2018

The Network Dynamics Of Social Influence In The Wisdom Of Crowds

Joshua Becker
University of Pennsylvania, joshua.becker@asc.upenn.edu

The Network Dynamics Of Social Influence In The Wisdom Of Crowds

Abstract

Research on the wisdom of crowds is motivated by the observation that the average belief in a large group can be accurate even when group members are individually inaccurate. A common theoretical assumption in previous research is that accurate group beliefs can emerge only when group members are statistically independent. However, network models of belief formation suggest that the effect of social influence depends on the structure of social networks. We present a theoretical overview and two experimental studies showing that, under the right conditions, social influence can improve the accuracy of both individual group members and the group as a whole. The results support the argument that interacting groups can produce collective intelligence that surpasses the collected intelligence of independent individuals.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Communication

First Advisor

Damon Centola

Keywords

collective beliefs, collective intelligence, forecasting, social network, team decision-making, wisdom of crowds

Subject Categories

Communication | Organizational Behavior and Theory | Quantitative, Qualitative, Comparative, and Historical Methodologies

THE NETWORK DYNAMICS OF SOCIAL INFLUENCE IN THE WISDOM OF CROWDS

Joshua Becker

A DISSERTATION

in

Communication

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

____

Damon Centola
Associate Professor, Annenberg School
Associate Professor, School of Engineering
Faculty, Warren Center for Network and Data Sciences
University of Pennsylvania

Graduate Group Chairperson

____

Joseph Turow
Robert Lewis Shayon Professor of Communication
Associate Dean for Graduate Students
University of Pennsylvania

Dissertation Committee
Damon Centola, Associate Professor of Communication, Associate Professor School of Engineering
Sandra Gonzalez-Bailon, Assistant Professor of Communication
Michael X. Delli Carpini, Walter H. Annenberg and Professor of Communication

To my partner, Shaina

and my advisor, Damon

both of whom challenged me like nobody else

and showed remarkable patience

and to my parents

who surrounded me with books

ACKNOWLEDGMENTS

This dissertation was made possible by the Wharton Risk Center Russell Ackoff Fellowship at the University of Pennsylvania, and by the Annenberg School for Communication at the University of Pennsylvania. I would also like to acknowledge invaluable intellectual contributions by Devon Brackbill, who co-authored Chapter 2, and all of the Network Dynamics Group at the University of Pennsylvania.

ABSTRACT

THE NETWORK DYNAMICS OF SOCIAL INFLUENCE IN THE WISDOM OF CROWDS

Joshua Becker

Damon Centola

Research on the wisdom of crowds is motivated by the observation that the average belief in a large group can be accurate even when group members are individually inaccurate. A common theoretical assumption in previous research is that accurate group beliefs can emerge only when group members are statistically independent. However, network models of belief formation suggest that the effect of social influence depends on the structure of social networks. We present a theoretical overview and two experimental studies showing that, under the right conditions, social influence can improve the accuracy of both individual group members and the group as a whole. The results support the argument that interacting groups can produce collective intelligence that surpasses the collected intelligence of independent individuals.

TABLE OF CONTENTS

ACKNOWLEDGMENTS .... **III**

ABSTRACT .... **IV**

LIST OF TABLES .... **VIII**

LIST OF ILLUSTRATIONS .... **IX**

PREFACE .... **X**

CHAPTER 1: THE WISDOM OF CROWDS **1**

Abstract **1**

Introduction **1**

Statistical Foundations for the Wisdom of Crowds **5**
The Diversity Prediction Theorem 5
Correlation Framework for the Wisdom of Crowds 6
Types of Estimation Problems 7
Independence and the Wisdom of Crowds 8

Maximizing the Accuracy of Aggregation Estimates **10**
Aggregation Methods 10
Triangulating Multiple Data Types 11
Prediction Markets 13
Diversity and Group Composition 14

The Social Dynamics of Collective Beliefs **15**
Herding and Groupthink 17
The Delphi Method 18
Numeric Estimates in Networks 19
Binary Beliefs in Networks 21
Conditions of Information Exchange 22

Relationship to Other Forms Collective Intelligence **23**
Problem Solving 24
Coordination 25

Conclusion **26**

CHAPTER 2: NETWORK DYNAMICS OF SOCIAL INFLUENCE IN THE WISDOM OF CROWDS **28**

Abstract **28**

Introduction **28**

Social Influence and the Wisdom of Crowds **30**

Formal Model of Social Influence **34**

Theoretical Predictions **35**

Experimental Design **36**
Estimation Tasks. 38
Experimental Analysis 38

Results **39**
Social Influence in Decentralized Networks 39
Comparison with Control Condition 44
Social Influence in Centralized Networks 46
Robustness 48

Discussion **48**

Appendix A: Estimation Tasks and Subject Experience **49**

Appendix B: Theoretical Predictions for Social Influence in the Wisdom of Crowds **51**
Model Definition 51
Social Influence in Decentralized Networks 55
Effect of Centralization on Group Accuracy 55
Robustness to Variation in Density and Average Degree 56

Appendix C: Supplemental Figures **57**

CHAPTER 3: FROM EFFICIENT MARKETS TO EFFECTIVE INVESTORS – INDIVIDUALS AND THE WISDOM OF CROWDS **63**

Abstract **63**

Introduction **64**

The Wisdom of Independent Crowds **67**

Social Influence and the Wisdom of Crowds **69**

The Wisdom of Financial Crowds **71**

Methods **74**
Experimental Design 75
Subject Experience. 77
Analysis 77

Results **78**
Effect of Social Information on Group Beliefs 79
Effect of Social Information on Individual Accuracy 80
Who benefits from the wisdom of crowds? 81
How does social influence shape forecast accuracy? 83

Discussion **84**
Conclusion 86

Appendix A: Theoretical Predictions for the Effect of Social Influence **87**

Appendix B: Experimental Design Details **89**

CONCLUSION **91**
Social Influence in Decentralized Networks. 91
Centralization. 93
Limiting Assumptions and Next Steps. 95

BIBLIOGRAPHY **98**

LIST OF TABLES

Tables in Chapter 3

Table 1. Change in individual error as a function of response to social information

Table A1. List of forecasting tasks.

LIST OF ILLUSTRATIONS

Figures in Chapter 2

Figure 1. Effect of social influence on group accuracy

Figure 2. Correlation between revision magnitude and individual error

Figure 3. Correlations with changes in group mean.

Figure A1. Screenshot of the experimental interface.

Figure A2. Theoretical effect of social influence on the group mean.

Figure A3. Theoretical effect of social influence on the group median.

Figure A4. Theoretical effect of network centralization.

Figure A5. Theoretical effect of network density.

Figure A6. Theoretical effect of average degree.

Figure A7. Power test based on data reported in previous experiments.

Figures in Chapter 3

Figure 1. Effect of social influence on individual accuracy.

Figure 2. Effect of social influence based on each individual's initial accuracy.

Figure A1. Theoretical effect of social influence on individual accuracy.

Figure A2. Screenshot of the experimental interface.

PREFACE

Research on the “wisdom of crowds” is motivated by the observation that the average belief of a large group of people can be surprisingly accurate, even when the group is composed of members who are individually inaccurate. Due to the theoretical importance of informational diversity, a key premise in prior research on the wisdom of crowds has been that groups will be accurate only when their members are statistically independent (Hong, Du, Wang, Fan, & Xu, 2016; Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Nofer & Hinz, 2014). Intuitively, this claim is based on the idea that people who are independent have errors who are independent, and therefore those errors will cancel out in large numbers. When people influence each other, conformity pressure can produce herding dynamics (Banerjee, 1992) and groupthink (Janis, 1982), undermining the wisdom of crowds (Surowiecki, 2004).

In contrast with this common theoretical assumption, this dissertation argues that a careful examination of the theoretical principles underpinning the wisdom of crowds suggests that social influence will not always undermine group accuracy. Instead, I argue that the effect of social influence on belief accuracy is contingent on the structure of social networks, and that social influence can improve belief accuracy when mediated through structured information exchange. In order to test this theory, we conduct two experiments using an online platform in which subjects completed estimation tasks both before and after observing the belief of others. In each experiment, we also study a control group, in which subjects revise their beliefs without social information. This design allows us to compare independent belief formation with social belief formation, thus controlling for any individual learning process. The first experiment uses general trivia tasks, and the analysis focuses on accuracy at the group level. The second experiment studies a task of practical interest—financial forecasting—and the analysis focuses on individual accuracy.

This dissertation begins with a review of research on the wisdom of crowds, presented in **Chapter 1**. In order to present a comprehensive description of group belief formation, this first chapter not only reviews previous research but also draws on the results of the experiments described in this dissertation, placing them in context with the broader literature. In this chapter, I argue that research on the wisdom of crowds can be described in terms of two theoretical orientations. Research on “collected intelligence” follows a crowdsourcing paradigm which aggregates the collected beliefs of many independent individuals, using psychological and statistical principles to extract the most accurate estimate possible. In contrast, research on “collective intelligence” studies the social processes of belief formation, and shows how groups can process information in ways that cannot be reduced to a statistical aggregation of independent individuals. I argue on this basis that statistical models of belief formation cannot be fully defined using the distribution of individual member characteristics alone, which assumes that individuals are independent conditional on those characteristics. Though this approach often serves as a useful approximation, a complete explanation of group belief formation must also account for the social network structure. Network models of belief formation demonstrate that no one person’s belief can be predicted without knowing the beliefs of their peers, whose beliefs in turn depend on their peers, and so on, such that the entire network shapes a single individual’s belief.

Following this review, **Chapter 2** describes an experiment designed to test the effect of social influence on the accuracy of the average belief in a group. This study describes novel theoretical results showing how social influence can improve the accuracy of the average belief in a group, but only when groups are embedded in social networks where each individual is connected to an equal number of others (a “decentralized” social network). In contrast, when groups are characterized by the presence of disproportionately influential members (a “centralized” social network), the accuracy of the group depends only on the accuracy of those highly influential “central” individuals.

One key principle of the wisdom of crowds is that the accuracy of the average belief is mathematically distinct from the average accuracy of individual beliefs—i.e., the error of the average is not equivalent to the average error. This statement, of course, reflects the core principle of crowd wisdom—groups can be accurate even when individuals are inaccurate. However, this statement also means that special care must be taken not only to understand how social influence impacts group-level (average) beliefs, but also to understand how social influence impacts individual beliefs. **Chapter 3** presents an analysis designed to specifically examine the role of social influence on individual accuracy. This chapter also tackles another limitation of the experiment described in Chapter 2: while Chapter 2 describes an experiment in which subjects complete simple trivia-like estimation tasks (e.g., counting the number of gumballs in a jar), the experiment in Chapter 3 tests the wisdom of crowds in a context of practical importance: financial forecasting.

Although each chapter in this dissertation is intended to stand alone as an academic paper (and thus some theoretical repetition should be expected) they also work in concert to form a coherent whole, a single thesis. Taken together, the research presented in this dissertation advances our understanding of the wisdom of crowds in several ways. Chapter 2 builds on a model of the wisdom of crowds (DeGroot, 1974) which predicts that social belief formation is governed by network structure. As described in Chapter 1, this model in its simplest form predicts that decentralized networks will converge directly on the mean of independent beliefs, such that social influence neither helps nor harms group accuracy. However, we found empirically that individuals who were more accurate also were less responsive to social influence. Simulations described in Chapter 2 predict that this process will lead to a systematic increase in the accuracy of the average belief in a group, when that group is embedded in a decentralized network. Thus, chapter 2 not only provides an empirical test of network theories of social influence, but also contributes a novel and empirically motivated theoretical prediction: social influence can improve the accuracy of group beliefs.

The results described in Chapter 3, however, do not replicate this finding: in this experiment, groups appear to converge directly on the mean of independent beliefs. While we cannot rule out the possibility of Type II error (i.e., the effects may simply be too weak to observe given the inherent noise associated with financial forecasting, or the observation that groups were accurate to begin with) the results presented in Chapter 3 ultimately serve to highlight an incredibly robust principle of the wisdom of crowds: individuals are nearly always going to become more accurate as a result of information exchange, even when the average belief does not change appreciably. In simulation, we find that increased accuracy of individuals is a robust phenomenon that is expected to hold under a wide range of empirically plausible assumptions. As discussed in both Chapter 1 and Chapter 3, the theoretical prediction that social influence can improve the accuracy of individuals follows very directly from the statistical principles of the wisdom of crowds. Nonetheless, this theoretical prediction is remarkably absent from prior research on the wisdom of crowds, and thus offers a novel and important theoretical contribution.

Taken together, the theoretical and empirical results presented in this dissertation provide a crucial correction to previous research on the wisdom of crowds, which commonly assumed that social influence undermines the wisdom of crowds. In contrast with this common assumption, the results presented here demonstrate that social influence in structured networks can improve belief accuracy at both the group level and the individual level.

CHAPTER 1: THE WISDOM OF CROWDS

Abstract

The wisdom of crowds is an empirical phenomenon that occurs when the average belief in a large group is accurate even when individuals are inaccurate. Although large groups are not guaranteed to produce accurate beliefs, statistical principles guarantee that the group belief will always be more accurate than the belief of a randomly selected individual. In this review, we describe research on the wisdom of crowds, observing that theoretical and empirical studies can be grouped into two general paradigms. Under the "collected intelligence" paradigm, researchers collect data from individuals, ideally independent, in order to generate aggregated crowd-sourced estimates. Under the "collective intelligence" paradigm, researchers examine how social processes shape belief formation in ways that cannot be explained solely in terms of the statistical distribution of independent individuals. This latter body of research has found that groups can actively process information to produce estimates that are more accurate than the aggregated beliefs of independent individuals. We argue that the wisdom of crowds represents a distinct social phenomenon, and we distinguish research on belief accuracy from social learning, collective problem solving, and coordination in networks.

Introduction

Research on the wisdom of crowds is motivated by the observation that the average belief in a large group of people can be surprisingly accurate, even when group members are individually inaccurate. This remarkable observation was made unexpectedly by Galton in 1907, who expected to find the ignorance of crowds, and the wisdom of crowds and has since been observed in a wide range of topic domains including financial investment (Kelley & Tetlock, 2013; Nofer & Hinz, 2014), medical diagnoses (Kurvers et al., 2016; Wolf, Krause, Carney, Bogart, & Kurvers, 2015), visual search tasks (Juni & Eckstein, 2017), geopolitical forecasting (Atanasov et al., 2016a; Mellers et al., 2014), and sports betting (Herzog & Hertwig, 2011; Peeters, 2018). Although the wisdom of crowds is an empirically motivated phenomenon, the ability for groups to generate accurate judgements can be theoretically explained by a set of basic statistical principles. One key principle is known as the "crowd beats average" law (Page, 2007) and can be derived readily from the well-known bias-variance decomposition for a statistical estimator. While these principles do not promise any minimum accuracy for group beliefs, they do provide a mathematical guarantee that the crowd belief, measured as the mean of individual beliefs, will be always more accurate than the belief of an average individual.

Early efforts to replicate this finding measured simple laboratory tasks such as estimating the weight of objects (Bruce, 1935; Gordon, 1924) and identifying the temperature in a classroom (Knight, 1921 as cited by Klugman, 1945). Belief aggregation has since shown a great deal of potential for practical application. Wolf et al (2015) tested several methods to pool the judgements of radiologists classifying mammography screenings, finding that no matter what aggregation rule was used, the pooled judgement provided both a greater number of true positives (correct identifications of cancer) and fewer false negatives (incorrect categorizations as healthy) than a randomly selected individual. These findings have been replicated with other classification tasks, including other medical procedures (Wolf et al., 2015) and are expected to hold for more general tasks such as surveillance and search-and-rescue (Juni & Eckstein, 2017). Sports fans (Peeters, 2018) and amateur financial analysts (Nofer & Hinz, 2014) have also been found to demonstrate the wisdom of crowds, and it is likely that aggregated beliefs will continue to be found useful for many types of estimates.

One particular area of interest is the prediction of future events. In 1944, Klugman (1947) surveyed soldiers to estimate when armistice would be reached in World War II. They found that the mean estimate for armistice was more accurate than 85% of individual estimates. Aggregated beliefs have become a common tool for forecasting the outcome of democratic elections (Murr, 2015), with so-called “citizen forecasts” outperforming even formal statistical models when properly calibrated (Graefe, 2014). In addition to geopolitical forecasting, estimate aggregation has become a widely adopted method for economic forecasting. Economic forecasts can range from macro-level predictions about metrics such as GDP (Jansen, Jin, & de Winter, 2016; Mellers et al., 2014) to firm-specific forecasts for such metrics as earnings and revenue (Drogen & Jha, 2013). It is important to note that in the context of economic forecasting, predictions are not only based on subjective estimates but are also based on formal statistical models. Clemen (1989) notes that researchers have reached nearly unanimous agreement that the combination of estimates—whether formal models, subjective beliefs, or a combination of the two—produces more accurate estimates than any individual or model alone. Whether subjective human judgement alone can outperform formal models depends on circumstances. Jansen et al. (2016) found that under typical market conditions where statistical models can be well calibrated with reliable data, formal statistical models outperform human judgement. However, they argue that human judgement offers better predictions than formal models under atypical conditions such as economic crises, when the assumptions of formal models break down.

A common explanation for the wisdom of crowds is that each individual forms some belief about the state of the world, plus or minus some individual error term, such that collections of independent individuals have independent errors that cancel out in large groups (Hogarth, 1978; L. Hong & Page, 2009). One prominent research paradigm, building on this statistical premise, attempts to identify how to best use crowd-sourced beliefs to produce accurate forecasts and other estimates. This research tests different ways to combine individual estimates into a single wisdom-of-crowd estimate, including methods such as weighting beliefs according to respondents confidence (Koriat, 2012; Kurvers et al., 2016; Prelec, Seung, & McCoy, 2017) or prior accuracy (Budescu & Chen, 2014; Mannes, Soll, & Larrick, 2014).

In contrast with this collected intelligence paradigm, in which each individual independently contributes their perspective, a growing body of evidence has shown that groups can actively process information to produce beliefs that are more accurate than those of independent individuals. Research in this second paradigm shows how collective intelligence can produce accurate beliefs through social processes that are irreducible to collections of independent individuals. Research on social influence has not only been helpful in refining the accuracy of aggregation methods (Kao et al., 2018; Madirolas & de Polavieja, 2015), but is also important to understanding situations where the outcome of interest is endogenous belief formation, rather than an estimate generated by an outside party. Even when the ability for an outside aggregator to form accurate beliefs is of little interest, the population distribution of beliefs may be inherently meaningful, as in organizational behavior and democratic processes.

This article will begin by reviewing formal theoretical models of group accuracy and discussing the notion of independence in the wisdom of crowds. Statistical models of the wisdom of crowds do allow for correlation between individual beliefs, but nonetheless are always consistent with the assumption that individuals are conditionally independent. Conditionally independent beliefs may be correlated, for example, if two people share the same information source; those people are assumed to be independent conditional on that shared information (as in multilevel regression). In contrast, network models of social influence show that a person's belief at any point in time cannot be fully explained without accounting for the beliefs of their peers and the influence those peers have on them. After laying out this statistical framework, we review research on estimates based on the collection of many individual beliefs. This research can be described in terms of a statistical or machine learning framework: researchers identify and optimize a function that maps data to an estimate. Following this section, we review theoretical models and empirical research on the effect of social influence on belief accuracy. Finally, we close with a discussion of related models of collective intelligence, defining the wisdom of crowds against processes such as coordination and problem solving.

Statistical Foundations for the Wisdom of Crowds

The ability for groups (either groups of people or groups of models) to produce judgements whose accuracy exceeds that of any individual can be explained by a model in which each individual estimate is composed of the true value, plus an error term:

$$x_i = (\theta + \varepsilon_i) \tag{1}$$

where x_i is the judgement of the i^{th} individual, θ is the true value to be estimated, and ε_i is the error of the i^{th} individual (Hogarth, 1978; L. Hong & Page, 2009). Under the assumption that ε_i is distributed identically and independently for all individuals, and that $E[\varepsilon]=0$, then the expected error of the mean—that is, collective error—will decrease as the group size increases, and approach the true value. However, the assumption that $E[\varepsilon]=0$ is unlikely to hold in the vast majority of empirical contexts. In most circumstances, people's beliefs are subject to common cognitive biases, which lead to systematic errors, such that a population will tend to regularly under- or over-estimate the true value (Kahneman & Tversky, 1977; Kao et al., 2018).

The Diversity Prediction Theorem. Fortunately, a well-known statistical regularity proves that the aggregated belief within a large population will always offer an accuracy benefit regardless of the presence of systematic bias (L. Hong & Page, 2008; Page, 2007):

$$E[\bar{x} - \theta]^2 = E[(x_i - \theta)^2] - E[(x_i - \bar{x})^2]\ . \tag{2}$$

Equation 2 is well known in mathematical statistics, where it shows that the bias (error) of any estimator (in this case, the error of the group average, indicated by the leftmost term) is equal to the mean squared error (average individual error—the first term on the right) minus the variance (belief diversity—the second term on the right). This statement has recently become known as the "diversity prediction theorem" (Page, 2007) because it shows that the accuracy of the group estimate depends on both the accuracy of individual estimates as well their variance, or "diversity." One important implication of this statement is that when individual error is held constant, group error decreases as diversity decreases.

The diversity prediction theorem also provides one of the fundamental principles of the wisdom of crowds, what Page (2007) terms the "crowd beats averages" law. This law is based on the fact that Equation 2 leads immediately to the inequality

$$E[\overline{x} - \theta]^2 < E[(x_i - \theta)^2] \tag{3}$$

which states that the error of the average estimate in a group (of any size) will always be lower than the average error obtained from any individual estimate alone. Thus, the average belief in a group will always be more accurate than a randomly selected individual belief. It should be noted that this identity applies to any sample belief distribution, and is not true only in expectation. It should also be noted that this identity does not require individuals to be independent.

Correlation Framework for the Wisdom of Crowds. A very similar principle, drawing on statistical principles commonly used to measure psychometric instrument reliability, describes the advantages of aggregation in terms of correlation (Hogarth, 1978). Where equation 1 assumes that every individual belief is drawn from the same distribution, it may also be the case the individual beliefs come from different distributions. This variation between individual belief distributions may emerge if individuals are exposed to different signals, for example. In the context of forecasting, such variation may occur if individuals have access to different data regarding the predicted outcome. This variation may emerge even if individuals have access to the same data, but interpret that data differently, reaching different conclusions (L. Hong & Page, 2008, 2009).

Together, these concepts imply a generative model in which each person's belief is drawn from some distinct distribution given by some particular instantiation of equation 1. The value $E[\varepsilon_i]$

for each distribution determines a person's expected accuracy. In equation 2, expected error for each individual is expressed in terms of mean squared error, but this concept can also be expressed in terms of correlation with the true value. This "reliability" metric (Hogarth, 1978) for accuracy asks: to what extent does a person's prediction correlate with the true value? The notion of correlation emerges from the assumption that an individual belief is drawn from some underlying probability distribution. This correlation can be measured in practice when people make repeated predictions over many trials (for example, predicting many future events).

In this conceptual framework, there are three correlations of interest: the reliability of the group estimate, expressed as the correlation between the group mean and the true value; the average reliability of the individual estimators; and the correlation between each individual response. These three terms are related with the following equation given by Hogarth (1978):

$$\rho_{y\bar{x}} = \frac{\bar{\rho}_{yx}}{\bar{\rho}_{x_i x_j}}. \tag{4}$$

This equation states that group reliability is equal to the ratio between the average individual reliability and the correlation between individual estimates. Though equation 4 requires slightly different assumptions and presents a different form than equation 2, it offers a very similar statement. Equation 4 states that as average individual reliability (error) is held constant, and inter-individual correlation is increased—thus increasing the similarity of individual estimates—group reliability (error) will decrease. In other words, the more individual beliefs are correlated, the less advantage the group estimate offers (Hogarth, 1978).

Types of Estimation Problems. Equations 2 and 4 make no assumption about the type of random variable to be estimated, or the conditions under which those beliefs are formed, meaning they apply to any kind of formal model or wisdom of crowds estimate, including those generated through social information processing. This generality means that the wisdom of crowds principle—that the error of the group will always be lower than the error of the average individual—holds for any type of belief that might be empirically useful. However, it is nonetheless important to distinguish theoretically between two basic types of estimation task. In the vocabulary of machine learning or statistical modeling, the two primary types of estimation are regression tasks, in which data (such as the information available to an individual) is mapped to a numeric estimate; and classification tasks, in which data is mapped to a single binary or categorical outcome. This distinction is important because both empirical and theoretical research on the wisdom of crowds occasionally yield divergent results depending on the which of the two estimation tasks is being studied. For example, as will be discussed in more detail below, empirical findings for numeric beliefs point to a very general correlation between accuracy and self-reported confidence or response to social influence (Becker, Brackbill, & Centola, 2017; Madirolas & de Polavieja, 2015; Moussaïd, Kämmer, Analytis, & Neth, 2013). However, the correlation been accuracy and confidence for binary classification tasks depends on whether common knowledge supports the correct or incorrect belief (Koriat, 2012; Prelec et al., 2017). Further, while theoretical models of numeric estimates predict that groups can always benefit from information exchange in networks where everyone is equally influential, models of binary belief exchange predict that the effect of social influence depends on the initial majority belief (Mossel & Tamuz, 2017).

Independence and the Wisdom of Crowds. When individual beliefs are assumed to be conditionally independent, the group belief distribution can be described as a function of the population distribution of individual characteristics such as demographic characteristics, information exposure, expertise, and cognitive model. For example, a stock forecast might be determined by the framing of the local news, the presence (or absence) of professional training, and the mental framework the forecaster uses to map information to a belief. In terms of the model described above, these factors all shape the conditionally independent probability distributions from which each individual draws when making a forecast. The crowd is, when people are conditionally independent, a crowd of distributions. This "collected intelligence" paradigm implies that the observed belief distribution can be described as a function of all the potentially conditioning variables: $X = \Psi(I)$ where X is the vector of individual beliefs, I is the matrix of individual characteristics, and Ψ is a function mapping individual characteristics to individual beliefs.

While theoretical accounts of beliefs as conditionally independent generate a useful approximation for a machine learning approach to the wisdom of crowds, these accounts do not fully explain the formation of collective beliefs. To account for social influence, network models describe group beliefs as a function of the initial distribution of beliefs as well as a weighted, directed network of influence (Becker et al., 2017; DeGroot, 1974; Golub & Jackson, 2010). The effect of social influence means that no person's belief can be defined independently (even in expectation), even if all of their individual characteristics are known, because each person's belief depends not only on the structure of the network in which they are embedded, but also on their precise location within that network. This "collective intelligence" paradigm holds that the group belief distribution must be defined as a function both of the distribution of individual characteristics and also the network adjacency matrix: $X = \Psi(I,A,t)$ where A is the network adjacency matrix and t is an index for time, which is necessary due to the dynamic nature of belief formation in networks.

One implication of the collected intelligence paradigm, i.e. the assumption of conditional independence, is that one sample of N people is equivalent in expectation to any other sample of N people when the two samples are drawn from the same population: the only difference is sampling error. However, the effective sample error of a group of N people embedded in a network is larger than the sample error of N conditionally independent people. This difference emerges from the fact that two identical groups can diverge as a result of social influence. Practically, this additional variance means that researchers must take extra care when calculating standard errors in statistical analysis. Conceptually, this additional variance means that two groups can generate radically different beliefs even when they are composed of individuals who are in every measurable way identical. The potential for two groups of identical individuals to produce different collective outcomes is a key principle of collective intelligence versus collected intelligence. When considering collective processes, one must not only account for the distribution of individual characteristics, but must also account for distinctly group-level properties, most notably network structure. Nonetheless, the assumption of crowds as collections of independent individuals provides a useful approximation, as indicated by the large body of research on crowdsourced estimate accuracy.

Maximizing the Accuracy of Aggregation Estimates

There are many different approaches to generating an estimate based on the collection of individual beliefs, all of which share a methodological orientation with machine learning: parameterize an aggregation function from a set of training data, and then apply it to new data to form estimates or predictions. As a wisdom of crowds estimator, the sample mean is one of many possible functions. The mean is a useful theoretical metric due both to its conceptual simplicity and it's mathematical tractability; the equations above, for example, are convenient in part because they apply equivalently to both numeric estimates and classification problems. In practice, however, there are many different ways to aggregate group beliefs. Galton (1907) studied the median, arguing that "According to the democratic principle of 'one vote one value,' the middlemost estimate expresses the vox populi, every other estimate being condemned as too low or too high by a majority of the voters." Immediately following Galton's publication, Hooker (1907) raised the question of whether the mean was a better function. A substantial amount research on the wisdom of crowds seeks to identify optimal functions for mapping crowd-sourced data to estimations.

Aggregation Methods. Ultimately, The optimal aggregation method depends on the properties of the underlying belief distribution (Kao et al., 2018). Although it is unlikely that there is a single function that's optimal under all circumstances, certain general principles have nonetheless been found to produce more accurate aggregated estimates. For the aggregation of a single distribution of estimates, with no additional information, Kao (2018) provides evidence that task-specific biases can be identified and corrected once certain parameters of the belief distribution are known. Estimates can be further improved by weighting aggregated beliefs more heavily towards those individuals who demonstrated prior accuracy (Mannes et al., 2014). Budescu and Chen (2014) note, however, that individual accuracy alone will not always identify those members who make the greatest contributions to crowd accuracy. For example, if a forecaster only offers predictions for future events that are easy to predict, they will have an apparently high accuracy. In order to provide a better method for weighting individual contributions to accuracy, Budescu and Chen weight forecasters by their relative contribution, which asks: how much more accurate is the crowd estimate with this forecast added to the pool? This weights the resulting estimate more heavily towards those individuals who are accurate when everyone else is inaccurate, and places less weight on forecasters that are accurate only in situations when everyone is accurate.

Welinder, Branson, Perona, and Belongie (2010) go one step further and weight beliefs according to three relevant characteristics for each contributor. Most notably, they distinguish between bias (the tendency to under- or over-estimate in numeric estimates, or the tendency toward false-positives or false-negatives in classification) and "competence," which reflects an individual's consistency and is measured by standard deviation. These distinctions are important because a person who is consistently wrong in a classification is in fact a remarkably useful data source—the true classification will always be the opposite of what they say. In addition to measuring competence and bias, they also estimate contributors' expertise, which allows individual weights to vary across topic. This characteristic recognizes that an individual may have high competence and low bias on some tasks, but low competence and high bias on other tasks.

Triangulating Multiple Data Types. In addition to weighting individuals by accuracy, estimates can be improved by the collection of supplementary information such as reported confidence. Koriat

(2012) studied binary choice questions and asked subjects to report confidence along with their answers. They found that people who rate themselves as more confident are also more likely to be correct, but that this effect is reversed when people's beliefs are misled by common knowledge (a finding replicated by Prelec et al., [2017]). These studies thus show that the accuracy of confidence-weighted estimates can vary by task type, and research on medical decision-making has found that confidence-weighting can indeed be useful in practically relevant contexts, as it can be used to improve the accuracy of cancer diagnoses (Kurvers, 2016). Experiments on social influence suggest that the relationship between confidence and accuracy is not limited to classification estimates: for numeric estimates, accurate individuals tend to revise their beliefs less in the presence of social information (Becker et al., 2017; Madirolas & de Polavieja, 2015). By observing how individuals response to social information, a weighted estimate can be generated using a process similar to that of confidence-weighted estimates (Madirolas & de Polavieja, 2015).

For classification problems, Prelec, Seung, and McCoy (2017) offer a solution to the inconsistent relationship between confidence and accuracy. Prelec et al. asked subjects to provide, in addition to their estimate, a statement about what they think other respondents answer will be. When an answer receives more votes than responses would predict, that answer is termed the "surprisingly popular answer." Prelec et al. derive an argument from Bayesian probability theory which predicts that the surprisingly popular answer is the most likely to be correct. Drawing on datasets testing trivia knowledge, fine art valuation, and medical diagnoses, they find empirically that this method provides estimates that are more accurate than either majority vote or confidence-weighted voting methods.

Triangulating the wisdom of crowds through supplementary data need not be limited to survey data. The increasingly popular use of "digital trace data" for research (Lazer & Radford, 2017) has provided wisdom-of-crowds researchers with an array of data sources. In one study, Nofer and Hinz (Nofer & Hinz, 2014) found that target stock prices posted in an online community could be aggregated to produce forecasts that were more accurate than expert analysts. Bollen, Mao, and Zeng (2011) further show how digital content can be used for forecasting even when the content itself is not an explicit forecast. They found that the emotional content of Twitter posts can predict shifts in the Dow Jones Industrial Average several days in advance. While Bollen et al studied only very broad market movements, Chen, De, Hu, and Hwang (2014) conducted a similar form of text analysis on user posts to an online social network devoted to stock market discussion. They argued that this data offered valuable predictive power for market returns on specific assets even after controlling for information contained in traditional media sources.

Prediction Markets. One of the most popular methods for generating accurate crowd-sourced forecasts is the use of prediction markets, in which individuals place a monetary bet on possible outcomes for a future event. Prediction markets have been reviewed extensively elsewhere (Arrow et al., 2008; Tziralis & Tatsiopoulos, 2012; Wolfers & Zitzewitz, 2004) and thus we will discuss them only briefly here. In one common form, individual participants in a prediction market make a bet between $0 and $1 on whether a future event will occur, earning $1 if their bet is correct and $0 if their bet is incorrect. Strictly speaking, participants in a prediction market are not socially independent, because they observe the market price of a contract before making a purchase. However, this influence does not appear to play a practically significant role, and theoretical models can explain market prices based on the independent distribution of beliefs about the probability that an event will happen. Wolfers and Zitzewitz (2006) derive a theoretical model of price equilibrium from a distribution of independent beliefs which shows how the market price can reflect the mean belief. Empirical evidence supports this model, showing both that market prices reflect mean beliefs and also that market prices are predictive of future events, such that an event trading at 60 cents will occur approximately 60% of the time (Cowgill & Zitzewitz, 2015; Wolfers & Zitzewitz, 2004)

While prediction markets typically involve very large populations, generating the necessary market liquidity, these markets have also been found to provide an effective forecasting tool in more limited contexts such as corporate decision-making (Cowgill & Zitzewitz, 2015). Although efficient and popular, prediction markets are not necessarily the optimal method for generating crowd-sourced estimates. In one study that examined the generation of estimates for a wide range of geopolitical outcomes, Atanasov et al (2016) found that prediction markets produced more accurate estimates than simple belief aggregation, but that calibrated and weighted belief aggregation produced the more accurate estimate than prediction markets.

Diversity and Group Composition. In addition to identifying optimal aggregation functions, another important factor for the optimization of crowdsourced estimates is group composition—e.g., deciding whom to include on your team. Equations 2 and 4 above are commonly interpreted to mean that the wisdom of crowds is maximized when groups are composed of individual whose beliefs are uncorrelated or have a high variance—i.e., are “diverse” (H. Hong, Du, Wang, Fan, & Xu, 2016; Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Nofer & Hinz, 2014). Hogarth (1978) used equation 4 to show formally that the advantage of large groups decreases as the correlation of individual beliefs increases (ceteris paribus). Similarly, Equation 2 indicates that an increase in similarity (ceteris paribus) is accompanied by an increase in group error. Since social influence has been found to increase the similarity between individual beliefs (Asch, 1951; Deutsch & Gerard, 1955), these mathematical principles are often interpreted (with inspiration from Surowiecki’s popular 2004 book) to mean that the best crowd estimates are obtained when individual contributors form their beliefs in isolation (H. Hong et al., 2016; Lorenz et al., 2011; Nofer & Hinz, 2014). However, as we will discuss below, the diversity prediction theorem also shows how social influence can paradoxically increase belief accuracy, not in spite of but because of the increased similarity between individuals.

While diversity is commonly measured as statistical variance, Page (2007) offers a more subtle argument on the role of diversity in group belief formation. Page argues that variance alone is not sufficient for diversity, but instead that what’s important is the diversity of cognitive models individuals use to form their beliefs (Economo, Hong, & Page, 2016). The idea of diverse cognitive models is illustrated by the distinction between “interpreted signals” and “generated signals” (L. Hong & Page, 2009). A generated signal is one in which an individual’s estimate or prediction is fully determined by the signal they receive about the true state of the world, which is generated with some random error as shown in equation 1. When signals are independent, the bias of the crowd estimate is equal to the expected error of the signal. Variation in individual beliefs is explained by either random noise, or exposure to different signals. An interpreted signal model, in contrast, allows for every individual in a population to receive the same signal—i.e., the same information—but nonetheless reach a different estimate or prediction, due to differing interpretations of that information. Conditional variation is explained by conditional use of different cognitive models, operating on the same received information.

It is in this sense of interpreted models that Page (2007) argues for the importance of diversity in a population, emphasized by their description of a crowd of people as a “crowd of models.” Page points out that even if individuals are socially independent—such that they do not influence each other directly—their beliefs will nonetheless be correlated (and share the same expected error) if they all use the same predictive model. In order to obtain a greater diversity of beliefs, Page advocates for assembling groups whose members use diverse predictive models.

The Social Dynamics of Collective Beliefs

Because social influence has been found to increase similarity between individual beliefs (Asch, 1951; Cialdini & Goldstein, 2004; Sherif, 1935), the variance component of the diversity prediction theorem has been invoked to argue that wise crowds must consist of individuals who are socially independent, i.e. individuals who do not influence each other (H. Hong et al., 2016; Lorenz et al., 2011; Nofer & Hinz, 2014). However, the critical limitation of this argument is the “ceteris paribus” assumption—that is, group error decreases as diversity decreases ceteris paribus, holding constant individual accuracy. However, the effect of a change in diversity measured as variance (in equation 2) or correlation (in equation 4) depends critically on the concurrent change in individual accuracy. Importantly, social processes that decrease diversity can also paradoxically *increase* average individual accuracy. As an extreme but illustrative example, suppose that every individual after deliberation adopts the average of the group's initial, independent belief distribution. By virtue of equation 2, average individual error will necessarily decrease—because individual error will become equal to group error. Thus, in order to fully understand the effect of social influence on group accuracy, it is necessary to consider change in both diversity and individual error.

One key formalization that provides a general framework for studying the effect of social influence is network theory. Although not all belief formation processes take place over an explicitly defined social network, a network model can often still be used to describe the effect of social influence. For example, in a committee discussion it is frequently the case that everyone observes everyone else, thus obviating the need for a network model to describe patterns of information flow. However, despite the unstructured communication, it may nonetheless be the case that some members are more influential on other members, and the resulting belief distributions can be calculated using an adjacency matrix representing the weighted, directed network of influence (DeGroot, 1974).

If a scholar simply asks, "is social influence helpful or harmful?" then they may find apparent controversy within research on the wisdom of crowds. However, evidence from theoretical models (Becker et al., 2017; Golub & Jackson, 2010; Mossel & Tamuz, 2017), observational case studies (Janis, 1982), and experimental tests (Becker et al., 2017) can all be brought into alignment by reference to social network theory. Evidence consistently indicates that social information processing can improve beliefs in "egalitarian" (Mossel, Sly, & Tamuz, 2015) or "decentralized" (Becker et al., 2017) networks, which are defined as networks in which every individual is equally influential. The ability for social influence to *either* improve *or* undermine belief accuracy can be illustrated by a comparison of centralized networks to decentralized networks. In a highly centralized social influence network, one or a small number of people are highly influential and everyone else is relatively uninfluential, either due to the structure of information exchange (the influential group is more widely observed) or the structure of influence (e.g., the influential group is more persuasive). As a result, the effective source of information shaping group beliefs is a small group rather than a large group. Since not everyone's opinion contributes, the resulting collective belief represents the wisdom of the few and not the wisdom of the crowd: the group does not use of all the information held by its members. In a decentralized network, where everyone is by definition equally influential, then the collective belief represents the combined opinion of all members. Thus network structure can determine whether social influence improves or undermines group accuracy (Becker et al., 2017).

Herding and Groupthink. The diversity prediction theorem is not the only source of concern regarding the potential risks of social influence, and such concerns are further motivated by the social psychological concept of "groupthink" (Janis, 1982). In developing the theory of groupthink, Janis conducted a series of case studies on decision-making in organizations, attempting to identify principles that might lead groups of otherwise intelligent people to make disastrous decisions. One potential problem they identified is pressure for group cohesion, which can lead individuals to suppress novel information if it appears to disagree with previously established group beliefs. Another risk is caused by "directive leadership," where a single individual directs group beliefs at the expense of information held by individual members. Although these mechanisms are quite different than the principle of statistical independence, they nonetheless highlight the importance of informational diversity—groups only make fully informed decisions if they use all the available information.

Formal models of herding also show how social influence can lead to widespread inaccuracy in a population. Banerjee (1992) developed a simple model in which each person has both a private signal (which differs for each individual) and a public signal (shared for each individual) about two different choices, and must determine which choice provides a higher payoff (thus, a binary classification problem). The key property of this model is that individuals make sequential decisions, and each individual observes the decision made by individuals earlier in the sequence. In a population where one person has an incorrect private signal, and all the other people have correct signals, the entire population will be led astray if the one person with the incorrect signal is the first person to make a decision—even when each person is acting rationally according to Bayesian principles (Banerjee, 1992). Because that first person informs the second person, who then makes an incorrect decision, this initiates an informational cascade (Bikhchandani, Hirshleifer, & Welch, 1992) that leads the entire population to make an incorrect decision, even if most individuals would make the correct decision independently. Like Janis' research on Groupthink (1982), however, the dangers of herding are also consistent with the expected benefits of decentralized networks: in a sequential interaction, the first person is more influential than the rest of the population.

The Delphi Method. Given the potential risks of social influence, an important area of research has been the identification of methods that allow groups to actively process information while avoiding herding and Groupthink. One early approach was termed the "Delphi" method (an homage to the Oracle of Delphi), a process which emerged out of military research designed to allow panels of experts to generate unbiased forecasts (Dalkey & Helmer, 1963). The method is simple: estimates are first obtained from experts by way of interview or written questionnaire; each expert's belief is then shared with the other experts via a mediator or written information exchange, thus avoiding the risks associated with direct interpersonal communication (such as normative pressures); and each expert is then asked to provide a revised belief.

In general, the Delphi method has provided more accurate estimates than independent individuals, but research on this process has produced inconsistent results (for a review, see Rowe & Wright, 1999). This inconsistency can be traced to methodological limitations leading to ambiguity on two crucial points. First, the precise method by which information was exchanged varied across studies. It is thus difficult to know why many but not all studies found that the Delphi method improved accuracy, or to determine what method of information exchange was most effective. Second, these studies frequently lacked sufficient experimental control, limiting causal inference. When comparing "staticized" or "nominal" groups (i.e., groups of independent individuals) with interacting groups, it was common practice for staticized groups to answer the question only once, while interacting groups had the opportunity to reconsider and revise their beliefs (see, e.g., Gustafson, Shukla, Delbecq, & Walster, 1973). Thus it is impossible to determine from these studies whether the improved accuracy resulted from truly social learning effects or whether they can be explained by individual learning effects. Contemporary research, however, has begun to combine formal models of social influence with carefully controlled experiments to provide a more thorough understanding of the effect of social influence on the wisdom of crowds.

Numeric Estimates in Networks. A core feature of Banerjee's herding model is that individuals make decisions once, sequentially. In contrast, DeGroot (1974) studies a simple model of opinion formation in which each individual begins with an independently generated belief, and then revises their belief by combining their own initial belief with the beliefs of some subset of the population. This subset can either represent the set of individuals who are observable to them (as in a sparse communication network) or, alternatively, the set of peers who are influential (i.e., a person may observe everybody in the population, but be influenced only by a subset). DeGroot showed that under very general conditions, a group following this process indefinitely will asymptotically converge on a single, shared belief which is a weighted mean of the group's initial, independent belief distribution. The weight that each person contributes is determined by their eigenvector centrality in the network adjacency matrix. Although DeGroot was not studying accuracy specifically, this model is readily adapted to examine the effect of social influence on the wisdom of crowds (Becker et al., 2017; Golub & Jackson, 2010).

Under the assumption that the sum of each person's outgoing ties (i.e., the total influence a person has on others) is identical, then each individual is equally central and thus contributes equally to the final group mean. Under such circumstances, the group will converge on the mean of independent beliefs, and the mean itself will not be changed by social influence. Thus if individuals in a population are all equally influential, social influence will improve average individual accuracy without undermining the wisdom of crowds—since the initial group error is necessarily less than or equal to initial average individual error, and they become equivalent asymptotically. This prediction is supported by the results of several experimental studies, including a reanalysis by Farrell (2011) of the data provided by Lorenz et al (2011), and several other studies (Becker et al., 2017; Becker & Centola, 2018; Gürçay, Mellers, & Baron, 2015).

The prediction that decentralized networks converge toward the mean assumes that individuals all respond equally to social influence, or that the response to social influence is distributed identically throughout a population (Becker et al, 2017). However, in another reanalysis of the experimental data produced by Lorenz et al (2011), Madirolas & de Polavieja (2015) found that individuals who hold beliefs closer to the truth also revise their answers less in response to social information, and thus the response to social information is correlated with beliefs. This finding is consistent with the observation by Koriat (2012) and Prelec et al (2017) that individuals who express greater confidence in their beliefs are also, on average, more accurate. This psychological tendency has collective implications. Becker et al (2017) note that individuals who place greater weight on their own belief are more central weighted network of social influence, since the diagonal of the adjacency matrix contributes to centrality—the diagonal indicates network "self-ties" and represents the influence of a person's initial belief on their revised belief. In both simulation and experiment, Becker et al. find that when populations are embedded in networks where everybody has an equal number of ties, the mean belief improves as it is drawn towards the belief of more accurate individuals.

Binary Beliefs in Networks. While research on the collected wisdom of crowds has examined both classification tasks (Koriat, 2012; Kurvers et al., 2016; Prelec et al., 2017) and numeric estimates (Galton, 1907; Klugman, 1947; Nofer & Hinz, 2014), research on social influence has largely focused on numeric estimates. Network models of categorical belief formation have not received the theoretical attention of the DeGroot model, but available research suggests that social influence has fundamentally different effects on categorical beliefs than on numeric estimates.

One popular model known as the "voter" model (Castellano, Fortunato, & Loreto, 2009) assumes that individuals can hold one of two beliefs, and that people update their beliefs asynchronously. With each update, a voter randomly selects a neighbor and adopts their opinion. One variation of this model assumes that each individual adopts the behavior of the majority of their peers. In both of these cases, as one might expect from symmetry, the consensus belief of the population is determined by the initial distribution of beliefs. As a result, consensus (if it emerges) is likely to reflect the initial majority (Castellano et al., 2009; Mossel & Tamuz, 2017). In this respect, models of binary beliefs are comparable to the DeGroot model, which predicts that groups in decentralized networks will converge on the mean of independent beliefs. However, as noted above, people sharing numeric estimates do not all respond equally to social information (Becker et al., 2017; Madirolas & de Polavieja, 2015). As also noted above, for binary beliefs the correlation between accuracy and confidence appears to hold only when the majority is already correct. Thus while inaccurate numeric beliefs appear to self-correct in a population, inaccurate binary beliefs may in fact become more pronounced.

In the course of developing this article, we have been unable to find any experimental tests or theoretical models (besides the voter model and its variants) studying the effect of social influence on the accuracy of binary beliefs in groups in simple information exchange. By simple information exchange, we mean a scenario in which: individuals first form some estimate based on a signal about the state of the world (i.e., available information), then learn about the estimates of others, and then revise their estimates (as in, e.g., the Delphi method). This process of simple information exchange is important because it reflects group processes in which a people discuss available information with negligible introduction of new information. While microeconomic models of social learning (Acemoglu & Ozdaglar, 2011; Golub & Sadler, 2017) have been extensively studied to understand binary belief formation in networks (Bala & Goyal, 1998; Banerjee, 1992; Bikhchandani et al., 1992; Krafft et al., 2016; Mossel, Sly, & Tamuz, 2015), these social learning models typically assume that individual decisions are accompanied by behavior which leads to new information being revealed.

Conditions of Information Exchange. While Janis' theory of Groupthink was developed based on observational studies of deliberation and information exchange in all its empirical richness, the theories and experimental studies discussed above assume that opinions and communication are limited to simple, unidimensional numeric beliefs. These studies provide compelling evidence to indicate that social influence can improve the accuracy of group beliefs as long as information exchange occurs in a controlled fashion. An important question to ask is what method of information exchange produces the greatest increase in group accuracy.

To address this question, Gustafson et al. (1973) compared information exchange via written feedback with information exchange via group discussion. Surprisingly, they found that they found that written feedback offered no improvement over independent estimates, while verbal discussion generated a significant increase in accuracy as compared with independent estimates. The finding that written feedback offered no accuracy benefit contradicts the more recent finding that written feedback improves estimate accuracy (Becker & Centola, 2018, 2018; Farrell, 2011; Gürçay et al., 2015). One explanation for this discrepancy is that Gustafson et al's null result is due to insufficient statistical power (Becker et al., 2017). Another possibility is that variation in the conditions under which information exchange takes place can moderate the effect of social influence on belief accuracy, which would explain both why verbal feedback was beneficial (in contrast with with Janis' [1982] observations of Groupthink) and also why written feedback did not improve estimate accuracy. The potential for verbal deliberation to improve estimate accuracy under controlled conditions has been echoed by Navajas et al. (2018) who found that small deliberating groups produced estimates more accurate than large independent groups, although they did not compare deliberating groups with groups exchanging written information. Navajas et al. also surveyed respondents after they provided their revised estimates, and found that providing explicit arguments to justify their beliefs was an important component of discussion, suggesting that verbal exchange allows people to share information that cannot be carried through numeric reports alone.

Taken together, these numerous experiments and case studies provide strong evidence that information exchange between group members *can* allow groups to actively process the information held by their members, but also that information exchange will not always lead to accuracy improvements. Given the currently inconsistent findings on the relative benefits of numeric feedback versus verbal discussion, an important area for future research will be identifying which forms of social influence allow groups to maximize the benefits of information exchange in order to generate accurate collective beliefs.

Relationship to Other Forms Collective Intelligence

This article has focused on the accuracy of group beliefs, reflecting the vast majority of research on the wisdom of crowds. However, the term "wisdom of crowds" has occasionally been applied more broadly to a range of collective processes such as crowdsourcing (Celis, Krafft, & Kobe, 2016; Kittur & Kraut, 2008; Shi, Teplitskiy, Duede, & Evans, 2017). While these processes fall under the broad umbrella term of "collective intelligence" (Malone & Bernstein, 2015; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010), collective belief formation can be distinguished from other collective processes along a few key dimensions. The first dimension, as discussed above in reference to social learning, is whether or not the group is simply aggregating and processing available information, as in the wisdom of crowds, or whether social dynamics impact information collection as well, as in microeconomic theories of social learning. Another dimension of analysis is the outcome of interest. In collective problem solving, for example, outcomes are measured by an objective payoff function that cannot be reduced to accuracy. A third dimension is the social nature of the task, as in coordination behavior, which is inherently social and does not reflect a task that can be done by independent individuals. While these dimensions are not intended to be exhaustive, they serve to further define the idea of the wisdom of crowds by providing contrast against these other types of collective intelligence processes.

Problem Solving. A related problem in collective intelligence is the ability for groups to identify solutions to problems. Early research on problem solving focused on brainstorming, which is a process in which individuals collectively list as many ideas as possible to solve a particular problem (Stroebe, Nijstad, & Rietzschel, 2010). As with research on the wisdom of crowds, research on problem solving has been characterized by an emphasis on the role of social influence. Research on brainstorming has identified some conditions under which social influence inhibits idea generation, and some conditions under which interacting groups can produce more ideas than equivalently sized groups of independent individuals. While brainstorming research focuses only on the number of ideas generated, Lazer & Friedman (2007) model group problem solving with an emphasis on the quality of the ideas generated. They formally model problem solving as exploration in “NK Space” (Kauffman, 1993) which characterizes a multidimensional search for solutions to any particular problem and can be parameterized to represent simple or complex problems. Where group accuracy depends chiefly on the centralization of information exchange networks (Becker et al. 2017), problem solving depends primarily on the structural efficiency of information exchange networks, with the counterintuitive finding that increased efficiency leads to herding on suboptimal solutions (Lazer & Friedman, 2007). This type of model can be applied to many crowd-sourcing problems, such as FoldIt and GalaxyZoo (Khatib et al., 2011). Wikipedia, a prominent crowd-sourcing platform, has occasionally been described as a “wisdom of crowds” phenomenon (Kittur & Kraut, 2008; Niederer & Van Dijck, 2010; Shi et al., 2017) but is better described in terms of brainstorming and problem solving.

Using Lazer & Friedman's (2007) model, problem solving can be formally distinguished from belief formation in two ways. One distinction is the ability for estimating groups to produce optimal (accurate) outcomes by averaging two suboptimal (inaccurate) estimates, a fundamental principle of the wisdom of crowds. No such concept exists for problem solving: there is no sense in which "errors cancel out," and two suboptimal solutions cannot be combined with any guarantee of forming an optimal solution. Second is the effect of increasing group sizes. As the number of individuals contributing to a crowd-sourced estimate increases, the group average will converge on the expected bias as measured in Equation 1, with a diminishing marginal return: an infinitely large group will not yield a perfectly accurate answer. In contrast, increasing the size of problem solving groups will increase indefinitely the likelihood of hitting upon the optimal solution. As the number of individuals in Lazer & Friedman's model increases, the probability that the optimal solution is found approaches certainty.

Coordination. Another important process in collective intelligence is coordination behavior (Lewis, 1969), which drives social processes such as language conventions (Centola & Baronchelli, 2015) and resource usage (Hardin, 1968). In one classic problem, people must decide whether or not to go to a popular bar (Casti, 1996). If too many people attend the bar, it is too crowded to be fun; if too few people attend, there are not enough peers to sustain a party. One fundamental coordination problem is represented by pure coordination games, which are a widely studied game theoretic model used to characterize social and technological conventions. Pure coordination games capture behaviors with positive externalities—i.e., where the payoff to a behavior increases monotonically as other people adopt the solution (Young, 1993). Examples include language conventions, in which a word is only useful if it can be understood, and social media platforms, which only hold value if other people use them. Theoretical models have argued that, like problem solving and estimation tasks, the probability of optimal coordination depends on network structure. However, the factors that promote optimal coordination differ from the factors that promote optimal belief accuracy or solution generation. While evolutionary models initially argued that the best solution will spread in infinite time (Blume, 1993; Kandori, Mailath, & Rob, 1993; Young, 1993), subsequent analyses found that only sparse, highly clustered networks favored the ability for optimal solutions to spread (Ellison, 1993, 2000; Montanari & Saberi, 2010). In contrast, populations can remain stuck in suboptimal equilibria for long periods of time in highly dense or random networks.

While the empirical and theoretical characteristics of crowd estimates and coordination problems differ on many dimensions, one of the most fundamental differences lies in the payoff structure. In belief formation, the accuracy of any individual's estimate is independent from the behavior of other group members. For coordination behavior, not only is the payoff of a strategy dependent on the behavior of other group members, but the process itself is fundamentally undefined for an individual acting in isolation.

Conclusion

The defining characteristic of research on the wisdom of crowds as compared with other models of collective intelligence is an interest in the accuracy of group beliefs given some information set. In network models of belief formation, the assumption of fixed private signals distinguishes the wisdom of crowds from microeconomic models of social learning, in which belief updating involves the continual observation of new information. The focus on accuracy distinguishes the wisdom of crowds from models of coordination and collective problem solving, in which beliefs are assumed to be accurate, with other features of social influence forming the locus of attention. However, even as the wisdom of crowds is a distinct phenomenon that can be studied in isolation from other collective intelligence processes, collective belief formation is also a fundamental process that underlies many different behaviors and decisions. When solving complex problems or making coordination decisions, for example, individuals within groups must form accurate beliefs about the payoff of different strategies under consideration. One important area for future research on the wisdom of crowds and collective intelligence more broadly is to understand how these different process interact to form decisions that may require estimation, coordination, and problem solving simultaneously.

With regard to belief accuracy and the wisdom of crowds in particular, the two research paradigms described in this review each reflect distinct methodological and theoretical perspectives as well as distinct practical goals. Research on the aggregation of independent beliefs is distinct in that it adopts a machine learning perspective and is commonly aimed at allowing outside parties to aggregate individual beliefs to produce crowd sourced estimates. In contrast, research on collective belief formation is less focused on the optimization of crowdsourced estimates, and instead seeks to understand the endogenous processes through which organizations and societies can process information without the assistance of a centralized aggregation mechanism. While this process itself can be harnessed to generate accurate estimates for an external decision-maker, it also provides cautious but optimistic guidelines to promote optimal collective intelligence in groups.

Despite the growing popularity of research on the wisdom of crowds, many processes of collective belief formation remain unexplored, including both the difference between various forms of communication and also the effect of social influence on binary belief formation. Nonetheless, a large body of research points consistently to one common factor in the formation of accurate group beliefs: informational diversity. From the diversity prediction theorem to decentralized network structures, research on the wisdom of crowds has found that collective beliefs will be most accurate when groups take full advantage of all the information held by each individual member.

CHAPTER 2: NETWORK DYNAMICS OF SOCIAL INFLUENCE IN THE WISDOM OF CROWDS

Abstract

Since Galton's discovery of the "wisdom of crowds" theories of collective intelligence have suggested that in order for a group to be accurate, individuals within the group must be either independent, with uncorrelated beliefs, or diverse, with negatively correlated beliefs. Previous experimental studies argued that social influence undermines the wisdom of crowds, showing that individual estimates became more similar when subjects observed each other's beliefs, reducing diversity without a corresponding increase in group accuracy. In contrast, we find general network conditions under which social influence can improve group estimates, even as individual opinions become more similar. We present theoretical predictions and large scale experimental results showing that in decentralized communication networks, group estimates become more accurate as a result of exposure to social information. We also present results showing that in centralized networks, the influence of central individuals dominates the collective estimation process, and group estimates are as likely to increase in error as they are to become more accurate.

Introduction

Research on crowdsourcing (Sunstein, 2006), prediction markets (Wolfers & Zitzewitz, 2004), and financial forecasting (Nofer & Hinz, 2014) has found that the aggregated judgements of many individuals can be more accurate than the judgements of individual experts. This phenomenon, known as the "wisdom of crowds" (Surowiecki, 2004) has been observed to occur in areas ranging from medical decisions (Kurvers et al., 2016; Wolf et al., 2015) and geopolitical predictions (Mellers et al., 2014; Sjöberg, 2009) to sports betting (Herzog & Hertwig, 2009). However, harnessing the wisdom of crowds remains a challenge: previous theorists have argued that group estimates can be accurate only as long as individuals within the group do not communicate (Surowiecki, 2004; Lorenz et al, 2011).

Statistical explanations for the wisdom of crowd suggest that group estimations will be accurate as long as individuals within the group contribute estimates that are sufficiently diverse, so that individual errors cancel out leaving an accurate group estimate. To illustrate this point, Page (2007) developed the "diversity prediction theory," which states the following:

$$\overbrace{(c-\theta)^2}^{\text{collective error}} = \overbrace{\frac{1}{N}\sum_{i=1}^{N}(B_i-\theta)^2}^{\substack{\text{average}\\ \text{individual error}}} - \overbrace{\frac{1}{N}\sum_{i=1}^{N}(B_i-c)^2}^{\text{diversity}}$$

where B is the vector of estimates by N individuals; c is the collective estimate, defined as the group mean $\bar{B}$; and θ is the true value. The first term in the equation, on the left, represents the squared error of the crowd estimate; the second term in the equation represents the average squared error of individuals; and the third term in the equation represents diversity, which takes a form nearly identical to the common formula for statistical variance.

This mathematical identity (identical in form to the standard partitioning for the sum of squares in any predictive model) highlights two important points. First, the diversity prediction theorem shows that the error of the group mean is guaranteed to be smaller than the average error of individual estimates. This equation does not require that the group mean will always be more accurate than any individual estimate, but it does establish minimum expectations for the accuracy of crowd estimates. Second, the diversity prediction theorem shows that diversity, which is comparable to statistical variance, is a crucial component of crowd wisdom. In situations where every individual is accurate (low average error), there will be low diversity but high collective accuracy, because individuals hold similar beliefs, and those beliefs are true. However, in situations where individuals tend to be very inaccurate – precisely those scenarios in which collective wisdom is most beneficial – the collective estimate will only be accurate if diversity is very high.

Social Influence and the Wisdom of Crowds

One major risk faced by decision-making bodies stems from the individual tendency toward conformity (Janis, 1972; Surowiecki, 2004; Sunstein, 2006). In perhaps the most famous experimental study of conformity, Asch (1951) demonstrated that subjects were willing to provide an obviously false response on a question of factual accuracy after being able to observe that everyone else in the study disagreed with their initial belief. The observation that people will adjust their beliefs to become more similar to social referents has been replicated in a number of studies with a range of experimental designs, using both discrete choice and continuous response tasks (Deutsch & Gerard, 1955; Jenness, 1932; Lorenz et al., 2011; Myers & Bishop, 1971; Sherif, 1935).

The tendency toward conformity presents a challenge for organizations and other groups that want to take advantage of the wisdom of crowds. In his pioneering set of case studies on group decision-making, Janis (1972) identified "Groupthink" as the cause behind some of the most catastrophic group decisions in history, such as the infamously botched Bay of Pigs invasion. While the process of Groupthink is characterized by several mechanisms, they all are driven by a need for group cohesion. In particular, the pressure towards conformity serves to reduce the diversity of information available within a group. This emerges not only from the conformity effects described above, where individuals over time gradually hold increasingly similar beliefs, but also due to a tendency for group norms to suppress the introduction of new information which appears to deviate from the beliefs of the group as a whole.

The case studies described by Janis (1972) and Surowiecki (2004) demonstrate that conformity can lead groups to neglect the diversity of information contained within their members, but this anecdotal evidence is not sufficient to support a general claim that the wisdom of crowds can only emerge when individuals are independent (Surowiecki, 2004). To develop a more comprehensive theory of communication and collective intelligence, it is necessary to distinguish between two forms of social influence: normative influence and informational influence (Deutsch & Gerard, 1955).

Normative influence is driven by the goal of conformity for conformity's sake (e.g., to avoid sanction) and is marked by an outward compliance toward group beliefs, but does not necessarily entail a change in private beliefs (Deutsch & Gerard, 1955). While outward compliance can ultimately engender a change in a person's true belief, due to the need for cognitive consistency (Festinger & Carlsmith, 1959) or the dynamics of self-perception (Bem, 1967) these cognitive processes are quite distinct from the effects of informational influence. While normative influence shapes outward conformity through a desire for group cohesion, informational influence directly shapes a person's true beliefs. This occurs because the beliefs and behaviors of peer referents may be rationally interpreted as a valid signal indicating the best decision. In the case of judgements where there is a true answer, people recognize their own potential for error, and so descriptive social norms – that is, beliefs about what other people believe – offer real informational value, because of their perceived potential to improve accuracy (Cialdini & Goldstein, 2004).

Experimental evidence suggests that both normative and informational influence lead to decreased diversity of beliefs within a group. In a replication of the original Asch (1951) conformity study, Deutsch and Gerard (1955) compared the effects of social influence under two different conditions. In the "normative influence" condition, subjects exchanged information about beliefs while being able to see each other; in the "informational influence" condition, subjects exchanged information anonymously. Deutsch and Gerard found that both conditions induced conformity on beliefs, but that anonymous interaction produced significantly weaker effects. These results open the door for the possibility that the dangers posed to collective decisions may be limited to the context of normative pressure, and may not appear when social influence is limited to informational influence.

Despite the potential for loss of diversity, Gustafson et al (1973), offers suggestive evidence that social influence can improve the accuracy of group estimates when interactions are limited to information exchange. Remarkably, this experiment showed that estimates generated after group discussion were actually more accurate than estimates by independent individuals! However, because independent individuals submitted only one estimate, while the groups completed the task multiple times, it is not possible to determine whether their results were truly due the effect of social influence, or simply the result of individual learning. Moreover, the groups were very small, and without a clear mechanism, it is difficult to generalize these results to large networks.

In the most comprehensive study to date, Lorenz et al (2011) conducted an experiment in which groups of 12 individuals each completed a series of estimation tasks on matters of factual accuracy. In each task, the subjects were asked to provide an initial response and then revise their answer several times. In set of trials, they observed the beliefs of other subjects between revisions. In a control condition, the subjects revised their beliefs several times, but were not exposed to social information. With this design, Lorenz et al were able to directly assess the effect of informational influence on the accuracy of the mean estimate within a group, while controlling for the effect of individual learning.

In contrast with Gustafson et al (1973), Lorenz et al (2011) found no evidence that social influence improved the accuracy of collective estimates. However, they observed that informational influence reduces the diversity of estimates within a group. The authors argued on this basis that social influence undermines the wisdom of crowds through the range reduction effect, in which a greater number of individual beliefs is required to bracket the true answer; and the “social influence effect” which refers to "the fact that social influence diminishes diversity in groups without improving its accuracy” (p.9022).

Lorenz et al argue that social influence decreases diversity, but that diversity is required for the wisdom of crowds to be effective, and thus social influence undermines the wisdom of crowds. However, a careful consideration of the diversity prediction theorem suggests that social influence nonetheless can improve the wisdom of crowds. First, the diversity prediction theorem indicates that individuals may become more accurate. In order for group diversity to decrease, one of the other two terms in the equation must change – either collective error or average individual error. If group accuracy does not change, then average individual error must decrease. A subsequent re-analysis of the publicly available data from the Lorenz et al (2011) study confirms this hypothesis, that individual error did in fact decrease (Farrell, 2011) a finding which has since been replicated by other researchers (Gürçay et al., 2015).

However, the finding by Farrell (2011) still leaves the effect of social influence on collective accuracy undetermined. Depending on how much individual error was reduced in comparison with the decrease in diversity, it remains possible that the collective error either increased, decreased, or remained unchanged.

To address this question, we build on a formal model of social influence (DeGroot, 1974) to test the effect of informational influence on collective accuracy. Under the most general assumptions, group beliefs are expected to converge on the group mean. As a result of the skewed as distributions observed in previous studies (Galton, 1907; Lorenz et al, 2011) this model predicts that median estimate should become more accurate as the group converges toward the mean of independent estimates. By calibrating this model with publicly available data from previous experiments, we find that the results reported by Lorenz et al did indicate Type II error, due to insufficient statistical power (see Appendix figure A7).

In addition to predicting that the median estimate will improve under the most general assumptions, this model also predicts that group mean itself will become more accurate as a result of informational influence. In research on the aggregation of estimates by independent individuals, self-reported confidence has been shown to correlate with accuracy (Koriat, 2012). At the same time, individuals participating in group decisions are expected to exhibit a "self-weighting" effect, such that their contribution to collective decisions is moderated by self-perceptions regarding their own expertise (Gustafson et al, 1973). If self-weight is correlated with accuracy, then social influence can produce a reliable increase in the accuracy of collective estimates.

Formal Model of Social Influence

To study the effect of social influence on the wisdom of crowds, we adopt DeGroot's (1974) model of belief formation. In this model, each person *i* in a group of *N* people begins with some independent initial estimate, $B_{0,i}$. After observing the estimates of others, people revise their belief by taking a weighted average of their own initial belief and the beliefs of others in the social network. DeGroot studies the formation of collective beliefs in a weighted, directed influence network, where entries in the network adjacency matrix indicate both whether *i* observes *j*, as well as the strength of influence that *j* has on *i*'s belief. DeGroot showed that when groups revise their beliefs repeated indefinitely, they will asymptotically converge on a weighted mean of the initial independent beliefs. Each person's independent belief contributes a weight equivalent to their eigenvector centrality in the weighted, directed *influence network*. This result indicates that individuals can be more influential by being more central in the communication network as a whole, as well as by being more influential on those in their immediate ego network.

To study the effect of network topology independently from the effect of variation in individual persuasiveness, we define the effect of social influence on a binary, undirected *communication network*. To isolate the effects of informational influence, we assume that subjects place equal weight on all of the beliefs they are able to observe. Thus, each individual's revision can be expressed as a weighted sum of their own estimate and the average estimates of network neighbors:

$$B_{t+1,i} = \alpha_i \times B_{t,i} + (1-\alpha_i) \times \overline{B}_{t,j\in N_i}$$

where $B_{t,i}$ indicates the response of subject *i* at time *t*; α_i indicates the self-weight a subject places on their own initial estimate; (1 - α_i) indicates the weight they place on the average estimate of their network neighbors; and $\overline{B}_{t,j\in N_i}$ indicates the average estimate of subject *i*'s network neighbors at time *t*. The group revision process is fully determined by three parameters: the vector of initial beliefs B; the vector of self-weight, α; and a binary adjacency matrix A, which defines who can observe whom in the communication network.

Theoretical Predictions

The results obtained by DeGroot (1974) indicate that if everyone is equally central in the influence network, than a group will converge on the arithmetic mean of initial, independent beliefs. This "decentralization" (Freeman, 1978) can be achieved when individuals are all equally connected in the communication network and also equally influential on their immediate peers. We find that convergence towards the mean holds more generally under the assumption that self-weight (α) is independently, identically distributed (i.i.d.) throughout the population. In previous studies, the distribution of independent estimates was found to demonstrate a skew such that median belief not only underestimated the true value, but was also lower than the mean belief. As a result, convergence toward the mean has the effect of improving the accuracy of the median estimate.

Thus, one possibility is that self-weight is distributed i.i.d., and groups thus converge towards the mean, improving both individual estimates (Farrel, 2011; Gurcay et al, 2015) as well as the median estimate, an important indicator of the wisdom of crowds (Galton, 1907). However, if self-weight is not distributed i.i.d., then individuals who systematically place more weight on their own belief will also have a stronger influence on collective belief. The importance of self-weight is due to the fact that each individual's eigenvector centrality in the influence network is determined in part by self-weight, which is represented as a "self-tie" in the weighted influence network. People who weight themselves more highly will, correspondingly, give less weight to social information, thus decreasing the centrality of their peers and making themselves more central in the weighted social influence network.

Previous studies have observed a correlation between self-reported confidence and accuracy (Koriat, 2012). One important possibility, then, is that self-weight is not i.i.d., but is instead correlated with belief. The model described above predicts that when error and self-weight are negatively correlated (so that people with more error have less self-weight), social influence in decentralized networks will increase group accuracy.

While beliefs in decentralized networks may thus be shaped by heterogeneity in the individual response to social information, group beliefs in centralized networks are overwhelmingly determined by the beliefs of those individuals most central in the communication network. Even when accuracy is correlated with self-weight, highly centralized networks only improve when the central individuals hold a belief that pulls the group toward truth.

Experimental Design

We recruited 1,360 participants from the World Wide Web to take part in a series of estimation challenges. Subjects were randomized either to one of two experimental social network conditions, or to a control condition. In all conditions, participants were prompted to complete estimation tasks and were awarded a monetary prize based on the accuracy of their final estimate. In the network conditions, participants were placed into either a decentralized network, in which everyone had equal connectivity, or a centralized network, in which a highly connected central member had a disproportionate number of connections.

Each social network contained 40 subjects. Within each network, all subjects were simultaneously shown the same image prompt (e.g., a plate of food) and asked to estimate a numerical quantity (e.g., the caloric content). There were three rounds for each estimation task. In Round One, participants provided an independent estimate based on the prompt. In both network conditions, participants were then shown the average estimate of the peers directly connected to them in their social network, and prompted to submit their answers again in Round Two. Subjects were then shown the average of their peers' revised estimates, and prompted to submit a final estimate in Round Three. Thus, for each question, participants provided one independent estimate and two estimates after exposure to social information, for a total of three estimates per question. Subjects were not provided with any information about their social networks, which ensured that the subject experience was identical across the two network conditions.

Subjects who were randomized to the control condition were not placed into social networks, but were instead given the opportunity to answer the same questions without being exposed to social influence. These participants were still given the opportunity to revise their initial answer two times, providing a total of three independent estimates per question. All control participants observed the same sets of questions in the same order as participants embedded within the experimental networks. More generally, the subject experience in the control condition was identical to that of subjects in other conditions, except that participants were not given any social information.

Each experimental trial of the study consisted of an identical set of questions provided to one decentralized network (40 individuals) and one centralized network (40 individuals). For each set of questions that was asked in the experimental trials, we also collected responses from 40 independent individuals in the control condition, who collectively formed a "control group" for that set of questions. Each subject participated only once in our study—either in one network condition or in one control group—such that every network condition and control group was comprised of a unique set of 40 individuals.

Because subjects in the network conditions were not statistically independent, all analyses of collective estimates in the network conditions were conducted at the group level. Additionally, because each network completed multiple estimation tasks within an experimental trial, we cluster our main analysis at the trial level such that each network provides a single, independent observation. In total, we conducted 13 experimental trials, comprising 520 subjects in each network condition (1,040 experimental subjects in total). In 6 of the experimental trials, subjects answered

4 questions in each trial, where each question set was unique. In the remaining 7 trials, subjects answered 5 questions in each trial, using 2 unique questions over repeated trials. In total, this produced 8 unique question sets.

Control groups were conducted corresponding to each unique question set, producing 8 control groups, each of size 40 (320 control subjects in total). Because subjects in the control groups were independent from each other, fewer overall subjects were required for the control analyses. Nevertheless, for proper comparison with the experimental conditions, we still conducted our control trials with subjects in groups of 40, and conduct our analyses at the group level.

Estimation Tasks. To ensure that our findings are robust to variations in the distribution of estimates, we conducted two sets of experimental trials, using questions that generate distributions with different shapes. In six of the thirteen trials, subjects were given count-based questions (e.g., “how many candies are in this jar?”). Because these are zero-bounded on the left and unbounded on the right, count-based questions generate highly skewed distributions (Lorenz et al, 2011), in which the median is expected to improve even if the mean remains unchanged. In the remaining trials, we asked participants to provide responses to percentage based question (e.g., “what percentage of people in this photograph are wearing hats”). These responses are constrained to fall between zero and one hundred, and did not produce any systematic skew in the distribution of estimates. For a detailed description of the estimation tasks, see Appendix B.

Experimental Analysis. We measure the cumulative effect of social influence on collective judgments by comparing the initial estimates of each group (i.e., in Round One of our study), with the final estimates of each group after two rounds of revision (i.e., in Round Three). For our main experimental outcomes, cluster-robust estimates are generated by testing differences in the average error and average change in error across all estimation tasks completed by each group.

For analyses where we examine correlations across each separate estimation task completed by each group, we obtain cluster-robust estimates by using cluster-robust error and adding a groups-specific intercept to the regression model.

For results where we report percent change, all comparisons are made between final estimates (i.e., Round Three) and independent estimates (i.e., Round One), so that percent change is measured as the magnitude of the change in the estimate divided by the magnitude of the initial estimate. To facilitate comparisons across different estimation tasks of different scales (i.e., some questions have true answers over 1000, while some questions have true answers under 100) we normalize all estimates, dividing them by the standard deviation of the independent responses for each question. All reported changes in error are therefore measured in terms of the distance of each estimate from the truth, represented as the number of standard deviations (s.d.) away from the true answer (comparable to a z-score).

Results

Social network structure significantly affected the wisdom of crowds. We found both that decentralized networks showed the predicted increase in collective accuracy, and that centralized networks exhibited the predicted bias toward the beliefs of central individuals. We begin our analysis by confirming that in the independent round (i.e., Round One of all trials) groups exhibited the wisdom of crowds. Consistent with earlier studies (1,6-9,14), we found that, on average, both the mean and the median of each group's estimate was more accurate than the majority of its members. In the results that follow, we analyze how social influence affected the trajectory of group estimates in each of the network conditions.

Social Influence in Decentralized Networks. Social influence dramatically reduced the diversity of group estimates. As shown in Figure 1D, two rounds of revision significantly narrowed the standard deviation of responses (N=13 trials, P<0.001, Wilcoxon signed rank test), producing a 43% reduction in the average standard deviation between Round One and Round Three. This sizable reduction in diversity replicates the main finding from previous experimental research on social influence in the wisdom of crowds (14).

However, this reduction in diversity did not undermine the wisdom of crowds. Rather, consistent with previous research (29,30), we found that social influence in decentralized networks produced significant improvements in individual accuracy. Across all 13 trials with decentralized networks, average individual error was significantly lower in Round Three than it was in Round One, decreasing by 23% on average (N=13 trials, P<0.001, Wilcoxon signed rank test). In addition to these individual level improvements, we also found that the average error of each group's median estimate was significantly lower in Round Three (0.67 s.d.) than in Round One (0.76 s.d.) (N=13 trials, P<0.001, Wilcoxon signed rank test), resulting in a 12% decrease in average error, as shown in Figures 1A and 1C.

In our analysis of how social influence produced these group-level improvements in the median, our initial expectation was that self-weights were independently and identically distributed within each network. On this assumption, the DeGroot (20) model predicts that social influence in decentralized networks can improve the group median by pushing it towards the mean of the group's independent estimate, which is not expected to change. Remarkably, however, we found that, on average, each group's mean estimate also became more accurate. After two rounds of exposure to social influence, the average error of the group mean at Round Three (0.62 s.d.) was significantly lower than at Round One (0.69 s.d.) (N=13 trials, P<0.01, Wilcoxon signed rank test), resulting in a 10% reduction in the average error of the group mean. These findings can be explained with the DeGroot model by observing that individuals' self-weights were not identically distributed in the population.

Figure 2 shows that across all network conditions the magnitude of an individual's revisions from Round One to Round Three was significantly correlated with the magnitude of their initial error (N=4340 estimates by 1040 subjects, ρ=0.41, 95% CI [0.39, 0.43], P<0.001, Analysis of

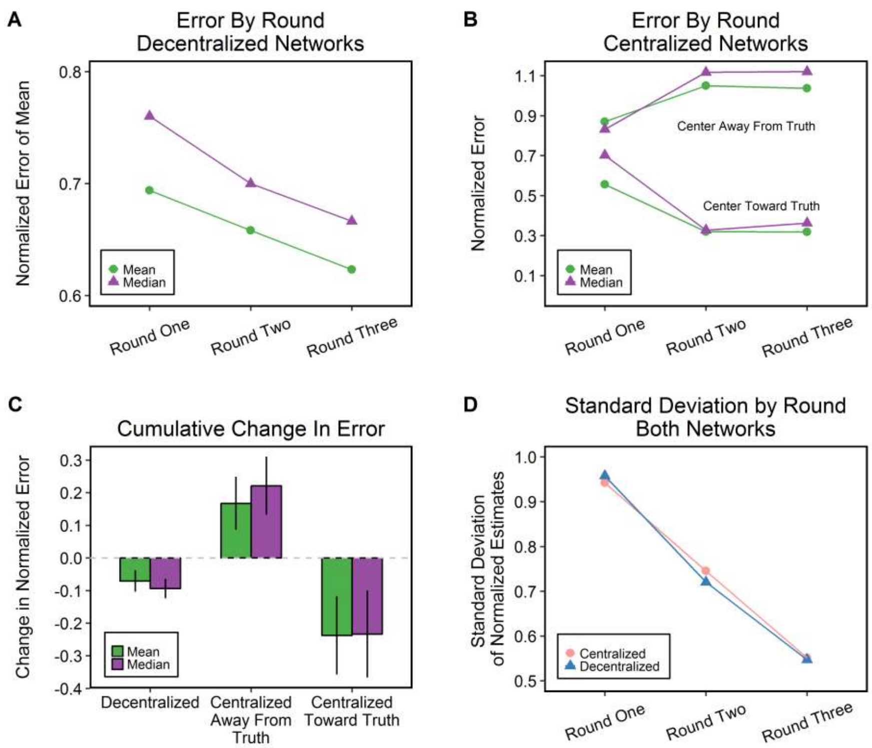

Fig. 1. Effect of social influence on group accuracy. Average error and standard deviation in 13 experimental trials for each network condition. **(*A*)** In decentralized networks, both the mean and the median became more accurate over two rounds of social influence. **(*B*)** In centralized networks, the effect of social influence on the accuracy of the group mean and group median was determined by the initial estimate of the central node. Results are conditioned on whether the central node was in the direction of truth relative to the group estimate. **(*C*)** Total change from Round One to Round Three with bootstrapped 95% error bars, indicating that changes shown in panels A and B are significant. Both the mean and median of estimates in decentralized networks became more accurate (N=13, P<0.01 for mean, P<0.001 for median). For centralized networks, the mean and median became less accurate when the central node provided an estimate in the opposite direction of truth (N=13, P<0.01 for both mean and median). Both the mean and median became more accurate when the central node provided an estimate in the direction of truth (N=12, P<0.01 for the mean and median). **(*D*)** In both network conditions, the standard deviation (i.e., diversity of opinions) decreased significantly after each round of revision (N=13, P<0.001 for both conditions).

Covariance). Because each individual completed multiple estimation tasks, we measure this relationship between individual accuracy and revision magnitude after controlling for correlation between estimates by the same individual. The results (Figure 2) show that initially accurate individuals made smaller revisions to their estimates, while initially inaccurate individuals made larger revisions. Consistent with the DeGroot model, one explanation for this revision pattern is that individuals who were more accurate had greater self-weight in their revisions than individuals who were less accurate. This explanation is consistent with the observed behavior, however our analysis also needs to account for the observation that individuals who were more accurate also had estimates that were closer to their observed neighborhood average. Consequently, the positive correlation between error and revision magnitude may be due to the fact that subjects whose initial estimates were farther from their neighborhood average were inclined to make larger revisions, rather than to the fact that more accurate individuals had a stronger self-weighting.

To control for this potentially confounding effect, we measured the partial correlation between error and revision magnitude, while holding constant the distance between the subject's initial estimate and the initial neighborhood estimate. The inset in Fig. 2 shows that even with this statistical control, more accurate individuals still made smaller revisions to their estimates than less accurate individuals (N=4340 estimates by 1040 subjects, ρ=0.25, 95% CI [0.22, 0.28], P<0.001, Analysis of Covariance). This suggests that accurate individuals placed more weight on their own estimates and less weight on social information. By contrast, less inaccurate individuals had a lower self-weight, and were more influenced by social information. For clarity, we refer to this partial correlation between accuracy and self-weight as the "revision coefficient."

As discussed above, each individual's social influence weight in the network is determined in part by their self-weight, so that individuals who place more weight on their own estimate are also more influential in the collective estimate. When considered in the context of our theoretical model, the correlation shown in Fig. 2 indicates that more accurate individuals had a larger social influence weight in the network, which can pull the group estimate toward a more accurate mean. These analyses suggest a direct positive relationship between the average revision coefficient

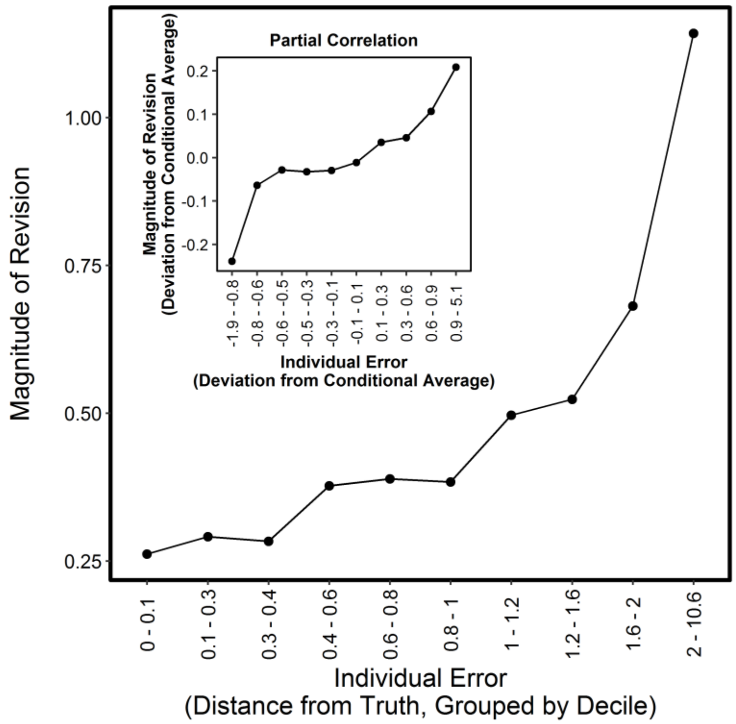

Fig 2. Correlation between revision magnitude and individual error. Each point in the main figure shows the average size of individuals' revisions from Round One to Round Three for individuals located in each decile of the distribution of individual error (i.e., average distance from zero error). Measured for N=4340 estimates provided by 1040 individuals assigned to one of 13 decentralized networks or 13 centralized networks. This figure shows a positive "revision coefficient," such that individuals with greater error in their initial estimates made significantly larger revisions. Controlling for correlation between estimates by the same individual (SI Appendix), we find a positive correlation between individual error and individual revision magnitude (N=4340, ρ=0.41, 95% CI [0.39, 0.43], P<0.001). **Inset:** On the y-axis, positive values indicate larger revisions than would be expected based on the distance between an individual's estimate and their neighborhood estimate. On the x-axis, positive values indicate greater initial error than would be expected given the distance between an individual's estimate and their neighborhood estimate. After controlling for the distance between each individual's initial estimate and the average estimate of their neighborhood, there is still a significant correlation between individual error and individual revision magnitude (N=4340, ρ=0.25, 95% CI [0.22, 0.28], P<0.001).

among the members of a group and the expected improvement in the accuracy of the group mean. Fig. 3A shows, for decentralized networks, the correlation between the improvement in the group mean for each question, and the group's revision coefficient for that question, for each of the 59 group estimation tasks completed in decentralized networks. Because each group completed multiple estimation tasks, these analyses control for correlations across multiple estimates made by the same group.

Consistent with theoretical expectations, the correlation shown in figure 3A indicates that in decentralized networks, groups with higher revision coefficients also exhibited larger improvements in group accuracy (N=59 estimation tasks, ρ=-0.71, 95% CI [-0.82, -0.56]. By contrast, figure 3B shows that centralized networks (as discussed below) exhibited no significant correlation between a group's average revision coefficient and a change in group accuracy (N=57 estimation tasks, ρ=-0.16, 95% CI [-0.33, 0.10]).

Figure 3A indicates that, in decentralized networks, the greater the correlation between individual accuracy and self-weight, the more likely it is that the group mean will improve. Additional simulation analyses (see Appendix figure A2) show that in decentralized networks a positive revision coefficient is sufficient to produce increases in group accuracy consistent with our empirical findings. Notably, across all experimental trials, the average revision coefficient for all subjects was positive suggesting that in very large populations with decentralized networks, social influence is likely to generate consistent improvements in the accuracy of the group mean.

Comparison with Control Condition. These improvements in both the mean and the median, as well as the accuracy of individuals' estimates, all contrast with the results from the control condition (i.e., without social influence). Subjects in the control condition were able to revise their answers several times, but were not provided any information about the estimates of other participants. Between Round One and Round Three, groups in the control condition showed only a small (3%) decrease in average standard deviation, which was significantly smaller than the reduction in diversity in decentralized networks (43%) and centralized networks (42%) (N=21, 13 experimental

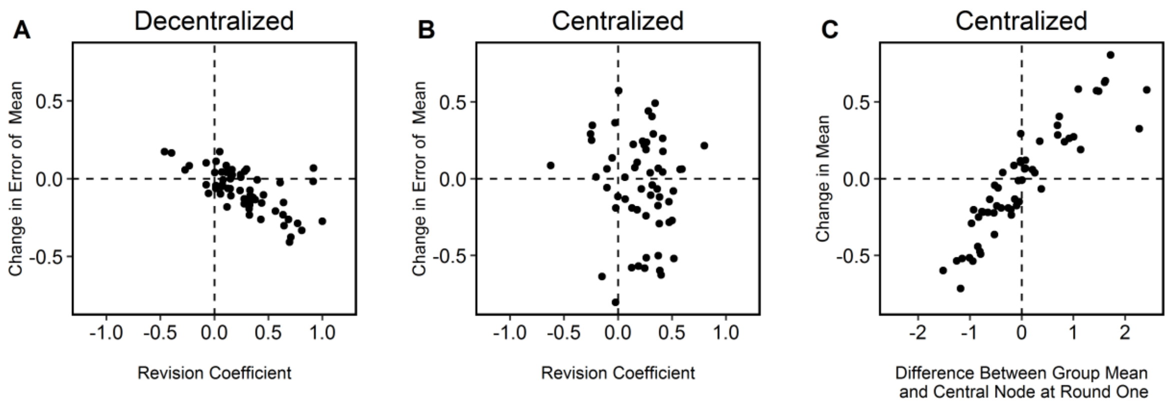

Figure 3. Correlations with changes in group mean. Shown are all 59 estimation tasks completed over 13 experimental trials. In centralized networks, two estimation tasks are omitted where the central node did not provide any response. Decentralized networks show all 59 estimation tasks. **(*A*)** In decentralized networks, the "revision coefficient" for each group estimate – i.e., the partial correlation for all members of a network between individuals' accuracy and their revision magnitudes on a given estimation task – is highly correlated with the change in the error of the group mean (*N*=59, ρ=-0.71, 95% CI [-0.84, -0.51]). On estimation tasks in which groups exhibited larger revision coefficients, they showed significantly greater improvements in the accuracy of the group mean. **(B)** By contrast, in centralized networks, there was no significant correlation between the revision coefficient and the change in group mean (*N*=57, ρ=-0.16, 95% CI [-0.33, 0.10]). **(*C*)** In centralized networks, the change in the group mean is strongly correlated with the behavior of the central node. The difference between the initial group estimate and the initial estimate of the central node is highly correlated with the change in the group's estimate (*N*=57, ρ=0.92, 95% CI [0.88, 0.95]). When central node has an estimate larger than group mean, the group mean typically increased; when the central node is below the group mean, the group mean typically decreased.

trials and 8 control trials, $P<0.001$ for both comparisons, Wilcoxon rank sum test). The opportunity for revision produced a small (3%) decrease in average individual error even in the absence of social information ($N=8$ control trials, $P<0.001$, Wilcoxon signed rank test). However, this improvement was significantly smaller than the 23% improvement by individuals in decentralized networks ($N=21$, 13 experimental and 8 control trials, $P<0.001$, Wilcoxon rank sum test). Moreover, in the control condition, these individual improvements produced no significant changes in the accuracy of either the group mean ($P>0.94$) or the group median ($P>0.64$). These results indicate that the improvements in collective judgment observed in decentralized networks are not explained by independent learning effects, but are due to the network dynamics of social influence.

Social Influence in Centralized Networks. In each centralized network, one randomly selected participant was given disproportionate exposure to the rest of the network by being given many more network ties than other. Because these central individuals had more network ties than other individuals, they had much greater weight in the resulting network of social influence. As expected, the diversity of estimates in centralized networks (shown in Fig. 1D) significantly decreased after social influence ($N=13$ trials, $P<0.001$, Wilcoxon signed rank test), reducing the average standard deviation by 42%. However, averaged over all trials, social influence in centralized networks did not reliably improve either the group mean ($P>0.63$) or the group median ($P>0.78$). Instead, as predicted by the DeGroot model, the effects of social influence were determined by the initial estimates of the central individual.

To analyze these effects, we divided the group estimates in centralized networks into two categories, based on the initial estimate of the central nodes. In one category ("center toward truth") the influence of the central node is expected to increase the accuracy of the group mean. This category includes estimates in which the central node was more accurate than the group mean, and also estimates in which the central node was less accurate, but was on the opposite side of the truth from the group mean. For instance, if the true value is 100 and the group mean is

90, a central node with an estimate of either 105 (more accurate) or 120 (less accurate) will pull the group toward the truth. The second category ("center away from truth") includes trials in which the estimate of the central node pulled the group mean away from the truth. For instance, if the estimate of the central node is instead 70. This analytical strategy was used to identify the effects of social influence on both the group mean and the group median, as reported below.

All 13 trials produced responses to at least one question in which the central individual was away from truth relative to the group estimate, while only 12 trials produced responses where the central individual was towards truth. Accordingly, our analyses for each category use N=13 trials and N=12 trials, respectively. As shown in Fig. 1B and 1C, when the central individual's estimate was toward truth, the average error of the group mean after social influence (0.32 s.d.) was 43% lower than the average error of the group mean before social influence (0.56 s.d.), producing a significant increase in group accuracy (N=12 trials, P<0.01, Wilcoxon signed rank test). Correspondingly, the same analysis for the median showed that the error of the median also decreased significantly by 48% in these group estimations from Round One (0.70 s.d.) to Round Three (0.36 s.d.) (N=12 trials, P<0.01, Wilcoxon signed rank test). Similarly, when the central individual provided an estimate that was away from truth, social influence increased the error of the group mean by 19% and the error of the median by 32% (Fig. 1B and 1C), significantly reducing the accuracy of both the mean and the median of estimates (N=13 trials, P<0.01 for both comparisons, Wilcoxon signed rank test).

Figure 3C shows the effects of the central node on the collective estimate for each of the 57 estimation tasks in which the central node offered a response. As above, because each group completed multiple estimation tasks, these analyses control for correlations between multiple estimations made by the same group. The positive slope in Fig. 3C (N=57 estimation tasks, ρ=0.92, 95% [0.88, 0.95]) indicates that the group estimates in centralized networks moved toward the initial belief of the central individual – i.e., higher estimates by the central node made the group mean increase, while lower estimates made the group mean decrease.

Robustness. To conclude our analyses, we examined the robustness of our theoretical and experimental findings under variations in the network parameters, such as average degree, graph density, and population size. Model simulations predict that graph density and average degree have no effect on the results observed in our experimental study. However, we found that the effects of social influence on the wisdom of crowds are significantly strengthened with larger population sizes. Our analyses indicate that recent small group studies arguing that social influence undermines the wisdom of crowds (even in a decentralized network) (Lorenz et al, 2011) were insufficiently statistically powered to identify the improvements in collective accuracy that we found. Simulations based on the publicly available data from these studies show that Type II error can explain the negative findings from previous experiments (see Appendix figure A3).

Discussion

Our study differs in several respects from previous work on the network dynamics of collective intelligence. Unlike research on social coordination (Centola & Baronchelli, 2015; Dall'Asta, Baronchelli, Barrat, & Loreto, 2006; Judd, Kearns, & Vorobeychik, 2010) and group problem solving (Bavelas, 1950; Lazer & Friedman, 2007; Shore, Bernstein, & Lazer, 2015), our study does not consider situations where social interaction is necessary for groups to achieve a collective outcome. Instead, we identify how the network dynamics of social influence can affect collective estimation tasks in situations where social influence has been predicted to have a negative effect on the quality of group judgments (Baddeley, 2010; Janis, 1982; Lorenz et al., 2011; Myers & Bishop, 1971; Sunstein, 2006; Surowiecki, 2004).

Our finding that groups have the ability to generate accurate estimates even in the presence of social influence has useful implications for the design of several kinds of collective decision processes. As described in previous studies (Lorenz et al., 2011), if social influence did indeed undermine the wisdom of crowds, then democratic institutions and organizational decision procedures could be improved by preventing people from communicating during a voting process.

Based on these ideas, commercial and non-profit organizations have implemented automated aggregation tools in order to collect individuals' independent beliefs in ways that minimize the information exchanged between them (Bonabeau, 2009). Our findings argue against this approach to aggregation. In contrast, we have shown how social learning in networks can amplify the influence of accurate individuals, leading to both individual and collective judgments that are more accurate than those which could typically be obtained by independent aggregation alone. We therefore anticipate that process interventions within political discussion settings (Fishkin & Luskin, 2005) and organizational decision contexts (Green, Armstrong, & Graefe, 2007) may benefit more from approaches that manage communication networks, rather than approaches that attempt to increase independence in the aggregation process.

Appendix A: Estimation Tasks and Subject Experience

The experimental interface is shown in Figure A1, indicating Round 1 as an example. For each question, participants first provided an independent response without any social information (Round 1). Then, subjects in network conditions were shown the average (mean) response of their network neighbors, and were given a chance to revise their answer (Round 2). This design ensured that users in both network conditions received identical user experiences. This second step was repeated, providing a total of three answers from each individual for a given question, and the screenshot of the final chance to revise after social exposure (Round 3). Subjects had one minute to provide each response, such that each round lasted a maximum of one minute. This entire process was repeated for four unique count-based or five unique percentage-based questions, resulting in a total of twelve count-based or fifteen percentage-based responses from each subject.

In trials where we provided count-based estimation tasks, each group completed four tasks. We conducted 6 independent experimental trials of this kind of task, with four questions each, producing a total of 24 count-based estimations by decentralized networks, and 24 count-based estimations by centralized networks. We used a unique question set for each trial, yielding 6 unique question sets. To create independent control groups for each question set, we ran 6 independent control groups, each with 40 individuals, producing 24 control group estimations.

In trials where we used percentage-based estimation tasks, each group completed five estimation tasks. We conducted 7 independent experimental trials of this kind of task, with five questions each, producing a total of 35 percentage-based estimations by decentralized networks, and 35 percentage-based estimations by centralized networks. We used 2 unique question sets, which were randomly assigned across trials. One set was used in three of the trials, the other was used in four of the trials. To create independent control groups for each question set, we ran 2 independent control groups, each with 40 individuals, producing 10 control group estimations. Because control groups are composed of statistically independent individuals, we only require a single control group for each question set to compare to the experimentally replicated trials. In total, we observed 59 estimations by decentralized networks, 59 estimations by centralized networks, and 34 estimations by control groups.

For count-based estimation tasks, we used four different image prompts: (A) a picture of food, asking participants to estimate the number of calories; (B) a bowl of coins, asking participants to estimate the number of coins; (C) a jar of candies, asking participants to estimate the number of candies; and (D) a picture showing several consumer goods, asking participants to estimate the total cost of all the items. A unique set of image prompts based on these four categories was used for each of the six trials in which subjects completed count-based estimation tasks.

For percentage-based questions, we used five estimation tasks, each with its own image prompt: (A) an image of dots of two colors; (B) a crowd of people (C) a crowd of people holding umbrellas; (D) the numbers 1 through 10 repeated many times in different colors; and (E) a rectangle with a dark purple and a light purple segment. Two questions were created for each image prompt, one asking subjects to estimate a percentage-based parameter, and the other asking for the complement of that parameter. For example, in image prompt (B) showing a crowd of people, in three trials subjects were asked "What percentage of people in this photograph are wearing hats?" and in four trials subjects were asked "What percentage of people in this photograph are not wearing hats?" By using question sets that asked participants to estimate either a value or its complement, we obtained a wider variety of estimate distributions.

Appendix B: Theoretical Predictions for Social Influence in the Wisdom of Crowds

To model the effect of social influence on group accuracy, we first estimate the self-weight (α_i) each individual placed on their own belief for each estimate in our experimental study. We then measure the relationship between error and this self-weight. We then use the coefficient from this regression in our simulation, adding a variable noise term to allow us to continuously vary the strength of the correlation between -1 and +1. Our simulation results show that a strong positive correlation between accuracy and self-weight is sufficient to generate an improvement in collective accuracy, while a strong negative correlation leads to an increase in error. When this correlation is zero, the group converges on the mean in decentralized networks. In contrast, consensus beliefs in centralized networks are determined almost entirely by the network structure and the distribution of individual beliefs. Even when there is a strong positive or negative correlation between accuracy and self-weight, collective beliefs after social influence are largely determined by the belief of central individuals.

*Model Definition***.** To identify theoretical expectations for the effect of social influence on the wisdom of crowds, we use agent-based simulations to model the change in group mean and median under a range of assumptions. In particular, we vary several parameters: network structure, including centralization, density, and average degree; initial opinion distribution shape, including normal (symmetrical) and log-normal (asymmetrical); the accuracy of the collective estimate prior to social influence; the correlation among individuals between error and revision magnitude; and the decay in individual responsiveness to social influence (i.e., the increase in self-weight) over time.

As described in the main text, our model of collective judgments builds on DeGroot's (1974) formalization of local information aggregation, in which an agent *i* updates their estimate, $R_{t,i}$, after being exposed to the estimates of their network neighbors, $\overline{R}_{t,j \in N_i}$. We define an agent's revision process with three components: their own estimate; the estimates of network neighbors; and "self-weight," or the amount of weight they place on their own estimate relative to the estimates of their network neighbors. Each agent responds to social information by adopting a weighted mean of their own estimate and the estimates of their neighbors, according to the rule:

$$R_{t+1,i} = \alpha_i \times R_{t,i} + (1-\alpha_i) \times \overline{R}_{t,j \in N_i}, \tag{1}$$

where the value $R_{t,i}$ indicates the response of agent *i* at time *t*; α_i indicates the self-weight an agent places on their own initial estimate; $(1\text{-}\alpha_i)$ indicates the weight they place on the average estimate of their network neighbors; and $\overline{R}_{t,j \in N_i}$ indicates the average estimate of agent *i*'s network neighbors at time *t*. Outcomes are therefore determined by three parameters: the communication network (i.e., who can observe whom), the distribution of independent estimates $R_{1,i}$ and the distribution of self-weights α_i.

At the population level, this model describes the dynamics of a group belief as a function of the distribution of initial, independent beliefs and an adjacency matrix defining a network of social influence. In this network, a tie from node A to node B is weighted and directed, and represents the amount of weight node A places on the belief of node B, where the sum of the outgoing tie weights for a single node *i* equals $(1 - \alpha_i)$.

That is, in the network adjacency matrix *A:*

$$A_{i,j} = \frac{1-\alpha_i}{k_i} \text{ whenever } \mathrm{i} \neq \mathrm{j} \tag{2}$$

and

$$A_{i,i} = \alpha_i \tag{3}$$

where k_i equals the total number of network neighbors who are observed by an agent. Since an agent's self-weight equals α_i, and the sum of outgoing ties equals (1-α_i), then we can define an agent's revised belief as a weighted mean of their initial unrevised belief and the average belief of network neighbors.

Using this model of revision, we can estimate the parameter α_i for a given revision by a given individual as a function of their initial estimate $R_{t,i}$, their social signal $\overline{R}_{j\in N_{i,t}}$, and their revised estimate $R_{t+1,i}$. Rearranging equation 1 above shows the solution

$$\alpha_i = \frac{R_{t+1,i} - \overline{R}_{t,j\in N_i}}{R_{t,i} - \overline{R}_{t,j\in N_i}} \tag{4}$$

and we use this equation to estimate self-weights in our empirical data. Because our theoretical model assumes that $0 \leq \alpha \leq 1$, we discard values that fall outside this range in the remainder of this analysis. Over 80% of estimated values fall between zero and one.

Using standard OLS regression, we estimate the following relationship between error and α in our experimental data:

$$\alpha_i = 0.74 - 0.05\varepsilon_i \tag{5}$$

where ε_i indicates the absolute value of the error for an estimate by agent *i*. For count-based response distributions, $\varepsilon_i = |\ln(R_i) - \ln(\text{truth})|$ where ln indicates the natural log function, while for symmetric distributions $\varepsilon_i = |R_i - \text{truth}|$ where R_i indicates the estimate by agent *i*. We use this relationship between error and self-weight in our simulations, so that each simulated agent's α is determined according to this empirically estimated model and their randomly generated estimate.

We simulate outcomes for two conditions, one in which initial responses (R_1) are sampled from a skewed distribution and one in which initial responses are sampled from a symmetric distribution. For the skewed distribution, we sample a log-normal distribution (shape parameters μ=6.1, σ=0.7; mean=600; s.d.=500) and for the symmetric distribution we sample a random normal distribution with equivalent mean and standard deviation (μ=600, σ=500). These parameters generate estimate distributions comparable to those observed in our experimental data, while allowing us to directly test for the effect of a skew. We test three levels of accuracy with respect to the group mean: underestimation (truth = mean + 150), overestimation (truth = mean - 150), and exactly accurate (truth=mean). In the case of the skewed distribution, this means that the median underestimates the true value whenever the mean underestimates the true value; and the median overestimates the true value when the mean overestimates the true value; and the median underestimates the true value when the mean is exactly accurate.

Each simulation is initialized by generating a random binary communication network, assigning each agent a belief according the estimate distribution as defined above, and assigning each agent a value for α from a random distribution as estimated above. A key parameter of interest is the strength of correlation between error and revision magnitude, as described above. To vary correlation, we generate a weighted combination[1] of the value for α as determined by equation 5 (ie, a degenerate random variable) and a random variable drawn from the empirical distribution of α. To generate a negative correlation between error and α_i (ie, a positive correlation between accuracy and α_i) we use the arithmetic complement of equation 5. Once initialized, simulations are deterministic: the vector of agent beliefs, the vector of α_i , the binary network adjacency matrix, and the number of rounds fully determine the outcome of a simulation. Using

[1] Variation in correlation is accomplished with the following algorithm: for each agent, we randomly draw an initial estimate from a distribution that matches our empirical data. Based on the error of this estimate, we generate two values. The first term, A_1, is a fixed determinate value generated according to equation 5 above. The second term, A_2, is random variable sampled from the generated distribution of A_1. The final value for α is defined as $\alpha = wA_1 + (1-w)A_2$ where w is a weight parameter that determines the strength of correlation between error and α. When w=1, correlation=1. When w=0, correlation=0. Since A_1 and A_2 have the same distribution, this varies the correlation between α and error while holding mean α constant.

this process, we simulate outcomes comparable to our experimental design, calculating the outcome after two revisions (three rounds).

*Social Influence in Decentralized Networks***.** To identify the general network dynamics of social influence, Fig. A2 and A3 show the effect of social influence under a range of assumptions about response distribution, group accuracy, and individual behavior for networks with N=1000 nodes. When self-weight is not correlated to accuracy (center point on the x-axis of each panel) the mean of the group is unaffected by social influence. When independent estimates follow a skewed distribution, the group median is drawn toward the mean (Fig. A3, Panels A-C). When more accurate individuals have a higher value for α (correlation > 0) the group mean also becomes more accurate (Fig. A3). When inaccurate individuals make smaller revisions than accurate individuals the group mean becomes less accurate (Fig. A3). Empirically, we found that accurate individuals tend to place more weight on their own beliefs. These simulated outcomes are consistent with our empirical finding (Fig. 3 in the Main Text) that in networks where this correlation was strongly positive, the collective belief became more accurate.

*Effect of Centralization on Group Accuracy***.** To test the effect of network centralization in networks of the same size as our empirical trials, we simulated outcomes for a continuous range of centralization while holding density and population size fixed in networks of size N=40 in a group that underestimates the true value. To illustrate the effect of the most central node's accuracy on the collective change in the group mean, we condition the results on whether the most central node held an initial belief that was in the direction of truth relative to the initial group mean, or whether the central node pulled the group away from truth at the initial round. When the most central member held a belief that represented a movement away from truth (Fig. A4, bottom set of points), the group mean after social influence decreased with centralization, leading to an increase in error. When the most central member held a belief that fell in the direction of truth (Fig. A4, top set of points), the group mean after social influence increased with centralization, leading to a decrease in error. Each panel in the figure shows one of three assumptions about individual error and movement: perfect negative correlation, no correlation, and positive correlation. In decentralized networks, the correlation between accuracy and self-weight determines the effect of social influence. While the potential for improvement is greatest when centralization is exactly equal to zero, this effect is robust to a small amount of centralization. However, as centralization increases, the wisdom of crowds is increasingly determined by the belief of the most prominent individual.

Robustness to Variation in Density and Average Degree. In contrast to the large effect of network centralization, network density (Fig. A5) and average degree (Fig. A6) have negligible effects on group accuracy. To test the robustness of our experimental results, we simulate outcomes in a range of conditions with networks of size N=40. The simulations hold centralization fixed at 0 and increase the number of ties in a random network with homogeneous degree. As density increases, there is a slight increase in the change in the median due only to an increase in the speed of convergence (Fig. A5). In the long run, density has no effect, as asymptotic results are determined only by the eigenvector of the adjacency matrix. The minimal short-term effect is not enough to account for the difference between networks in our empirical observations, and the effect of network centralization more than overcomes the small effect of density in our experimental trials,. Moreover, whatever effect density does have in our experimental outcomes, it acts in opposition to the effects of centralization, making our empirical estimation of the effect of centralization on the wisdom of crowds conservative.

The effects of average degree are similar to the effects of density in direction and magnitude for the change in the group mean (Fig. A6). The simulations hold density fixed at 0.05 (which is the density of the centralized network in our experimental trials), and keep centralization at 0 by using random networks with homogeneous degree distributions. Average degree is increased by increasing the population size, while holding density fixed. This procedure uses population sizes ranging from N=40 to N=1000. We also ran simulations holding density fixed at

0.1, which is the density of the decentralized network in our study, and the results were qualitatively similar.

Appendix C: Supplemental Figures

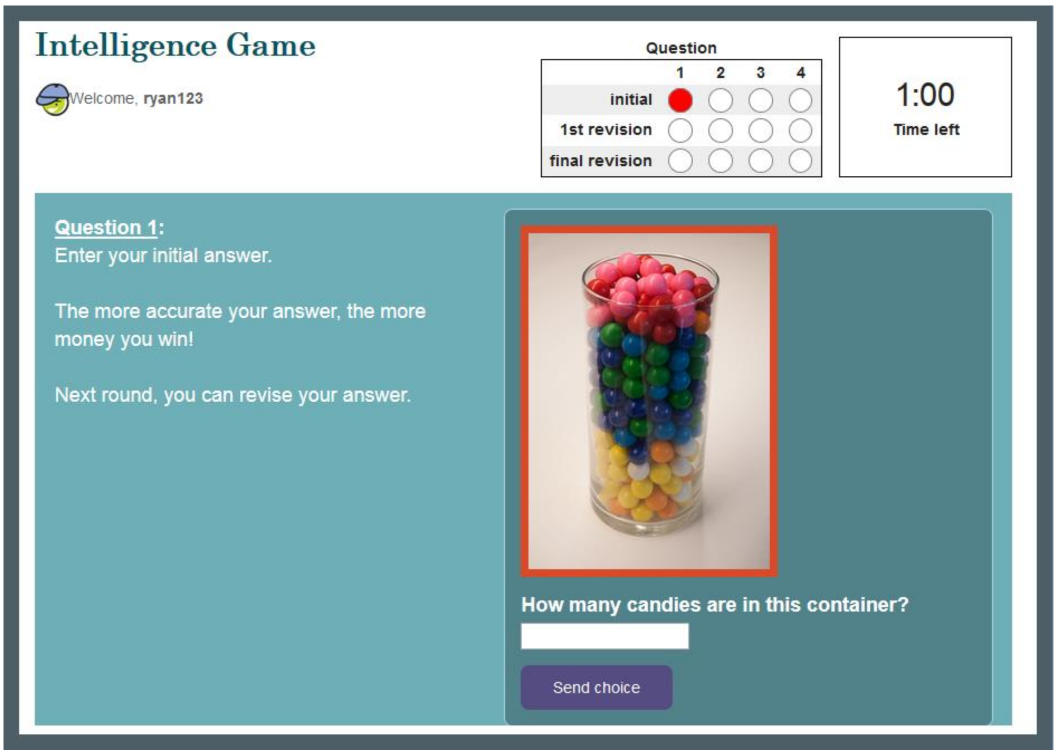

Figure A1. Screenshot of the experimental interface, Round 1.

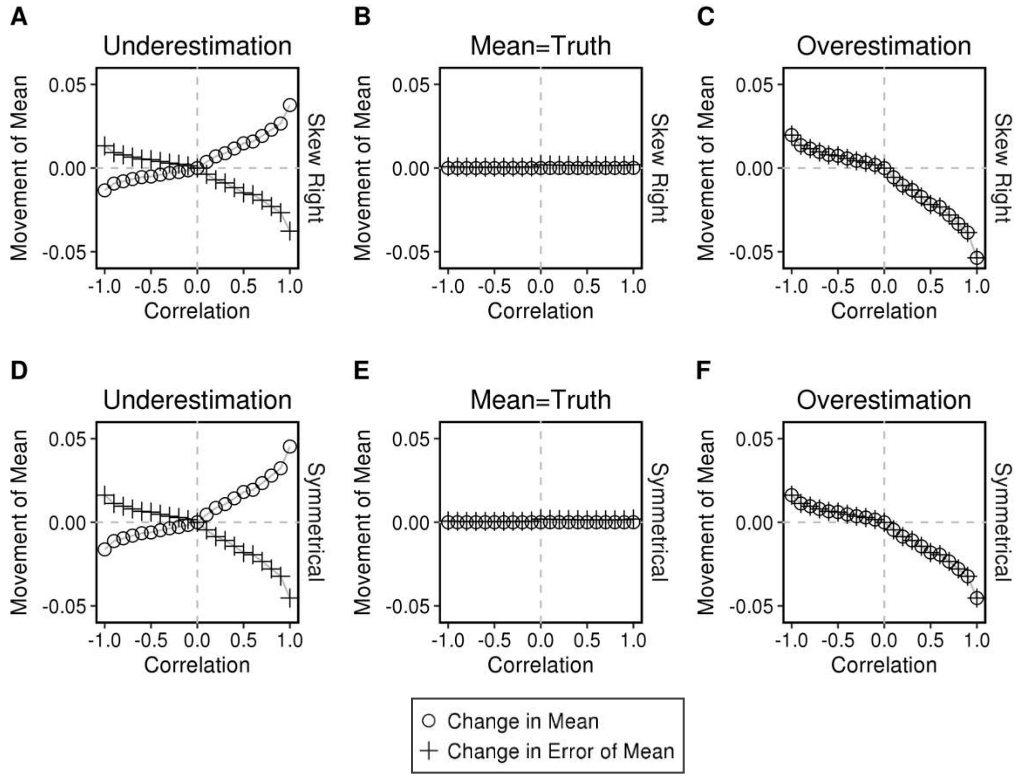

Figure A2. This figure shows the change in the mean and change in the error of the mean for simulated trials. The x-axis of each panel indicates the correlation between error and α_i. The y-axis for each panel indicates the change in the mean and the change in the error of the mean, as measured in units of standard deviation. The top row shows outcomes for a right-skewed (log normal) response distribution, and the bottom row shows a symmetrical (normal) distribution. In the left column, the mean of independent responses underestimates the truth by 0.5 standard deviations; in the center, the mean equals the truth; and in the right column, the mean overestimates the truth. Theoretical predictions are all consistent with our experimental results. When correlation is greater than zero (accurate individuals move less) the group mean always either improves or remains the same. When correlation equals zero, the group mean remains unchanged. N=1,000 nodes per network, 10,000 simulations per point.

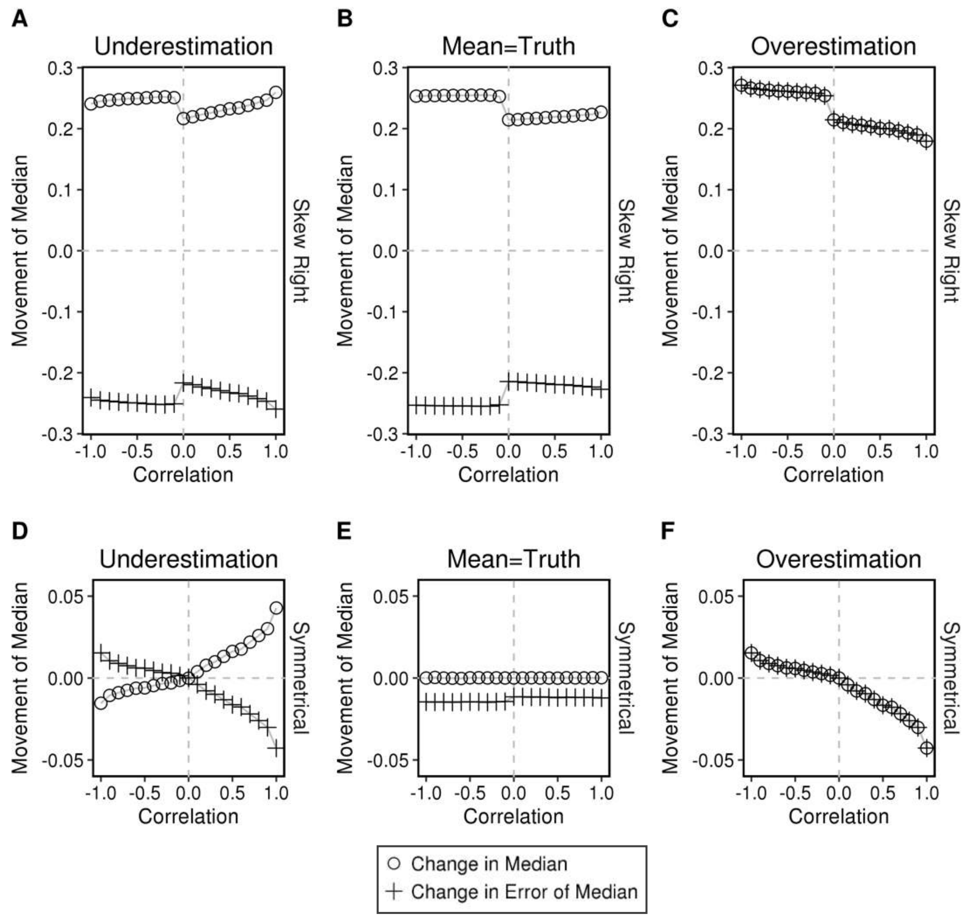

Figure A3. This figure shows the same model parameters as Fig. S9, but displays results for the median instead of the mean. The x-axis of each panel indicates the correlation between accuracy and α_i. The y-axis for each panel indicates the change in the mean and the change in the error of the mean, as measured in units of standard deviation. The top row shows outcomes for a right-skewed (log normal) response distribution, and the bottom row shows a symmetrical (normal) distribution. In the left column, the mean of independent responses underestimates the truth by 0.5 standard deviations; in the center, the mean equals the truth; and in the right column, the mean overestimates the truth. Theoretical predictions are all consistent with our experimental results. In the skew-right distribution, the median improves in most cases. Even in accurate symmetrical distributions, sample error leaves some room for improvement in the median, as shown in Panel E. N=1,000 nodes per network, 10,000 simulations per point.

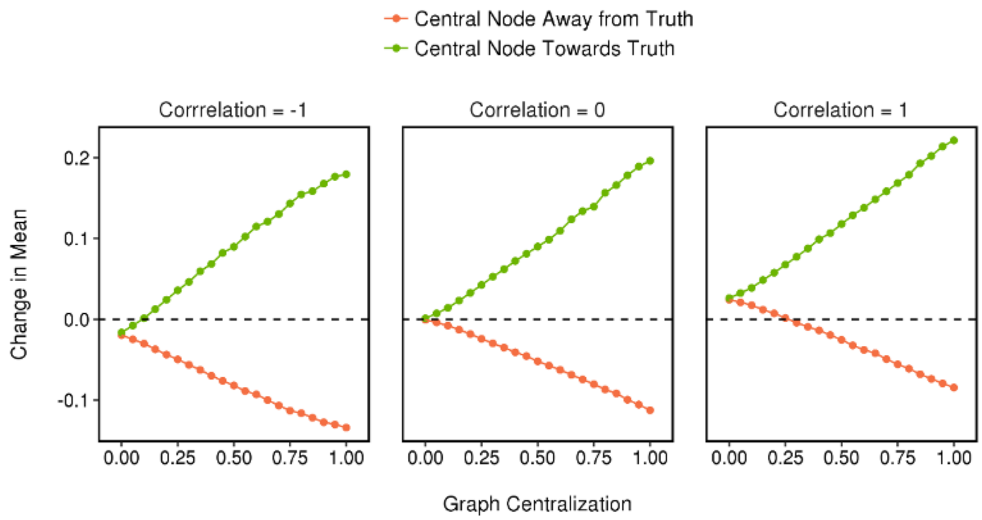

Figure A4. The effect of centralization in networks on the change in the group mean. These simulations reflect a group that underestimates the true value, and therefore an increase in group mean indicates a decrease in error. The influence of prominent nodes has a stronger effect on group beliefs than the correlation between accuracy and revision magnitude. Results are plotted based on whether or not the core node is in the direction of the true value (green) or away from the true value (red). Centralization cannot be controlled directly in network generating algorithms (see text), and results are plotted according to the resulting centralization score for each randomly generated network. N=40 nodes per network, 10,000 repetitions per point.

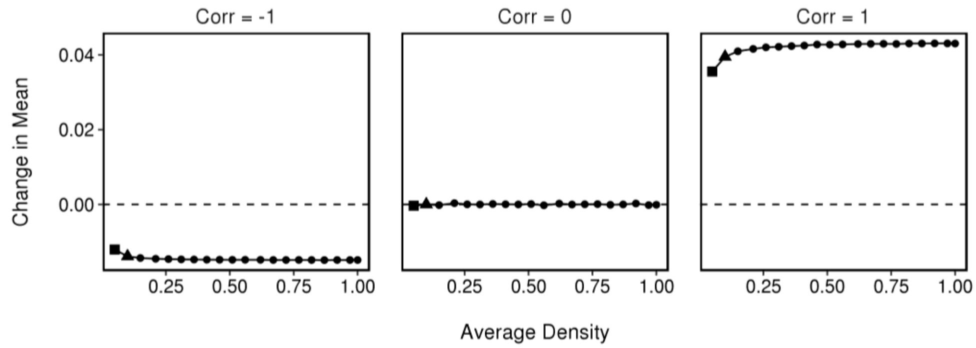

Figure A5. The effect of density in networks on the change in the group mean. These simulations reflect a group that underestimates the true value, and therefore an increase in group mean indicates a decrease in error. These networks vary only in average degree, holding network size (and centralization) constant. The densities of the two networks used in our study are noted in both panels: the centralized (box) and decentralized (triangle) networks. N=40 nodes per network, 10,000 simulations per point

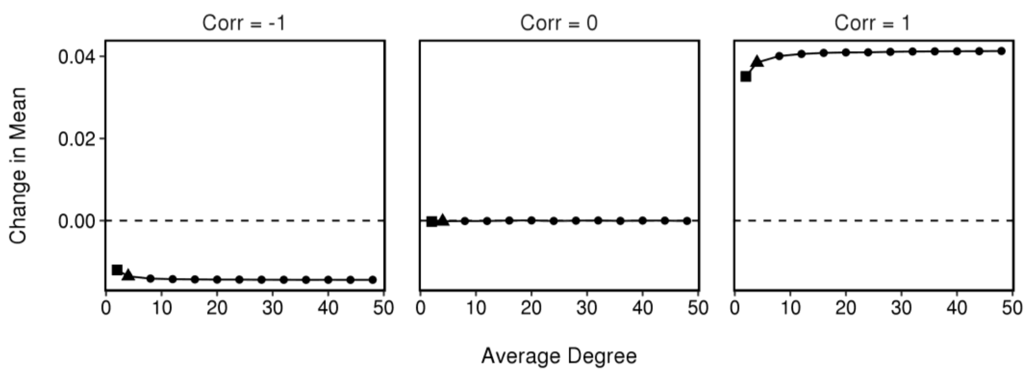

Figure A6. The effect of average degree in networks on the change in mean. These simulations reflect a group that underestimates the true value, and therefore an increase in group mean indicates a decrease in error. These networks are random networks where every node has the same degree varying N and degree concomitantly in order to hold density constant. Average degree for networks used in our study are noted in both panels: the centralized (box) and decentralized (triangle) networks. 10,000 simulations per point.

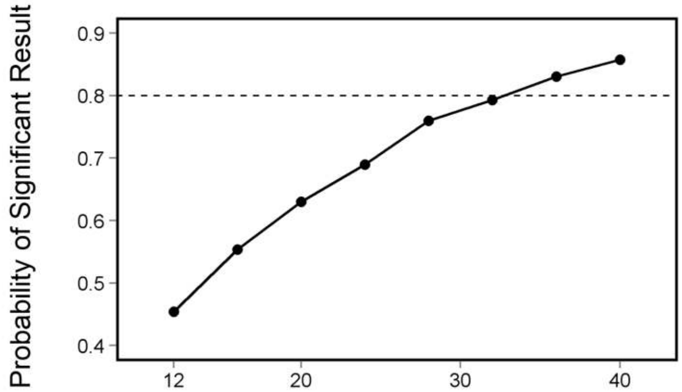

Figure A7. Points show the expected experimental power as a function of population size, with a dashed line indicating standard design with 80% probability of significant results. Because the wisdom of crowds emerges only in large groups, low population sizes are less likely to reliably demonstrate improvement as a result of influence. With the experimental design implemented by Lorenz et al, we estimate greater than 50% probability of type II error. Each point measures 1,000 bootstrapped p-values for data drawn from 10,000 simulated trials.

CHAPTER 3: FROM EFFICIENT MARKETS TO EFFECTIVE INVESTORS – INDIVIDUALS AND THE WISDOM OF CROWDS

> Three economists went hunting, and came across a large deer. The first economist fired, but missed by a meter to the left. The second economist fired, but missed by a meter to the right. The third economist didn't fire, but shouted in triumph: "We got it! We got it!"

Abstract

Research on the "wisdom of crowds" is based on the observation that the average belief in a group can be accurate even when individual members are inaccurate, a phenomenon that is useful to third-party aggregators but offers no benefit to group members as individuals. One striking example of the divergence between collective and individual accuracy is financial forecasting, in which retail (non-institutional) investors display the wisdom of crowds even as these same individuals lose money compared to benchmarks. While popular theoretical accounts argue that individuals must remain independent in order to preserve group accuracy, we test an alternative theory which predicts that individuals can learn from each other and benefit from the wisdom of crowds. We present the results of an experimental study in which subjects made financial forecasts before and after learning the beliefs of peers connected to them in a social network. We find that information exchange generates, on average, a 25% decrease in individual error while still preserving the wisdom of the crowd. These results show that structured social networks can improve individual performance while maintaining collective wisdom.

Introduction

It is common in social and organizational research to treat measurements of the group as representative of individuals within the group, and yet it can be the case that the average individual looks nothing like the average *of* individuals (Mozer, Pashler, & Homaei, 2008; Way, Morgan, Clauset, & Larremore, 2017). While the divergence between individual and group characteristics can sometimes lead to methodological problems such as the ecological fallacy, it is also a defining characteristic of research on collective intelligence, which studies groups as macro level phenomena that cannot be reduced to collections of individuals (DeDeo, 2014; Krafft et al., 2016; Woolley et al., 2010). One striking example of this paradigm can be found in the growing body of research on the "wisdom of crowds," which is motivated by the observation that the average belief in a large group can be remarkably accurate even when most group members are individually inaccurate (Galton, 1907; Page, 2007; Sunstein, 2006; Surowiecki, 2004). The remarkable ability for groups to generate accurate beliefs from inaccurate individuals has been observed in a wide variety of decision tasks including medical diagnoses (Kurvers et al., 2016; Wolf et al., 2015), sports betting (Herzog & Hertwig, 2011), visual search tasks (Juni & Eckstein, 2017), geopolitical forecasting (Atanasov et al., 2016b), and financial forecasting (Kelley & Tetlock, 2013; Nofer & Hinz, 2014).

The advantages of belief aggregation have long been known as a method to improve the practice of economic and financial forecasting (Clemen, 1989), and recent attention to the benefit of belief aggregation has spurred the development of a range of aggregation methods designed to harness the wisdom of crowds. These techniques include weighted averages accounting for past accuracy (Budescu & Chen, 2014; Mannes et al., 2014; Welinder et al., 2010), surveys which obtain supplementary information to triangulate accurate estimates (Koriat, 2012; Prelec et al., 2017), and machine learning techniques to harness digital trace data (Bollen et al., 2011; Nofer & Hinz, 2014; Peeters, 2018). These approaches are useful when an outside perspective enables large scale, independent belief aggregation, but the wisdom of crowds as an aggregate property offers no direct benefit to group members as individual decision-makers.

The divergence between group accuracy and individual accuracy is particularly salient in the case of financial beliefs. On the one hand, financial markets demonstrate a remarkable (if imperfect) ability to efficiently aggregate information and generate prices which reflect the underlying asset value (B. G. Malkiel, 2003; Shiller, 2003). The wisdom of financial crowds can be observed even when limiting analysis to retail (i.e., non-institutional) investors (Kelley & Tetlock, 2013), whose aggregated forecasts can outperform professional analysts (Nofer & Hinz, 2014). On the other hand, however, individual investors fare relatively poorly, and their portfolios consistently underperform compared to benchmarks (Barber, Lee, Liu, & Odean, 2009; Barber & Odean, 2000; B. Malkiel, 2016). As a result of this divergence, identifying whether individuals can benefit from the wisdom of crowds holds great practical importance.

The theoretical challenge facing individuals in the crowd stems from a common assumption in wisdom of crowds research, namely that groups generate the most accurate beliefs when they are composed of individuals who are statistically independent (H. Hong et al., 2016; Lorenz et al., 2011; Nofer & Hinz, 2014). This assumption is based on two key arguments. One argument emerges from Surowiecki's (2004) influential book *The Wisdom of Crowds*, which popularized the wisdom of crowds in contemporary research. This book, along with Janis' classic *Groupthink* (1982), argued that the dynamics of social influence generate group beliefs which do not reflect the full set of information held by individuals. Page (2007) offers a second, more formal argument based on the statistical properties of belief aggregation. Page's "diversity prediction theorem" draws attention to the importance of diversity (statistical variance) among individual beliefs in the generation of accurate group beliefs. Because social influence has been commonly observed to generate belief conformity (Asch, 1951; Deutsch & Gerard, 1955; Sherif, 1935; Cialdini & Goldstein, 2004)—i.e., reduced variance—the diversity prediction theorem has provided further support for the argument that social influence undermines the wisdom of crowds (Lorenz et al., 2011). Together, these two arguments are invoked to suggest that people must be kept socially independent to preserve the wisdom of crowds (H. Hong et al., 2016; Lorenz et al., 2011; Nofer & Hinz, 2014).

Understanding whether—and how—individuals can learn from each other without undermining the wisdom of crowds thus holds not only practical but theoretical significance. By challenging the assumption that statistical and social independence is not required for groups to maximize accuracy, we argue that the wisdom of crowds can emerge under much more general conditions than might be expected based on previous literature. The possibility that crowd wisdom might be robust to information exchange is not new to this paper, but we present a more general and less restrictive theoretical claim than previous work. Prior theoretical models have required that the response to social influence vary based on belief accuracy—such that individuals who are more accurate are less likely to respond to social influence—leading to the prediction that accuracy not only is robust to social influence but even can be improved at both the individual level (Gürçay et al., 2015) and the group level (Becker, Brackbill, & Centola, 2017; Madirolas & de Polavieja, 2015). However, these models require some mechanism whereby individuals know (implicitly or otherwise) their own accuracy. While a noisy correlation between accuracy and receptivity to social influence has been found to hold empirically for simple trivia questions (Becker et al., 2017; Gürçay et al., 2015; Madirolas & de Polavieja, 2015), it may not generalize to belief formation more generally and thus a more robust mechanism for social learning would be valuable for advancing research on the wisdom of crowds and understanding how groups can form accurate beliefs.

In contrast with these cognitively demanding models, we argue that social learning is a robust mechanism that can improve individual accuracy under a wide range of conditions, even the extreme case where the accuracy/receptivity relationship is reversed. Building on a formal model of opinion formation in networks (DeGroot, 1974), we turn the diversity prediction theorem on its head and argue that individuals in decentralized social networks will become more accurate after observing each other's beliefs *by virtue of becoming more similar*. We test this prediction with a laboratory experiment in which subjects make financial forecasts before and after being exposed to the beliefs of others in a social network.

The Wisdom of Independent Crowds

The ability for groups to form judgements whose accuracy exceeds that of any individual member can be explained by a model in which each individual estimate is composed of the true value, plus an error term:

$$x_i = (\theta + \varepsilon_i) \ ,$$

where x_i is the judgement of the i^{th} individual, θ is the true value to be estimated, and ε_i is the error of the i^{th} individual (Hogarth, 1978; L. Hong & Page, 2009). Under the assumption that ε_i is distributed identically and independently for all individuals, and that $E[\varepsilon]=0$, then the expected error of the group (sample) mean—that is, collective error—will decrease as the group size increases, and approach the true value. If, however, individuals belief are not independent—and thus error terms are correlated—then the advantage of group size is diminished, and the accuracy of the group is expected to decrease (Hogarth, 1978).

This model shows how it is that accurate collective beliefs can emerge when individuals are independent, such that a large diversity of estimates within the groups produces individual errors that cancel each other out. However, the assumption that $E[\varepsilon]=0$ is unlikely to be a reliable property of empirical belief distributions. In many circumstances, people's beliefs are subject to common cognitive biases, which lead to systematic errors, such that a population will tend to regularly under- or over-estimate the true value (Kahneman & Tversky, 1977). Fortunately, however, a well-known statistical regularity shows that the aggregated belief within a large population can offer an accuracy benefit even under these circumstances, regardless of the presence of systematic bias. Simply put, the error of the average belief in a group will always necessarily be lower than the average error of any individual belief:

$$E[\bar{x} - \theta\,]^2 = E[(x_i - \theta)^2\,] - \; E[(x_i - \bar{x})^2\,] \ .$$

This statement is well known in mathematical statistics, where it shows that the bias (error) of any estimator (in this case, the error of the group average, indicated by the leftmost term) is equal to the mean squared error (the average individual error, the first term on the right) minus the variance (group diversity, the second term on the right). This statement leads immediately to the inequality

$$E[\bar{x} - \theta]^2 < E[(x_i - \theta)^2]$$

which Page (2007) terms the "crowd beats averages" law. Thus one major implication of the bias-error-variance decomposition for an estimator is that the error of the average estimate in a group (of any size) will always be lower than the average error obtained from any individual estimate alone.

This inequality has long been known as an explanation for the value of belief aggregation (Hogarth, 1978; Zajonc, 1962; Zarnowitz, 1984). More recently, equation 2 has become known as the "diversity prediction theorem" (Lorenz et al., 2011; Page, 2007) which emphasizes a second major result of this statement: as the diversity (variance) increases, and average individual error (mean squared error) is held constant, then group error (estimator bias) will decrease. This poses a potential problem for organizational settings where communication is unavoidable, since social influence between individuals has been found to increase the similarity of people's beliefs (Asch, 1951; Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955; Sherif, 1935), thereby reducing the diversity of opinions in the group.

One interpretation of the diversity prediction theorem is that as the members of a group become more similar, their collective beliefs will become less accurate (H. Hong et al., 2016; Lorenz et al., 2011; Nofer & Hinz, 2014). Because social influence tends to increase similarity (Asch, 1951; Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955), a common assumption in aggregation methods for extracting the wisdom of crowds is that group accuracy is maximized when individuals are independent (H. Hong et al., 2016; Nofer & Hinz, 2014). By contrast, when group members are aware of each other's beliefs (through deliberation or some other form of information exchange), social influence is expected to generate herding effects that undermine the the accuracy of beliefs in groups (Baddeley, 2010; Lorenz et al., 2011; Surowiecki, 2004).

Social Influence and the Wisdom of Crowds

It is certainly possible for social influence to lead people in groups to make suboptimal decisions, as shown empirically through case studies of "groupthink" in organizations (Janis, 1982) and theoretically in microeconomic models of herding behavior (Banerjee, 1992; Bikhchandani et al., 1992). The phenomenon of groupthink, for example, is expected to diminish the accuracy of group decisions when group norms for cohesion prevent individuals from sharing information that runs counter to previously established group beliefs. Case studies on the wisdom of crowds have identified how group failure can result from the overly strong influence of group leaders, which can similarly prevent groups from using all the information held by their members (Surowiecki, 2004). However, these studies are not sufficient to show that social influence will *always* undermine group decisions, and even as they highlight the risks of social influence they also highlight the possibility that the negative effects of social influence can be avoided.

The present study builds on a formal model of belief formation in networks (DeGroot, 1974) which demonstrates how structural features of a population (e.g., the network topology) can determine whether social influence undermines group accuracy, or, conversely, whether it facilitates social learning while preserving group accuracy. In this model, each member of a group is assumed to start with some independent belief. Upon learning the beliefs of others connected to them in a social network (i.e., "network neighbors"), individuals adopt a weighted average that combines their own initial belief and the beliefs of their network neighbors. This process iterates over time, leading groups to reach a consensus—such that every member holds the same beliefs—in any connected network (DeGroot, 1974).

Crucially, the value of the eventual consensus—the point toward which group beliefs are drawn—and thus group accuracy depends on structure of the influence network (Becker et al., 2017; DeGroot, 1974). In this revision process, more central members (those observed by a greater number of peers) contribute more weight to the consensus belief than less central individuals[2] (DeGroot, 1974). These properties also appear in a number of related models with varying assumptions about the individual decision-making process (Acemoglu, Dahleh, Lobel, & Ozdaglar, 2011; Golub & Jackson, 2010; Mossel et al., 2015) and thus provide theoretical support for the empirical observation that overly influential leaders can undermine group decisions (Janis, 1982; Surowiecki, 2004). The important result for the present study is that "decentralized" networks, defined as those networks in which everyone is equally connected and thus equally central, will be drawn directly toward the mean of independent beliefs. Paradoxically, the diversity prediction theorem guarantees that if group members converge toward the mean of independent beliefs, then average individual error will decrease even as diversity decreases, not in spite of the increased similarity but because of it. As a result, this model offers an explanation for observational case studies showing that social influence undermines the wisdom of crowds (Janis, 1982; Surowiecki, 2004) while also identifying conditions under which social influence may allow individuals to benefit from the wisdom of crowds.

While a decentralized network can produce convergence to the mean (DeGroot, 1974), a decentralized topology for the binary communication network (who can observe whom) is not sufficient to guarantee that everyone is equally influential, since group beliefs are ultimately determined by the weighted network of social influence—not just who can observe whom, but how much they influence each other (Becker et al., 2017). For example, status and other meta-information can make certain people more influential than others. Even anonymous information exchange is not sufficient to ensure equal influence, because group members may place variable weight on their *own* belief, which alters their centrality in the resulting influence network[3]. In other

[2] Specifically, the asymptotic consensus is equal to a weighted mean of independent beliefs, where the weight for each person is equal to their eigenvector centrality in the influence network (DeGroot, 1974)
[3] In the model, the amount of weight each person places on their own belief is reflected as a "self-tie" in the network, i.e. the diagonal in the adjacency matrix, which contributes to eigenvector centrality.

words, when some individuals place more weight on their own belief and less weight on social information, they increase their relative influence in the social network. In the extreme case, one individual who never changes their belief would asymptotically determine the beliefs for an entire population, as others revise their belief gradually toward social information. Although this extreme situation is unlikely, the example illustrates how some individuals can become disproportionately influential in a network even when information exchange is anonymous and every person is observed by an equal number of others.

The present study tests this convergence-to-the-mean hypothesis—and its implications for improving individual performance—in the context of financial forecasting. Before studying this phenomenon empirically, we first tested the robustness of this hypothesis computationally by simulating the effect of social influence for a population embedded in a network in which everyone is equally connected, but where each individual places a variable amount of weight on social information. We found that despite the fact that some individuals may be more influential due to varying social weight, the group nonetheless converges on the mean of independent beliefs as long as this weight is independently, identically distributed (i.i.d) throughout the population (see Appendix). Another paradoxical property of the diversity prediction theorem is that individual accuracy can improve even when the group mean becomes less accurate, and we find in simulation that individuals can improve even under some conditions where this i.i.d. assumption is violated. As detailed in the Appendix, these predictions hold under a wide range of empirically plausible assumptions, offering a robust mechanism whereby groups individuals embedded in decentralized networks will benefit from social influence. The goal of this study is to test this formal theoretical expectation on the important task of forecasting market behavior.

The Wisdom of Financial Crowds

One of the most common and yet most controversial examples of the wisdom of crowds is the efficient markets hypothesis (B. Malkiel, 2016; Schijven & Hitt, 2012; Surowiecki, 2004). The efficient markets hypothesis takes several forms, all of which share the general claim that the current market price for an equity reflects the best possible price given available information (Fama, 1970). Informally, this theory states that if there were any information that could be used to predict a future increase (or decrease) in the fundamental value of the asset, it would already be incorporated into the price. Thus one implication of this hypothesis that it is impossible to “beat the market” by investing in something that can be expected to rise in value more than the market as a whole (B. Malkiel, 2016). Just as with the more general wisdom of crowds phenomenon, an efficient market does not require rational individuals. Individual investors may vastly overestimate or underestimate in the fundamental value of an asset, but those errors will cancel out in aggregate, revealing the best possible price (B. G. Malkiel, 2003; Shiller, 2003).

Given the general difficulty (or impossibility) of beating the market, even in the presence of occasional inefficiencies (B. G. Malkiel, 2003; Shiller, 2003), the best advice to any investor is a passive “buy and hold” strategy with a diversified investment portfolio, rather than the active selection and exchange of individual assets (B. Malkiel, 2016). However, despite the common advice that “active investment” is not a profitable strategy, active investment remains a popular pursuit among both institutional and retail investors (Barber et al., 2009; Chater, Huck, & Inderst, 2010; B. Malkiel, 2016). Due to the popularity of this do-it-yourself approach to portfolio design, retail investment in particular provides a clear example of the divergence between collective accuracy and individual inaccuracy. On the one hand, retail investors have been found to exhibit the wisdom of crowds. In one study, large scale analyses of investment recommendations posted to online investment communities were found to produce superior investments to those of professional analysts (Nofer & Hinz, 2014). Direct analysis of retail market orders have found that purchase behavior can predict future news sentiment, a proxy for the fundamental value of an asset (Kelley & Tetlock, 2013). On the other hand, however, retail investors consistently tend to lose money as compared with a buy-and-hold strategy (Barber et al., 2009; Barber & Odean, 2000; B. Malkiel, 2016). As a result, understanding how to improve the accuracy of individual investor beliefs is an important topic of research.

As described above, theoretical models of belief formation (Appendix) argue that two conditions are sufficient to allow individuals to learn from each other and thus improve their belief accuracy without undermining the wisdom of crowds. One condition is a decentralized network structure, which means that no single person or group of individuals can influence a disproportionately large number of people. In practice, many social networks have been observed to be highly "centralized" (Barabâsi et al., 2002; Eagle & Pentland, 2006; Ebel, Mielsch, & Bornholdt, 2002;; Liljeros, Edling, Amaral, Stanley, & Aberg, 2001) including influence networks in online investment platforms (Pan, Altshuler, & Pentland, 2012). As a result, experimental methods are essential for identifying the effects of social influence on financial beliefs under theoretically ideal conditions, since observational studies are constrained by the properties of existing social networks. The second condition required for social influence to improve belief accuracy is that individual responses to social influence are not negatively correlated with accuracy. While the first condition—decentralized networks—can be established by embedding individuals in intentionally structured social network networks, this second condition is an endogenous feature of investors as a population. The importance of the relationship between error and response to social influence is thus a key motivation for testing the effect of social influence on financial beliefs specifically, rather than drawing conclusions solely from prior research on trivia tasks.

In the context of financial forecasting, the second condition could be violated if certain types of investors are systematically less likely to make use of social information, and thus exert a greater amount of influence on group beliefs. Unlike the trivia questions studied in previous research (Becker et al., 2017; Gürçay et al., 2015; Lorenz et al., 2011; Moussaïd et al., 2013), financial beliefs are prone to a unique set of cognitive biases, such as the tendency for people to overestimate their ability to predict random events (Fisher & Statman, 2000; Langer & Roth, 1975). Another common bias is optimism, which can exacerbate overconfidence (Baker & Nofsinger, 2002) and tends to be asymmetrical—people are more likely to hold overly optimistic beliefs than overly pessimistic beliefs (Kahneman & Riepe, 1998). If it is the case that optimistic investors are more confident and thus less responsive to social influence, than the assumptions of our model will be violated, and group beliefs may systematically increase their estimates (regardless of the true value) rather than becoming reliably more accurate. Importantly, as noted above, it is possible for individual beliefs to become more accurate even when beliefs measured at the group level become less accurate (see Appendix). The goal of this study both to test the robustness of social learning for individuals and also to test whether individual financial beliefs can be improved without undermining accuracy as measured at the group level.

Methods

Our experimental design advances previous research in several ways, and is designed to test the theory that social influence is a robust mechanism that can improve the accuracy of individual beliefs even in the presence of cognitive biases and limited information exchange. The primary contribution of our study is a focus on individual level outcomes, in contrast with a focus in previous research on the accuracy of group level beliefs (Becker et al., 2017; Mellers et al., 2014; Navajas, Niella, Garbulsky, Bahrami, & Sigman, 2018), which can mask large variations in individual level accuracy (Mozer et al., 2008). One particular concern facing individuals in crowds is that if the group as a whole is pulled toward the group mean, it may be the case that individuals who are initially accurate become less accurate as a result of social influence. If this occurs, the improvement for the least accurate individuals would come at the cost of reduced accuracy for the most skilled investors. While one prior study (Gürçay et al., 2015) was designed to examine the effect social influence on individual accuracy, their analysis did not identify whether the benefits accrued to all group members or whether the improvement for the least accurate members came at a cost to the most accurate members. Moreover, this study examined only trivia questions, which do not reflect cognitive biases found in tasks such as financial forecasting.

Another limitation of this study (Gürçay et al., 2015) is that the design allowed participants to engage in detailed discussion and displayed their reported confidence level, which was based on their theory that social influence increases belief accuracy due to a correlation between confidence and accuracy. The presence of multiple types of information exchange make it impossible to distinguish the mechanisms that shaped individual beliefs, which limits replicability. In contrast, we study the effect of purely informational influence (Deutsch & Gerard, 1955) on belief accuracy. We test the theory that people can become more accurate simply by knowing what their peers believe.

One final limitation of previous research on individual accuracy is the use of experimental designs that allowed individuals to see the beliefs of the entire study population (Gürçay et al., 2015). However, individuals in organizational settings frequently interact with only a small subset of the population, and our simulations suggest that even this limited "local" influence (Centola, Willer, & Macy, 2005) can improve belief accuracy in networks where everyone is observed by an equal number of peers. To test minimal sufficient conditions to improve group accuracy, we used a web-based experimental platform in which individuals provided estimates before and after observing information about the numeric estimates of a limited number of peers in the social group, forming a random 4-regular graph of information exchange. To ensure that our results are robust against the biases that shape real-world beliefs, we study effect of social influence on the accuracy of financial forecasts.

Experimental Design. In order to test whether social influence can improve the accuracy of financial beliefs, we conducted a web-based experiment in which 1,286 individuals forecasted the future price of 55 exchange traded equities. For each forecast, subjects were randomized to either a "social" condition, in which they were able to observe the belief of other participants, or a "control" condition, in which they generated their forecast independently.

Subjects were recruited from Amazon Mechanical Turk. Prior to being assigned to an experimental session, all subjects completed a survey in which they reported their investment experience, their educational background, their personal portfolio size, and whether they were a professionally employed in finance. To ensure that our subject pool represented a population comparable to that of retail investors, only those subjects who reported at least 1 year of investment experience were invited to contribute their forecasts. Subjects were not informed that their eligibility depended on their survey response, and thus there was no incentive for false reporting. A total of 3,484 respondents completed the enrollment survey, of which 1,575 reported one or more years of investment experience. All 1,575 eligible respondents were invited to participate. Of these subjects, 760 reported between 1 and 5 years of investment experience, 275 reported 5 to 10 years, and 251 reported greater than 10 years of experience.

Data was collected over the course of 7 experimental sessions between November 26, 2017 and December 10, 2017. For a given experimental session, all eligible participants were sent a login link to access the experimental platform at a pre-scheduled time, in order to enable simultaneous participation. Simultaneous participation was important both to enable information exchange as well as to ensure that forecast accuracy was not impacted by exogenous variables such as changes in the underlying equity value. At the scheduled time, all subjects who were logged into the experimental platform were randomly assigned to an equity (i.e., what they were to forecast) and an experimental condition (social or control). Subjects were allowed to participate more than once, and were re-randomized for each experimental session. To incentivize accurate forecasts, subjects were rewarded based on the accuracy of each individual forecast.

A single trial consisted of 80 individuals forecasting the future value of a single equity. For each trial, 40 individuals were placed into a single, connected social network that allowed them to observe the forecasts of 4 network neighbors. This network topology was random and undirected. The remaining 40 individuals were assigned to the control condition, in which they provided estimates in isolation (without any social interactions). In total, we collected data for 55 trials, each of which was assigned a unique equity to forecast.

Subject Experience. Once a session began, each subject was prompted to forecast the closing price for an exchange-traded equity approximately 4 weeks after the date of the experimental session. (The exact forecast period varied slightly between sessions to adjust for dates on which the markets are closed). The complete list of equities is shown in the appendix (Table S1). To provide an informational environment comparable to that in which investors typically make decisions (and also to discourage subjects from leaving the page, and missing the forecast submission time) the experimental interface included a stock ticker (showing prices for the previous 5 days, 1 month, 1 year, and 5 years) and recent headlines related to the equity to be forecasted (see Appendix figure S1).

Subjects were provided 60 seconds to complete their first forecast ("Round 1"). They were then prompted to enter their forecast again, providing a second estimate, and were given another 60 seconds for this estimate ("Round 2"). Subjects were then prompted to provide a third forecast, and were given 60 seconds for this third and final forecast ("Round 3"). Subjects were rewarded based on the accuracy of their final forecast.

Between revisions, subjects in the social condition were shown the average answer of their 4 network neighbors. Subjects in the control condition were simply asked to provide 3 independent estimates. The subject experience in the social condition and the control condition was identical except for the presence of social information. Our primary outcome of interest is whether subjects in the social condition improved after being exposed to social information.

Analysis. Some users failed to provide a response in one or more Round, and our analysis is conducted only for those users that provided a response to all three Rounds, providing a total of 4,071 estimates for analysis, 2,041 from the social condition and 2,030 from the control condition. Following previous work on the wisdom of crowds (Lorenz et al., 2011; Madirolas & de Polavieja, 2015) we logarithm-transformed all prices (predicted and actual) prior to analysis. This transformation controls for the fact that prices are unbounded "on the right" (stock prices/estimates can be infinitely large) but bounded "on the left" (stock prices/estimates cannot go below zero) and is commonly applied in stock price analyses. For each estimate, we therefore measure error as $|\log(E) - \log(T)|$ where *E* indicates the estimate, *T* indicates the true value, and log indicates the natural logarithm. For small values, $\log(E) - \log(T) \approx \frac{E-T}{T}$ and thus error as reported here can be interpreted as comparable to percent error. While error is always positive and measures the average distance of an estimate from the true value, we also measure bias as $\log(E) - \log(T)$ which can be positive or negative and indicates whether estimates tend to systematically under- or over-estimate the true value.

Because subjects in the social condition are not statistically independent, all outcomes are measured at the trial level. For example, to assess the change in error, we measure the average error across all individual estimates for each trial, producing 55 outcomes for the social condition and 55 outcomes for the control condition. When measuring error as a function of initial accuracy, we first measured average error for each subgroup for each trial, and then averaged the results for each subgroup over the 55 independent trials. For all tests, statistical significance is determined by assuming that independence only exists at the group level, such that each condition of each trial constitutes a single independent data point. This conservative assumption accounts for the potential within-group correlation between individuals (e.g., if one group contains a particularly accurate member, then all group members benefit) and thus prevents any single trial from being overly influential in our analysis. To minimize parametric assumptions, all tests use a Wilcoxon signed rank test.

Results

We found that social information improved the accuracy of individual forecasts without any deleterious effects on the accuracy of the wisdom of crowds. We begin our analysis by measuring the wisdom of crowds in our subject population, by comparing the mean price forecast to the true price for each stock. Because subject beliefs are independent at Round 1, we combine the forecasts for the social and control conditions. As a benchmark for comparison, we first measure the extent to which the current price would have predicted the future price. We find that using the current price as the forecast had a significantly negative bias (P<0.05), consistent with the common model of market prices as a random walk with gradual upward movement. In contrast, the average bias of the group mean for the 55 forecasting tasks was not significantly different from zero (P>0.5) indicating that groups neither under-estimated nor over-estimated stock prices. This result suggests that the wisdom-of-crowds estimate effectively incorporates predictions about the general upward movement of the market. To quantify the relative accuracy of the wisdom of crowds against individual forecasts, we compare the mean error of individual estimates with the error of the mean estimate. We find that across the 55 estimation tasks, the mean estimate offers an error that was an average of 48% lower (95% C.I. [39%, 56%]) than the error of individual subjects' estimates. That is, consistent with the wisdom of the crowd hypothesis, the error of the average was lower than the average error of individuals.

Effect of Social Information on Group Beliefs. Social information significantly changed subjects' financial forecasts. As shown in Figure 1A, social influence decreased the mean standard deviation ($\bar{\sigma}$) of estimates by over 50% from Round 1 to Round 3 ($\bar{\sigma}_1$=0.11, $\bar{\sigma}_3$=0.05, P<0.001), indicating a significant increase in the similarity of participants' estimates. By contrast, there was no significant change in the similarity of individual estimates in the control condition ($\bar{\sigma}_1$=0.14, $\bar{\sigma}_3$=0.12, P>0.70). Despite the increased similarity among individual answers, we found that the collective error remained unchanged. We test this by measuring the absolute value of the error of the mean (i.e., the distance between the mean forecast and the true outcome) for each of the 55 estimates at Round 1 and Round 3. If collective financial forecasting were systematically influenced by one particular subgroup, then collective beliefs might deviate from the mean of independent beliefs. For example, if the most optimistic investors were also the least likely make use of social

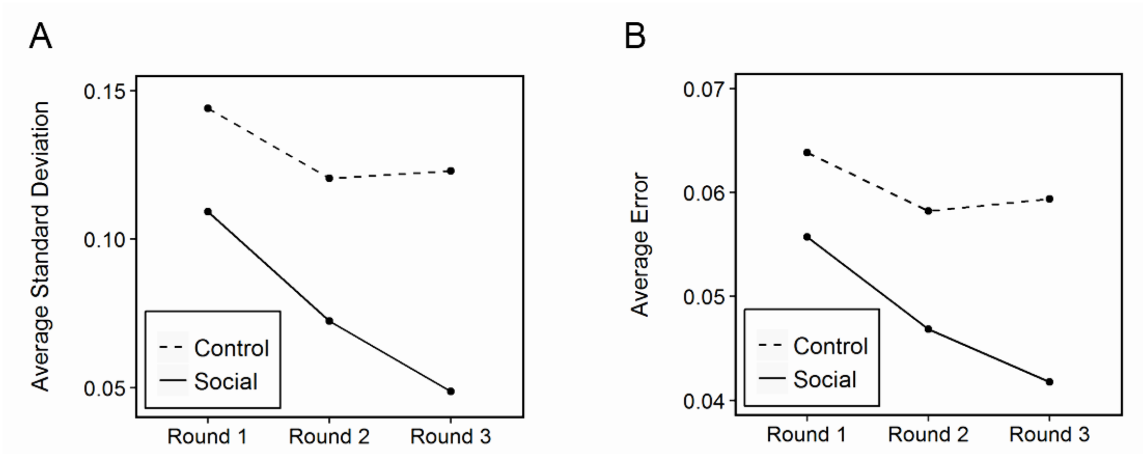

Figure 1. Standard deviation (left) and average individual error (right) at each round, averaged over 55 experimental trials.

information, then social influence would lead group beliefs to systematically overestimate the true value and thus increase in error. However, we find that the error of the mean ($\bar{E}$) was not significantly different at Round 1 versus Round 3 ($\bar{E}_1$=0.033, $\bar{E}_3$=0.030, P>0.38). This result indicates that as groups became more similar, they converged directly on the mean of independent beliefs.

Effect of Social Information on Individual Accuracy. The diversity prediction theorem indicates that if the standard deviation of individual estimates decreases without a change in the group mean, then individual error must necessarily decrease. To measure individual error while controlling for correlation between individuals in the same trial, we conduct a conservative analysis that measures outcomes only at the group level. We first measure the average individual error at Round 1, Round 2, and Round 3 for each of the 55 trials. We then calculate the average of these 55 measurements for each Round, shown in Figure 1B. We find that the average error ($\bar{\varepsilon}$) for each subject's forecast decreased significantly from Round 1 to Round 3 in the social condition ($\bar{\varepsilon}_1$=0.056, $\bar{\varepsilon}_3$=0.042, N=55, *P*<0.001), reducing the average error of each individual's forecast by 25%.

To explain this improvement in individual accuracy, one possibility is that the reduction in error was due to individual learning dynamics that might occur from having more time to spend on the forecast, which provides more time to study price charts and consider the news items present in the interface. To test for this, we also examine the change in error for the control condition. However, we found no significant difference in the control condition between error at Round 1 and error at Round 3 ($\bar{\varepsilon}_1$=0.064, $\bar{\varepsilon}_3$=0.059, N=55, P>0.98). We also directly compared the average change in the social condition with the average change in the control condition using a paired test that compares each condition for the 55 trials, and found that the improvement in the social condition was significantly larger than the improvement in the control condition (P<0.01). As a result, although there were no significant differences in forecast accuracy between conditions in Round 1, the final (Round 3) forecasts by individuals in the social condition were significantly more accurate than final forecasts by individuals in the control condition (P<0.001).

Who benefits from the wisdom of crowds? The less accurate a subject's forecast is at Round 1, the more likely it is that their peers are more accurate than them, and thus the more likely they are to observe social information that improves the accuracy of their forecast. At the same time, the most accurate subjects at Round 1 are also the most likely to observe social information that draws them away from the true value. Thus, one concern is that the collective improvement in accuracy benefits the least accurate group members but comes at a cost to the most accurate group members. While we found that the benefits of social information were indeed greatest for the individuals who were the least accurate at Round 1, we also found that social information did not harm the accuracy of individuals who were initially accurate.

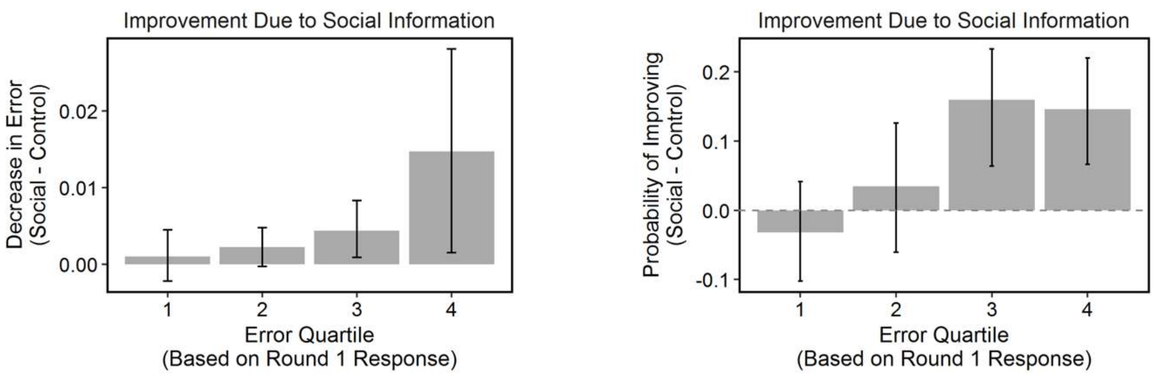

Figure 2. Social information increased accuracy for the least accurate individuals. Both panels show outcomes based on subject error at Round 1. Quartile 1 indicates the most accurate subjects, and Quartile 4 indicates the least accurate subjects. To control for variation in forecasting difficulty, we assign forecasts to quartiles separately for each trial. **Panel A:** We first measure the change in error from Round 1 to Round 3 for both the social and the control condition. Each bar shows the arithmetic difference in the change in error between the social condition and the control condition, with 95% confidence intervals based on a Wilcoxon signed rank test. A positive value indicates that error decreased more for the social condition. **Panel B:** We first measure the percentage of revisions that generate increased accuracy. We then measure the arithmetic difference in this percentage between the social and the control condition, shown with 95% bootstrapped confidence intervals. A positive value indicates that the social condition was more likely to improve.

To test the relative effect of social influence for accurate and inaccurate subjects, we divided subjects for each trial into quartiles (i.e., four equally sized groups) based on their error at Round 1, as shown in Figure 2A. If subjects revised their estimates randomly, then error for the least (or most) accurate subjects at Round 1 would be expected to decrease (or increase) as a result of regression to the mean. Therefore, to measure the effect of social information on belief revision for each quartile, we calculate the arithmetic difference between the change in error in the social condition and the change in error in the control condition. We find that for the most accurate subjects (Quartile 1), change in error is not significantly impacted by exposure to social information (*P*>0.53). However, for the least accurate subjects (Quartile 4), the increase in accuracy (decrease in error) is significantly larger for those subjects exposed to social information (*P*<0.05) than for subjects in the control condition.

How does social influence shape forecast accuracy? One possible explanation for this effect is that any revision after Round 1 improves belief accuracy, and that exposure to social influence simply prompts the least accurate individuals to reconsider their forecast, producing revisions that increase accuracy. However, social information had no effect on the likelihood (P) that subjects would revise their estimates, measured either for the group as a whole ($\bar{P}_{soc}$=78%, $\bar{P}_{ctrl}$=76%, *P*>0.27) or for the least accurate individuals (Quartile 4) ($\bar{P}_{soc}$=83%, $\bar{P}_{ctrl}$=83%, *P*>0.98).

While social information did not increase the likelihood of revision, it did increase the likelihood that subjects who made revisions would increase in accuracy, as shown in Figure 2B. Among the most accurate subjects (Quartile 1), only 24% of those who revised their answer were more accurate at Round 3 than Round 1; however this effect was the same in both the social influence and control conditions, and we therefore conclude that the quality of revisions for initially accurate subjects was not significantly impacted by the presence of social information ($\bar{P}_{soc}$=22%, $\bar{P}_{ctrl}$=25%, *P*>0.39). For the least accurate subjects (Quartile 4), however, social information significantly increased the probability that revisions would generate more accurate forecasts (*P*<0.001). In the social condition, 86% of revisions by the least accurate subjects resulted in more accurate final forecasts, while in the control condition only 74% of revisions by the least accurate subjects resulted in increased accuracy.

The way that individuals responded to social information mediated whether it had a positive effect on the quality of their forecasts. Some participants used social information as a positive signal, and moved toward it. However, others used the information as a negative signal, and moved

Use of Social Information	Revision 1 (Round 1 to 2)		Revision 2 (Round 2 to 3)	
Move Toward	-0.025	[-0.038, -0.01]	-0.025	[-0.036, -0.01]
Move Away	0.008	[0.004, 0.01]	-0.002	[-0.016, 0.01]

Table 1. Change in error as a function of response to social information. Brackets indicate 95% bootstrapped confidence intervals. Error decreased for those individuals who revised their answer towards social information, and increased for those individuals who revised their answer away from social information.

away from it. People who responded positively to social information showed significantly greater improvements as a result of revision. Table 1 shows the average change in error for the two types of responses to social information (moving towards or away) as measured for the revision between Round 1 and Round 2, as well as for the revision between Round 2 and Round 3. We find that individuals who revised their estimates in the direction of social information saw the largest improvement, significantly greater than individuals who revised their estimates away from information ($P<0.001$ for both the first revision and the second revision). In contrast, individuals who revised their answer away from social information saw a significant increase in error for the first revision ($P<0.001$) and no significant change for the second revision ($P>0.2$).

Discussion

While a common argument in prior research is that social influence undermines group decisions (Armstrong, 2006; Janis, 1982; Lorenz et al., 2011; Nofer & Hinz, 2014), our research contributes to a growing body of evidence arguing that managed communication can improve the accuracy of group beliefs (Becker et al., 2017; Golub & Jackson, 2010; Gürçay et al., 2015). A common tendency in previous research on the wisdom of crowds has been a focus on group level metrics of accuracy (Becker et al., 2017; Galton, 1907; Lorenz et al., 2011; Nofer & Hinz, 2014), despite the fact that groups can produce accurate estimates even as individuals make decisions based on inaccurate beliefs. In financial forecasting, this disparity is reflected in wise crowds and efficient markets composed of individuals who largely lose money due to opportunity cost. To identify conditions under which individuals can benefit from the wisdom of crowds, we studied financial forecasts to test a theory which predicts, with minimal assumptions, that individuals can learn through informational influence in decentralized networks.

Like any experiment, the design decisions that provided control also induced constraints that limit generalizability. The short time period of the experimental design meant that subjects had limited ability to produce carefully consider analyses of the fundamental value of the stock. However, available evidence suggests that few retail investors engage in such calculated behavior (Bhattacharya, Hackethal, Kaesler, Loos, & Meyer, 2012; Chater et al., 2010). Additionally, one benefit of studying financial forecasts, in contrast with trivia questions, is to test whether results are robust to the presence of common cognitive biases. Because increased reaction time is expected to exacerbate cognitive biases (Greenwald, McGhee, & Schwartz, 1998; Wright, 1974), this limitation suggests that our results are a conservative test of the ability for beliefs to improve accuracy in the face of cognitive biases.

Although we show that social influence *can* improve belief accuracy, our results don't mean that social influence will always be beneficial. The mechanisms described in case studies on Groupthink (Janis, 1982) derive from a number of social processes that are not reflected in anonymous information exchange limited to numeric signals. This experimental process was designed to test minimal sufficient conditions to improve belief accuracy, and our results should be interpreted with caution when considering social influence in day-to-day settings, where the influence network is shaped by normative pressures (Davis & Greve, 1997) and status effects (Burt, 1987). We anticipate that future research will be able to identify which constraints in our present design can be relaxed while allowing (and even improving) individual learning and maintaining the wisdom of crowds.

One important feature of our design which is consistent with social networks in practice is that our subjects was directly exposed to only a limited set of other people. Despite these limited information sets, those individuals who revised their beliefs to become more similar to peers saw greater gains in accuracy. One characteristic of such "sparse" networks is that even as individuals directly interact with a small subset of the population, they indirectly influence and are influenced by the entire population. This influence results from iterated revisions over time, by which a person's network of influence extends far beyond their immediate connections. As a result, our findings suggest that people can learn from the wisdom of crowds even without any centralized information exchange, due to the ability for information to diffuse through networks.

Conclusion. Despite the limitations imposed by our experimental design, the theory and results presented here support the argument that social networks can allow individuals to benefit from the wisdom of crowds under a wide range of conditions. In contrast with previous theories that require a correlation between accuracy and the use of social information (Becker et al., 2017; Gürçay et al., 2015), we argue that social influence is a robust mechanism which can reliably improve individual belief accuracy under a wide range of conditions. Although we found no evidence of a meaningful correlation between confidence and accuracy in this study (which would have been accompanied by an increase in accuracy of the mean [Becker et al., 2017]), we nonetheless found that individual beliefs improve, simply by virtue of becoming more similar.

As organizational decisions are increasingly mediated by digital platforms (Bailey, Leonardi, & Barley, 2012; Bonabeau, 2009; Cowgill & Zitzewitz, 2015), research on structured information exchange will become increasingly important to understanding how organizations can harness the wisdom of crowds. By presenting a simple and replicable method to improve the accuracy of beliefs formed by individuals in groups, our design using minimal numeric signals provides a benchmark expectation of improvement, against which different theories of moderating factors can be tested. Future work should not be limited to the study of strictly numeric informational influence, as the theoretical framework that we adopt to make our argument (DeGroot, 1974) can be used to describe the effect of social influence on belief accuracy for a wide range of mechanisms. The processes that shape influence such as status and expertise can be reflected in the tie strength between individuals. Thus, the dynamics of belief formation in contexts such as face-to-face meetings (Janis, 1982; Surowiecki, 2004) and informal organizational networks (Krackhardt & Hanson, 2001) can both be described in terms of the weighted, directed network of social influence that we study here. In contrast with statistical research aimed at optimizing the aggregation of independent estimates by external observers, our study highlights the ways in which social networks can spread the wisdom of crowds to individuals within the organization itself. We hope that these results will thus push researchers past the question of whether social influence helps or harms group decisions, and towards an examination of how social influence can be most effectively harnessed to allow the wisdom of crowds to shape beliefs in organizations.

Appendix A: Theoretical Predictions for the Effect of Social Influence

To develop theoretical predictions for the effect of social influence on the accuracy of group beliefs, we use simulations to measure the effect of (a) correlation between accuracy and response to social influence, and (b) network centralization (Figure A1). In contrast with group accuracy, which improves only in decentralized networks, we find that individual accuracy can improve in centralized networks even when the group level (mean) belief becomes less accurate (Figure A1, right hand side). However, the effect of social influence in centralized networks depends entirely upon the accuracy of central nodes. We also find that individual improvement is sensitive to the correlation between accuracy and response to social influence. In decentralized networks where there is a strong negative correlation (such that accurate individuals revise their beliefs more, and inaccurate individuals revise their beliefs less) both group error and individual error increases. However, even this increase in individual error only occurs when group beliefs are initially inaccurate. When group beliefs are initially accurate, individual error decreases under all conditions. Taken together, these simulations suggest that the improvement of individual accuracy is a robust phenomenon under a wide (but not exhaustive) range of empirically plausible assumptions, even in cases where groups may become less accurate.

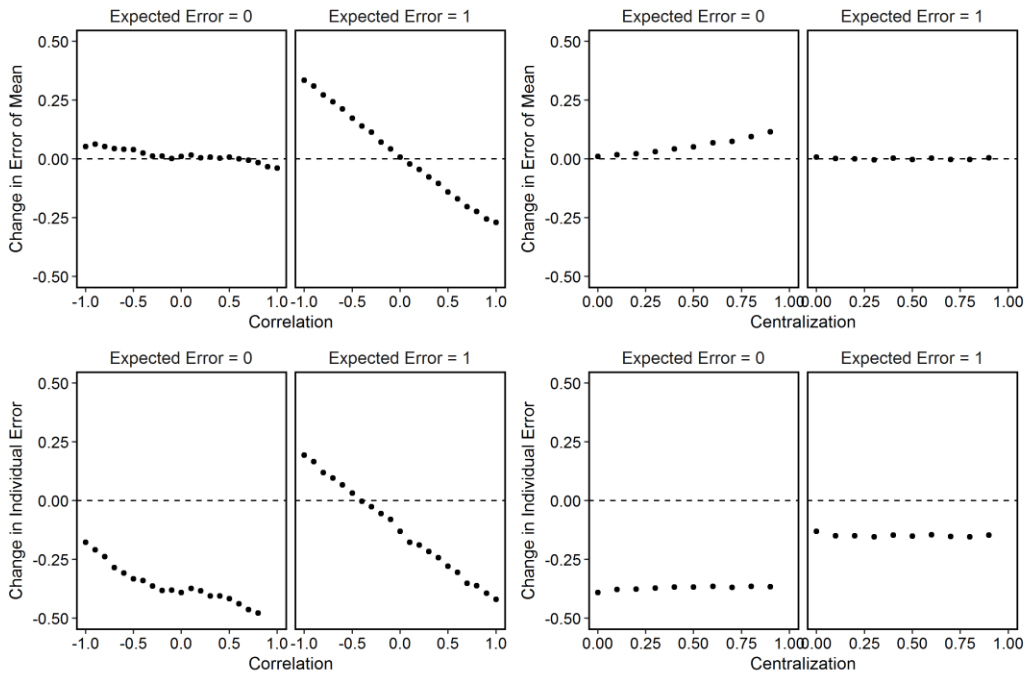

Figure A1. The top row shows the effect of social influence on the error of the mean belief. The bottom row shows the effect of social influence on average individual error. The left column shows the effect of social influence in decentralized networks as a function of the correlation between accuracy and the response to social influence. The right column shows the effect of social influence as a function of network centralization, when the correlation between accuracy and the response to social influence is zero. In each quadrant, the left panel shows the effect of social influence when groups are initially accurate, and the right panel shows the effect of social influence when groups are initially inaccurate. Simulations reflect experimental parameters, and thus study a population of N=40 making two revisions (three estimates) with k=4 neighbors per node in a decentralized network. Minimum 100 simulation repetitions per point. Initial beliefs are drawn from a random normal distribution with σ=1 and either μ=1 or μ =0.

Appendix B: Experimental Design Details

Figure A2. Screenshot of the experimental interface.

Equity	Trial Date
AAPL	11/26/2017
AMZN	11/26/2017
FB	11/26/2017
GOOG	11/26/2017
IBM	11/26/2017
IFLY	11/26/2017
INTC	11/26/2017
NFLX	11/26/2017
SPY	11/26/2017
TAN	11/26/2017
TWTR	11/26/2017
VNQ	11/26/2017
XBI	11/26/2017
CMCSA	11/29/2017
DELL	11/29/2017
EBAY	11/29/2017
GRUB	11/29/2017
HPQ	11/29/2017
NVIDIA	11/29/2017
PYPL	11/29/2017
SNAP	11/29/2017
TSLA	11/29/2017
DIS	12/2/2017
EXPE	12/2/2017
FORD	12/2/2017
XRT	12/2/2017
SBUX	12/2/2017
WFC	12/2/2017
YELP	12/2/2017

Equity	Trial Date
CSCO	12/3/2017
CITI	12/3/2017
NXP	12/3/2017
ORCL	12/3/2017
QCOM	12/3/2017
SONY	12/3/2017
TRIP	12/3/2017
CARZ	12/9/2017
FDN	12/9/2017
PFE	12/9/2017
VZ	12/9/2017
VOD	12/9/2017
XSD	12/9/2017
XTH	12/9/2017
AET	12/10/2017
AAL	12/10/2017
BOFA	12/10/2017
CVS	12/10/2017
DAL	12/10/2017
FOX	12/10/2017
GE	12/10/2017
JPM	12/10/2017
M	12/10/2017
NVS	12/10/2017
SINA	12/10/2017
TWX	12/10/2017

Table A1. List of forecasting tasks.

Conclusion

Social Influence in Decentralized Networks. The research inspiring this dissertation shared a common theoretical expectation for the wisdom of crowds: that individuals must remain independent in order to form accurate judgements. In contrast, this dissertation has shown that decentralized networks can allow social learning to increase the accuracy of beliefs within groups, even when the average belief remains unchanged. Theoretical results presented in Chapter 3 predict that the benefit to individuals is a robust expectation that can hold under a wide range of plausible conditions, even where the average belief becomes less accurate. Moreover, our results indicate that social influence in decentralized networks can increase the accuracy of the average belief itself—social influence can improve the wisdom of crowds estimate.

Together, chapter 2 and 3 demonstrate that the positive effects of social influence on belief accuracy hold across two distinct types of estimation tasks. While chapter 2 demonstrates the potential for social influence to improve the accuracy of arithmetic estimates, chapter 3 demonstrates how social networks can alleviate a tension intrinsic to financial markets. Equilibrium theories such as the efficient markets hypothesis argue that the aggregated decisions of many people produce market prices that accurately reflect the future value of assets, even as those markets are marked by enormous inequality in the distribution of profits. Moreover, retail investors tend to lose money compared to standard benchmarks, highlighting a stark divergence between collective efficiency and individual performance. By allowing individuals to learn from each other, Chapter 2 showed how social influence can reduce the inequality of belief accuracy between individuals. These results suggest that, when these beliefs are translated into decisions, social influence has the potential to reduce the inequality of success within financial markets by distributing gains more equally. In contrast with intuitive concerns that social influence in financial markets can lead to herding dynamics, the results presented in Chapter 2 suggest that social influence can improve individual performance without undermining market efficiency.

Remarkably, the results presented in Chapter 2 show more than just individual improvement as groups converge toward an accurate wisdom-of-crowds estimate. My analysis of arithmetic estimates found that, unexpectedly, the mean belief itself became more accurate. This result is not predicted by the DeGroot model under the assumption that responsiveness to social influence is identically distributed throughout a population. Instead, explaining the increase in accuracy of the group mean requires accounting for an empirical correlation between individual accuracy and individual response to social information. Notably, although the overall improvement in group accuracy was not statistically significant for financial forecasting tasks, both arithmetic estimation tasks and financial forecasting tasks demonstrated a positive correlation between accuracy and response to social information. These results suggest that the qualitative dynamics of collective intelligence are comparable for both arithmetic estimates and financial forecasting. The null result for financial forecasting is likely explained by the high initial accuracy, which left limited room for improvement and thus reduced the statistical power of detecting a decrease in the accuracy of the average belief.

There are several hypothetical explanations for this correlation between accuracy and response to social information. Survey-based research on estimate accuracy discussed in Chapter 1 suggests that individuals who are more accurate are also more confident, and that individuals who are more confident are less responsive to social influence. However, this explanation begs the question: why are individuals who are more accurate also more confident? One simple possibility is that some individuals put more cognitive effort into the task at hand (e.g., counting gumballs or studying relevant information for a financial forecast) which produces greater accuracy while also limiting the amount of attention available for social information. A second possibility is that some individuals have greater skill for the task at hand and also, consciously or subconsciously, are aware of their greater skill—as a result, the rational behavior would be to place less weight on social information. Yet another possibility is that innate estimation ability correlates with social intelligence: if people who have greater ability are also less likely to attend to social information due to personality characteristics, this correlation between accuracy and response to social information will occur.

Any or all of these explanations may be true, but the crucial factor for the collective dynamics of group belief formation is not why the correlation occurs but that the correlation occurs at all—that people who are more accurate make smaller revisions. Although the theoretical model presented in this dissertation indicates that the reason for this correlation should not matter, it remains possible that the different cognitive models may yield different collective dynamics—just as the initially unexpected heterogeneity between subjects produced qualitatively different collective outcomes. Thus future psychological research may focus specifically on how the use of social information relates to accuracy by identifying mediating factors (such as confidence or innate ability). To that end, these results show how research on collective dynamics can inform psychological research, just as psychological research can inform theoretical models of collective behavior. The results of this psychological research can then be analyzed with theoretical models to form predictions for how different behavioral patterns within a population can produce different collective outcomes.

Centralization. Even as this dissertation identified the potential benefits of social influence in decentralized networks, my theoretical model and empirical evidence also identified the potential risks of social influence in centralized networks. When a social network is characterized by the presence of one or a small number of highly central nodes, the effect of social influence depends entirely on the accuracy of those central individuals: accurate central nodes will increase both group-level and individual-level accuracy, and inaccurate central nodes will cause groups and individuals to become less accurate.

The risks created by network centralization provide a key insight to into the wisdom of crowds in practice. In the experiment presented in Chapter 2, centralization was generated by intentionally structured social networks. In practice, centralization can emerge endogenously, and central nodes may not be randomly placed. For example, individuals may attain influential positions due to status or perceived expertise. However, such individuals are not necessarily more likely to be accurate. For example, in a study of an online social networks devoted to investment discussion, Pan et al. (2012) found that individual popularity ranking is only loosely coupled to portfolio performance. That is, the actual success of the investors is largely irrelevant to their influence. Research on political influence networks have similarly found that people's estimation of their peers' political knowledge is only loosely coupled with the actual expertise of those peers (Klofstad, 2009). As a result, it is likely that highly influential central nodes are frequently no more likely than any other member of the population to hold accurate beliefs, generating effects similar to those described in Chapter 2 of this dissertation.

Thus one important area for future research is to identify empirically whether influential individuals are likely to hold more accurate beliefs. It is important to note that a correlation between accuracy and influence is not sufficient to eliminate the risks associated with centralization. Even where such a correlation exists, a noisy relationship can still undermine the wisdom of crowds as compared with decentralized networks. The greater the accuracy of the population as a whole, the stronger the correlation between accuracy and influence must be in order for central individuals to improve the beliefs of the population as a whole.

The effect of network centralization also offers an important guideline for harnessing the wisdom of crowds even in structured communication networks, especially since it is increasingly common for organizations to design platforms with the explicit goal of generating accurate economic and geopolitical forecasts (Bonabeau, 2009; Cowgill & Zitzewitz, 2009; Drogen & Jaha, 2013). However, one frequent practice in such contexts (e.g., on such forecasting platforms as Vetr.com and the IARPA forecasting challenge) is to allow individuals to view aggregated information on the estimates of other participants before providing their own estimate. Although this process is not explicitly targeted at producing social networks, the chronological ordering effects of participants entering the system does create an emergent centralization: those individuals who provide their estimates earlier in the process will be more influential than those individuals providing estimates later in the process. When this influence is mapped onto a directed social network, the individuals who provide earlier estimates will have a greater centrality, and the network as a whole will be centralized.

This dissertation draws attention to the importance of network structure and opens the door for future research to identifying those ways in which social information exchange can shape network structure as a result of both intentional and unintentional processes such as status and platform design.

Limiting Assumptions and Next Steps. Both the theoretical model and experimental design presented in this dissertation make several key assumptions. One assumption key to interpreting the experimental results presented in Chapter 2 and Chapter 3 is that allowing subjects to observe the average belief of their peers will produce comparable results to showing subjects each individual estimate by each peer. To some extent, this assumption is supported by prior empirical research. Lorenz et al. (2011) utilized both a full information condition and an aggregate information condition, producing qualitatively comparable results at the group level. Madirolas and Polavieja (2012) further analyze the data provided by Lorenz et al. (2011) and find that individual-level behavior in the full information condition can be consistently explained under the assumption that individuals make use of social information by adopting a weighted mean that combines their own initial belief with aggregated social information—precisely the model described in chapter 2. However, while these results show that behavior is largely consistent with the assumption that aggregated information is equivalent to detailed information, it is not definitive proof. Moreover, these results may not generalize to situations outside the laboratory, especially where factors such as social identity may lead individuals to place different weight on information coming from different classes of peers.

One key direction for future research is to understand empirically how people make use of social information in practice. After identifying empirical behavioral practices, this research will be able to make use of the general theoretical model presented in Chapter 2 to understand how such individual behavior impacts group belief formation. In political belief formation, for example, one common theory is that partisan bias leads people to make differential use of social information when it is presented by copartisans versus members of competing political ideology. Current research addressing this possibility is exploring the effect of providing partisan indicators along with numeric signals, finding that partisan indicators do not qualitatively change the effects of social information but do change the magnitude of effects: individuals make more use of social information in the absence of partisan signals (Guilbeault, Becker, & Centola, 2018). This research is also exploring how individuals respond to social information when people are shown detailed estimation that includes not only individual beliefs but also identity information about the individuals holding those beliefs.

Another key design feature of this dissertation is the explicit and sole use of numeric signals as a mechanism for conveying information about peer beliefs. In practice, however, social information can be exchanged through open discussion that includes not only numeric signals but also complex arguments, status signals, and other detailed information. One critical for future research is identifying whether such detailed discussion inhibits or enhances the wisdom of crowds. One possibility is that detailed discussion allows individuals to exchange status information which leads to unintentional influence centralization, thus undermining the wisdom of crowds. However, another possibility is that detailed discussion allows individuals to exchange factual and logical arguments which enhance decision-making. Discussion may also allow people to exchange confidence signals, enabling the group to collectively place more weight on accurate beliefs.

For each of the open questions discussed in this conclusion, the continued use of digital platforms as a mechanism for experimental research will play a vital role in determining how researchers and practitioners can relax the assumptions underlying this dissertation. By allowing careful experimental control over subject interactions, digital platforms will allow researchers to relax assumptions iteratively, gradually increasing the complexity of interactions as well as the type of social information exchanged. The high fidelity data collection enabled by digital platforms will also allow researchers to closely analyze and model the social processes that unfold in different types of interactions. Because the theoretical model presented in Chapter 2 provides a general framework for describing influence networks, all of these iterations can be directly compared to the findings presented in this dissertation.

Communication in practice is increasingly mediated by digital platforms, and the precise control enabled by web-based tools can provide an advantage not only in experimental design but also in implementation. Even with minimal modification, the experimental protocol used in chapters 2 and 3 can be readily adapted for implementation in decision-making contexts. For example, current research is investigating how this protocol can be applied in the context of mobile technology to support medical decision-making by physicians (Centola, Becker, Aysola, & Zhang, 2018). As I push forward to address more complex questions, such as the role of argumentation and other detailed discussion, I will investigate the principles of managed communication to understand how digital technology can be optimally designed to enable social networks of all types to grow their collective intelligence and harness the wisdom of crowds.

BIBLIOGRAPHY

Acemoglu, D., Dahleh, M. A., Lobel, I., & Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, *78*(4), 1201–1236.

Acemoglu, D., & Ozdaglar, A. (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, *1*(1), 3–49.

Armstrong, J. S. (2006). How to Make Better Forecasts and Decisions: Avoid Face-to-face Meetings. *Foresight: The International Journal of Applied Forecasting*, (5).

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., … Nelson, F. D. (2008). The promise of prediction markets. *Science-New York Then Washington-*, *320*(5878), 877.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Groups, Leadership, and Men. S*, 222–236.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., … Mellers, B. (2016a). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science*, *63*(3), 691–706.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., … Mellers, B. (2016b). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science*, *63*(3), 691–706.

Baddeley, M. (2010). Herding, social influence and economic decision-making: socio-psychological and neuroscientific analyses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 281–290.

Bailey, D. E., Leonardi, P. M., & Barley, S. R. (2012). The Lure of the Virtual. *Organization Science*, *23*(5), 1485–1504.

Baker, H. K., & Nofsinger, J. R. (2002). Psychological biases of investors. *Financial Services Review*, *11*(2), 97.

Bala, V., & Goyal, S. (1998). Learning from neighbours. *The Review of Economic Studies*, *65*(3), 595–621.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, *107*(3), 797–817.

Barabâsi, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, *311*(3), 590–614.

Barber, B. M., Lee, Y.-T., Liu, Y.-J., & Odean, T. (2009). Just how much do individual investors lose by trading? *Review of Financial Studies*, *22*(2), 609–632.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, *55*(2), 773–806.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, *22*(6), 725–730.

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., & Centola, D. (2018). From Efficient Markets to Effective Investors - Individuals and the Wisdom of Crowds. *Unpublished Manuscript.*

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183.

Bhattacharya, U., Hackethal, A., Kaesler, S., Loos, B., & Meyer, S. (2012). Is Unbiased Financial Advice to Retail Investors Sufficient? Answers from a Large Field Study. *The Review of Financial Studies*, *25*(4), 975–1032.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 992–1026.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, *5*(3), 387–424.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bonabeau, E. (2009). Decisions 2.0: The power of collective intelligence. *MIT Sloan Management Review*, *50*(2), 45.

Bruce, R. S. (1935). Group judgments in the fields of lifted weights and visual discrimination. *The Journal of Psychology*, *1*(1), 117–121.

Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science*, *61*(2), 267–280.

Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 1287–1335.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*(2), 591.

Casti, J. L. (1996). Seeing the light at El Farol: a look at the most important problem in complex systems theory. *Complexity*, *1*(5), 7–10.

Celis, L. E., Krafft, P. M., & Kobe, N. (2016). Sequential Voting Promotes Collective Discovery in Social Recommendation Systems. Retrieved from https://arxiv.org/abs/1603.04466

Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proceedings of the National Academy of Sciences*, *112*(7), 1989–1994.

Centola, D., Becker, J., Aysola, J., & Zhang, J. (2018). Improving Medical Decision Making through the Wisdom of Crowds. *In Preparation*.

Centola, D., Willer, R., & Macy, M. (2005). The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms1. *American Journal of Sociology*, *110*(4), 1009–1040.

Chater, N., Huck, S., & Inderst, R. (2010). Consumer decision-making in retail investment services: A behavioural economics perspective. *Report to the European Commission/SANCO*.

Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591–621.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583.

Cowgill, B., & Zitzewitz, E. (2015). Corporate prediction markets: Evidence from google, ford, and firm x. *The Review of Economic Studies*, rdv014.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458–467.

Dall'Asta, L., Baronchelli, A., Barrat, A., & Loreto, V. (2006). Nonequilibrium dynamics of language games on complex networks. *Physical Review E*, *74*(3), 036105.

Davis, G. F., & Greve, H. R. (1997). Corporate elite networks and governance changes in the 1980s. *American Journal of Sociology*, *103*(1), 1–37.

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.

DeMarzo, P. M., Vayanos, D., & Zwiebel, J. (2003). Persuasion Bias, Social Influence, and Unidimensional Opinions. *The Quarterly Journal of Economics*, *118*(3), 909–968.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*(3), 629.

Drogen, L. A., & Jha, V. (2013). Generating Abnormal Returns Using Crowdsourced Earnings Forecasts from Estimize.

Eagle, N., & Pentland, A. S. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*(4), 255–268.

Ebel, H., Mielsch, L.-I., & Bornholdt, S. (2002). Scale-free topology of e-mail networks. *Physical Review E*, *66*(3), 035103.

Economo, E., Hong, L., & Page, S. E. (2016). Social structure, endogenous diversity, and collective accuracy. *Journal of Economic Behavior & Organization*, *125*, 212–231.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica: Journal of the Econometric Society*, 1047–1071.

Ellison, G. (2000). Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *The Review of Economic Studies*, *67*(1), 17–45.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Farrell, S. (2011). Social influence benefits the wisdom of individuals in the crowd. *Proceedings of the National Academy of Sciences*, *108*(36), E625–E625.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*(2), 203.

Fisher, K. L., & Statman, M. (2000). Cognitive biases in market forecasts. *Journal of Portfolio Management*, *27*(1), 72–81.

Fishkin, J. S., & Luskin, R. C. (2005). Experimenting with a democratic ideal: Deliberative polling and public opinion. *Acta Politica*, *40*(3), 284–298.

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.

Golub, B., & Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149.

Golub, B., & Sadler, E. D. (2017). Learning in social networks.

Gordon, K. (1924). Group Judgments in the Field of Lifted Weights. *Journal of Experimental Psychology*, *7*(5), 398.

Graefe, A. (2014). Accuracy of vote expectation surveys in forecasting elections. *Public Opinion Quarterly*, *78*(S1), 204–232.

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.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.

Guilbeault, D., Becker, J., & Centola, D. (2018). Bipartisan Social Networks Reduce Political Bias in the Interpretation of Climate Trends. *Under Review.*

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.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, *162*(3859), 1243–1248. https://doi.org/10.1126/science.162.3859.1243

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231–237.

Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision Making*, *6*(1), 58.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, *21*(1), 40–46.

Hong, H., Du, Q., Wang, G., Fan, W., & Xu, D. (2016). Crowd Wisdom: The Impact of Opinion Diversity and Participant Independence on Crowd Performance.

Hong, L., & Page, S. E. (2008). Some microfoundations of collective wisdom. In H. Landemore & J. Elster (Eds.), *Collective Wisdom* (pp. 56–71). Cambridge University Press.

Hong, L., & Page, S. E. (2009). Generated and interpreted signals. *Journal of Economic Theory*, *144*(5), 2174–2196.

Hooker, R. H. (1907). Mean or median. *Nature*, *75*(1951), 487.

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.

Jenness, A. (1932). The role of discussion in changing opinion regarding a matter of fact. *Journal of Abnormal and Social Psychology*, *27*(3), 279–296.

Judd, S., Kearns, M., & Vorobeychik, Y. (2010). Behavioral dynamics and influence in networked coloring and consensus. *Proceedings of the National Academy of Sciences*, *107*(34), 14978–14982.

Juni, M. Z., & Eckstein, M. P. (2017). The wisdom of crowds for visual search. *Proceedings of the National Academy of Sciences*, *114*(21), E4306–E4315.

Kahneman, D., & Riepe, M. W. (1998). Aspects of investor psychology. *Journal of Portfolio Management*, *24*(4), 52–+.

Kahneman, D., & Tversky, A. (1977). *Intuitive prediction: Biases and corrective procedures*. DTIC Document.

Kandori, M., Mailath, G. J., & Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica: Journal of the Econometric Society*, 29–56.

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. *BioRxiv*, 288191.

Kauffman, S. A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA.

Kelley, E. K., & Tetlock, P. C. (2013). How wise are crowds? Insights from retail orders and stock returns. *The Journal of Finance*, *68*(3), 1229–1265.

Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popović, Z., & Baker, D. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, *108*(47), 18949–18953.

Kittur, A., & Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: quality through coordination. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 37–46). ACM.

Klugman, S. F. (1945). Group Judgments for Familiar and Unfamiliar Materials. *The Journal of General Psychology*, *32*(1), 103–110. https://doi.org/10.1080/00221309.1945.10544484

Klugman, S. F. (1947). Group and individual judgments for anticipated events. *The Journal of Social Psychology*, *26*(1), 21–28.

Knight, H. C. (1921). *A comparison of the reliability of group and individual judgments* (PhD Thesis). Columbia University (1921?).

Koriat, A. (2012). When are two heads better than one and why? *Science*, *336*(6079), 360–362.

Kossinets, G., & Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science*, *311*(5757), 88–90.

Krackhardt, D., & Hanson, J. R. (2001). 16 informal networks: The company behind the chart. *Creative Management*, 202.

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.

Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, *32*(6), 951.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, *52*(4), 667–694.

Lazer, D., & Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, *43*, 19–39.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge: Harvard University Press.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Aberg, Y. (2001). The web of human sexual contacts. *Nature*, *411*(6840), 907–908.

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.

Malkiel, B. (2016). *A Random Walk Down Wall Street*. New York: W.W. Norton and Company.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59–82.

Malone, T. W., & Bernstein, M. S. (2015). *Handbook of collective intelligence*. MIT Press.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, *107*(2), 276.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., … others. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, *25*(5), 1106–1115.

Montanari, A., & Saberi, A. (2010). The spread of innovations in social networks. *Proceedings of the National Academy of Sciences*, *107*(47), 20196–20201.

Mossel, E., Sly, A., & Tamuz, O. (2015). Strategic learning and the topology of social networks. *Econometrica*, *83*(5), 1755–1794.

Mossel, E., & Tamuz, O. (2017). Opinion exchange dynamics. *Probability Surveys*, *14*, 155–204.

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLoS One*, *8*(11), e78433.

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science*, *32*(7), 1133–1147.

Murr, A. E. (2015). The wisdom of crowds: Applying Condorcet's jury theorem to forecasting US presidential elections. *International Journal of Forecasting*, *31*(3), 916–929.

Myers, D. G., & Bishop, G. D. (1971). Enhancement of dominant attitudes in group discussion. *Journal of Personality and Social Psychology*, *20*(3), 386.

Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., & Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, *2*, 1.

Niederer, S., & Van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, *12*(8), 1368–1387.

Nofer, M., & Hinz, O. (2014). Are crowds on the internet wiser than experts? The case of a stock prediction community. *Journal of Business Economics*, *84*(3), 303–338.

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. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)* (pp. 203–209). IEEE.

Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, *34*(1), 17–29.

Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, *541*(7638), 532–535.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375.

Schijven, M., & Hitt, M. A. (2012). The vicarious wisdom of crowds: Toward a behavioral perspective on investor reactions to acquisition announcements. *Strategic Management Journal*, *33*(11), 1247–1268.

Sherif, M. (1935). A study of some social factors in perception. *Archives of Psychology (Columbia University)*. Retrieved from http://psycnet.apa.org/psycinfo/1936-01332-001

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. (2017). The Wisdom of Polarized Crowds. *ArXiv Preprint ArXiv:1712.06414*.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, *17*(1), 83–104.

Shore, J., Bernstein, E., & Lazer, D. (2015). Facts and figuring: An experimental investigation of network structure and performance in information and solution spaces. *Organization Science*, *26*(5), 1432–1446.

Sjöberg, L. (2009). Are all crowds equally wise? A comparison of political election forecasts by experts and the public. *Journal of Forecasting*, *28*(1), 1–18.

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.

Tziralis, G., & Tatsiopoulos, I. (2012). Prediction markets: An extended literature review. *The Journal of Prediction Markets*, *1*(1), 75–91.

Way, S. F., Morgan, A. C., Clauset, A., & Larremore, D. B. (2017). The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences*, 201702121. https://doi.org/10.1073/pnas.1702121114

Welinder, P., Branson, S., Perona, P., & Belongie, S. J. (2010). The multidimensional wisdom of crowds. In *Advances in neural information processing systems* (pp. 2424–2432).

Wolf, M., Krause, J., Carney, P. A., Bogart, A., & Kurvers, R. H. J. M. (2015). Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PloS One*, *10*(8), e0134269.

Wolfers, J., & Zitzewitz, E. (2004). *Prediction markets*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w10504

Wolfers, J., & Zitzewitz, E. (2006). *Interpreting prediction market prices as probabilities*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12200

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.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology*, *59*(5), 555.

Young, H. P. (1993). The evolution of conventions. *Econometrica: Journal of the Econometric Society*, 57–84.

Zajonc, R. B. (1962). A note on group judgements and group size. *Human Relations*, *15*(2), 177–180.

Zarnowitz, V. (1984). The accuracy of individual and group forecasts from business outlook surveys. *Journal of Forecasting*, *3*(1), 11–26.