

**Network Structures of Collective Intelligence:
The Contingent Benefits of Group Discussion**

Joshua Becker

Kellogg School of Management, Northwestern University

Northwestern Institute on Complex Systems, Northwestern University

Correspondence to: joshua.becker@kellogg.northwestern.edu

WORKING PAPER

Last updated: August 31st, 2019

ABSTRACT

Research on belief formation has produced contradictory findings on whether and when communication between group members will improve the accuracy of estimations such as economic forecasts, medical diagnoses, and job candidate assessments. While some evidence suggests that carefully mediated processes (i.e., the “Delphi method”) produce more accurate beliefs than unstructured discussion, others argue that unstructured discussion outperforms mediated processes. Still others argue that independent individuals produce the most accurate beliefs. This paper shows how network theories of belief formation can resolve these inconsistencies. Emergent network structures of influence—even in groups with no apparent structure, such as committees—interact with the pre-discussion belief distribution to moderate the effect of communication on belief formation. As a result, communication sometimes increases and sometimes decreases the accuracy of the average belief in a group. The effects differ for mediated processes and unstructured communication, such that the relative benefit of each communication format depends on both group dynamics as well as the statistical properties of pre-interaction beliefs. These results resolve contradictions in previous research and offer practical recommendations for teams and organizations.

****NOTE**** *This draft is a work in progress. This manuscript presents a re-analysis of past experiments by myself and others. I am currently collecting data to replicate the empirical findings presented here. I welcome feedback at all levels: joshua.becker@kellogg.northwestern.edu*

Replication data and code available at:

<https://github.com/joshua-a-becker/emergent-network-structure>

Network Structures of Collective Intelligence: The Contingent Benefits of Group Discussion

Forming accurate beliefs is critical to making good decisions. The fundamental paradox in group decision-making is that interaction between group members is an integral part of the decision-making process, and yet social influence dynamics pose the risk of undermining decision quality through processes such as herding (Da & Huang, 2019; Lorenz, Rauhut, Schweitzer, & Helbing, 2011) and groupthink (Janis, 1982). As a result, a great deal of research has sought to understand why group decision-making sometimes fails (Janis, 1982) and what practices, if any, can allow interacting groups to produce accurate beliefs (Green, Armstrong, & Graefe, 2007). The processes studied by this research are frequently described as “collective intelligence” processes (Bonabeau, 2009; Malone & Bernstein, 2015; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010), which refers to the tendency for groups to produce macro-level behaviors that cannot be reduced to a sum of individual level characteristics (DeDeo, 2014; Krafft et al., 2016).

Broadly speaking, decision-making is a complex behavior that involves many distinct processes including idea generation, idea evaluation, and idea selection (Davis, 1973). Each of these individual processes has been the subject of extensive research. For example, idea generation has been studied through research on brainstorming (Stroebe, Nijstad, & Rietzschel, 2010) and is subject to unique concerns such as the exploration-exploitation tradeoff (Lazer & Friedman, 2007; March, 1991). The present paper focuses specifically on the formation of accurate beliefs, which is important to idea evaluation—e.g., assessing the expected payoff of an investment. As a basic process, belief formation plays a fundamental role in many kinds of decision-making.

Belief formation has been subject to academic interest in a wide range of domains. One of the most commonly studied processes of belief formation is forecasting, which is a critical component of strategic decision-making and can include generating sales projections (Cowgill & Zitzewitz, 2015), predicting the success of an advertising campaign (Hartnett, Kennedy, Sharp, & Greenacre, 2016), or estimating future macro-economic indicators (Jansen, Jin, & de Winter, 2016). Forecasting is also a common task in crowdsourcing, in which organizations take advantage of the potential benefits collective intelligence without the risks of groupthink by aggregating the beliefs of a large number of independent individuals. For example, amateur forecasters contribute their predictions about publicly traded companies on the web-based platform Estimize, which

produces earnings and revenue forecasts that regularly beat wall-street analysts (Da & Huang, 2019; Drogen & Jha, 2013). Beyond forecasting, belief formation plays a role in more quotidian decisions, highlighting a wide range of contexts in which collective intelligence can be valuable. One critical process in most organizations is hiring, which can be quite costly if not done well (Abbasi & Hollman, 2000). Hiring decisions involve evaluating candidates on such factors as performance, cultural fit, and the likelihood of remaining employed (or achieving tenure). Another decision suited to the benefits of collective intelligence is medical decision-making, which requires estimating the probability of a particular pathology diagnosis (Pauker, 1982).

While these decisions certainly can (and often are) made by individuals, teams play an increasingly important role in decisions that were once dominated by individuals, such as medical decision-making (Christensen, Abbott, Gretchen, & Chapman, 2003) and even hedge fund management (Massa, Reuter, Zitzewitz, & Stanford, 2006). Importantly, the rise of teams offers enormous potential benefit—indeed, the core motivation of collective intelligence theory is that groups can produce better (e.g., more accurate) decisions than individuals (Krafft et al., 2016; Sunstein, 2006; Woolley et al., 2010). For example, studies on crowdsourcing have found that the aggregated beliefs of multiple physicians can outperform even the most skilled individuals (Kurvers et al., 2016), and that aggregate beliefs of amateur investors can outperform expert analysts (Chen, De, Hu, & Hwang, 2014; Drogen & Jha, 2013).

Although decision-making can be a complex process that involves the synthesis of many different sources and types of information, a common ingredient in many decisions is one or more critical numeric values. For example, a marketing team may need to estimate the sales volume for a potential new product; the success of a hiring decision depends on the probability that a candidate will remain in the position; strategy decisions are influenced by forecasts about a wide range of economic indicators. These decision tasks all benefit from accurate estimates of a numeric value. One characteristic of numeric estimation tasks is that they can be described with very general statistical models (Becker, Brackbill, & Centola, 2017; Csaszar & Eggers, 2013; Hogarth, 1978; Hong & Page, 2008; Kao et al., 2018; Lamberson & Page, 2012). This paper will therefore focus very broadly on theoretical and empirical research examining numeric estimation tasks with the expectation that these insights will inform a wide range of decision-making processes in teams and organizations.

The fundamental question that this paper addresses is the following: when do communication practices such as committee discussion improve the accuracy of group beliefs, and when are the benefits of collective intelligence most effectively harnessed by aggregating the contributions of independent individuals?

TO GROUP OR NOT TO GROUP?

One key argument for the benefits of group decision-making over individual decision-making is the expectation that that no single individual is likely to provide an exactly correct numeric estimate. Both statistical principles (Hogarth, 1978; Page, 2007) and empirical data (Atanasov et al., 2017; Galton, 1907; Ven & Delbecq, 1974) suggest that when group members pool their beliefs, errors cancel out in aggregate, allowing groups to generate more accurate beliefs (and thus produce better decisions) than individuals acting in isolation. In fact, the “crowd beats average” principle (Page, 2007)—commonly known as the “variance bias tradeoff” in statistical estimation theory—provides a mathematical guarantee that the average belief in a group will be more accurate (in terms of squared error) than a randomly selected individual. The intuitive interpretation of these principles is that when people have diverse information and perspectives, errors “cancel out” and produce accurate group beliefs.

However, groups in practice often fail to take advantage of all the diversity of information held by their individual members. Example of how group decision-making can fail are found in case studies describing “Groupthink” dynamics (Janis, 1982) in organizational decisions. In these studies, norms favoring group cohesion produced conformity pressures which prevented individuals from sharing information that contradicted existing group beliefs (Janis, 1982; Surowiecki, 2004). Importantly, social influence can harm belief accuracy even in the absence of normative pressure. Herding models have shown that strictly rational individuals who would produce accurate decisions independently can produce inaccurate decisions when their beliefs are formed sequentially (Banerjee, 1992). This process has been shown empirically to impact financial crowdsourcing. For example, early versions of Estimize allowed contributors to see the estimates of other forecasters before providing their own forecast, reducing overall accuracy; when they revised their design to require independent contributions, accuracy increased (Da & Huang, 2019).

One especially popular strategy for taking advantage of a group’s collected intelligence while minimizing the risks associated with social influence is the “wisdom of crowds” approach,

also known as “crowdsourcing.” This approach is simple and yet powerful: people simply collect the independent beliefs of a large number of individuals, thus harnessing the statistical benefits of group beliefs without exposing the decision to the risks of herding or groupthink. This approach is reflected in the revised Estimize platform (Drogen & Jha, 2013) as well as other crowdsourcing efforts by firms such as Google (Cowgill & Zitzewitz, 2015), Ford (Cowgill & Zitzewitz, 2015), Dell (Bayus, 2013), and Best Buy (Dvorak, 2008). More advanced statistical methods can improve upon simple aggregation with weighting methods that, for example, try to identify people who are consistently more accurate (Budescu & Chen, 2014).

However, while the wisdom of crowds strategy can be extremely effective, it is limited in two respects. First, it is frequently infeasible or undesirable to prevent contributors from interacting, as when a decision must be made by people who naturally interact in the course of their work. Second, even if social interaction is potentially avoidable, strategies that harness only the collected intelligence of independent individuals fail to take advantage of the potential *benefits* of group interaction. As some researchers argue, properly structured communication processes can allow groups to produce even more accurate beliefs than could be obtained by independent individuals (Becker et al., 2017; Dalkey & Helmer, 1963; Gustafson, Shukla, Delbecq, & Walster, 1973; Ven & Delbecq, 1974).

The Delphi Method

Motivated by the potential benefits of group interaction, some researchers have studied the “Delphi” method as a process that can theoretically harness the collective intelligence¹ of interacting groups while mitigating the risks associated with social influence dynamics (Dalkey & Helmer, 1963). Although the exact method varies widely across implementations (Green et al., 2007; Humphrey-Murto & de Wit, 2019), the Delphi method generally involves allowing decision-makers to share limited information through a facilitator. Some versions allow participants to write down motivating arguments, while others are limited to the exchange of numeric estimates. In laboratory studies utilizing this method, researchers have studied topics ranging from the trivial, such as guessing average height/weight ratios (Gustafson et al., 1973) to the more serious, such as optimizing military campaigns (Dalkey & Helmer, 1963). However, most experiments utilizing the Delphi do not explicitly measure accuracy (e.g. by using questions with no known true answer)

¹ Though the phrase “collective intelligence” has been introduced relatively recently, the spirit is the same.

as their focus is on other factors such as consensus formation (Dalkey & Helmer, 1963) and the perceived ability to contribute diverse perspectives without repercussion (Ven & Delbecq, 1974).

Critically, however, those experiments which did actually measure accuracy yielded inconsistent and at times contradictory results (Hastie, 1986; Rowe & Wright, 1999). In the experiments, people are typically asked to complete numeric estimates before and after some form of group information exchange, allowing the researchers to assess the effect of communication on belief accuracy. Gustafson (1973) found that open discussion produced the most accurate estimates while the Delphi method produced the least accurate estimates, with independent individuals falling in the middle. Findings by Gough (1975; as cited by Hastie, 1986) also found that open discussion outperformed the Delphi method, but found that both methods outperformed independent judgements. Still others (Larache & Moinpour, 1983) found that the Delphi method outperformed both independent individuals and unstructured discussion.

There are several possible explanations for these inconsistencies. One possible explanation that must be considered is statistical sampling error. In addition to the studies cited above, several researchers (Fisher, 1981; Boje & Murnighan, 1982; Snizek 1990) found no statistically significant difference between different types of groups. Coupled with the fact that the analytic strategies varied considerably across studies—leaving the possibility that researchers were choosing those metrics which were most favorable to their argument—these null results suggest the potential for Type I error in those papers that did find differences. However, if this is indeed the case, then researchers are as uninformed as ever about whether group discussion is helpful, harmful, or completely neutral when it comes to group belief accuracy. The present paper argues that the (potential) benefits of social interaction are not a statistical fluke, reducing uncertainty by conducting an analysis across four independently conducted experiments (with a replication in progress as of this writing).

Another possible explanation emerges from the observation that each of these experiments simply studied different estimation tasks, including both different categories (i.e., trivia vs. economic estimation) and different questions within categories. This explanation is supported by one report (Brockhoff, 1975) that the Delphi method outperformed unstructured discussion for some tasks, while unstructured discussion outperformed Delphi for other tasks. However, Brockhoff offers no explanation for this divergence, which is a strictly empirical observation. This explanation therefore simply begs the theoretical question: why does unstructured discussion

sometimes produce the most accurate beliefs, Delphi interaction at other times, and independent individuals still other times?

In order to explain why task characteristics matter, the present paper argues that the network structure of social influence—even in groups with no apparent network structure, such as committees where everyone communicates directly—interacts with the statistical properties of belief formation. This analysis will show that unstructured communication is sometimes better, and sometimes worse, than limited numeric information exchange. This analysis will also show that unstructured discussion is sometimes better, and sometimes worse, than independent individuals. However, limited numeric information exchange appears to consistently produce more accurate beliefs than independent individuals.

THE NETWORK DYNAMICS OF BELIEF FORMATION

This paper proposes the idea that network theory can explain why social interaction sometimes improves group accuracy and sometimes reduces accuracy, and in so doing resolve the apparently contradictory findings from prior research. Although this paper analyzes discussions in which there is no apparent network structure, as in a committee all sitting together around a table, the theory is motivated by the behavior of highly centralized networks, in which one subgroup or individual exerts a disproportionately large influence on belief formation. My key theoretical insight is based on the expectation that even when groups lack apparent network structure, e.g. when everyone can communicate with everyone, some people can nonetheless be more influential than others.

In networks dominated by a single individual, the effect of communication depends entirely on the position of that central individual's belief relative to the group's mean belief. Notably, the effect of communication in centralized networks does not depend simply on whether the central individual is more (or less) accurate than the group as a whole. As Becker et al. (2017) show empirically, groups can be pulled in the direction of the central individual without adopting that central individual's belief wholesale. What is important therefore is not how *accurate* the central individual is, but whether their belief falls on the same side of the mean as the true numeric value being estimated. If the central individual falls on the truth side—even if they are wildly inaccurate—the group can be pulled toward the truth and thereby become more accurate by virtue of being pulled toward the central individual. If the central individual falls on the opposite side of

the mean from truth, the group will be drawn toward that central individual's beliefs and become less accurate.

This dynamic means that network centralization will sometimes improve the accuracy of a group's mean belief, and sometimes harm the accuracy of a group's mean belief. Almaatouq (2019) studies this dynamic in simulation to show that the effect of social influence in centralized networks—whether it is helpful or harmful—depends on the initial belief distribution. In networks dominated by a single central individual (as studied by Becker et al., 2017) the probability that the group will improve is equal to the proportion of individuals on the truth side of the mean—i.e., the probability that the central node will fall on the truth side of the mean. The present paper argues that, if otherwise unstructured groups display emergent centralization as a result of endogenous variation in influence, then the effect of social influence will also vary according to initial belief distribution. In other words, even unstructured networks will generate emergent centralization causing them to behave like Almaatouq's (2019) and Becker et al.'s (2017) explicitly centralized networks.

The present analysis adopts the DeGroot (1974) model of social influence to formally express the notion of emergent belief centralization. DeGroot studies a model in which each individual begins with an independent belief, then observes the belief of some peers, and then revises their belief toward a weighted average of their own initial belief and the beliefs of their peers. DeGroot shows that if this process is repeated indefinitely, groups asymptotically converge on a single shared consensus belief². DeGroot further shows that the consensus belief is a weighted sum of initial beliefs, where each individual's contribution to group beliefs (i.e., their weight) is equal to their eigenvector centrality. Eigenvector centrality is a common measurement of network centrality that counts not only how influential a person is on their immediate neighborhood, but how influential their neighbors are, and so forth. (Intuitively: a person with only one friend can be highly influential if their one friend is the President of the United States.) In previous studies examining the wisdom of networked crowds, researchers assumed networks were “binary,” which means that two people either do or do not communicate (Becker et al., 2017; Noriega-Campero et al., 2018; Pan, Altshuler, & Pentland, 2012). However, DeGroot's model allows for non-binary “weighted” networks, in which a network tie reflects not only whether person A observes person

² Consensus requires certain conditions, but they are very broad: in an undirected network/graph, consensus will be reached as long as there is one giant component.

B, but also how much weight person A places on the belief of person B. Person A is then assumed to update their belief by taking a weighted average of all the beliefs of all their peers.

Resolving the Contradiction

Importantly, the mechanisms allowing influence to vary depend on communication format. Unstructured communication offers several mechanisms through which an individual can increase their influence including being persuasive, asserting status, or just being more talkative. This variation in influence leads to variation in eigenvector centrality, and thus variation in each person's contribution to group beliefs. In other words, unstructured communication can produce groups with a centralized emergent network, despite the fact that the explicit (binary) communication network—who communicates with whom—is perfectly decentralized. In contrast, when groups communicate via the Delphi method, i.e. anonymous exchange of numeric information, there is no opportunity to be more influential on other people. Delphi networks therefore remain relatively decentralized. As a result, comparing Delphi communication to unstructured discussion is like comparing centralized to decentralized networks.

Putting these theoretical pieces together, network theory can explain the variable effects of communication and resolve the apparent contradictions of prior research. As Almaatouq (2019) points out, centralized networks are sometimes better (more likely to produce increased accuracy) and sometimes worse than decentralized networks. This variation occurs because centralized networks are more likely to improve than decentralized networks when central individuals are likely to hold a belief in the direction of truth, but less likely to improve when the central individual is likely to hold a belief in the direction away from truth; decentralized networks are relatively stable in comparison. The variable benefits of centralization are the key factor in explaining why Delphi communication sometimes produced more accurate beliefs than unstructured discussion and sometimes produced less accurate beliefs. Because unstructured discussion groups behave like centralized networks, while Delphi groups behave like decentralized networks, unstructured discussion will sometimes perform better than Delphi and sometimes perform worse, and the relative benefit depends on the statistical properties of pre-interaction beliefs.

Hypotheses

If this theoretical argument is correct, one should expect the effect of unstructured discussion to depend on initial belief distribution, just as in explicitly centralized networks. This insight represents the main hypothesis of this paper:

***Hypothesis 1a:** The probability that unstructured discussion improves the accuracy of the mean belief in a group increases as a function of the proportion of individuals on the truth side of the mean. When the majority is on the truth side, the group is more likely to become more accurate; when the majority is away from truth, the group is more likely to become less accurate.*

However, because Delphi networks will remain relatively decentralized and therefore demonstrate relatively stable performance (due to the lack of opportunities for people to differentiate themselves) there will be no effect of initial belief distribution. This insight leads to the secondary hypothesis of the paper, and the resolution of prior contradictory findings:

***Hypothesis 1b:** When the majority of a group is on the truth side of the mean, unstructured discussion will be more likely to improve the accuracy of the mean belief than numeric exchange (Delphi exchange). When the majority of the group is on the opposite from the truth, unstructured discussion will be less likely to improve the accuracy of the mean belief than numeric exchange.*

In this paper, testing for the effect of initial belief distribution is the primary method for identifying the effects of emergent network structure.

While the main evidence for the role of centralization is the effect of initial belief distribution, this analysis also directly measures emergent centralization. Network centralization can be quantified as the extent to which one or a small number of individuals possess disproportionate influence—i.e., are highly central (Freeman, 1978). In other words, centralization is a measurement of how evenly distributed are centrality scores, as when Gini coefficient is used

to measure the relative equality of an income distribution. By theory, centralization is expected to be the reason that initial belief distribution matters, leading to the following hypothesis:

***Hypothesis 2:** More centralized discussions will show a stronger effect of initial belief distribution than less centralized discussions. Centralization will therefore increase the probability of improvement for groups where the majority is on the side of truth, and decrease the probability of improvement when the majority is on the opposite side. When groups display no centralization, initial belief distribution will have no effect.*

METHODS

To illustrate the explanatory power of this theory, this paper presents a reanalysis of data from previous experiments that measured belief accuracy in groups before and after interaction. This reanalysis tests for the effects of emergent network centralization. This analysis uses four previously published studies (Becker et al., 2017; Becker, Porter, & Centola, 2019; Gürçay, Mellers, & Baron, 2015; Lorenz et al., 2011). These datasets were all made publicly available through the initial publications.

Detailed methods can be found in the initial publications, and each study follows a similar procedure. Subjects were asked to complete estimation tasks (e.g. visual estimation, trivia questions, and political facts) before and after a exchanging information via a computer mediated communication process. An example of visual estimation task is an image of a jar of gumballs where subjects are asked to estimate how many gumballs are in the jar. An example of a trivia question is estimating the length of the border of Switzerland. An example of a political fact is asking subjects to estimate the number of undocumented immigrants living in the United States.

In three of the studies (Lorenz et al, 2011; Becker et al, 2017; Becker et al, 2019) subjects only exchanged numeric estimates. These studies therefore represent a method equivalent to a digitally mediated version of the Delphi method. Lorenz et al. (2011) allowed 5 rounds of revision (1 independent estimate and 4 socially influenced estimates) while Becker et al. (2017; 2019) allowed 3 rounds of revision (1 independent estimate and 2 socially influenced estimates). In contrast, Gürçay et al allowed subjects to engage in continuous, unstructured discussion via a computer chat interface, so that subjects provided only two answers, a pre-discussion and a post-

discussion estimate (1 independent estimate and 1 socially influenced estimate). The data from Gürçay et al. is missing chat transcripts from 5 groups, and therefore those trials are omitted from analyses where the chat transcripts are necessary (i.e., measuring emergent centralization).

Measurements

In each experiment, groups completed multiple estimation tasks, and I define a single experimental trial as one group completing one estimation task. For each trial, I measure the accuracy of the average initial belief and the average final belief. I measure accuracy as the distance to the truth—i.e., the absolute value of the arithmetic difference between the truth and the mean belief. My main outcome metric is a simply binary outcome: did the mean belief become more accurate?

This paper examines two main predictor variables. In order to reflect the effect of the initial (pre-influence) belief distribution, I measure the initial proportion of individual estimates on the truth side of the mean, henceforth ϕ . Recall that for networks dominated by a single individual, ϕ indicates the probability that social influence will improve the accuracy of the mean belief when beliefs are uncorrelated with centrality. This measurement is defined as follows: Each individual belief falls either above the mean, or below the mean. Similarly, the truth (correct answer) falls either above the mean or below the mean. I define ϕ as the proportion of individual beliefs that are on the same side of the mean as the truth.

The second predictor variable is network centralization. According to the theoretical argument given above, the reason that ϕ matters is because ϕ represents the probability that any given individual will pull the group towards the true answer. Under the hypothesis that groups display emergent centralization even in unstructured networks, this value therefore represents the probability that a central individual—who exerts disproportionately large influence on group beliefs—will pull the group toward the answer. In previous research, network centralization was manipulated by directly controlling who could observe whom. However, the present paper is concerned with emergent centralization as a result of endogenous communication dynamics. I therefore measure emergent centralization based on the dynamic behavior of a group.

While many factors can impact individual influence, the most observable factor is talkativeness. That is, I assume that people who talk more (i.e., send more chat messages) are more influential on group beliefs. I therefore use the number of chat messages sent by each person

as a proxy for their network centrality. Following standard network metrics (Badham, 2013) I measure network centralization as the Gini coefficient on individual centrality (i.e., talkativeness) scores. This can be interpreted in the same way as the Gini coefficient on, e.g., income distributions: when one person does all the talking, they possess all the influence, and the network is highly centralized. When everybody talks the same amount, influence is distributed evenly, and the network is decentralized. Because there are other ways that people can be influential (e.g., through persuasive arguments or by asserting expertise) this measurement is necessarily imperfect, as it does not reflect all possible ways in which a person can be influential. Therefore, if this analysis does show a relationship between emergent centralization and group belief formation, it serves as a conservative test of my theoretical argument.

Control Variables

Previous work has shown that in decentralized networks, group beliefs will become more accurate when there is a correlation between individual accuracy and adjustment, such that people with greater accuracy (i.e., smaller) make smaller adjustments (Becker et al., 2017). Because this value is strongly predictive of improvement in empirical data, I include this metric as a control variable in all analyses.

Following previous work (Adjodah, Chong, Leng, Krafft, & Pentland, 2017; Becker et al., 2019; Kao et al., 2018; Madirolas & de Polavieja, 2015) I measure adjustment to social influence, α , as the proportion of the distance closed between a person's initial independent belief and their observed³ social information:

$$\alpha_i = \frac{x_{i,pre} - x_{i,post}}{x_{i,pre} - \bar{x}}$$

where $x_{i,pre}$ and $x_{i,post}$ indicate the pre- and post-influence beliefs for individual i , and \bar{x} indicates the mean of observed beliefs (i.e., that person's peers). This value is then reverse coded, so that it represents stubbornness, i.e. centrality:

$$stubbornness = 1 - \alpha_i$$

When this value is 0, an individual completely adopts social information, and when this value is 1, that person does not revise their answer at all. In theoretical models this value falls only between

³ For the data provided by Becker et al. (2019), the exact social information is not recorded in the replication dataset, and I approximate this value as the mean of the group belief, which is equivalent in expectation.

0 and 1, but in empirical data, this value can fall outside of that range if a subject moves away from or overshoots the average belief of their peers.

Using this metric, I can control for the correlation between stubbornness and error: a negative correlation indicates that stubborn people have lower error (higher accuracy) and that groups are expected to improve (Becker et al., 2017). A positive correlation indicates that stubborn people have higher error (lower accuracy) and that group accuracy is expected to decrease.

Identification

My main statistical test is a logistic regression predicting the likelihood that a group will improve based on ϕ , communication format, and emergent centralization. Because each group (i.e., set of individuals) in each experiment completed multiple estimation tasks, I use hierarchical logistic regression with outcomes clustered by group. This test includes a control for error/stubbornness correlation. To test the underlying mechanistic role of emergent centralization, I also conduct a hierarchical logistic regression predicting the probability of improvement as a function of the Gini coefficient on talkativeness. This test includes a control for the error/stubbornness correlation as well as the total number of chat messages sent.

RESULTS

This analysis begins by testing the main effect predicted by the theory of emergent network structure, that the effect of social influence is moderated by initial belief distribution even in apparently unstructured networks. As with explicitly centralized networks, I expect the emergent centralization of unstructured discussion to draw group beliefs in the direction of the majority opinion. To test this “majority effect,” this analysis reports a simple outcome for each estimation task by each group: was the mean belief closer to the truth after discussion?

I first report results for unstructured discussion. As shown in Figure 1, I find that the probability of improvement in unstructured discussion is substantially explained by ϕ , or the proportion of individuals in the direction of truth. Since each group completed multiple estimation tasks, I statistically test this effect using hierarchical logistic regression with outcomes nested in groups, finding that the probability of improvement in unstructured discussion significantly increases with ϕ ($P < 0.001$). Critically, this analysis shows that variation in initial belief distribution can *completely reverse* the effects of unstructured discussion: when $\phi > 1/2$ (majority

in the direction of truth) a majority of discussion become more accurate ($P < 0.001$), and when $\phi < 1/2$ (majority opposite truth) a majority of discussion groups become less accurate ($P < 0.01$).

In contrast with unstructured discussion, the moderating effect of initial belief distribution is substantially weaker for numeric information exchange. This outcome is consistent with the expectation that the effect of centralization in numeric exchange will be substantially weaker than the effects of centralization in unstructured discussion. While we do see a slight positive trend in logistic regression, it is only marginally significant ($P < 0.11$).

--- insert Table 1 about here ---

One empirical challenge in identifying the effect of ϕ is that as ϕ increases, the correlation between accuracy and adjustment also increases, creating a confounding collinearity. Table 1 shows the logistic regression results for this basic test, with and without a control for the error/stubbornness correlation. This analysis shows that once the error/stubbornness correlation is controlled, even the marginal effect of ϕ in numeric exchange disappears ($P > 0.90$). While this null result should not necessarily be interpreted to mean that ϕ has no impact on numeric exchange, the statistical weakness of this effect—given the large sample size of $N=299$ trials—nonetheless indicates that the effect of ϕ is small enough that it does not make a practical difference for numeric exchange. In contrast, the effect of ϕ remains strong in unstructured discussion even after controlling for the error/stubbornness correlation ($P < 0.01$). These findings suggest that unstructured discussion displays a emergent centralization, while numeric exchange does not.

--- insert Figure 1 about here ---

The critical implication of this finding, as can be seen visually in Figure 1, is that unstructured discussion for some belief distributions is more likely to increase accuracy than numeric exchange; while for other belief distributions, numeric exchange is more likely to increase belief accuracy than unstructured discussion. Comparing only those cases where a majority is on the truth side, i.e. $\phi > 1/2$, there is a nominal advantage to unstructured discussion: 67% of unstructured discussions improved as compared with 63% improvement in numeric exchange, though the two groups are statistically indistinguishable. However, the difference gets larger as ϕ

increases. For those trials in which $\phi > 0.75$, this analysis shows that unstructured discussion improved belief accuracy in 84% of trials, as compared with a 69% improvement rate in numeric exchange ($P < 0.03$, logistic regression). Thus for those belief distributions in which a majority of people are in the direction of truth, unstructured discussion is more likely to improve the accuracy of the mean belief than numeric exchange. In contrast, numeric exchange shows a stark advantage when a majority of people are on the opposite side of truth: numeric exchange improved belief accuracy in 58% of trials where $\phi < 1/2$, compared with only 33% improvement for unstructured discussion ($P < 0.01$, logistic regression).

Emergent Centralization in Unstructured Discussion

The remainder of this analysis tests for the underlying mechanism hypothesized to produce the effect of belief distribution: emergent centralization. For a given discussion (i.e., a single estimation task) by a single group, I measure emergent talkativeness centralization as the Gini coefficient on the number of chat messages sent by each person. If emergent centralization is the reason that initial belief distribution matters, then my theoretical model predicts that initial belief distribution will have a stronger effect for more centralized networks. To test this effect, I conduct a hierarchical linear regression on outcomes for the trials in which subjects engaged in unstructured discussion. This regression includes an interaction term between ϕ and centralization as well as controls for the error/stubbornness correlation and the total number of chat messages sent. Although this regression does not produce a statistically significant interaction term, the overall effect is in a direction consistent with the hypothesized mechanism. Therefore, while I cannot reject the null hypothesis that Gini centralization has no interaction with ϕ , the results are nonetheless consistent with predictions based on the underlying theoretical model. (Moreover, it is worth noting that as of this writing, the replication dataset shows nearly identical results. However, I have not finished collecting the full amount of the pre-registered sample size and therefore do not report those results here.)

To visually represent the interaction between emergent network centralization and initial belief distribution, Figure 2 shows the model-fitted probability of improvement as a function of emergent centralization (Gini coefficient) for four levels of ϕ . This interaction shows that, when $\phi < 1/2$ —i.e., when the majority is away from truth—the probability of accuracy improvement decreases as the group becomes more centralized. This is consistent with the expectation that

centralization amplifies the effects of the initial belief distribution. In contrast, when $\phi > 1/2$ —i.e., when the majority is toward truth—centralization increases the probability of improvement, again consistent with the expectation that centralization amplifies the initial belief distribution. Another way of interpreting this figure is to interaction is the observation that the gap between trials at different levels of ϕ , or the effect of initial belief distribution, becomes larger as groups become more centralized.

--- insert Figure 2 about here ---

In understanding the effect of network centralization, an important question to ask is: who are the central individuals? Previous research has shown a correlation between self-reported confidence and accuracy (Heath & Gonzalez, 1995; Madirolas & de Polavieja, 2015) and we therefore examine individual accuracy to determine whether people who are more talkative are also more accurate. If there were a similar correlation between talkativeness and belief accuracy (or more precisely, the likelihood of falling on the truth side of the mean) one would expect emergent centralization to reliably improve group belief accuracy. I found, however, neither a numerically meaningful nor statistically significant correlation between accuracy and talkativeness.

--- insert Table 2 about here ---

Table 2 shows the coefficients on a hierarchical linear regression (to control for within-group correlation) comparing the relationship between percent-error and the number of chat messages sent for each individual. In order to allow the coefficient to be interpreted as a correlation coefficient (ranging -1 to 1) both the predictor variables and outcome variables are mean-centered and normalized by the standard deviation. In order to compare error sizes across trials with different very different scales, I take the logarithm of the percent error, setting minimum error at 1% since the logarithm is not defined for zero. The relationship between accuracy and talkativeness centrality is effectively zero. (Consistent with previous research, I found a strong and consistent correlation between error and response to social information. As shown in Table 2, the strength of this correlation is nearly identical for both Delphi exchange and unstructured

discussion.) Importantly, the belief of the most talkative individual predicts whether or not the group will improve: the mean belief in a group became more accurate in 63% of trials where the most talkative person held a belief on in the direction of truth, but only 45% of trials where the most talkative person is on the opposite side of truth ($P < 0.01$, proportion test).

DISCUSSION

One important goal of this paper was to resolve the apparent contradictions in previous research comparing numeric exchange (“Delphi method”) with unstructured discussion. This analysis shows that despite surface-level contradictions, each of the claims from previous research—numeric exchange is better than Delphi method, Delphi method is better than numeric exchange, independent estimates are optimal—is consistent within a broader pattern of results showing the effect of the initial belief distribution. The analysis here showed that unstructured discussion outperforms numeric exchange for certain initial belief distributions, and Delphi method outperforms unstructured discussion for other belief distributions. Although it is not possible in hindsight to be certain that this explanation is the correct explanation for prior results, an interaction between task characteristics (belief distribution) and emergent network structure is nonetheless a *sufficient* explanation.

Empirically, the analysis presented here showed that the variable benefits of unstructured discussion are due to a “majority effect” wherein the mean belief is drawn in the direction of the majority. Thus, the benefits (or risks) of unstructured discussion depend on initial belief distribution. While this result may seem intuitive—the idea that majorities dominate group discussion is hardly surprising—it is an advancement from prior theoretical arguments, which predicted that networks in which everyone is equally connected (decentralized) would simply converge toward the mean of independent beliefs (Becker et al., 2019; Hélène Landemore & Page, 2015). Moreover, this effect appears to be unique to unstructured discussion. If the majority opinion does have any effect when groups interact via numeric exchange, it is small enough that it was difficult to detect in the present analysis, suggesting that the majority effect is not in fact a universal characteristic of information exchange in any practical sense, despite its intuitive appeal.

The primary theoretical goal of this analysis was to show that social network theory plays an important role in explaining belief formation even when groups don’t have an explicitly defined network structure. Because network centralization here is an endogenous feature of group

interaction, and not experimentally controlled, it is difficult to make a strict causal argument about the role of emergent centralization in unstructured discussion. However, the clear empirical relationship between talkativeness centralization and the majority effect—groups with greater emergent centralization showed a stronger majority effect—provides strong supporting evidence for the theoretical explanation that emergent network structure determines group belief formation. This finding is also consistent with previous research by Woolley et al (2010) showing that groups who displayed more equal turn-taking achieved higher performance ratings on a variety of tasks. Although Wooley et al. offered no theoretical explanation for this empirical finding, the argument made here is consistent with more general research showing that network centralization can undermine group performance in a variety of domains (Becker, 2019; Becker et al., 2017; Rulke & Galaskiewicz, 2000).

To Group or Not to Group? Implications for practice

As Almaatouq (2019) points out—and as this analysis also shows—network centralization can sometimes be beneficial, if the majority is in the right direction. Here, though, the analysis shows that centralization is *rarely* beneficial. Unstructured discussion only showed a clear advantage for cases where $\phi > 0.75$. When $0.75 > \phi > 0.5$, the two formats were roughly equivalent. And when $\phi < 0.5$ —when the majority was away from truth—numeric exchange showed a clear advantage. As a result, if a decision-maker in practice is uncertain what statistical properties their task will demonstrate, numeric exchange (Delphi method) is a safer choice. This analysis therefore offers a simple recommendation: limit communication to numeric exchange. When numeric exchange is not feasible, it may therefore be desirable to avoid discussion altogether and simply aggregate independent estimates. However, discussion is often a natural or even necessary component of group decision-making; in such an event, it may at least be possible to aggregate independent beliefs and compare them with post-discussion beliefs to make an informed decision.

Notably, because the real risk of communication is associated with unequal participation, one solution is simply to encourage participants to all contribute equally! In the long run, such facilitated discussion—falling somewhere between unstructured discussion and strictly mediated numeric exchange—may offer the greatest benefits. However, as a potential intervention, the efficacy of facilitated discussion may depend on which context-specific problems the Delphi method is invoked to solve. One problem that the Delphi method is intended to solve is the effect

of status—anonymity ensures that nobody’s beliefs are given priority. In a context where status is a problem, facilitated conversation may not be sufficient, but might be combined with anonymity (e.g. using a digital platform) to ensure equal participation and equal contribution to collective beliefs formation. Ultimately, the goal as demonstrated by statistical models of the wisdom of crowds is to ensure that the group’s final belief fully reflects all the diversity of constituent members. This analysis shows that unstructured discussion is inherently antithetical to that goal due to emergent centralization; however, emergent centralization is not the only risk.

One limitation of this analysis is the relatively anemic version of the Delphi method used in the experiments analyzed here. In these experiments, the only information subjects could observe was the numeric belief of other subjects. While this simple implementation is consistent with the experimental research forming the basis for prior (contradictory) claims about the accuracy benefits of the Delphi method (e.g. Gustafson et al., 1973; Larreche & Moinpour, 1983), some applications of the Delphi method allow participants to share not only numbers, but details and arguments. However, there are no data available (to my knowledge) that explicitly measures the accuracy benefits of argument-sharing. While the present analysis therefore highlights several key considerations for decision-making in practice, additional empirical evidence is needed to understand how decisions in practice fit into the general statistical and network model analyzed in this paper.

Conclusion

While this re-analysis can help resolve contradictions in previous research, the nature of the laboratory experiments analyzed here preclude broad-sweeping recommendations about practice. However, two key principles stand out: know your group, and know your task. The principle “know your group” refers to the fact that two key group factors moderate the relative benefits of discussion: the error/stubbornness correlation (do the accurate people know who they are?) and emergent centralization (do some people dominate conversations?). The principle “know your task” refers to the fact that the relative effect of discussion also depends on task characteristics—the statistical properties of pre-discussion belief distributions. While no general principle is yet clear to indicate which decision tasks will have what properties, this variation suggests that groups should be able to calibrate best practices that meet their particular needs. However, the potential to calibrate strategies to both groups and tasks depends on several as-yet untested hypotheses.

Nonetheless, the general statistical model presented here—showing how previously contradictory results are in fact fully consistent when considered in a broader perspective—provides a path forward for understanding how collective intelligence dynamics can operate (or fail) to generate accurate beliefs, estimates, and forecasts.

REFERENCES

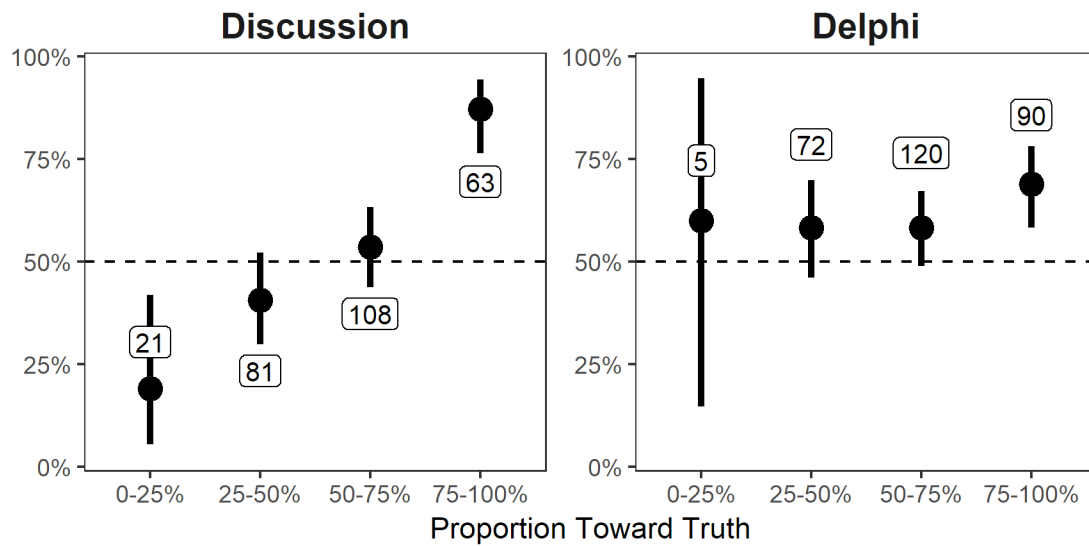
- Abbasi, S. M., & Hollman, K. W. (2000). Turnover: The real bottom line. *Public Personnel Management*, 29(3), 333–342.
- Adjodah, D., Chong, S. K., Leng, Y., Krafft, P., & Pentland, A. (2017). Large-Scale Experiment on the Importance of Social Learning and Unimodality in the Wisdom of the Crowd. *ArXiv Preprint ArXiv:1712.10284*.
- Almaatouq, A. (2019). *A chapter from the dissertation of Abdullah Almaatouq which is not yet published or publicly available*.
- Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., ... Mellers, B. (2017). Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls. *Management Science*, 63(3), 691–706. <https://doi.org/10.1287/mnsc.2015.2374>
- Badham, J. M. (2013). Commentary: Measuring the shape of degree distributions. *Network Science*, 1(2), 213–225.
- Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797–817.
- Bayus, B. L. (2013). Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science*, 59(1), 226–244.
- Becker, J. (2019). The Network Dynamics of Social and Technological Conventions. *Draft Manuscript*.
- Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences*, 114(26), E5070–E5076.
- Becker, J., Porter, E., & Centola, D. (2019). The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences*, 201817195.
- Bonabeau, E. (2009). Decisions 2.0: The power of collective intelligence. *MIT Sloan Management Review*, 50(2), 45.
- Brockhoff, K. (1975). The Performance of Forecasting Groups in Computer Dialogue and Face-to-face Discussion. In H. Linstone & M. Turnoff (Eds.), *The Delphi Method. Techniques and Applications* (Addison-Wesley). London.
- Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science*, 61(2), 267–280.
- Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, 27(5), 1367–1403.

- Christensen, C., Abbott, A. S., Gretchen, B., & Chapman, F. A. (2003). Team Medical Decision Making. In *Decision making in health care: Theory, psychology, and applications* (pp. 267–285).
- Cowgill, B., & Zitzewitz, E. (2015). Corporate prediction markets: Evidence from google, ford, and firm x. *The Review of Economic Studies*, rdv014.
- Csaszar, F. A., & Eggers, J. P. (2013). Organizational Decision Making: An Information Aggregation View. *Management Science*, 59(10), 2257–2277.
- Da, Z., & Huang, X. (2019). Harnessing the Wisdom of Crowds. *Management Science*, *Forthcoming*.
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458–467.
- Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review*, 80(2), 97–125.
- DeDeo, S. (2014). Group minds and the case of Wikipedia. *Human Computation*, 1 (1), 2014. *Human Computation*, 1.
- DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121.
- Drogen, L. A., & Jha, V. (2013). *Generating Abnormal Returns Using Crowdsourced Earnings Forecasts from Estimote*.
- Dvorak, P. (2008, September 6). Best Buy Taps 'Prediction Market'—WSJ. *Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/SB122152452811139909>
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215–239.
- Galton, F. (1907). Vox populi (The wisdom of crowds). *Nature*, 75, 450–51.
- Gough, R. (1975). The effect of group format on aggregate subjective probability distributions. *Utility, Probability, and Human Decision Making*.
- Green, K. C., Armstrong, J. S., & Graefe, A. (2007). Methods to elicit forecasts from groups: Delphi and prediction markets compared. *Foresight: The International Journal of Applied Forecasting*, (8), 17–20.
- Gürçay, B., Mellers, B. A., & Baron, J. (2015). The power of social influence on estimation accuracy. *Journal of Behavioral Decision Making*, 28(3), 250–261.
- Gustafson, D. H., Shukla, R. K., Delbecq, A., & Walster, G. W. (1973). A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organizational Behavior and Human Performance*, 9(2), 280–291.
- Hartnett, N., Kennedy, R., Sharp, B., & Greenacre, L. (2016). Marketers' Intuitions about the Sales Effectiveness of Advertisements. *Journal of Marketing Behavior*, 2(2–3), 177–194.
- Hastie, R. (1986). Experimental evidence on group accuracy. *Decision Research*, 2, 129–157.
- Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes*, 61(3), 305–326.

- Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, 21(1), 40–46.
- Hong, L., & Page, S. E. (2008). Some microfoundations of collective wisdom. In Helene Landemore & J. Elster (Eds.), *Collective Wisdom* (pp. 56–71). Cambridge University Press.
- Humphrey-Murto, S., & de Wit, M. (2019). The Delphi method—More research please. *Journal of Clinical Epidemiology*, 106, 136–139.
- Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (Vol. 349). Houghton Mifflin Boston.
- Jansen, W. J., Jin, X., & de Winter, J. M. (2016). Forecasting and nowcasting real GDP: Comparing statistical models and subjective forecasts. *International Journal of Forecasting*, 32(2), 411–436.
- Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., ... Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of the Royal Society, Interface*, 15(141). <https://doi.org/10.1098/rsif.2018.0130>
- Krafft, P. M., Zheng, J., Pan, W., Della Penna, N., Altshuler, Y., Shmueli, E., ... Pentland, A. (2016). Human collective intelligence as distributed Bayesian inference. *ArXiv Preprint ArXiv:1608.01987*.
- Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., ... Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, 201601827.
- Lamberson, P. J., & Page, S. E. (2012). Optimal forecasting groups. *Management Science*, 58(4), 805–810.
- Landemore, Hélène, & Page, S. E. (2015). Deliberation and disagreement: Problem solving, prediction, and positive dissensus. *Politics, Philosophy & Economics*, 14(3), 229–254.
- Larreche, J.-C., & Moinpour, R. (1983). Managerial judgment in marketing: The concept of expertise. *Journal of Marketing Research*, 20(2), 110–121.
- Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, 52(4), 667–694.
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025.
- Madirolas, G., & de Polavieja, G. G. (2015). Improving collective estimations using resistance to social influence. *PLoS Computational Biology*, 11(11), e1004594.
- Malone, T. W., & Bernstein, M. S. (2015). *Handbook of collective intelligence*. MIT Press.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87.
- Massa, M., Reuter, J., Zitzewitz, E., & Stanford, G. S. B. (2006). *The rise of teams in fund management*. Working paper, INSEAD.

- Noriega-Campero, A., Almaatouq, A., Krafft, P., Alotaibi, A., Moussaid, M., & Pentland, A. (2018). The Wisdom of the Network: How Adaptive Networks Promote Collective Intelligence. *ArXiv Preprint ArXiv:1805.04766*.
- Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.
- Pan, W., Altshuler, Y., & Pentland, A. (2012). Decoding social influence and the wisdom of the crowd in financial trading network. *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*, 203–209. IEEE.
- Pauker, S. G. (1982). Prescriptive approaches to medical decision making. *American Behavioral Scientist*, 25(5), 507–522.
- Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, 15(4), 353–375.
- Rulke, D. L., & Galaskiewicz, J. (2000). Distribution of Knowledge, Group Network Structure, and Group Performance. *Management Science*, 46(5), 612–625.
- Stroebe, W., Nijstad, B. A., & Rietzschel, E. F. (2010). Chapter four-beyond productivity loss in brainstorming groups: The evolution of a question. *Advances in Experimental Social Psychology*, 43, 157–203.
- Sunstein, C. R. (2006). *Infotopia: How many minds produce knowledge*. Oxford University Press.
- Surowiecki, J. (2004). *The wisdom of crowds*. Anchor.
- Ven, A. H. V. D., & Delbecq, A. L. (1974). The effectiveness of nominal, Delphi, and interacting group decision making processes. *Academy of Management Journal*, 17(4), 605–621.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.

Figure 1. Relationship between ϕ (proportion toward truth) and the probability that communication improves the accuracy of a group's mean belief.

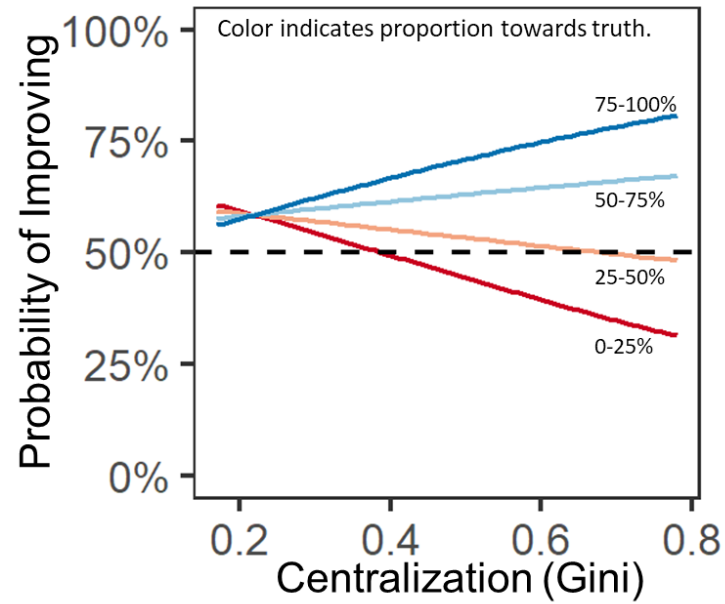


Notes: Numbered captions indicate sample size for each point. Each datapoint is an experimental trial.

Table 1. Statistical relationship between ϕ (proportion toward truth) and the probability that communication improves the accuracy of a group's mean belief.

	% Error (Logged)			
	Discussion		Delphi	
Talkativeness	0.01 (0.03)		0.02 (0.03)	
Stubbornness		-0.25*** (0.03)	-0.25*** (0.03)	-0.25*** (0.03)
Intercept	-0.66*** (0.20)	-0.04 (0.21)	-0.09 (0.22)	-0.03 (0.21)
Observations	2,099	2,099	2,099	2,847
Log Likelihood	-4,005.26	-3,978.94	-3,981.23	-5,652.97
Akaike Inf. Crit.	8,018.53	7,965.88	7,972.45	11,313.93
Bayesian Inf. Crit.	8,041.12	7,988.48	8,000.70	11,337.75
Note:	+p<0.1 *p<0.05 **p<0.01 ***p<0.001			

Notes: Each datapoint is an experimental trial.

Figure 2. Interaction between ϕ (proportion toward truth) and talkativeness centralization

Notes: Figure shows fitted values based on regression model. Each datapoint is an experimental trial.

Table 2. Relationship between accuracy and centrality

	% Error (Logged)			
	Discussion		Delphi	
Talkativeness	0.01 (0.03)		0.02 (0.03)	
Stubbornness		0.23*** (0.03)	0.23*** (0.03)	0.23*** (0.03)
Intercept	-0.66*** (0.20)	-1.20*** (0.20)	-1.24*** (0.22)	-1.21*** (0.20)
Observations	2,099	2,099	2,099	2,847
Log Likelihood	-4,005.26	-3,974.91	-3,977.22	-5,648.75
Akaike Inf. Crit.	8,018.53	7,957.81	7,964.44	11,305.50
Bayesian Inf. Crit.	8,041.12	7,980.41	7,992.69	11,329.32
<i>Note:</i>		* p<0.05; ** p<0.01; *** p<0.001		

Notes: Each datapoint is an individual person.