cognitively similar or dissimilar led to no significant differences in social influence between the similarity and dissimilarity conditions. Our supplemental study (N=200) highlights a core mechanism through which latent cognitive differences lead to social influence—exposure to novel insights. The findings from these studies contribute to research on social influence, collective intelligence, and cognitive diversity in groups.

Contributions

Social Influence

Social influence research typically examines the consequences of being exposed to an actor or set of ideas relative to the state of not being "treated" (Salganik, Dodds, and Watts 2006; Liu, King, and Bearman 2010; Centola 2021). In contrast, *all* participants in our studies were confronted with the ideas of people who disagreed with them on a substantive issue. Yet these exposure events varied considerably in their efficacy. Our results therefore emphasize the importance of understanding not

just whether exposure occurred but also its potential potency (i.e., based on the latent cognitive differences between the influence source and target) and its context (i.e., whether those differences are observable).

Indeed, insights from this study can help reconcile some of the conflicting findings in prior work that has examined the conditions under which people are influenced to change their minds after being exposed to opposing points of view. For example, two of the studies reviewed above report that such exposure led to greater subsequent polarization of views (Wojcieszak 2011; Liu and Srivastava 2015). In both contexts, the interactions between individuals occurred face-to-face, thereby exposing individuals to identity-related information about one another and perhaps also enabling them to draw inferences about their latent cognitive similarities and differences. In contrast, in the study by Guilbeault and colleagues (2018), people were apt to be influenced by their peers and thus improved the accuracy of their forecasts in the experimental condition in which they had no knowledge about the identity or latent cognition of the information source. The study by Balietti and colleagues (2021) represents an intermediate case in which participants were exposed to an opposing viewpoint but also given information not relevant to the policy choice about their degree of similarity with the argument source. In other words, they were given superfluous identity-relevant information but not told anything about their similarity or dissimilarity to the argument source on latent cognitive dimensions relevant to the policy choice. The fact that such exposure led to a decrease in polarization suggests that access to information about incidental forms of similarity may still have allowed latent cognitive differences that were relevant for the focal issue to remain obscured and thus to fuel social influence. Future work in this vein would benefit from assessing the extent to which a given research design directly or indirectly exposes participants to the latent cognition of the argument source.

More generally, this study has implications for network models of social influence (Marsden and Friedkin 1993; Friedkin 2006; Becker, Brackbill, and Centola 2017). Such models typically involve the specification of a weight matrix, the elements of which represent the influence pattern in the network. The construction of this weight matrix typically involves choices such as the degree to which an actor's attributes shape the degree of influence from others, the mechanism

of social influence (e.g., communication or social comparison), which alters exert influence on an actor and which do not, and the magnitude of influence exerted by others (Leenders 2002). Our findings suggest that latent cognitive differences may prove to be a powerful predictor of influence magnitude—so long as they are not observable by the influence target.

Finally, we make an empirical contribution to social influence research by introducing a novel experimental paradigm that can be readily adapted to understanding how cognitive diversity relates to social learning on just about any issue of interest. To do so, researchers must simply identify the schemas that are most salient for a given issue, reconfigure the schema elicitation task by choosing a different set of potentially associated concepts, measure cognitive distances between people, expose people to the arguments of opposing others who vary in cognitive similarity, and manipulate the context of this exposure. We see potential in this paradigm's ability to reveal the conditions under which people are most likely to moderate their stance on other polarized topics—such as as climate change, policing reform, and vaccine mandates—that are more emotionally charged than hypothetical scenarios of the kind we used in this study.

Collective Intelligence

Collective intelligence research has focused on the consequences of cognitive differences for forecast accuracy (Surowiecki 2005; Becker, Brackbill, and Centola 2017)—for example, predictions about future stock prices (Chen, De, Hu, and Hwang 2014) or the likelihood of climate catastrophe (Aminpour, Gray, Jetter, Introne, Singer, and Arlinghaus 2020; Guilbeault, Becker, and Centola 2018). Although these studies aim to understand how accurate predictions ultimately translate into more effective decisions (Surowiecki 2005; Matzler, Strobl, and Bailom 2016; Page 2019), accuracy does not automatically beget higher quality decisions (Becker and Smith 2021; Frey and van de Rijt 2021). Findings from the present study demonstrate the link between cognitive differences and potentially better decisions that arise when people engage with and learn from opposing others.

Collective intelligence research has also highlighted a variety of ways in which identity-based homophily can blunt social learning—for example, when political identities are publicly displayed

(Mäs, Flache, and Kitts 2014; Bail, Argyle, Brown, Bumpus, Chen, Hunzaker, Lee, Mann, Merhout, and Volfovsky 2018; Guilbeault, Becker, and Centola 2018; Guilbeault and Centola 2020). We add to this understanding by identifying a novel mechanism—mere knowledge of latent cognitive similarities and differences—that can account for why people may fail to learn from one another even when information about social identities is unavailable. It remains to be explored the specific cues people use to infer the presence of latent cognitive similarities and differences in the absence of overt identity indicators or through what means such signals can be dampened.

More generally, our results lend support for the view that one of the reasons people reject the ideas of people with different identities is that they associate incongruous identities with fundamentally different ways of seeing the world. In the context of U.S. politics, for example, the reasons that liberals and conservatives dismiss each other's ideas may go beyond the simple mechanics of identity signaling; rather, it may involve deeply rooted assumptions about how each side construes substantive issues. We suspect that identities may simply represent a cognitive shortcut for understanding others' schematic associations and how they relate to one's own. This naturally raises the question of whether it is possible to induce people to become more receptive to novel arguments when identity differences are present. In this direction, a recent study referenced above found that liberals and conservatives became more receptive to counter-partisan arguments from their political out-group when non-political dimensions of identity similarity between the source and recipient of the argument (e.g., hobbies and lifestyle) were highlighted (Balietti, Getoor, Goldstein, and Watts 2021). It remains unclear the extent to which these findings generalize to the study of novelty and social influence in non-political settings such as the context examined in this study, revealing an important direction for future research.

Cognitive Diversity in Groups

This study also adds to a growing literature that has identified distinct facets of cognitive diversity in groups and new ways of measuring it—ranging from mental models (Carley and Palmquist 1992) to cognitive styles (Aggarwal and Woolley 2019) to schemas of poverty (Hunzaker and Valentino

2019) to linguistic categories (Guilbeault, Baronchelli, and Centola 2021). Although our approach in this study focused on cognitive differences at the dyadic level, it can be readily extended to measure cognitive diversity in larger social groups. Doing so would open up the possibility of using schema-based measures as a means to constructing groups that are well-suited to a given set of tasks. For example, rather than simply assembling groups randomly to support social learning (Hong and Page 2004), it may prove more effective to assemble groups of problem solvers on the basis of latent cognitive dissimilarity—without making them aware that this is the basis of assignment.

Similarly, Lix and colleagues (2022) demonstrate that groups that have a capacity to modulate their levels of discursive diversity—that is, the degree to which group members express their ideas in semantically divergent ways—achieve success by matching their levels of discursive diversity to changing task requirements. Yet it remains unclear how one would assemble a group that possesses this capability. Findings from the present study suggest a possible approach: identifying the schemas that are most relevant for a group's task requirements, measuring those schemas across a set of prospective group members, and then selecting a subset of individuals such that their schemas are mostly convergent on some concepts and mostly divergent on others. We conjecture that groups with a balanced portfolio of relevant schemas on which they are aligned and not aligned might be especially effective at navigating the competing pressures of ideation and coordination.

Our findings also challenge some prevailing assumptions in the literature on group cognitive diversity and performance. For example, perspective taking—that is, seeking to understand in a non-judgmental way others' viewpoints and the underlying rationale behind their stances—is widely viewed as a mechanism that can support mutual learning and creative production in groups (Grant and Berry 2011; Hoever, Van Knippenberg, Van Ginkel, and Barkema 2012). Yet, insofar as perspective taking leads people to focus on the latent cognitive similarities and differences that exist between themselves and others, our results suggest that it could also undermine social influence by activating a subtle similarity-attraction mechanism.

Our study also underscores that simply assembling a diverse collection of individuals into a group may be insufficient for realizing the benefits of diversity. Attending to the *process* by which

people understand and engage with their differences may be key to unlocking the value of group cognitive diversity. For example, a management team that is debating whether to merge with a competitor or an academic department that is deciding between two finalists for a faculty position may benefit from having individuals first write out their stance and supporting arguments and then read anonymized versions of everyone else's views. Even if such a decision process did not lead to consensus, it might yield greater mutual understanding.

In general, group contexts vary in the degree to which members are naturally aware of the latent cognitive similarities and differences that exist between them. When social identities are salient, for example, people are likely to make strong assumptions about their degree of cognitive overlap with others. Yet when social identities are either not salient or not strongly institutionalized, people may be more open to learning from one another when they engage in more superficial or transactional interactions with others that minimize their chances of becoming aware of others' different ways of thinking. We leave to future research the task of identifying the kinds of interactions that lead people to draw inferences about others' cognition and the conditions that shape the accuracy of these perceptions.

Limitations and Future Research

This study has certain limitations, which point to avenues for future research. First, although a virtue of our experimental paradigm is that it allows us to isolate the effects of obscured versus observable cognitive differences on social influence independent of the potentially confounding roles of social identity and other forms of categorical membership, this "clean" design comes at the cost of reduced external validity. For example, in most real-world settings, latent cognitive differences are only partially obscured. It remains to be explored how cognitive diversity relates to social learning in hybrid contexts in which some facets of latent cognitive difference are observable, while other dimensions are not.

Second, our study focused on social influence that arises from exposure to people holding opposing views. Yet people are also capable of being influenced by others who agree with them

about a polarizing issue—for example, by considering new arguments that lead them to commit more strongly to their viewpoint or by seeing discordant arguments that lead them to soften their stance. We speculate that such shifts in the strength of one's views are more likely to occur when a person is exposed to the ideas of people who share the same viewpoint and who are cognitively dissimilar rather than cognitively similar. We leave the testing of this hypothesis to future work.

Finally, although it was relatively easy in our study to identify a schema (i.e., about leaders) that was salient to the substantive issue of interest (i.e., leadership decision-making scenarios), we recognize that contentious issues vary in the degree to which they activate a single schema versus multiple schemas. For issues such as our motivating example of new venture growth strategies, multiple schemas such as those related to "cooperation," "control," and "risk" may be activated simultaneously. To account for this possibility, future research can adapt our experimental paradigm by identifying a broader set of salient schemas that matter for a given issue, configuring and implementing multiple schema elicitation tasks to assess these different schemas, and computing distances across all relevant schemas. Doing so would not only yield more robust measures of cognitive similarity or dissimilarity but would also allow researchers to examine how variance in cognitive diversity across different schemas might matter for social influence.

Conclusion

In his 1998 Nobel Peace Prize lecture[11], John Hume, one of the primary architects of the Northern Ireland peace process, remarked: "All conflict is about difference....Difference is the essence of humanity....The answer to difference is to respect it." Our study echoes Hume's sentiment: Social influence between opposing others does arise from engagement with ideas of others who are cognitively different. Yet, perhaps ironically, the way to respect differences in the course of easing conflicts might be to not call too much attention to them.

[11] https://www.nobelprize.org/prizes/peace/1998/hume/lecture/

References

Abrams, Dominic and Michael A Hogg. 1990. "Social Identification, Self-categorization and Social Influence." *European Review of Social Psychology* 1:195–228.

Aggarwal, Ishani and Anita Williams Woolley. 2019. "Team Creativity, Cognition, and Cognitive Style Diversity." *Management Science* 65:1586–1599.

Amabile, Teresa M., Regina Conti, Heather Coon, Jeffrey Lazenby, and Michael Herron. 1996. "Assessing the Work Environment for Creativity." *Academy of Management Journal* 39:1154–1184.

Aminpour, Payam, Steven A Gray, Antonie J Jetter, Joshua E Introne, Alison Singer, and Robert Arlinghaus. 2020. "Wisdom of Stakeholder Crowds in Complex Social—Ecological Systems." *Nature Sustainability* 3:191–199.

Bail, Christopher A, Lisa P Argyle, Taylor W Brown, John P Bumpus, Haohan Chen, MB Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences* 115:9216–9221.

Baldassarri, Delia and Peter Bearman. 2007. "Dynamics of Political Polarization." *American Sociological Review* 72:784–811.

Balietti, Stefano, Lise Getoor, Daniel G Goldstein, and Duncan J Watts. 2021. "Reducing Opinion Polarization: Effects of Exposure to Similar People with Differing Political Views." *Proceedings of the National Academy of Sciences* 118:e2112552118.

Becker, Joshua, Douglas Guilbeault and Ned Smith. 2021. "The Crowd Classification Problem: Social Dynamics of Binary-Choice Accuracy." *Management Science* 0:1–52.

Becker, Joshua, Devon Brackbill, and Damon Centola. 2017. "Network Dynamics of Social Influence in the Wisdom of Crowds." *Proceedings of the National Academy of Sciences* 114:E5070–E5076.

Boutyline, Andrei and Laura Katherine Soter. 2020. "Cultural Schemas: What They Are, How

to Find Them, and What to Do Once You've Caught One." *American Sociological Review* 86:728–758.

Byrne, Donn. 1997. "An Overview (and Underview) of Research and Theory Within the Attraction Paradigm." *Journal of Social and Personal Relationships* 14:417–431.

Carley, Kathleen and Michael Palmquist. 1992. "Extracting, Representing, and Analyzing Mental Models." *Social Forces* 70:601–636.

Centola, Damon. 2021. *Change: How to Make Big Things Happen*. New York City, NY: Little, Brown and Company.

Centola, Damon. 2022. "The network science of collective intelligence." *Trends in Cognitive Sciences* .

Centola, Damon, Douglas Guilbeault, Urmimala Sarkar, Elaine Khoong, and Jingwen Zhang. 2021. "The reduction of race and gender bias in clinical treatment recommendations using clinician peer networks in an experimental setting." *Nature communications* 12:1–10.

Cerulo, Karen A, Vanina Leschziner, and Hana Shepherd. 2021. "Rethinking Culture and Cognition." *Annual Review of Sociology* 47:63–85.

Chemers, Martin. 2014. *An Integrative Theory of Leadership*. New York, NY: Psychology Press.

Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media." *The Review of Financial Studies* 27:1367–1403.

Cialdini, Robert B and Noah J Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55:591–621.

Cialdini, Robert B and Brad J Sagarin. 2005. "Principles of Interpersonal Influence." .

Collins, Allan M and Elizabeth F Loftus. 1975. "A Spreading-activation Theory of Semantic Processing." *Psychological Review* 82:407.

Converse, Sharolyn, JA Cannon-Bowers, and E Salas. 1993. "Shared Mental Models in Expert Team Decision Making." In *Individual and Group Decision Making: Current Issues*, volume

221, pp. 221–46. Hillsdale.

d'Andrade, Roy G. 1995. *The Development of Cognitive Anthropology*. Cambridge University Press.

de Vaan, Mathijs, David Stark, and Balazs Vedres. 2015. "Game Changer: The Topology of Creativity." *American Journal of Sociology* 120:1144–1194.

DeGroot, Morris H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69:118–121.

DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology* 120:1473–1511.

DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23:263–287.

Feinberg, Matthew and Robb Willer. 2015. "From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence?" *Personality and Social Psychology Bulletin* 41:1665–1681.

Fiske, Susan T, Amy JC Cuddy, and Peter Glick. 2007. "Universal Dimensions of Social Cognition: Warmth and Competence." *Trends in Cognitive Sciences* 11:77–83.

Flache, Andreas and Michael W Macy. 2011. "Small Worlds and Cultural Polarization." *The Journal of Mathematical Sociology* 35:146–176.

French, John RP. 1956. "A Formal Theory of Social Power." *Psychological Review* 63:181.

Frey, Vincenz and Arnout van de Rijt. 2021. "Social Influence Undermines the Wisdom of the Crowd in Sequential Decision Making." *Management Science* 67:4273–4286.

Friedkin, Noah E. 2006. *A Structural Theory of Social Influence*. Number 13. Cambridge University Press.

Friedkin, Noah E and Eugene C Johnsen. 2011. *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*, volume 33. Cambridge University Press.

Gauchat, Gordon and Kenneth T Andrews. 2018. "The Cultural-Cognitive Mapping of Scientific Professions." *American Sociological Review* 83:567–595.

Goldberg, Amir. 2011. "Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined." *American Journal of Sociology* 116:1397–1436.

Goldberg, Amir and Sarah K Stein. 2018. "Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation." *American Sociological Review* 83:897–932.

Grant, Adam M and James W Berry. 2011. "The Necessity of Others is the Mother of Invention: Intrinsic and Prosocial Motivations, Perspective Taking, and Creativity." *Academy of Management Journal* 54:73–96.

Guilbeault, Douglas, Andrea Baronchelli, and Damon Centola. 2021. "Experimental Evidence for Scale-induced Category Convergence Across Populations." *Nature Communications* 12:1–7.

Guilbeault, Douglas, Joshua Becker, and Damon Centola. 2018. "Social Learning and Partisan Bias in the Interpretation of Climate Trends." *Proceedings of the National Academy of Sciences* 115:9714–9719.

Guilbeault, Douglas and Damon Centola. 2020. "Networked Collective Intelligence Improves Dissemination of Scientific Information Regarding Smoking Risks." *PloS one* 15:e0227813.

Hoever, Inga J, Daan Van Knippenberg, Wendy P Van Ginkel, and Harry G Barkema. 2012. "Fostering Team Creativity: Perspective Taking as Key to Unlocking Diversity's Potential." *Journal of Applied Psychology* 97:982.

Hogg, Michael A and John C Turner. 1987. "Intergroup Behaviour, Self-stereotyping and the Salience of Social Categories." *British Journal of Social Psychology* 26:325–340.

Hogg, Michael A, John C Turner, and Barbara Davidson. 1990. "Polarized Norms and Social Frames of Reference: A Test of the Self-Categorization Theory of Group Polarization." *Basic and Applied Social Psychology* 11:77–100.

Hong, Lu and Scott E Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101:16385–16389.

Hunzaker, MB Fallin. 2016. "Cultural Sentiments and Schema-consistency Bias in Information Transmission." *American Sociological Review* 81:1223–1250.

Hunzaker, MB Fallin and Lauren Valentino. 2019. "Mapping Cultural Schemas: From Theory to Method." *American Sociological Review* 84:950–981.

Leenders, Roger Th AJ. 2002. "Modeling Social Influence through Network Autocorrelation: Constructing the Weight Matrix." *Social Networks* 24:21–47.

Liu, Christopher C and Sameer B Srivastava. 2015. "Pulling Closer and Moving Apart: Interaction, Identity, and Influence in the US Senate, 1973 to 2009." *American Sociological Review* 80:192–217.

Liu, Ka-Yuet, Marissa King, and Peter S Bearman. 2010. "Social Influence and the Autism Epidemic." *American Journal of Sociology* 115:1387–1434.

Lix, Katharina, Amir Goldberg, Sameer Srivastava, and Melissa A Valentine. 2022. "Aligning Differences: Discursive Diversity and Team Performance." *Management Science* .

Macy, Michael W, James A Kitts, Andreas Flache, and Steve Benard. 2003. "Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure." In *Dynamic Social Network Modeling and Analysis*, edited by R. Breiger, K. Carley, and P. Pattison, pp. 162–173. Cornell University Working Paper Series.

Mancini, Anna, Antonio Desiderio, Riccardo Di Clemente, and Giulio Cimini. 2022. "Self-induced consensus of Reddit users to characterise the GameStop short squeeze." *Scientific reports* 12:13780.

Mark, Noah P. 2003. "Culture and Competition: Homophily and Distancing Explanations for Cultural Niches." *American Sociological Review* pp. 319–345.

Marsden, Peter V. 1981. "Introducing Influence Processes into a System of Collective Decisions." *American Journal of Sociology* 86:1203–1235.

Marsden, Peter V and Noah E Friedkin. 1993. "Network Studies of Social Influence." *Sociological Methods & Research* 22:127–151.

Mäs, Michael, Andreas Flache, and James A Kitts. 2014. "Cultural Integration and Differentiation in Groups and Organizations." In *Perspectives on Culture and Agent-based Simulations*, pp. 71–90. Springer.

Matzler, Kurt, Andreas Strobl, and Franz Bailom. 2016. "Leadership and the Wisdom of Crowds: How to Tap into the Collective Intelligence of an Organization." *Strategy & Leadership* .

McDaniel, Michael A and Deborah L Whetzel. 2005. "Situational Judgment Test Research: Informing the Debate on Practical Intelligence Theory." *Intelligence* 33:515–525.

Miles, Andrew, Raphaël Charron-Chénier, and Cyrus Schleifer. 2019. "Measuring Automatic Cognition: Advancing Dual-process Research in Sociology." *American Sociological Review* 84:308–333.

Mohr, John W. 1998. "Measuring Meaning Structures." *Annual Review of Sociology* 24:345–370.

Moyer, Daniel, Samuel Carson, Thayne Dye, Richard Carson, and David Goldbaum. 2015. "Determining the influence of Reddit posts on Wikipedia pageviews." In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 9, pp. 75–82.

Page, Scott. 2019. *The Diversity Bonus*. Princeton University Press.

Pelled, Lisa Hope, Kathleen M. Eisenhardt, and Katherine R. Xin. 1999. "Exploring the Black Box: An Analysis of Work Group Diversity, Conflict and Performance." *Administrative Science Quarterly* 44:1–28.

Peus, Claudia, Susanne Braun, and Dieter Frey. 2013. "Situation-Based Measurement of the Full Range of Leadership Model—Development and Validation of a Situational Judgment Test." *The Leadership Quarterly* 24:777–795.

Postmes, Tom, S Alexander Haslam, and Roderick I Swaab. 2005. "Social Influence in Small Groups: An Interactive Model of Social Identity Formation." *European Review of Social Psychology* 16:1–42.

Rivera, Mark T, Sara B Soderstrom, and Brian Uzzi. 2010. "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms." *Annual Review of Sociology* 36:91–115.

Rosenberg, Seymour, Carnot Nelson, and PS Vivekananthan. 1968. "A Multidimensional Approach to the Structure of Personality Impressions." *Journal of Personality and Social Psychology* 9:283.

Salganik, Matthew J, Peter Sheridan Dodds, and Duncan J Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311:854–856.

Srivastava, Sameer B and Mahzarin R Banaji. 2011. "Culture, Cognition, and Collaborative Networks in Organizations." *American Sociological Review* 76:207–233.

Surowiecki, James. 2005. *The Wisdom of Crowds*. Anchor.

Taylor, Sean J, Lev Muchnik, Madhav Kumar, and Sinan Aral. 2022. "Identity effects in social media." *Nature Human Behaviour* pp. 1–11.

Tenenbaum, Joshua B, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. 2011. "How to Grow a Mind: Statistics, Structure, and Abstraction." *Science* 331:1279–1285.

Turner, John C. 1991. *Social Influence.* Thomson Brooks/Cole Publishing Co.

Turner, John C, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. 1987. *Rediscovering the Social Group: A Self-Categorization Theory.* Basil Blackwell.

Turner, John C and Katherine J Reynolds. 2011. "Self-Categorization Theory." *Handbook of Theories in Social Psychology* 2:399–417.

Turner, John C, Margaret S Wetherell, and Michael A Hogg. 1989. "Referent Informational Influence and Group Polarization." *British Journal of Social Psychology* 28:135–147.

Wojcieszak, Magdalena. 2011. "Deliberation and Attitude Polarization." *Journal of Communication* 61:596–617.

Wood, Michael Lee, Dustin S Stoltz, Justin Van Ness, and Marshall A Taylor. 2018. "Schemas and Frames." *Sociological Theory* 36:244–261.

Wood, Wendy. 2000. "Attitude Change: Persuasion and Social Influence." *Annual Review of Psychology* 51.

Zhao, Xiaoquan, Andrew Strasser, Joseph N Cappella, Caryn Lerman, and Martin Fishbein. 2011. "A measure of perceived argument strength: Reliability and validity." *Communication methods and measures* 5:48–75.

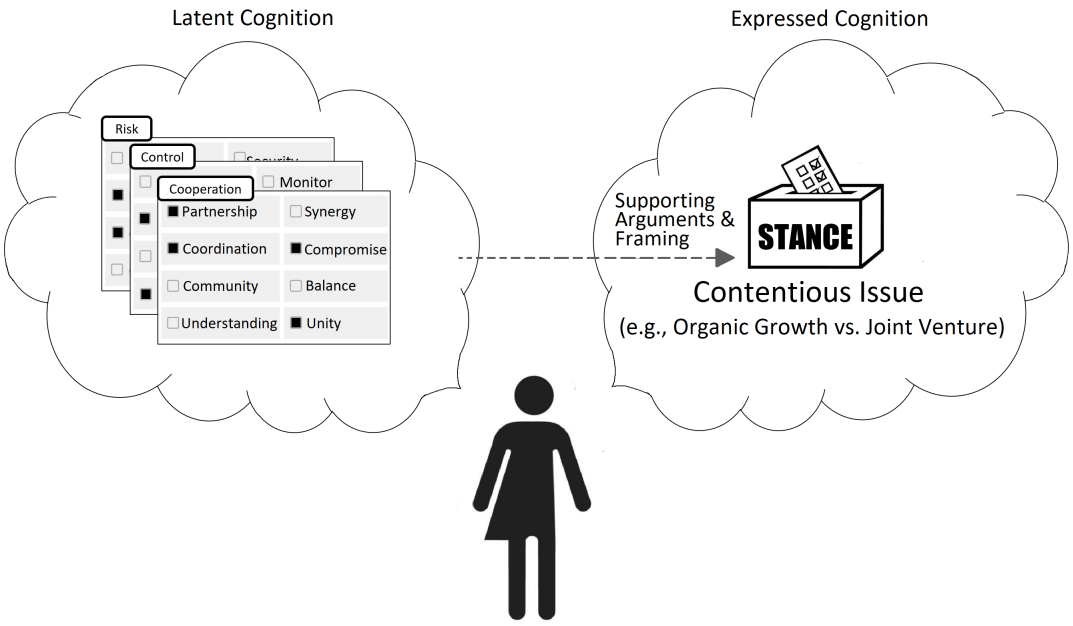

Figure 1: A stylized representation of the distinction between latent and expressed cognition. The left thought bubble highlights latent cognition in the form of one's schematic associations about focal concepts (here "risk," "control," and "cooperation,") that pertain to a substantive issue such as whether to pursue organic growth or growth through a joint venture. The right thought bubble shows expressed cognition in the form of one's stated stance on the issue and the ways in which one frames supporting arguments that are expressed to others. One's stance, supporting rationale, and the framing of arguments are all informed by one's schematic associations.

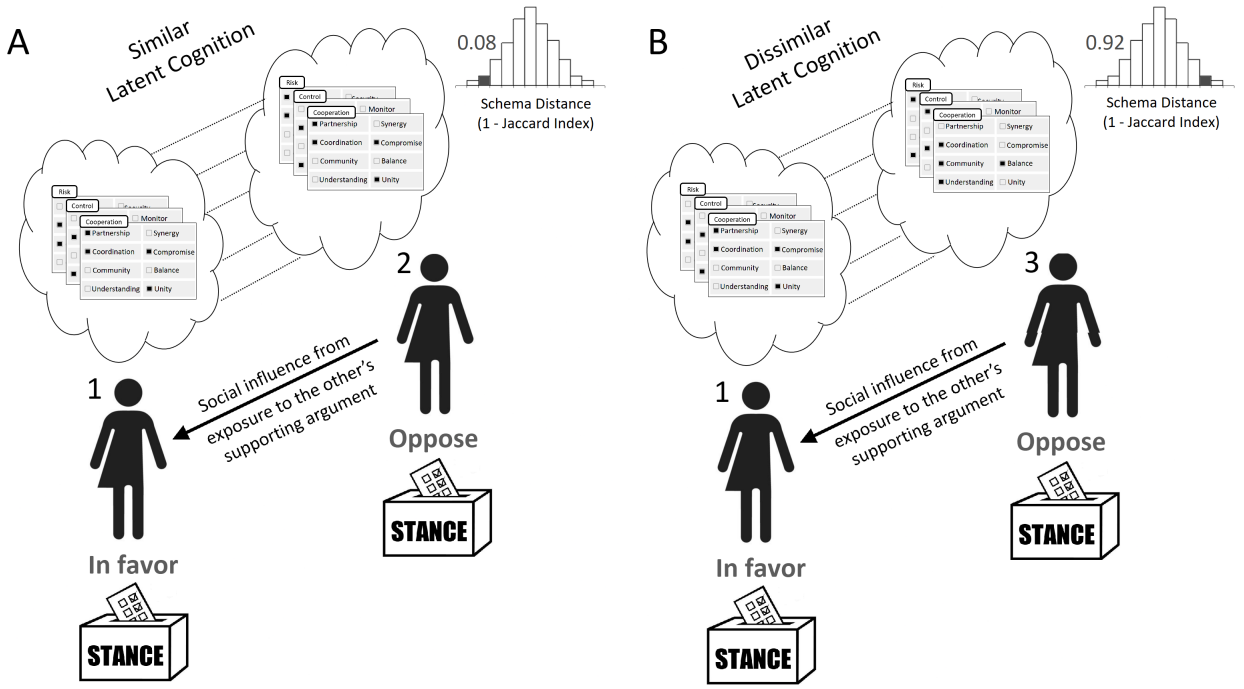

Figure 2: A stylized representation of the two empirical conditions of interest: Panel A, which depicts social influence between disagreeing peers who have similar schemas of "risk," "control," and "cooperation;" and Panel B, which depicts social influence between disagreeing peers who have dissimilar schemas of "risk," "control," and "cooperation."

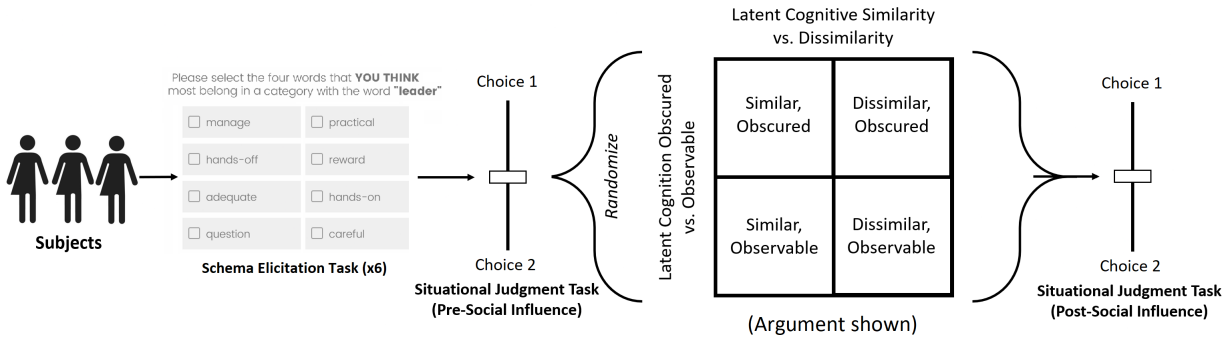

Figure 3: Overview of the main study procedure.

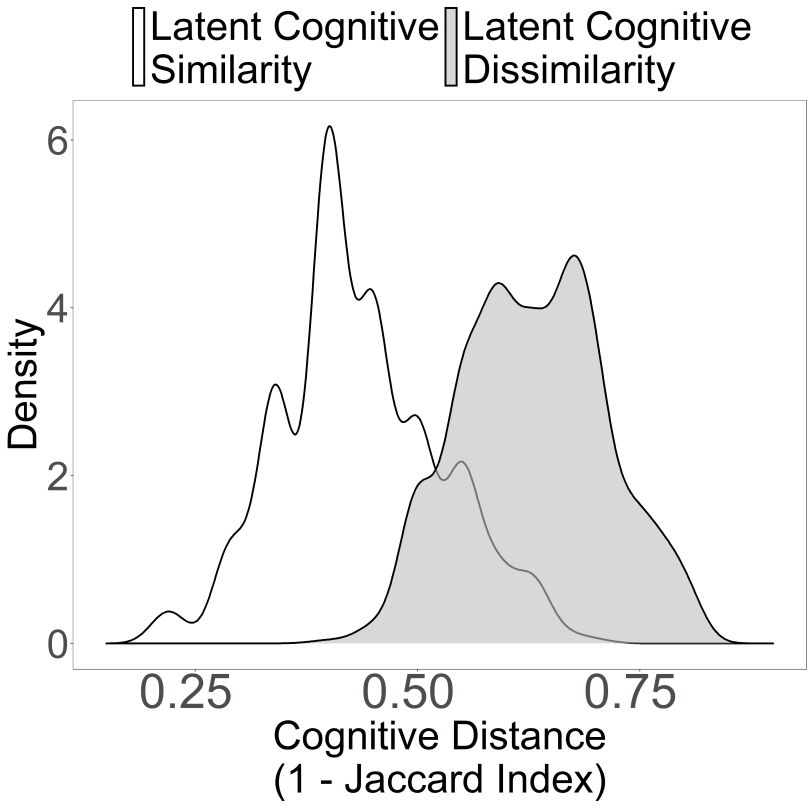

Figure 4: The distribution of cognitive distance between a main study participant and the Reference Sample participant who produced the argument that was presented to that participant. The distribution in white corresponds to the latent cognitive similarity conditions, whereas the one in grey corresponds to the latent cognitive dissimilarity conditions. (The observed versus observable conditions are collapsed.)

Table 1: Schema Elicitation Task: Words Used to Assess Leadership Schemas

Social perception category (Fiske et al., 2006)	**Selected associated concepts**	**Set Valence**
Good-intellectual; Good-social	Important, Reliable, Cautious, Practical, Scientific, Serious, Artistic, Reserved	Positive Set
Bad-intellectual; Bad-social	Foolish, Impulsive, Superficial, Dishonest, Irresponsible, Boring, Vain, Wasteful	Negative Set
Good-intellectual; Bad-social	Stern, Cold, Moody, Shrewd, Critical, Dominating, Humorless, Persistent	Mixed Set 1
Bad-intellectual; Good-social	Modest, Tolerant, Helpful, Sentimental, Warm, Sociable, Happy, Humorous	Mixed Set 2
Neutral	Hands-on, Hands-off, Question, Reward, Careful, Responsible, Manage, Adequate	Neutral Set 1
Neutral	Delegate, Discipline, Monitor, Network, Social, Control, Recognizable, Influence	Neutral Set 2

Table 2: Logistic Regression of Opinion Change When Latent Cognition is Obscured (Hypothesis 1)

	Coefficients	Odds Ratio
Latent Cognitive Dissimilarity	0.92***	2.52***
	(0.26)	
Argument Fixed Effects	Included	Included
Constant	-0.99	0.36
	(0.53)	
N	506	
R^2	0.15	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: A logistic regression predicting whether participants in the conditions in which latent cognition was obscured changed their mind about the SJT depending on whether they were exposed to an argument from a cognitively dissimilar versus similar individual from the Reference sample. The model includes argument fixed effects.

Table 3: OLS Regression of the Magnitude of Opinion Change When Latent Cognition is Obscured (Hypothesis 1)

	Coefficients	CI
Latent Cognitive Dissimilarity	6.72*	1.04 - 12.40
	(2.89)	
Argument Fixed Effects	Included	
Constant	21.44***	9.03 - 33.86
	(6.32)	
N	506	
R^2	0.05	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. An OLS regression predicting the extent to which participants in the conditions where latent cognition was obscured updated their stance in the direction of the opposing view depending on whether they were exposed to an argument from a cognitively dissimilar versus similar individual from the Reference Sample. The model includes argument fixed effects..

Table 4: Logistic Regression of Opinion Change on the Full Sample (Hypothesis 2)

	Coefficients	Odds Ratio
Latent Cognitive Dissimilarity	0.73**	2.08**
	(0.23)	
Latent Cognition Observable	0.56**	1.75**
	(0.12)	
Latent Cognitive Dissimilarity x Latent Cognition Observable	-0.98***	0.37***
	(0.29)	
Argument Fixed Effects	Included	
Constant	-1.08**	0.34**
	(0.36)	
N	1000	
R^2	0.08	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: A logistic regression predicting whether participants changed their mind about the SJT by experimental condition. The model includes argument fixed effects.

Table 5: OLS Regression of the Magnitude of Opinion Change on the Full Sample (Hypothesis 2)

	Model 1	Model 2
Strength of Initial Stance		0.42***
		(0.06)
Latent Cognitive Dissimilarity	4.97	4.68
	(2.64)	(2.59)
Latent Cognition Observable	4.05	3.78
	(2.40)	(2.35)
Latent Cognitive Dissimilarity x Latent Cognition Observable	-8.54**	-8.29*
	(3.35)	(3.28)
Argument Fixed Effects	Included	Included
Constant	19.41***	5.04
	(4.36)	(4.81)
N	1000	1000
R^2	0.02	0.06

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. An OLS model predicting the extent to which participants updated their stance toward the opposing view by experiment condition. Argument fixed effects are included. Model 1 does not control for the strength of participants' initial stance, while Model 2 includes this control.

Appendix A: Selecting The Situational Judgment Test

Procedure

To identify topics on which participants would disagree, we considered a range of existing situational judgment tests (SJTs)—vignettes of managerial dilemmas that have been shown to elicit a variety of interpretations about how one should act (Peus, Braun, and Frey 2013; McDaniel and Whetzel 2005). Most SJTs are associated with many categorical choices, which complicates the process of identifying opposing views given that some choices may be complementary or orthogonal rather than in direct opposition to one another. To address this issue, we pretested responses to a range of SJTs with self-identified organizational leaders on the Prolific platform. From this broader set, we identified an SJT (described in greater detail below) that elicited two dominant and largely opposing choices from participants (McDaniel and Whetzel 2005; Peus, Braun, and Frey 2013). We then adapted this SJT to give participants a binary choice between the two most popular choices that emerged in our pretesting. In addition, rather than simply presenting participants with a categorical choice, we instead presented them with a slider scale representing how strongly they favored a given choice. Moving the slider to the far left indicated strong support for one choice, while moving it to the right right indicated strong support for the other. Moving the slider to the middle indicated ambivalence between the two. The scale ranged from -50 (the strongest possible position in favor of Option 1) to 50 (the strongest possible position in favor of Option 2). Next, we pretested this modified version of the SJT with a separate Prolific sample. Our modified design was successful in eliciting polarized responses among participants from our pretest, as well as participants from our main study (see Figure A1).

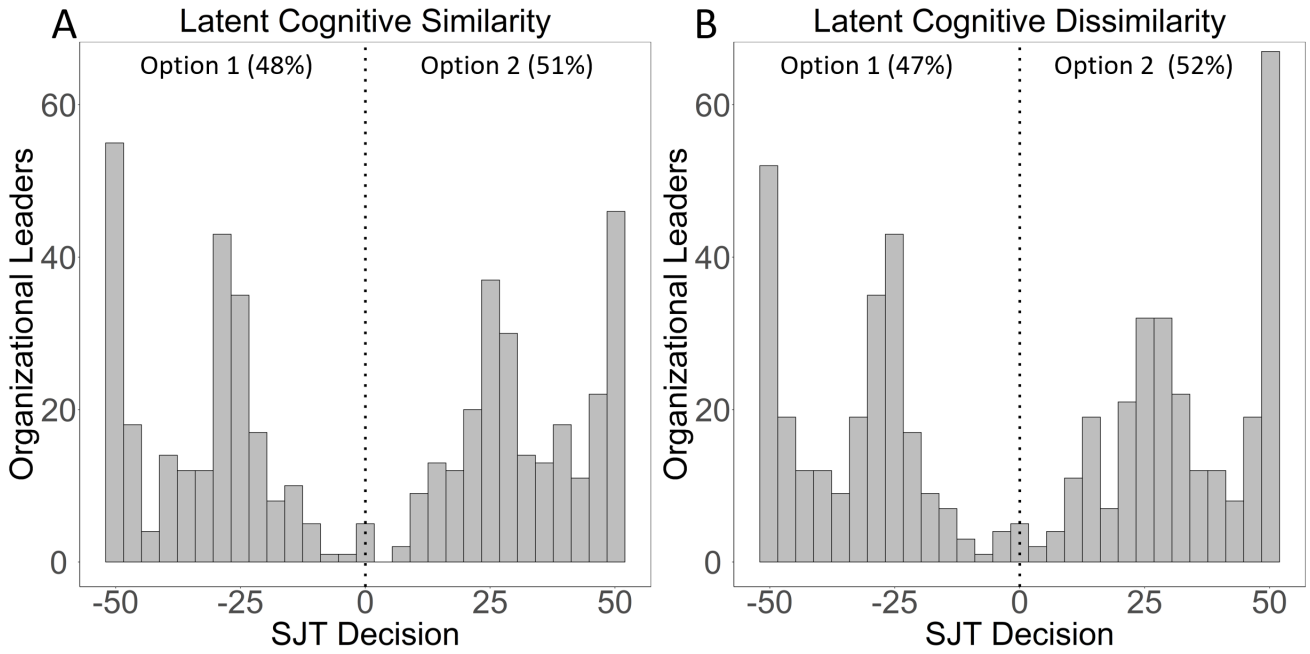

Figure A1. The baseline distribution of responses the selected SJT across the latent cognitive similarity and latent cognitive dissimilarity conditions. Data includes responses from both the latent cognition obscured and the latent cognition observable conditions.

Figure A1 further illustrates that our randomization procedure worked as intended. There was no significant difference in the initial decisions made by participants in the latent cognitive similarity versus latent cognitive dissimilarity conditions when latent cognition was obscured (p = 0.89, Wilcoxon Rank Sum Test); likewise, there was no significant difference in the initial decisions made by participants in the latent cognitive similarity versus latent cognitive dissimilarity conditions when latent cognition was observable (p = 0.50, Wilcoxon Rank Sum Test). Furthermore, there was no significant difference in the initial decisions made by participants in the latent cognition observable versus latent cognition obscured condition when participants were assigned to the latent cognitive similarity condition (p = 0.58, Wilcoxon Rank Sum Test), nor was there a significant difference in the initial decisions made by participants in the latent cognition observable versus latent cognition obscured conditions when participants were assigned to the latent cognitive dissimilarity condition (p = 0.99, Wilcoxon Rank Sum Test). These analyses indicate that any differences in the outcomes across conditions are unlikely to be driven by chance differences in the initial distribution of participants' responses to the SJT across conditions.

Importantly, because our selected SJT reliably produced a bimodal distribution of responses, we could readily match participants in a given study with arguments developed by participants from our Reference Sample who made the opposite choice about the same SJT, allowing us to see if this exposure led focal participants to update their choice when given an opportunity to revisit the SJT. We describe the SJT in greater detail below.

BAT Validation

To validate that our BAT task effectively detects cognitively relevant differences in how participants evaluate arguments relating to the SJT, we conducted a supplemental analysis using our Reference sample in which we examined the correlation between the cognitive dissimilarity between argument raters using our word association task and the dissimilarity of their ratings for each argument. This analysis included pairwise comparisons between the ratings of all 200 participants across all of the arguments they rated, yielding 337,108 data points. We evaluated this relationship using an OLS regression that predicts the Euclidean distance between participants' argument ratings as a function of their cognitive dissimilarity, while including argument fixed effects and clustering standard errors by the specific pair of participants being compared. Memo Figure 1 depicts the results.

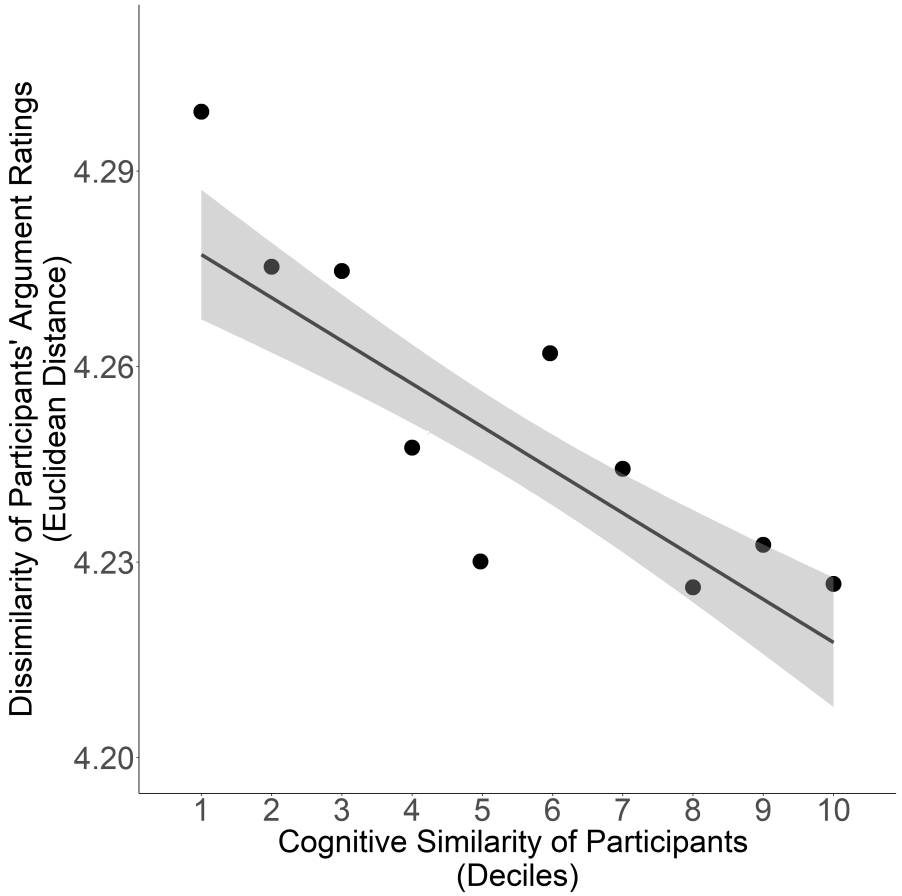

Figure A2. The correlation between the cognitive similarity between participants in our Reference sample and the degree of dissimilarity in their argument ratings. Cognitive similarity is measured by the Jaccard index between participants' responses on the schema elicitation task; the trend line displayed reflects the linear trend across all individual data points, and the data points reflect the average results aggregated into deciles of cognitive similarity. Dissimilarity in participants' argument ratings is identified by calculating the euclidean distance between participants' ratings of an argument across all features: novelty, interestingness, informativeness, convincingness, logicality, and emotionality.

We find that cognitive similarity is significantly and negatively correlated with an increase in the Euclidean distance between participants' argument ratings (β= -0.33, SE=0.02, t=-12.09, p<0.0001), while controlling for fixed effects by argument, SJT, SJT choice, and participant. The general linear trend characterizing this correlation is displayed in figure A2. These same results hold with equally high levels of statistical significance if dissimilarity in participants' argument ratings is measured alternatively using cosine similarity, or more coarsely using hamming distance,

which measures how many rating dimensions along which participants provide the same responses. This analysis helps to validate that the word association task deployed in this study does, indeed, identify meaningful cognitive differences between participants.

The Selected Situational Judgment Test

Participants were presented with the following vignette: "A member of your department has been employed for three years, but his original project expired after two years. Thus, as the team leader, you assigned him to a new job. However, in the last months you noticed that the employee regularly shows up in the office quite late and does not work longer than absolutely necessary. In his current project, the employee achieves very little progress.

As the team leader, what would you do in this situation?

To indicate your choice, move the slider toward the option you prefer. Indicate the strength of your preference by moving the slider closer to the option you prefer. The closer you move the slider to -50, the more you prefer Option 1. The closer you move the slider to 50, the more you prefer Option 2. You must indicate a preference to proceed."

Choices:

Option 1. Ask the employee in a personal conversation why he only makes little progress in his project at present and offer to support him with the further project design by providing specific feedback.

Option 2. Point out to the employee how important his full commitment is to you. Openly communicate your criticism of the employee's current work ethics, but emphasize that you highly valued his performance on former projects.

Examples of Opposing Arguments Used in Main Study

Table A1 includes examples of the arguments used as social stimuli in our studies. Table A1 shows two arguments in support of each of the decision options for the SJT.

SJT (Sample Arguments)
Option 1
-As a leader, it is imperative to maintain an open and healthy relationship with those who work under you. In this context, it is clear that the employee has some issues regarding motivation/work ethic and requires adjustment in order to continue effectively. By doing this in an open and supportive manner, the employee should feel capable of expressing the factors that have impacted their performance. This in conjunction with the offer of support will provide them with the opportunity to improve and succeed. Though there is no guarantee their behavior will change, such a course will at least afford the opportunity, which is what's best both for the employee and the company. Hence a personal conversation is needed.
-If this employee only recently started coming in late and underperforming, there must be a reason for it. It's important to talk to the employee to understand why there is a disconnect between them and their assignment. Perhaps the assignment is not a right fit and the employee and company could benefit from the employee being moved to a different project. If the issue is personal and unrelated to the job, it's important to know that as well, since, in that instance, moving the employee to a different assignment may not change anything. A company's success depends on the success of their employees. While it's important to reinforce the necessity of high quality work, that work should not come at the expense of the employees' wellbeing if the company wants to succeed in the long run.
Option 2
-As his manager, a personal conversation would be inappropriate. It could show favoritism. Pointing out how he is failing to meet expectations helps him understand what needs to change. Expressing that his earlier accomplishment did not go unnoticed could prevent him from feeling like he is not a part of the team.
-I chose this option since I think that it is important to voice your criticisms/constructive feedback to the employee so that they know what exactly they need to change to become an effective employee. I do not think that it is a good idea to sugar coat the situation because the employee may never know what went wrong in their performance. However, I also think that it is incredibly important to point out the things that the employee is doing well or has done well in the past to increase motivation. Otherwise, they might feel like they are being reprimanded and being looked down upon. The goal is to motivate not to tear down.

Table A1. A sample of the arguments for each decision option for the SJT to which participants were exposed.

Appendix B: Validation of Argument Randomization

Our randomization algorithm was designed to ensure that each argument was equally likely to appear in the cognitively similar or dissimilar condition for each SJT. To achieve this design, we used the schema elicitation task data associated with participants from our Reference Sample and

simulated the probability of each argument being shown across each condition. We found that if our algorithm was designed to randomly select one of the top two most similar arguments in the latent cognitive similarity condition or one of the top two most dissimilar arguments in the latent cognitive dissimilarity condition, then each argument would be approximately equally likely to appear in each condition. Here we confirm that this randomization design succeeded. For each argument, we measured the fraction of participants in each condition who were exposed to it. We then used paired statistical comparisons to compare the probability of an argument appearing in one condition versus the other. We find that there was no significant difference in the probability of an argument appearing in the latent cognitive similarity versus latent cognitive dissimilarity condition when latent cognition was obscured ($p = 0.93$, Wilcoxon Signed-Rank Test). Similarly, there was no significant difference in the probability of an argument appearing in the latent cognitive similarity and latent cognitive dissimilarity condition when latent cognition was observable ($p = 0.93$, Wilcoxon Signed-Rank Test).

Appendix C: Supplemental Study to Test Mechanisms

Supplemental Study: Sample and Procedure

The goal of this supplemental study was to test whether our proposed mechanism—argument novelty—mediates the relationship between cognitive dissimilarity and social influence. We tested this pre-registered hypothesis using a different sample than used in our main study to avoid potential confounds that could have been introduced by using the same sample. For example, if we had asked participants in our main study to rate opposing arguments after already having made the decision to change their mind, it might have induced them to rate the argument more favorably than they would have on their own. Similarly, if we had asked our main study participants to rate opposing arguments before determining whether to change their mind, their choice to stick to their original position or deviate from it might have been influenced by the desire to be internally consistent.

The sample included unique set of 200 self-identified organizational leaders from Prolific. They had the following demographic profile: their mean age was 29.8 ($\sigma = 9.8$) years. 39.8%

identified as Female. 5.3% identified as Black or African American; 4.0% as Multiracial; 3.2% as Asian; 14.5% as Hispanic/Latino; and 63.7% as White (with the remaining responding that they "prefer not to say").[12]

The procedure for this supplemental study followed the design of our main study but with three notable exceptions: (1) rather than giving participants a chance to update their decision about a given SJT, we instead exposed participants to *all* of the opposing arguments and asked participants to rate each argument on how novel, interesting, informative, convincing, logical, and emotional they found it (using a 5-point Likert scale); (2) after evaluating the arguments, participants were asked how likely they would be to change their decision (without actually being given the opportunity to do so); and (3) to establish the robustness of our mechanism, we also included a second SJT that had been pre-tested and validated in the same manner as the SJT used in our main study. With respect to (1), argument features were selected based on the validated set of features and Likert measures outlined in Zhao et al.'s (2011) review paper, which summarizes prior methods for how to assess argument strength. The arguments were presented one-by-one and in random order. Figure C1 provides a schematic representation of this study design.

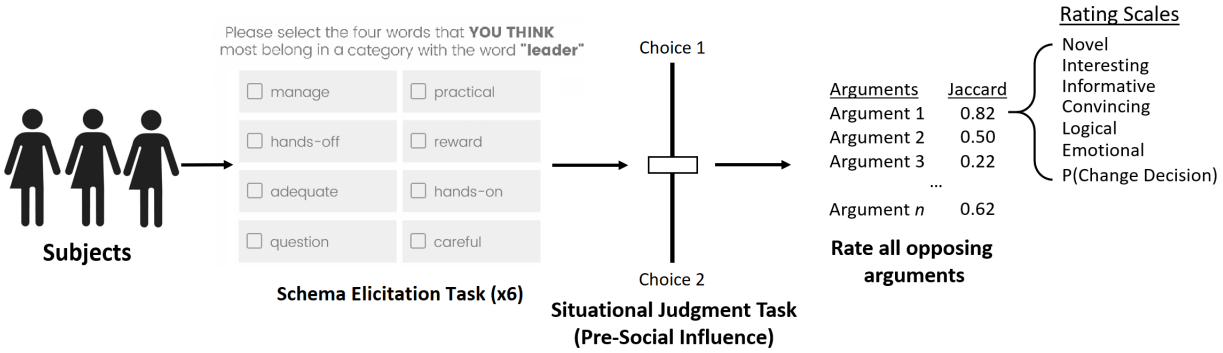

Figure C1. Schematic overview of the procedure for the supplemental mechanism study.

[12] 17.7% listed their highest level of education as High School Diploma (or lower); 10.4% as some College; 47.5% as Bachelor's Degree; 9.6% as Master's Degree; and 14.5% as MD or PhD. The participants reported the following number of years at work: less than 1 year 4.8%; 1-5 years 45%; 5-10 years 27.7%; 10-15 years 10.4%; 15-20 years 5.6%; and 20+ years 6.5%. Their sectors of employment were as follows: Legal 3.4%; Arts 3.8%; Agriculture & Food 4.2%; Medicine 5.4%; Finance 7.9%; Government 6.4%; Architecture & Construction 8.6%; Education 6.6%; Hospitality & Tourism 8.6%; Business and Administration 8.5%; and Information Technology (IT) 21.3%. The remaining participants did not specify an industry sector.

Supplemental Study: Results

Consistent with expectations, we find that the relationship between cognitive similarity and perceived novelty is negative and significant ($p < 0.001$, r = -0.06). We also find that cognitive similarity is significantly and negatively associated with participants' self-reported willingness to change their mind ($p < 0.001$, r = -0.09), lending further support for Hypothesis 1. Finally, as anticipated, we show a strong positive relationship between perceived argument novelty and a participant's self-reported willingness to change her recommended decision ($p < 0.001$, $r = 0.35$).

To formally test our proposed novelty mechanism, we conducted a mediation analysis. Figure C2 depicts these results. While cognitive dissimilarity to an argument's source is positively and significantly related to participants' assessments of whether they would change their decision, its predictive power is significantly decreased when including the mediator of perceived argument novelty. Argument novelty is significantly and positively related to cognitive similarity and significantly and positively related to participants' assessments of whether they would change their decision (Sobel test; $p < 0.05$). A structural equation model, shown in Figure C2, finds a significant indirect effect of cognitive dissimilarity on a participant's self-reported willingness to change decision via the mediator of perceived argument novelty. The bootstrapped unstandardized indirect effect (aggregated across 1000 bootstrapped samples) was statistically significant at .14 ($p = 0.001$). The proportion of the effect mediated (0.19) was statistically significant ($p = 0.001$). Thus, we find support for our proposed mechanism of argument novelty.

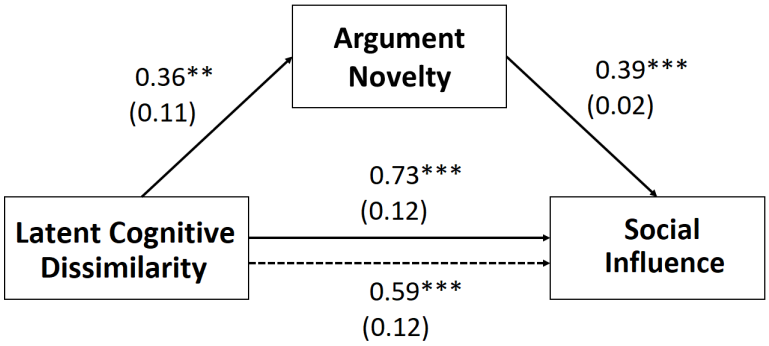

Figure C2. This analysis tests whether the novelty rating of an argument mediates the effect of latent cognitive dissimilarity on social influence. Model includes fixed effects for argument and SJT.

Table C1 shows that these results are robust to controlling for other argument rating dimensions, as well as argument and SJT fixed effects, and with standard errors clustered by participant. (Recall that this supplemental study included assessments of opposing arguments for two different SJTs and that a given participant rated all opposing arguments for each SJT.) Specifically, the degree of cognitive similarity between the focal participant and the author of the argument continued to significantly and negatively predict willingness to change opinion (β = -0.36, SE=0.13, $p < 0.001$). Furthermore, Table C1 shows that argument novelty continues to significantly predict a greater willingness of participants to change their opinion, controlling for a range of argument features, as well as the cognitive similarity between the focal participant and the author of the argument ($\beta = 0.14$, SE=0.04, $p < 0.001$). It is worth noting that ratings of how convincing and informative an argument was perceived to be were also significantly predictive of opinion change. This finding aligns with prior work, which shows that the ratings of positive attributes of arguments tend to be correlated (Zhao, Strasser, Cappella, Lerman, and Fishbein 2011).

Table C1: OLS Regression of the Participants' Willingness to Change Opinion as a Function of Argument Features

	Coefficients
Cognitive Similarity (Jaccard)	-0.36**
	(0.13)
Novel	0.14***
	(0.04)
Interesting	0.04
	(0.04)
Convincing	0.39***
	(0.03)
Informative	0.10**
	(0.03)
Logical	0.05
	(0.03)
Emotional	0.03
	(0.03)
SJT Fixed Effects	Included
Argument Fixed Effects	Included
Constant	-0.16
	(0.42)
N	3603
R^2	0.34

Standard errors in parentheses

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Note: An OLS regression predicting the extent to which participants reported a willingness to change their decision as a function of argument features, while controlling for argument and SJT fixed effects and clustering standard errors at the participant level.