Social Influence Undermines the Wisdom of the Crowd in Sequential Decision Making

Frey, Vincenz; Van de Rijt, Arnout

Published in:
Management Science

DOI:
10.1287/mnsc.2020.3713

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Social Influence Undermines the Wisdom of the Crowd in Sequential Decision Making

Vincenz Frey*, Arnout van de Rijt

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 67, No. 7, July 2021, pp. 4273–4286
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Social Influence Undermines the Wisdom of the Crowd in Sequential Decision Making

Vincenz Frey,[a,*] **Arnout van de Rijt**[b,c]

[a] Department of Sociology, University of Groningen, 9712 TG Groningen, Netherlands; [b] Department of Political and Social Sciences, European University Institute, 50014 Florence, Italy; [c] Department of Sociology, Utrecht University, 3584 CH Utrecht, Netherlands

*Corresponding author

Contact: v.c.frey@rug.nl, https://orcid.org/0000-0001-5665-9697 (VF); a.vanderijt@eui.eu, https://orcid.org/0000-0002-4208-3452 (AvdR)

Received: August 21, 2018
Revised: November 7, 2019; February 3, 2020
Accepted: May 6, 2020
Published Online in Articles in Advance: October 8, 2020

https://doi.org/10.1287/mnsc.2020.3713

Abstract. Teams, juries, electorates, and committees must often select from various alternative courses of action what they judge to be the best option. The phenomenon that the central tendency of many independent estimates is often quite accurate—"the wisdom of the crowd"—suggests that group decisions based on plurality voting can be surprisingly wise. Recent experimental studies demonstrate that the wisdom of the crowd is further enhanced if individuals have the opportunity to revise their votes in response to the independent votes of others. We argue that this positive effect of social information turns negative if group members do not first contribute an independent vote but instead cast their votes sequentially such that early mistakes can cascade across strings of decision makers. Results from a laboratory experiment confirm that when subjects sequentially state which of two answers they deem correct, majorities are more often wrong when subjects can see how often the two answers have been chosen by previous subjects than when they cannot. As predicted by our theoretical model, this happens even though subjects' use of social information improves the accuracy of their individual votes. A second experiment conducted over the internet involving larger groups indicates that although early mistakes on easy tasks are eventually corrected in long enough choice sequences, for difficult tasks wrong majorities perpetuate themselves, showing no tendency to self-correct.

History: Accepted by Yuval Rottenstreich, decision analysis.
Funding: V. Frey benefited from funding from the H2020 European Research Council's project DisCont [Grant 694262]. A. van de Rijt was supported by the National Science Foundation, Division of Social and Economic Sciences [Grant SES-1340122].
Supplemental Material: Data and the online appendices are available at https://doi.org/10.1287/mnsc.2020.3713.

Keywords: wisdom of the crowd • social influence • sequential decision making • information cascades • team performance

1. Introduction

The wisdom of the crowd is the phenomenon that when many individuals independently estimate a quantity, the central tendency will be rather accurate—closer to the truth than most individual estimates and often even closer to the truth than experts (Surowiecki 2004, Sunstein 2006, Sjöberg 2009, Davis-Stober et al. 2014). This surprising accuracy of group judgments can be understood as resulting from individual errors canceling out (Hogarth 1978, Hong and Page 2004, Larrick and Soll 2006, Keuschnigg and Ganser 2016). Different individuals make different mistakes, because they hold diverse perspectives and bring in different areas of expertise. When individual mistakes are distributed symmetrically around the truth, the aggregation of many assessments produces an estimate close to the truth. Condorcet's jury theorem (Condorcet 1785, Baker 1975) provides a theoretical foundation for crowd wisdom in binary decision situations where individual voters are more likely to vote for the right option than for the wrong option. The theorem states that the majority rule (the median choice) will almost always produce the right verdict in a jury consisting of a large number of independent voters.

The wisdom of the crowd has been observed in a wide range of settings, in estimating weights of objects (Galton 1907, Wagner and Suh 2014), in political forecasts (Sjöberg 2009, Murr 2015), climate-related events (Hueffer et al. 2013), physician diagnostics (Kurvers et al. 2016), and economic forecasts (Kelley and Tetlock 2013, Budescu and Chen 2014, Nofer and Hinz 2014). The principle has been applied by practitioners across fields (Surowiecki 2004) and used in electronic technologies and online platforms that collect and aggregate opinions, for example, for organizational decision making (Spann and Skiera 2003, Armstrong 2006, Cowgill and Zitzewitz 2015).

The focus of this article is on scenarios in which individuals do not form their opinions independently but, as is plausible in many real-world settings, are influenced by those of others (Cialdini and Goldstein 2004). We ask whether such social influence enhances or impedes crowd wisdom. Social influence will be at play in decision processes in various small groups, such as organizational teams, administrative boards, hiring committees, military units, and juries. With the rise of digital technologies and online services that provide social information on a large scale, social influence may increasingly also affect decisions and behaviors of large crowds and whole societies.

Recent experiments suggest that social influence enhances the wisdom of crowds (Lorenz et al. 2011, Lorenz et al. 2015, Becker et al. 2017, Friedkin and Bullo 2017; for theoretical results, see Ganser and Keuschnigg 2018). In these experiments, all subjects first stated their independent estimates of some quantity—for example, the number of calories in a given meal—and they could then revise their estimates after receiving information about the estimates of others. In the experiment of Becker et al. (2017), with several rounds of feedback and adjustment, participants generally revised their estimates in the direction of the crowd, consistent with a social influence mechanism, and the revisions *improved* crowd wisdom: The group median and mean moved significantly toward the truth. Becker et al. (2017) furthermore show that the findings of an earlier study with such a setup (Lorenz et al. 2011) are also consistent with a wisdom-enhancing effect of social influence, although that study originally reported no significant impact on collective accuracy.[1] In a more recent experiment, Lorenz et al. (2015) also find social influence to increase collective accuracy, with the median of final estimates often being significantly closer to the truth than the median of initial estimates, after different modes of numeric feedback and discussion via a chat function. Finally, Friedkin and Bullo (2017) ask subjects to calculate a probability, finding that compared with initial estimates, the fraction of correct answers increased after deliberation in small groups. Friedkin and Bullo (2017) do not report whether the group median or mean improved as a result of social influence, but this was likely the case given the increased prevalence of correct answers.

Becker et al. (2017; see also Friedkin and Bullo 2017) provide a theoretical account of these findings, building on the DeGroot learning model (DeGroot 1974), where individuals update their prior estimates as a weighted mean of others' estimates. If less accurate individuals are more strongly influenced than more accurate individuals—which Becker et al. found and which was also observed by Madirolas and de Polavieja (2015) in the data of Lorenz et al. (2011)—then DeGroot learning produces the result that social influence improves the group's median estimate.

In the present study, we qualify this emergent claim that social influence promotes crowd wisdom. The claim stands at odds with the conventional theoretical result that social influence renders groups *less* accurate. The conventional argument is that because social influence leads to positively correlated opinions, it undermines the usefulness of the opinion brought in by each additional group member. Collective accuracy deteriorates when opinions are correlated (Hogarth 1978; Grofman et al. 1983; Clemen and Winkler 1985; Ladha 1992, 1995; Broomell and Budescu 2009; Davis-Stober et al. 2014). We reconcile this tension between classic theory and recent experimental work, arguing that an initial stage in which individuals first make independent estimates prevents the deleterious effect of social influence. Such an independent start is assumed in the DeGroot model and was incorporated in the design of each of the experiments discussed above. An independent start limits social influence as everybody at least initially contributes an independent estimate.[2] Moreover, the median of these independent estimates must—because of the wisdom of the crowd!—provide a "wise anchor" that is close to the truth and around which the ensuing social influence process can converge.[3]

A broad range of real-world scenarios will lack such an independent start. Instead, choices are made one at a time rather than all at once, implying that assessments that are contributed to the group early can influence later ones. For example, if a group of managers or professors decides which of two strategies to follow or which job candidate to hire, they will typically *not* reach a decision through a process that starts with everyone contributing a truly independent assessment. There will rather be a discussion in which views are shared sequentially. This suppresses the development of independent opinions because those who express their views later may recruit further evidence favoring the popular opinion and focus on aspects that others have emphasized (Tversky and Kahneman 1974, Mussweiler et al. 2000). In addition to organizational decision making, other examples of such sequential decision making include jury deliberations (Sunstein 2000; Henrich 2015, p. 138), the adoption of competing technologies (Arthur 1989), consumer choices (Salganik et al. 2006), collective belief in news items (Shao et al. 2016, Allcott and Gentzkow 2017), and crowdfunding (van de Rijt et al. 2014, Alpern and Chen 2017, Polzin et al. 2018).[4]

We examine the impact of social influence on the wisdom of the crowd in situations in which group members state sequentially which of two alternatives they deem correct. Each group member first learns how often the two alternatives were chosen by previous group members and then makes his or her own choice. Our theoretical model captures the intuition that under social influence mistakes by those who speak first may lead those who come later to make the same mistakes, such that an unfortunate start can propagate and lead to a wrong majority. The phenomenon that early mistakes can lead those who come later to err is known from models of information cascades (Banerjee 1992, Bikhchandani et al. 1992). As in these models, individuals can improve their own assessments when adjusting them on the basis of others' earlier assessments: Because of the wisdom of crowds, it is likely that the majority of earlier assessments is more accurate than one's own assessment. Hence, rational individuals may follow the crowd to improve their individual accuracy, but by doing so they may undermine the accuracy of the crowd's verdict. This idea of a "rational herd" is central in the literature on social learning and information cascades (Chamley 2004, Rauhut et al. 2011). In the next section, we develop a computational model that applies this notion to the context of the wisdom of the crowd. The predictions of this model—that social influence enhances individual accuracy but undermines group accuracy—are tested in a laboratory experiment (Section 3) and an online experiment (Section 4), and we discuss implications of our findings for decision making in groups (Section 5).

2. Model

In our model, individuals $1 \leq i \leq n$ make a binary choice C_i between two alternatives A and B, of which only one is correct. $C_i = 1$ if individual i's choice is correct, and $C_i = -1$ if false. Choices are made sequentially. We define the difficulty of the choice task d as the probability that an independently choosing individual is wrong. We assume that independent assessments are more likely right than wrong ($d < 0.5$).[5]

We consider two scenarios, which correspond to the two conditions in our human subject experiments. In the *independent scenario*, individuals choose independently, knowing nothing about the choices of those who came before. In the *social influence scenario*, each individual observes the total number of times alternatives A and B were chosen by preceding individuals. Individuals do not know the order in which past choices were made. For the social influence scenario, we assume that an individual's tendency to choose option A tends to increase in the relative share of previous votes for A. We formulate individual choice within the common logistic choice framework:

$$Prob(C_i = 1) = \left(1 + \frac{d}{1-d} e^{-s_i \frac{\sum_{j<i} C_j}{i-1}}\right)^{-1}. \qquad (1)$$

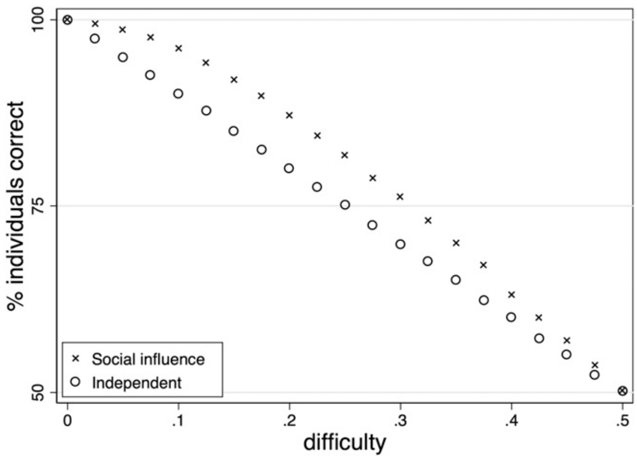
In Equation (1), s_i represents the susceptibility of individual i to social influence. For the independent scenario, we assume $s_i = 0 \; \forall \; i$. For the social influence scenario, we assume that individuals exhibit varying levels of susceptibility to social influence (compare Sewell 2018, p. 135). We assume that average susceptibility to social influence is positive; but by assuming a normal distribution of $s_i \sim N(\mu > 0, \sigma > 0)$, we allow a portion of individuals to be counterconformists ($s_i < 0$) who are influenced toward the alternative that has less support. In sum, Equation (1) implies that the chance of a correct choice decreases in d and—in the influence condition—increases in the relative share of previous choices of the correct answer.

We now derive two hypotheses from simulations of this model. For the social influence scenario, we present results for an average susceptibility to influence of $\mu = 3$ with an individual-level variance of $\sigma = 1$. We first focus on small groups of 12 individuals (the median group size in our laboratory experiment) and later extend the analysis to groups of 100 individuals (the size of "large groups" in the online experiment). We simulate the model for 10,000 groups in each scenario (with social influence or without) and for each of 21 difficulty levels ($d = 0.000, 0.025, 0.050, \ldots 0.500$).

First, we find that social influence increases individual accuracy. Because of the wisdom of the crowd, the more popular answer is more likely correct than a single individual's assessment. Therefore, the pull toward the more popular answer is more likely to correct an otherwise wrong answer than to lead an individual astray from the right answer. This simulation result is illustrated in Figure 1. The horizontal axis measures the level of difficulty ranging from $d = 0.0$ (everyone knows the answer) to $d = 0.5$ (independent, individual assessments are a coin flip). The vertical axis measures the fraction of individuals choosing correctly. The fraction of correct choices decreases with difficulty in the independent scenario (circles) as well as in the social influence scenario (crosses). If choices are made independently, the expected proportion of individuals choosing correctly simply equals $1 - d$, by the definition of difficulty d (if $s_i = 0 \; \forall \; i$, Equation (1) reduces to $Prob(C_i = 1) = 1 - d$). In the presence of social influence ($s_i \sim N(3,1)$), the proportion of correct choices is higher. Figure 1 thus shows that regardless of the difficulty of the choice at hand, individuals perform better if they have information about the choices of others and tend to follow the crowd than if they choose independently. These results consistently obtain in simulations with varying $\mu > 0$.

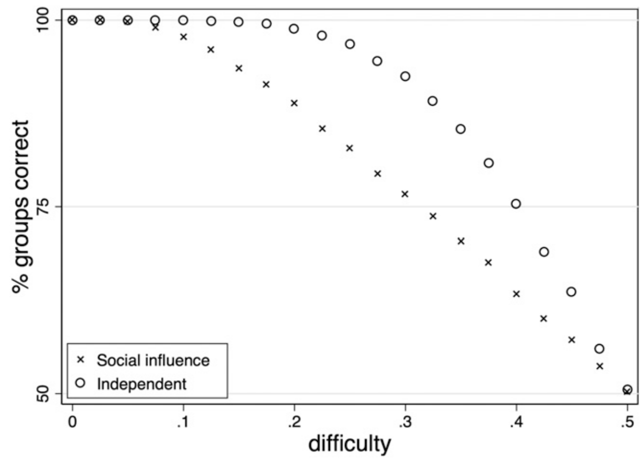
Hypothesis 1. Social influence *increases* the probability that an individual chooses correctly.

Figure 1. Predicted Percentages of Correct Choices as a Function of Difficulty, with Social Influence ($s_i \sim N(3,1)$) and Without Social Influence ($s_i = 0 \ \forall \ i$)



Note. Shown are averages of 10,000 runs for groups of size 12, for each difficulty level and for both scenarios.

Figure 2. Predicted Percentages of Correct Majorities as a Function of Difficulty, with Social Influence ($s_i \sim N(3,1)$) and Without Social Influence ($s_i = 0 \ \forall \ i$)



Note. Shown are averages of 10,000 runs for groups of size 12, for each difficulty level and for both scenarios.

The effect of social influence completely reverses as we shift the focus from performance at the individual level to performance at the level of the group. We define a group as wise if at least half of the group members choose the correct alternative. We thus use the simple majority rule, which is prevalently used in theory and practice and has been found to often perform about as well as more complex aggregation rules (Kerr and Tindale 2004, Hastie and Kameda 2005). Moreover, the majority rule naturally captures how go/no-go decision are commonly made by groups in binary choice environments. Figure 2 shows the relationship between the presence of social influence and the simulated probability of a correct majority, at different levels of task difficulty d. Consistent with the jury theorem (Condorcet 1785, Baker 1975), groups of independently choosing individuals are almost always wise when choice difficulty is not too high (circles). As task difficulty approaches impossible (0.5), the probability of an incorrect majority becomes substantial because the simulation results in Figure 2 pertain to groups of only 12 individuals. In the social influence scenario, groups are more often wrong at every level of difficulty (crosses). This happens because of occasional false starts, in which the first subjects make the wrong choice. These false starts are prolonged when consecutive subjects are influenced toward the wrong but more popular option and may eventually produce wrong final majorities.

Hypothesis 2. Social influence *decreases* the probability that the majority choice is correct.

Hypotheses 1 and 2 predict opposite effects of social influence on individual and group wisdom. This happens because, in the model from which both are derived, social influence impacts both the median vote and the variance of the median vote. The expectation of the median vote is adjusted toward the correct answer, improving individual wisdom, while at the same time the standard error of the median is inflated, making a wrong majority more likely.

So far we have studied groups of modest size, consisting of 12 individuals. Does social influence also undermine the wisdom of the crowd in large groups? As it turns out, this depends on the strength of social influence and the difficulty of the choice. Under weak social influence, false starts are ultimately corrected in large enough crowds, regardless of difficulty. For homogeneous $s_i \sim N(\mu,0)$, it can be derived from Equation (1) that there exists for any task difficulty d a critical influence level $\mu^*(d)$ below which the probability of an individual following a wrong majority is always smaller than the relative size of that majority (Online Appendix D). As a result, wrong majorities shrink and become a minority after enough individuals have made a choice. Early mistakes do not propagate. This phenomenon of a self-correcting dynamic at low levels of feedback (van de Rijt 2019) has previously been observed in residential segregation models (Bruch and Mare 2006, van de Rijt et al. 2009) and information cascade experiments (Goeree et al. 2007). Above the critical social influence level, wrong majority answers approach a stable fraction. The critical influence level $\mu^*(d)$ cannot be expressed as a function of d in closed form, but it approaches 2 as d approaches 0.5. For less difficult questions, critical influence levels are higher (Online Appendix D).

Relaxing the assumption of homogeneous s and assuming finite groups of size 100 in simulations, we find that wrong majorities are increasingly common for higher degrees of social influence. Figure 3 shows for different levels of average social influence, μ, the percentage of correct majorities in these large groups. The results are averages of 10,000 runs at each difficulty level, with heterogeneity $\sigma = 1$. Under strong social influence ($\mu = 3$), groups frequently failed to reach a correct majority when choices are of intermediate difficulty (black crosses). Failures are more rare under weak social influence ($\mu = 2$; gray crosses) and never happen under independent choice, except when choices are very difficult (circles).

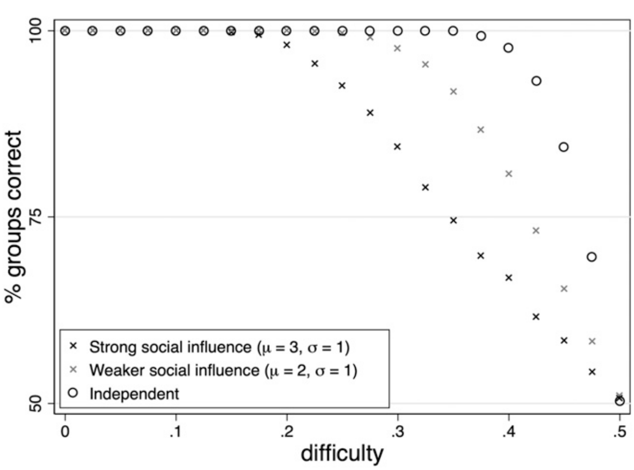
3. Laboratory Experiment

3.1. Design

In our experiment conducted at the Experimental Laboratory for Sociology and Economics at Utrecht University, we presented groups of 10–14 participants 30 questions.[6] Subjects answered the questions at separate computer stations and had no access to the internet for looking up correct answers. There were six questions in each of the five categories: visual, art, equations, history, and geometry. For each question, a subject had 20 seconds to choose between two answers, of which only one was correct. A subject could choose not to give an answer by letting this 20-second timer run out (a countdown was shown on the screen). Figure 4 shows one of the questions, and Online Appendix C provides an overview of all questions.

The members of a group answered a given question sequentially, one after the other. To avoid long waiting times, the 30 questions were placed in cyclic order. All subjects of all groups answered the questions in this order, but different subjects of the same group started at different positions in the cycle. In the *social influence condition*, the participants were truthfully informed about how many prior group members had opted for each answer, as illustrated in Figure 4. These popularity counts ("X times chosen") were not shown in the *independent condition*.

Figure 3. Predicted Percentages of Correct Majorities as a Function of Difficulty, with Strong Social Influence ($s_i \sim N(3,1)$), Weaker Social Influence ($s_i \sim N(2,1)$), and Without Social Influence ($s_i = 0 \ \forall \ i$)



Note. Shown are averages of 10,000 runs for groups of size 100, for each difficulty level and for each scenario.

Figure 4. One of the 30 Questions Asked of Laboratory Subjects

When did Germany invade Denmark?

□ 1939 (7 times chosen)

□ 1940 (2 times chosen)

We implemented an incentive scheme capturing typical organizational decision-making settings in which a team member's well-being depends predominantly on whether the team takes the correct decision but also on being among those who in hindsight voted for the correct decision. Accordingly, subjects were paid €0.10 per question that was answered correctly by a majority in their group and an additional €0.05 for each question they answered correctly individually. We return to the role of the incentive scheme in Section 4.3.

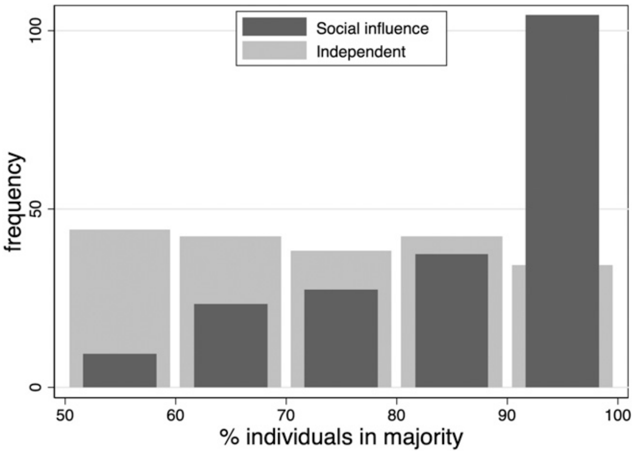
We conducted eight experimental sessions, involving in total 192 participants, predominantly undergraduate students. Each session had 21, 24, or 27 participants split randomly into two groups that differed in size at most by one participant. In each session, one of the two groups was assigned to the independent condition and the other to the influence condition.

3.2. Results

We focus our analysis on the 25 questions that were answered correctly by a majority of subjects in the independent condition, that is, the 25 questions with observed difficulty $d < 0.5$.[7] We report analyses for the other five questions in Online Appendix B. For the latter five questions, most subjects had the wrong intuition, so that our theory does not predict either a positive effect of social influence on individual wisdom, Hypothesis 1, nor a negative effect of social influence on crowd wisdom, Hypothesis 2 (see Endnote 5).

Figure 5 shows that our social influence manipulation had a substantial impact on decision making, leading to concentration on one of the two answers. In the influence condition, large majorities of 90%–100% were the most common outcome, whereas small majorities of 50%–60% were the least common. In the

Figure 5. Distribution of Majority Size by Experimental Condition



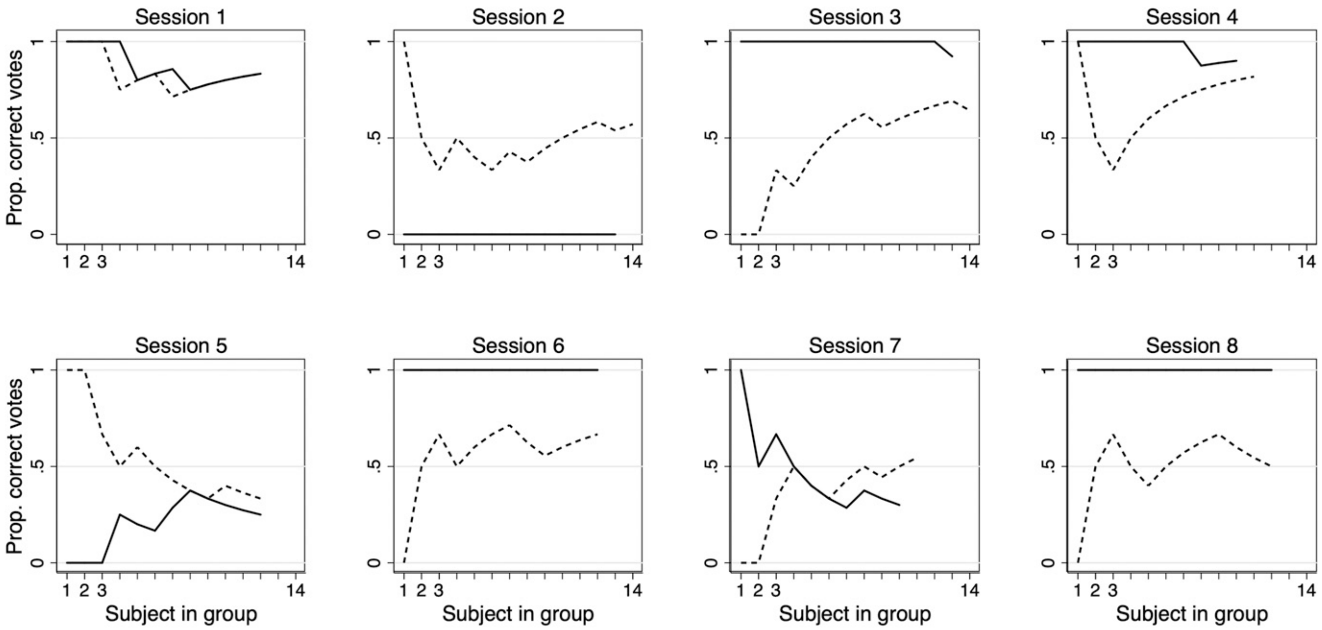
independent condition, majorities were about equally likely to be of any of the five sizes distinguished in Figure 5. This pattern suggests that the provision of social information led participants to herd around emergent majority answers.

The stickiness of early majorities and the emergence of larger majorities in the influence condition is illustrated in Figure 6, showing data from the question on the year of the German invasion of Denmark (see Figure 4). In most sessions, the proportion of correct answers in the independent condition (dashed lines) moved toward some intermediate value above 0.5 as more subjects cast their vote. This is to be expected, given an average proportion of correct answers in the independent condition of 0.61 (i.e., question difficulty $d = 0.39$). Providing information on previous group members' answers changed the dynamics quite visibly, leading to larger majorities and more wrong majorities in the influence condition (solid lines). In sessions 1, 3, 4, 6, and 8, the first subjects in the influence condition gave the correct answer ("1940"); this led almost all subsequent subjects to give the correct answer, too. Crucially, however, social influence could likewise lead to the propagation of wrong early majorities, with session 2 as an extreme example.

As predicted by Hypothesis 1, social influence improved individual accuracy in our laboratory experiment (Table 1). On average, 74.0% of the answers of a participant in the influence condition were correct. By contrast, in the independent condition only 71.2% of the answers were correct. A Mann–Whitney U test rejects the null hypothesis that the distribution of the percentages of correct answers per participant is identical across the influence condition and the independent condition ($n = 192$, $p = 0.040$). This evidence supports Hypothesis 1.

On the other hand, social influence undermined group performance, as predicted by Hypothesis 2. On the 25 questions with difficulty $d < 0.5$, the average percentage of correct majority answers per group was 89.5% in the independent condition but only 80.0% in the influence condition (Table 1). Figure 7 shows the percentage of correct majorities in the two conditions

Figure 6. Running Proportion of Correct Answers to the Denmark Question (Figure 4) in the Social Influence Condition (Solid Lines) and the Independent Condition (Dashed Lines), by Session



Note. Each line pertains to a single experimental group.

Table 1. Mean Levels of Individual Accuracy (% Correct Answers per Participant) and Collective Accuracy (% Correct Majority Answers per Group) in the Laboratory Experiment

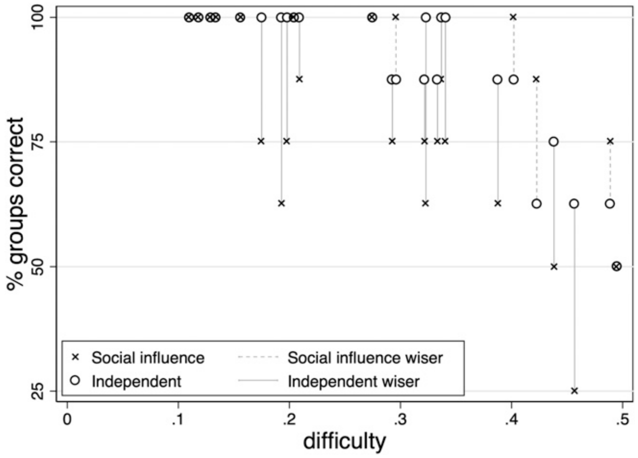
	Independent condition	Influence condition	*p*-value of difference
Individual accuracy	71.2	74.0	0.040[a]
Collective accuracy	89.5	80.0	0.014[b]

[a]Mann–Whitney U test.
[b]Wilcoxon signed-rank test (see main text).

separately for each question, arranged by question difficulty d (analogously to the display of simulation results in Figure 2). Seven of the easier questions were answered correctly by the majority in each group, regardless of the experimental condition. For 13 questions, we observed the predicted deleterious effect of social information on group accuracy: These questions produced more correct majorities in the independent condition than in the influence condition (solid lines). By contrast, only four questions produced more correct majorities in the influence condition than in the independent condition (dashed lines). Table A1 in Online Appendix A shows that, in seven of the eight sessions, the group in the influence condition had a correct majority answer to fewer questions than the group in the control condition. This group-level difference is significant in a Wilcoxon signed-rank test (n = 8, $p = 0.014$). Thus, the data from our laboratory experiment strongly support the hypothesis that social influence undermines crowd wisdom.

The data furthermore indicate that social information may undermine crowd wisdom also in groups that are larger than the 10- to 14-person groups observed in our laboratory. A logistic estimation of Equation (1) yields an estimate for the average susceptibility to social influence of 3.55 (95% confidence interval (CI): 2.96, 4.13), with a subject-level variance estimated at 3.29 (95% CI: 1.69, 6.38; see Table A2 in the online appendix). Our simulation analysis for such high and heterogeneous levels of social influence in large groups (Figure 3) suggests that wrong early majorities on difficult questions in the influence condition would in many cases have persisted across much longer strings of subjects.

Figure 7. Percentages of Groups with a Correct Majority per Question by Experimental Condition, Arranged by Question Difficulty



Notes. Each pair of a circle and a cross connected by a line pertains to one question. (x-axis values are jittered if two questions have the same observed difficulty d.)

4. Online Experiment

We additionally conducted an online experiment with participants recruited through Amazon's Mechanical Turk platform, rerouted to our study website. The purpose of the online experiment is twofold. First, we aimed to replicate the laboratory results in a different empirical environment and with a different subject population (on the composition of subject populations recruited via Mechanical Turk, see Huff and Tingley 2015, Levay et al. 2016).[8] Second, the internet provides the necessary scale to investigate under what conditions social influence undermines crowd wisdom also in larger groups and when false starts are ultimately corrected. The theoretical model suggests that early wrong majorities perpetuate if social influence is sufficiently strong and the question sufficiently difficult whereas they are eventually corrected otherwise (compare Figure 3, Online Appendix D).

4.1. Design

In the online experiment, we studied groups of both small size (15) and large size (100). We had 25 groups per group size in the influence condition as well as in the independent condition. The online experiment largely follows the protocol of our laboratory experiment, with a number of modifications intended to mitigate the threats to data quality posed by the use of subjects recruited through the internet.

First, we added three trivial attention-check questions (see Online Appendix C) and excluded participants who did not answer all of them correctly. We also excluded participants who gave overly fast answers (< one second) to more than three questions, excluding the attention checks; and we admitted only high-reputation participants who had at least a 95% approval rate on Mechanical Turk. Second, we used only questions in the categories visual, art, and geometry because we could not prevent participants from looking up correct answers to equations and history questions within the given 20 seconds. Third, although in the laboratory all members of a group were active simultaneously (answering different questions), only one person per group was active at a time in the online experiment. This allowed avoiding waiting times and keeping participants engaged. In addition, it allowed replacing excluded participants and removing the choices of an excluded participant from the popularity counts before the next group member was admitted. Fourth, to mitigate the risk that, after some questions, participants would begin to answer randomly and rapidly, we presented each participant with only 15 questions, rather than 30. Fifth, to discourage participants from participating multiple times using different accounts and in between trials looking up correct answers, we presented different groups with different subsets of 15 questions drawn from a larger set of 30 questions. (Online Appendix C provides an overview of all questions and the composition of the subsets of 15 questions.) Finally, we added "Abstain" as a possible answer in order to avoid that participants could make money by simply letting the 20 seconds timer run out on each question. Participants earned the $0.06 paid for a correct group majority only if they chose to abstain or gave an answer to the question. They earned an additional $0.03 per question that they answered individually correctly. There was an additional baseline payment of $0.30 for participation in the run with small groups, which was increased to $0.50 for the run with large groups, to ensure enough participants could be recruited to complete the long choice chains.

4.2. Results

The social influence treatment was also effective online.[9] In small as well as large groups, small majorities comprising less than 70% of the group members were more frequent if everyone had to give an independent answer than if information about earlier answers was provided. On the other hand, large majorities of more than 80% were more common with social influence than without (see Figure A1 in the online appendix).

Social information again helped individual participants at finding correct answers (Table 2). In small groups, participants in the independent condition answered on average 70.5% of the questions correctly, whereas those in the influence condition scored on average 71.6% of correct answers. In large groups, the difference was more pronounced, with 68.6% versus 74.7%. Mann-Whitney U tests reject the null-hypothesis of no difference in the percentage of correct answers per participant across the influence condition and the independent condition for large groups (n = 4,985, p = 0.000) but not for small groups (n = 748, p = 0.409).[10] In sum, the online experiment provides some further support for Hypothesis 1.

Social influence also tended to undermine the accuracy of group majorities. Figure 8(a) shows that for small groups, the percentage of correct majorities was higher in the independent condition than in the influence condition on 16 questions. By contrast, small groups in the influence condition outperformed those in the independent condition on only 4 questions. Overall, small groups had a correct majority answer in the independent condition to 91.5% of the questions, whereas their counterparts in the influence condition had a correct majority answer to only 82.0% of the questions (Table 2). A Mann-Whitney U test rejects the null-hypothesis of no difference in the percentage of correct majority answers per group (n = 50, p = 0.001). Thus, the online experiment provides further evidence for Hypothesis 2 in small groups in addition to the support obtained in the laboratory.

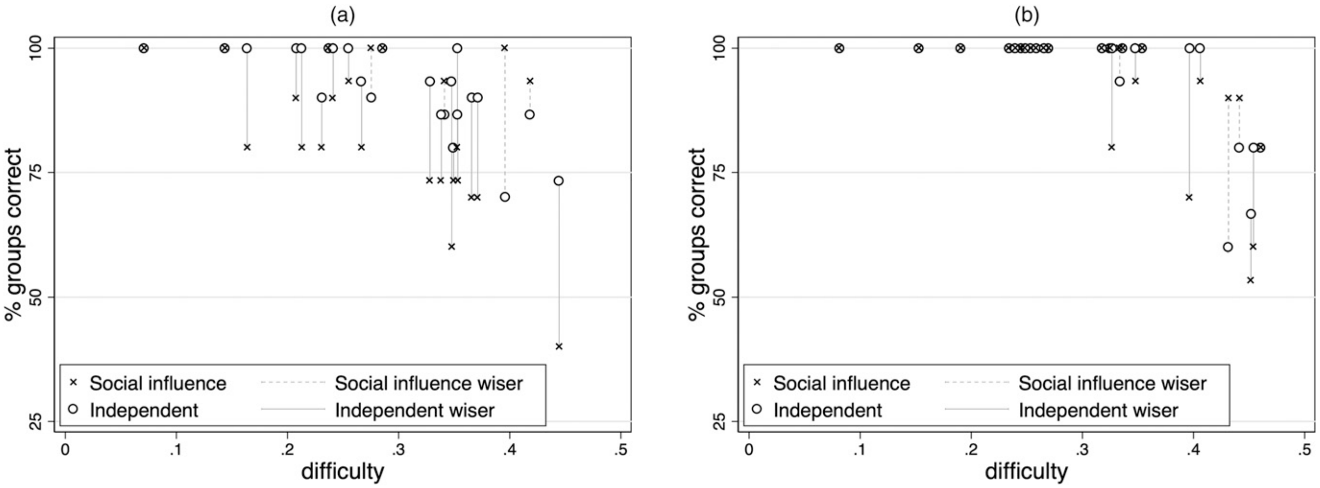
Using the data from 15-person groups in the influence condition for a logistic estimation of Equation (1) yields an estimate of the average susceptibility to social influence of $\mu = 2.36$ (95% CI = 1.91, 2.80; see Table A2, model (2), in the online appendix). This estimate is smaller than the one from the laboratory experiment, suggesting that only wrong early majorities on rather difficult questions might not get corrected over time in the influence condition in longer choice sequences. Figure 3 in Section 2 shows the model predictions for

Table 2. Mean Levels of Individual Accuracy (% Correct Answers per Participant) and Collective Accuracy (% Correct Majority Answers per Group) in the Online Experiment

	Independent condition	Influence condition	p-value of difference[a]
Small groups			
Individual accuracy	70.5	71.6	0.409
Collective accuracy	91.5	82.0	0.001
Large groups			
Individual accuracy	68.6	74.7	0.000
Collective accuracy	94.3	92.1	0.823

[a]Mann–Whitney U tests (see main text).

Figure 8. Percentages of Groups with a Correct Majority per Question by Experimental Condition, Arranged by Question Difficulty; (a) Small Groups (15 Participants); (b) Large Groups (100 Participants)



Notes. Each pair of a circle and a cross connected by a line pertains to one question. (x-axis values are jittered if two questions have the same observed difficulty *d*.)

100-person groups for a similar level of social influence ($\mu = 2$): In the independent condition, wrong majority answers are predicted to occur infrequently and exclusively for very difficult questions. The share of correct majority answers on relatively easy questions should be close to 100% also in the social influence condition, and a negative effect of social influence is only expected for rather difficult questions.

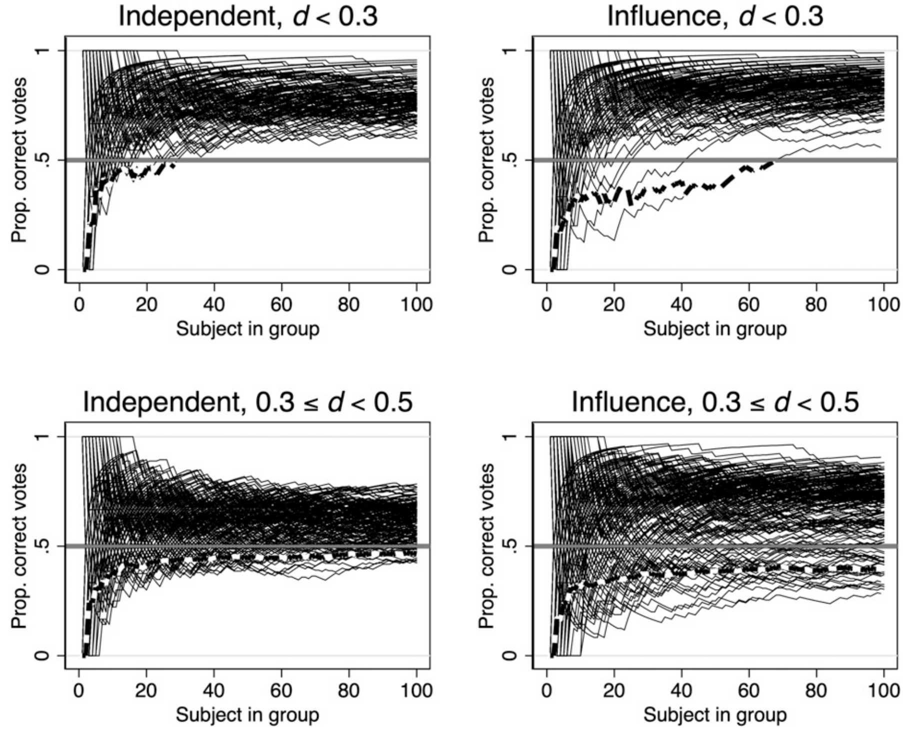
The pattern that Figure 8(b) shows largely tallies with these theoretical predictions. All 100-person groups had the correct majority answer to all questions of difficulty $d < 0.3$, irrespective of the information condition. On the more difficult questions with $0.3 \leq d < 0.5$, the groups in the independent condition outperformed those in the influence condition on six questions, whereas the reverse happened only on three questions. As social influence inhibited correct majorities only at high difficulty levels, the overall difference between independent and influence condition is less pronounced than for small groups; at 94.3% versus 92.1%, this difference is not statistically significant (n = 50, $p = 0.823$; Table 2).

We know from Condorcet's jury theorem that the wrong majorities in the independent condition in Figure 8(b) must eventually become minorities after enough subject choices. Were the wrong majority answers in the influence condition on a similar path of self-correction or had they instead reached a stable majority of false answers? To address this question, we plot in Figure 9 the evolution of the proportion of correct choices in each 100-person group on each question (thin lines) and the evolution of the proportion of correct choices among the wrong majorities (dashed lines), separately for the independent condition and the social influence condition and for easy and difficult questions. Panels (a) and (b) show that, on easy questions ($d < 0.3$), all false starts in the independent condition were corrected after about 30 group members cast their vote and also in the social influence condition, after about 70 choices.

On the other hand, for difficult questions ($0.3 \leq d < 0.5$), results suggest that in the social influence condition wrong majorities can persist. In the independent condition (Figure 9(c)), the share of correct answers in groups with a wrong majority increases as more and more subjects cast their vote (the dashed line moves upwards) and wrong majorities tend to correct (thin lines move upwards, crossing the 50% line). This is consistent with Condorcet's jury theorem and confirms that crowd wisdom increases in the number of independent judgements. By contrast, in the social influence condition (panel (d)), there is no noticeable correction of wrong majorities after about 50 participants. The proportion of correct answers among groups with a wrong majority (dashed line) is at 39.6% after 60 subjects as well as after 100 subjects and we see no tendency for the thin lines under the 50% line to move upward. This suggests that had the groups been even larger, wrong majority answers in the independent condition would have been corrected, whereas false majorities in the social influence condition would have perpetuated further.

What may explain that wrong majorities in the influence condition that persisted up to a certain point tended to remain at a stable level? We conjecture that information about the relative popularity of an answer may have a stronger pull if it is based on a larger absolute number of votes (cf. Mannes 2009). Confirming this explanation, we estimated the average strength of social influence among the first 50 subjects

Figure 9. Development of the Proportion of Correct Answers to a Question Within a Group as Consecutive Subjects Cast Their Vote (Thin Lines) and Development of the Proportion of Correct Answers Among Groups with a Wrong Majority (Dashed Lines) (Data from 100-Person Groups Only)



of each group in the influence condition to be at μ = 2.36 (95% CI = 2.06, 2.66) but at μ = 3.74 (95% CI = 3.31, 4.18) among the second 50 subjects of each group. Also, the inclusion of a term interacting "proportion correct votes" with "total number of votes" confirms that participants were more convinced by the popularity of an answer if that popularity was based on a larger absolute number of votes (Table A2, models (3)–(5), in the online appendix). This suggests that indeed the longer a false majority propagates, the less likely it will still be corrected.

4.3. Does Exclusively Rewarding Individual Accuracy Lead Majorities to Err More Often?

The counterproductive effect of social influence on crowd wisdom naturally raises the question, what can be done about it? One avenue for organizational intervention in team decision making is altering the rewards to team members for a correct team decision and for being among those who in hindsight voted for the correct decision. Our experiment used a mix of individual and group incentives, as this may mirror many real-world settings. What would happen if an organization only rewarded correct individual verdicts and completely eliminated individual team members' responsibilities for correct majority decisions?

Rational choice theory would suggest that herding in the influence condition would *increase* and, thus, also the risk of wrong majorities. Namely, under purely individual incentives, actors when in doubt should follow the majority, as this tends to result in correct answers, and they should neglect the potentially harmful effect of conformity on group accuracy. However, few may appreciate this counterintuitive relationship between conformity, individual accuracy, and group accuracy. Alternatively, one could argue that the exclusive incentivization of correct individual choices may nudge individuals to choose independently, ignoring others' choices. Herding in the influence condition and the risk of wrong majorities would then *decrease*.

To investigate this question and probe the effectiveness of such an organizational intervention, we ran in the online experiment an additional set of 50 groups of 15 participants in which participants were rewarded exclusively for individual accuracy ($0.11 per correct answer).[11] Participants again answered a higher percentage of questions correctly in the presence of information about previous group members' answers than in the absence of such information (72.8% versus 68.9%; Mann-Whitney U test: n = 750, p = 0.000).[12] However, paying participants only for the accuracy of their own answers made them significantly *less* susceptible to social influence, compared with the situation with a mix of individual and group incentives (Table A2, model (6), in the online appendix). Accordingly, there was a smaller difference in the percentage of correct majority answers per group between the independent condition and the influence condition (89.3% versus 86.0%); other than in the situation with mixed incentives, this difference was not statistically significant (Mann-Whitney U test: n = 50, $p = 0.291$). Nevertheless, we cannot reject the null-hypothesis that the effect of social influence on the percentage of correct majority answers per group is not moderated by the incentive scheme (factorial ANOVA; n = 100; $p = 0.133$).

5. Discussion and Conclusion

We conclude that groups seeking to make the wisest group decision face a *social dilemma*: Individuals are wiser when they let themselves be influenced by majorities who think differently, yet groups are wiser when composed of stubborn, independent voices.

The temptation of individual group members to conform has a rational basis. They can expect to increase *individual* performance by joining an emergent chorus supporting one course of action, even if their private inclination suggests another. After all, if most others think differently, they are probably right, so it pays to change one's mind. Unlike subjects in conformity experiments who are deliberately led astray from their correct intuitions by confederates (Asch 1955), in everyday settings following the crowd is often adaptive (see Krueger and Massey 2009 for a review). As we predicted theoretically, our experiments, too, suggest that people are better at answering questions correctly if they have information about the answers of others than if they have to choose independently. The wisdom of the crowd effect—the phenomenon that the majority of a group of voters is more likely to be correct than a single individual (Condorcet 1785)—can explain why social information is often valuable input for individual decision makers. Indeed, research suggests individuals could further improve individual performance if they revised their judgments more severely than they typically do when presented with social information (Mannes et al. 2012, Büchel et al. 2020).

At the same time, as we argued in this paper, independent opinions can be better input for *group* decisions than opinions that are informed by statements of others, even if the latter opinions tend to be more accurate. Condorcet's (1785) theorem presumes that individual votes are independent, and later theory development shows that the chance of a correct majority is reduced if votes are positively correlated (Hogarth 1978; Grofman et al. 1983; Ladha 1992, 1995). Our computational model reproduces this result, showing that in a sequential choice situation the wisdom of the crowd is undermined if individuals tend to adopt the views that others expressed before them, even though this increases the accuracy of individual voters. In our experiments, majorities were indeed more often wrong if information about previous answers was provided than if everyone had to answer independently. The experimental data furthermore indicate that, in sequential decision making, social influence can lead even large groups to wrong decisions, at least on difficult questions.

In team settings where individuals have full information on the sequence of previous choices, wrong early majorities could be less sticky than in our experiment where participants had only aggregate information (as may often be the case in large collectives). A deviation from an emergent consensus requires high confidence and can influence subsequent votes especially strongly. This phenomenon is known to render information cascades fragile (e.g., Gale 1996, p. 620), and full information on choice sequences could thus temper the wisdom-undermining effects. On the other hand, as our experiment only allowed for informational social influence among anonymous decision makers, the undermining effects might also be stronger in team settings where normative pressures to conform to emergent majority opinions among colleagues are added.

Our findings should extend to choice situations with n > 2 modal alternatives: A few early mistakes could lead to a wrong modal choice also if a third option is available. An open question is whether our empirical results extend to the estimation of continuous quantities, such as the projection of sales or a firm's monthly stock returns (Kelley and Tetlock 2013). Another important direction of further research involves the moderating effect of having better-informed voters vote first (e.g., Alpern and Chen 2017). On the one hand, early expert votes reduce chances of early mistakes, thus tempering the wisdom-undermining effect of social information. On the other hand, early expert votes may set a strong but flawed anchor as experts' opinions tend to be correlated because of similar backgrounds and shared information sources (Broomell and Budescu 2009, Sjöberg 2009), and their expertise and status has been shown to raise the social influence effects on later nonexpert voters (See et al. 2011; Henrich 2015, chapter 8; Büchel et al. 2020).

A key organizational challenge is how to incentivize team members so that teams reach wise majority decisions (Lichtendahl et al. 2013, Mann and Helbing 2017). The implication coming out of the present research is that teams should seek to prevent the natural inclination of individuals to conform to previously expressed opinions (for a similar conclusion see Bernstein et al. 2018). Organizations that wish to exploit crowd wisdom in team deliberations may do well to encourage independent expression. Our results could be interpreted in support of Armstrong's (2006) suggestion to avoid face-to-face meetings in order to achieve wise decisions in organizations. Our experiments suggest that organizations may be able to increase the likelihood of a correct majority verdict by reducing team members' stake in a correct majority decision. When incentives for group accuracy were eliminated in our experiments, so that subjects were only paid for choosing correctly, and not for achieving a correct majority, social influence effects became weaker.

Finally, organizations may wish to increase awareness of the social dilemma nature of group decision making under social influence. It is not at all straightforward that the seemingly innocuous behavior of suggesting the correct course of action to the group based on all available information may paradoxically make it more difficult for the group to identify the right choice. People tend to underestimate the value of independent judgments (Soll 1999, Yaniv et al. 2009), and they also tend to underestimate the degree to which they are influenced by others (Nolan et al. 2008). Only after the undermining impact of social influence on crowd wisdom is fully recognized can individuals and teams take steps to tie their hands and resist it.

Acknowledgments

The authors thank Loes Bouwman, Twan Huijsmans, and Fijnanda van Klingeren for assistance with running the laboratory experiment; Mike Claffey, Casper Kaandorp, and Stefan Wehrli for assistance with the online experiment; and Vincent Buskens, Jerker Denrell, Chengwei Liu, Tom Snijders, and four anonymous reviewers for helpful comments.

Endnotes

[1] The study from Lorenz et al. (2011) is titled "How Social Influence Can *Undermine* the Wisdom of Crowd Effect" (emphasis added). However, wisdom was not undermined through alteration of some aggregate measure as the median of opinions but instead by (1) narrowing the diversity of opinions; (2) reducing the range of opinions, often without enclosing the truth; and (3) increasing individuals' confidence in their estimates. Crowd wisdom did not change significantly because of social influence, despite the fact that social influence increased the accuracy of individual subjects (Farrell 2011, Rauhut et al. 2011).

[2] A recent experimental study by Minson et al. (2018) on decision making in dyads finds that compared with simple averaging of collaborators' independent assessments, discussion often aids dyadic accuracy if collaborators first make independent assessments, whereas it harms accuracy in the absence of independent initial assessments.

[3] In each of the above experiments, individuals acted as if aware of crowd wisdom, revising their estimates in the direction of this wise anchor. This is in line with studies that report how individuals use group judgments in the revision of their beliefs, although individuals tend to attribute too little weight to group judgments (Mannes 2009).

[4] A recent study by Jayles et al. (2017) investigates the wisdom of the crowd in sequential decision making, providing very limited social information to experimental subjects, namely, either the most recent response by another subject or the geometric mean of the previous three subject responses. Consistent with our argument, no social-influence-induced improvement in the wisdom of the crowd was found, except when a sufficient number of humans are substituted by bots who always give the correct answer.

[5] The results that we report for $d < 0.5$ reverse for $d > 0.5$. If $d > 0.5$, the majority choice in a large group of individuals is almost certainly wrong (compare Grofman et al. 1983, p. 264). That is, the wisdom of crowds becomes the "dumbness of crowds". For $d > 0.5$, our model shows that social influence mitigates the "dumbness of the crowd," whereas rendering individual votes less accurate (see Online Appendix B).

[6] The experiment was programmed in z-tree (Fischbacher 2007); subjects were recruited using the Online Recruitment System for Experimental Economics ORSEE (Greiner 2015), and the protocol was approved by the institutional review board of Stony Brook University (Committee On Research Involving Human Subjects (CORIHS)# C2015-3001-R1).

[7] The question difficulty d is calculated as the proportion of false answers to a question in the independent condition (not including instances in which a subject gave no answer at all). This calculation of d runs the risk of biasing results in favor of Hypothesis 2 (i.e., more correct majorities in the independent condition than in the influence condition) through misclassification of questions whose estimated and true difficulty levels lie on opposite sides of the $d = 0.5$ threshold value. Namely, by requiring the estimated $d < 0.5$, we exclude questions slightly above the $d = 0.5$ threshold on which respondents in the independent condition performed more poorly than expected because of random variation. We, therefore, verified that calculating difficulties d separately for each session i using only the data from the independent condition of the remaining sessions $j \neq i$ hardly changes which questions are excluded from the tests of our hypotheses and does not change the conclusions regarding our hypotheses under conventional significance levels. For ease of presentation, we work with the naive estimates in the main text and include the additional analysis in the replication files.

[8] A first trial of the online experiment revealed a lack of statistical power. We, therefore, conducted a second trial of the experiment with greater sample size and additional safeguards for data quality. In the second trial, we also changed questions that in the first trial were clearly too easy or too difficult, thus increasing their potential to differentiate crowd wisdom across conditions. Here we report exclusively results from the second trial.

[9] We again focus our analysis on questions with observed difficulty $d < 0.5$ and report results for questions whose difficulty unintentionally exceeded $d = 0.5$ in Online Appendix B. Five of the 30 questions had difficulty $d > 0.5$ in small groups as well as large groups, and one additional question is excluded only from the analysis for small groups (see Online Appendix C). Because different groups were presented different subsets of 15 questions, varying numbers of questions with difficulty $d < 0.5$ were presented in different groups (12.1 on average). This is accounted for by focusing on percentages of correct (majority) answers rather than absolute numbers. Question difficulty was calculated as the proportion of false answers among all "valid" answers in the independent condition, not taking into account choices on which subjects failed to give an answer within the given 20 seconds (0.4%) and choices of the abstain option (11.3%). Similar to the laboratory experiment (see Endnote 7), we verified that results do not change substantively if calculating question difficulty separately for groups presented with question subset i using only the data from the groups in the independent condition who were presented a different subset of questions $j \neq i$. This analysis is included in the replication package.

[10] For small groups, the test relies on 748 rather than 750 participants because two participants gave not a single valid answer. For large groups, n is 4,985 rather than 5,000 also because of a minor

programming issue that led to some groups having 98 or 99 rather than 100 participants.

[11] We are grateful to an anonymous reviewer suggesting this extension of our experiment.

[12] We again excluded five questions from the analysis as they had observed difficulty $d > 0.5$ (Online Appendix C).

References

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. NBER Working Paper No. w23089, National Bureau of Economic Research, Cambridge, MA.

Alpern S, Chen B (2017) The importance of voting order for jury decisions by sequential majority voting. *Eur. J. Oper. Res.* 258(3): 1072–1081.

Armstrong J (2006) How to make better forecasts and decisions: Avoid face-to-face meetings. *Foresight—Internat. J. Appl. Forecasting* (5):3–8.

Arthur WB (1989) Competing technologies, increasing returns, and lock-in by historical events. *Econom. J. (London)* 99(394):116–131.

Asch SE (1955) Opinions and social pressure. *Sci. Amer.* 193(5):31–35.

Baker KM (1975) *Condorcet: From Natural Philosophy to Social Mathematics* (University of Chicago Press, Chicago).

Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–817.

Becker J, Brackbill D, Centola D (2017) Network dynamics of social influence in the wisdom of crowds. *Proc. Natl. Acad. Sci. USA* 114(26):E5070–E5076.

Bernstein E, Shore J, Lazer D (2018) How intermittent breaks in interaction improve collective intelligence. *Proc. Natl. Acad. Sci. USA* 115(35):8734–8739.

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econom.* 100(5):992–1026.

Broomell SB, Budescu DV (2009) Why are experts correlated? Decomposing correlations between judges. *Psychometrika* 74(3): 531–553.

Bruch EE, Mare RD (2006) Neighborhood choice and neighborhood change. *Amer. J. Sociol.* 112(3):667–709.

Büchel B, Klößner S, Lochmüller M, Rauhut H (2020) The strength of weak leaders: An experiment on social influence and social learning in teams. *Experiment. Econom.* 23(2):259–293.

Budescu DV, Chen E (2014) Identifying expertise to extract the wisdom of crowds. *Management Sci.* 61(2):267–280.

Chamley C (2004) *Rational Herds: Economic Models of Social Learning* (Cambridge University Press, Cambridge, UK).

Cialdini RB, Goldstein NJ (2004) Social influence: Compliance and conformity. *Annual Rev. Psych.* 55:591–621.

Clemen RT, Winkler RL (1985) Limits for the precision and value of information from dependent sources. *Oper. Res.* 33(2):427–442.

Condorcet MJAN (1785) *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix* (Imprimerie Royale, Paris).

Cowgill B, Zitzewitz E (2015) Corporate prediction markets: Evidence from Google, Ford, and Firm X. *Rev. Econom. Stud.* 82(4): 1309–1341.

DeGroot MH (1974) Reaching a consensus. *J. Amer. Statist. Assoc.* 69(345):118–121.

Davis-Stober CP, Budescu DV, Dana J, Broomell SB (2014) When is a crowd wise? *Decision* 1(2):79–101.

Farrell S (2011) Social influence benefits the wisdom of individuals in the crowd. *Proc. Natl. Acad. Sci. USA* 108(36):E625.

Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Experiment. Econom.* 10(2):171–178.

Friedkin NE, Bullo F (2017) How truth wins in opinion dynamics along issue sequences. *Proc. Natl. Acad. Sci. USA* 114(43):11380–11385.

Gale D (1996) What have we learned from social learning? *Eur. Econom. Rev.* 40(3–5):617–628.

Galton F (1907) Vox populi. *Nature* 75(7):450–451.

Ganser C, Keuschnigg M (2018) Social influence strengthens crowd wisdom under voting. *Adv. Complex Systems* 21(6-7):1850013–1850013-23.

Goeree JK, Palfrey TR, Rogers BW, McKelvey RD (2007) Self-correcting information cascades. *Rev. Econom. Stud.* 74(3):733–762.

Greiner B (2015) Subject pool recruitment procedures: Organizing experiments with ORSEE. *J. Econom. Sci. Assoc.* 1(1):114–125.

Grofman B, Owen G, Feld SL (1983) Thirteen theorems in search of the truth. *Theory Decision* 15(3):261–278.

Hastie R, Kameda T (2005) The robust beauty of majority rules in group decisions. *Psych. Rev.* 112(2):494–508.

Henrich J (2015) *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton University Press, Princeton, NJ).

Hogarth RM (1978) A note on aggregating opinions. *Organ. Behav. Human Performance* 21(1):40–46.

Hong L, Page SE (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl. Acad. Sci. USA* 101(46):16385–16389.

Hueffer K, Fonseca MA, Leiserowitz A, Taylor KM (2013) The wisdom of crowds: Predicting a weather and climate-related event. *Judgment Decision Making* 8(2):91–105.

Huff C, Tingley D (2015) "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Res. Politics* 2(3):1–12.

Jayles B, Kim HR, Escobedo R, Cezera S, Blanchet A, Kameda T, Sire C, Theraulaz G (2017) How social information can improve estimation accuracy in human groups. *Proc. Natl. Acad. Sci. USA* 114(47):12620–12625.

Kelley EK, Tetlock PC (2013) How wise are crowds? Insights from retail orders and stock returns. *J. Finance* 68(3):1229–1265.

Kerr NL, Tindale RS (2004) Group performance and decision making. *Annual Rev. Psych.* 55:623–655.

Keuschnigg M, Ganser C (2016) Crowd wisdom relies on agents' ability in small groups with a voting aggregation rule. *Management Sci.* 63(3):818–828.

Krueger JI, Massey AL (2009) A rational reconstruction of misbehavior. *Soc. Cognition* 27(5):786–812.

Kurvers RH, Herzog SM, Hertwig R, Krause J, Carney PA, Bogart A, Argenziano G, Zalaudek I, Wolf M (2016) Boosting medical diagnostics by pooling independent judgments. *Proc. Natl. Acad. Sci. USA* 113(31):8777–8782.

Ladha KK (1992) The Condorcet jury theorem, free speech, and correlated votes. *Amer. J. Political Sci.* 36(3):617–634.

Ladha KK (1995) Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *J. Econom. Behav. Organ.* 26(3):353–372.

Larrick RP, Soll JB (2006) Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Sci.* 52(1):111–127.

Levay KE, Freese J, Druckman JN (2016) The demographic and political composition of Mechanical Turk samples. *SAGE Open* 6(1):1–17.

Lichtendahl KC Jr, Grushka-Cockayne Y, Pfeifer PE (2013) The wisdom of competitive crowds. *Oper. Res.* 61(6):1383–1398.

Lorenz J, Rauhut H, Kittel B (2015) Majoritarian democracy undermines truth-finding in deliberative committees. *Res. Politics* 2(2):1–10.

Lorenz J, Rauhut H, Schweitzer F, Helbing D (2011) How social influence can undermine the wisdom of crowd effect. *Proc. Natl. Acad. Sci. USA* 108(22):9020–9025.

Madirolas G, de Polavieja GG (2015) Improving collective estimations using resistance to social influence. *PLOS Comput. Biol.* 11(11):e1004594.

Mann RP, Helbing D (2017) Optimal incentives for collective intelligence. *Proc. Natl. Acad. Sci. USA* 114(20):5077–5082.

Mannes AE (2009) Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Management Sci.* 55(8):1267–1279.

Mannes AE, Larrick RP, Soll JB (2012) The social psychology of the wisdom of crowds. Krueger JI, ed. *Frontiers of Social Psychology: Social Judgment and Decision Making* (Psychology Press, New York), 227–242.

Minson JA, Mueller JS, Larrick RP (2018) The contingent wisdom of dyads: When discussion enhances vs. undermines the accuracy of collaborative judgments. *Management Sci.* 64(9):4177–4192.

Murr AE (2015) The wisdom of crowds: Applying Condorcet's jury theorem to forecasting US presidential elections. *Internat. J. Forecasting* 31(3):916–929.

Mussweiler T, Strack F, Pfeiffer T (2000) Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality Soc. Psych. Bull.* 26(9):1142–1150.

Nofer M, Hinz O (2014) Are crowds on the Internet wiser than experts? The case of a stock prediction community. *J. Bus. Econom.* 84(3):303–338.

Nolan JM, Schultz PW, Cialdini RB, Goldstein NJ, Griskevicius V (2008) Normative social influence is underdetected. *Personality Soc. Psych. Bull.* 34(7):913–923.

Polzin F, Toxopeus H, Stam E (2018) The wisdom of the crowd in funding: Information heterogeneity and social networks of crowdfunders. *Small Bus. Econom.* 50(2):251–273.

Rauhut H, Lorenz J, Schweitzer F, Helbing D (2011) Reply to Farrell: Improved individual estimation success can imply collective tunnel vision. *Proc. Natl. Acad. Sci. USA* 108(36):E626.

Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.

See KE, Morrison EW, Rothman NB, Soll JB (2011) The detrimental effects of power on confidence, advice taking, and accuracy. *Organ. Behav. Human Decision Processes* 116(2):272–285.

Sewell DK (2018) Heterogeneous susceptibilities in social influence models. *Soc. Networks* 52:135–144.

Shao C, Ciampaglia GL, Flammini A, Menczer F (2016) Hoaxy: A platform for tracking online misinformation. *Proc. 25th Internat. Conf. Companion World Wide Web (Montreal, Canada)* 745–750.

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

Soll JB (1999) Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psych.* 38(2):317–346.

Spann M, Skiera B (2003) Internet-based virtual stock markets for business forecasting. *Management Sci.* 49(10):1310–1326.

Sunstein CR (2000) Deliberative trouble? Why groups go to extremes. *Yale Law J.* 110(1):71–119.

Sunstein CR (2006) *Infotopia: How Many Minds Produce Knowledge* (Oxford University Press, Oxford, UK).

Surowiecki J (2004) *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business* (Little Brown, London).

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.

van de Rijt A (2019) Self-correcting dynamics in social influence processes. *Amer. J. Sociol.* 124(5):1468–1495.

van de Rijt A, Siegel D, Macy M (2009) Neighborhood chance and neighborhood change: A comment on Bruch and Mare. *Amer. J. Soc.* 114(4):1166–1180.

van de Rijt A, Kang S, Restivo M, Patil A (2014) Field experiments of success-breeds-success dynamics. *Proc. Natl. Acad. Sci. USA* 111(19): 6934–6939.

Wagner C, Suh A (2014) The wisdom of crowds: Impact of collective size and expertise transfer on collective performance. *47th Hawaii Internat. Conf. System Sci.* (IEEE, Piscataway, NJ), 594–603.

Yaniv I, Choshen-Hillel S, Milyavsky M (2009) Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *J. Experiment. Psych. Learn. Memory Cognition* 35(2):558–563.