Supporting Information:

The Network Dynamics of Social Influence in the Wisdom of Crowds

This appendix includes:

Extended Materials & Methods

Experiment Design. Each trial of the study consisted of an experimental design that compared baseline responses to socially influenced responses within each individual and group. As subjects came into the study, they were randomized to one of the groups, either to a decentralized network, a centralized network, or a control condition. If subjects were randomized to one of the two network conditions, then they were randomly assigned to one node in the network, and they maintained this position throughout the experiment. If subjects were assigned to the control condition, they did not receive any social information, and were simply given the chance to revise their responses in isolation. In each trial, all networks had the same number of subjects (*N*=40). By allowing individuals to make baseline, independent responses and then permitting them to make revisions after receiving social information, we can examine the effects of social influence on collective judgments. Our design also allows us to test how network centralization affects the wisdom of crowds.

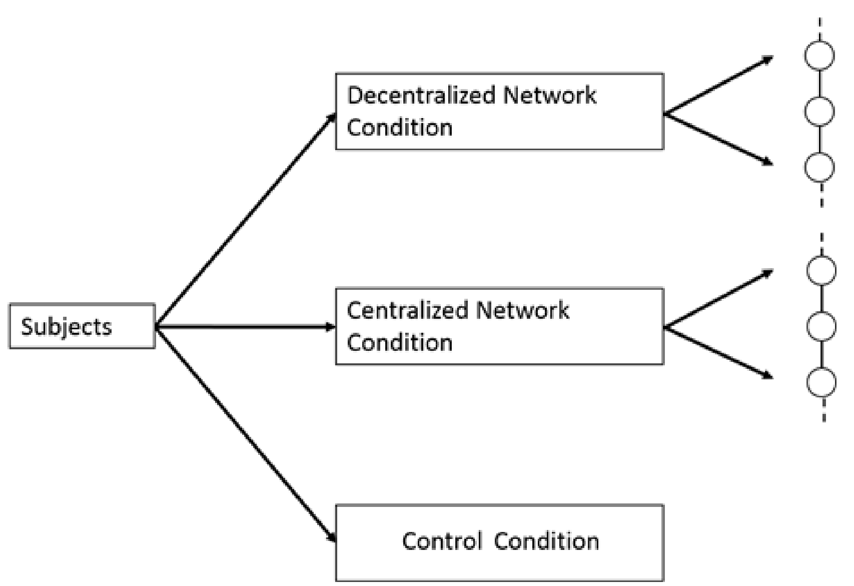


Fig. S1. Schematic of the experiment. Each subject is randomly assigned to a condition. If the condition contains a network, then the individual is randomly assigned to a single node within the network.

Subject Recruitment. Participants in our study were recruited via the World Wide Web to be players in an "Intelligence Game." Participants for questions soliciting count-based responses were recruited via online message boards advertising the opportunity to make money by participating in online experiments. Participants for questions soliciting percentage-based responses were recruited by sending an advertisement and a link for this experiment to participants who had previously completed a "HIT" on the Amazon Turk platform in which they registered to participate in studies with our research group. Upon arriving at the study website, participants viewed instructions on how to play the intelligence game, and waited while other subjects arrived. When a sufficient number of subjects arrived, all subjects were randomized to a condition and the trial would begin. Responses for count-based questions were collected over a 319-day period, July 28, 2015 through May 24, 2016 over which time online advertisements were posted to attract subjects to participate in the study. Responses for percentage-based questions were collected over a 4-day period from November 17, 2016 to November 20, 2016.

Subject Experience During the Experiment. To isolate the causal effect of social influence, the interface in each network condition was identical. For each question, participants first provided an independent response without any social information. The experimental interface of this independent round is depicted in Fig. S2. Then, subjects in network conditions were shown the average (mean) response of their network neighbors, and were given a chance to revise their answer. This design ensured that users in both network conditions received identical user experiences. The screenshot of the first chance to revise after social exposure (Round Two) is shown in Fig. S3. 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 Three) is shown in Fig. S4. 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. Subjects had one minute to provide each response, and the entire experiment lasted for twelve minutes. To motivate subjects, rewards were based on the accuracy of their answers. Subject payment was based on their error as a percent of the true value. Answers which were exactly correct earned the maximum payout ($2.50). Answers which were within 1% of true value received $1.25; within 10%, $1.00; within 15%, $0.75; within 20%, $0.35; within 30%, $0.25; within 90%, $0.15. Answers more than 90% from the true value did not earn any payment. This payment schedule was not observable to participants, who were only told "The more accurate your answers, the more you win!" This design allows us to examine the effect of social influence on independent beliefs. Participants in the control condition were provided with an identical interface, but without social information.

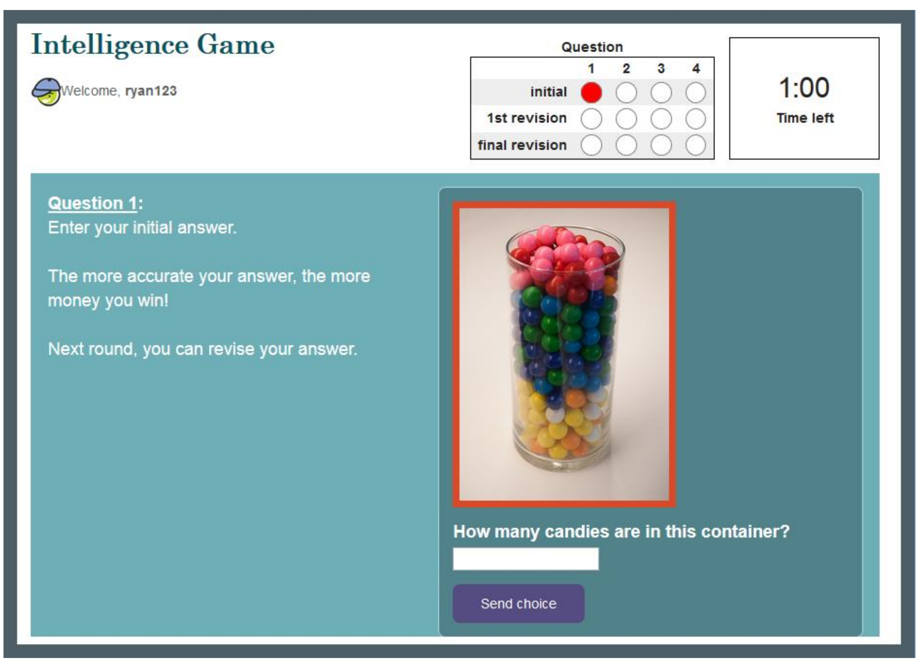


Fig. S2. Screenshot of the experimental interface in Round One (baseline).

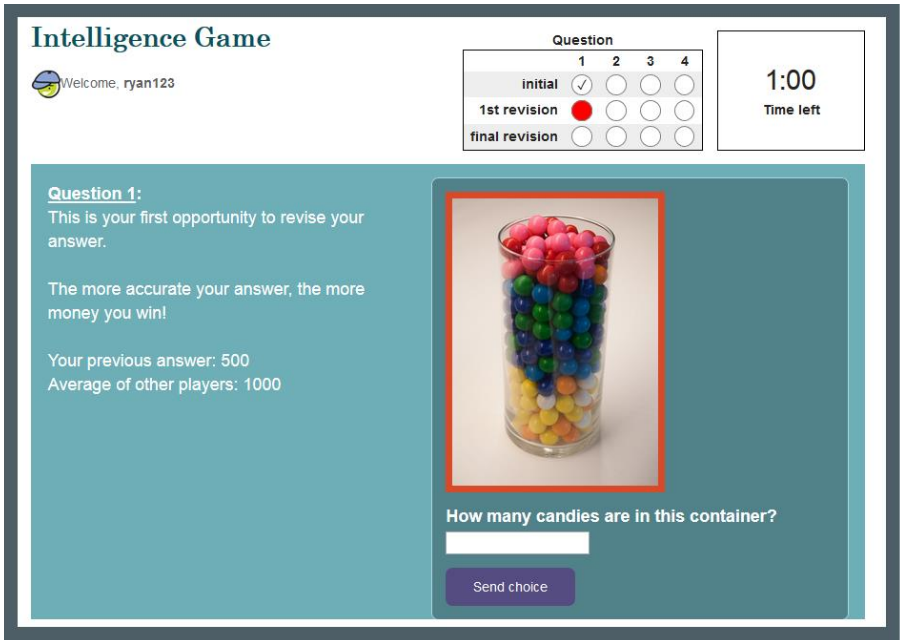


Fig. S3. Screenshot of the experimental interface in Round Two (first chance to revise after social exposure).

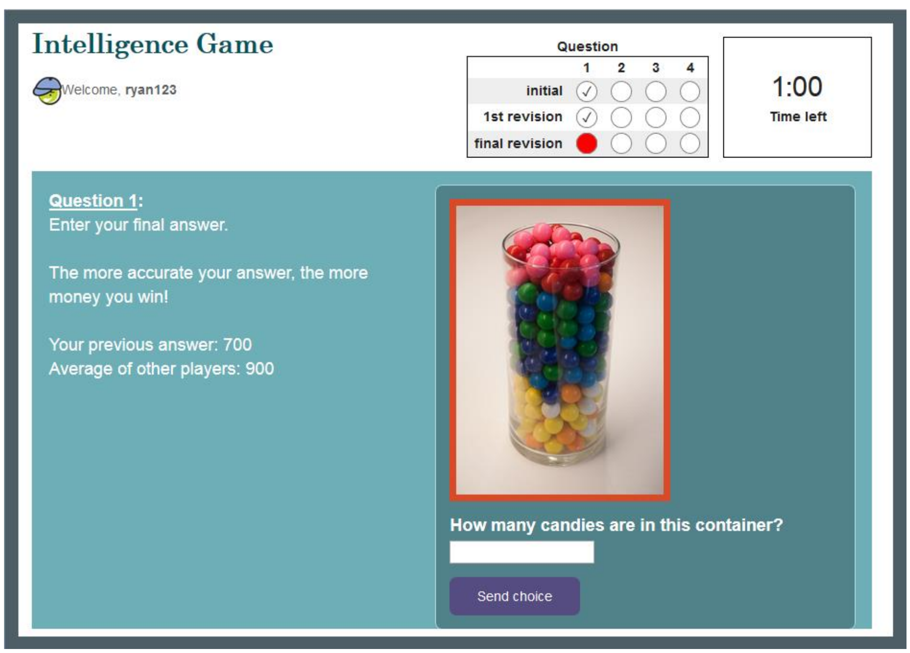


Fig. S4. Screenshot of the experimental interface in Round Three (final chance to revise after social exposure).

Estimation Tasks. Estimation tasks were taken from diverse domains, including health, economics, and arithmetic estimations. For the set of trials in which we provided users with count-based questions, we generated six unique question sets. We assigned a single centralized network, decentralized network, and control group to each question set, so that we obtained a single group estimate per condition for each question (see Ensuring Data Quality, below). For the set of trials in which we provided users with percentage based tasks, we provided each question simultaneously to a large number of unique groups (see Ensuring Data Quality, below), so that we obtained multiple group estimates for each question. Because count-based questions generate a consistent skew, percentage-based questions were introduced to ensure that our results were robust to variation in the skew of the distribution of independent estimates. To produce a wide variation of independent estimate distributions, half the groups estimated one parameter (eg, “what percent of people in this crowd are wearing hats”) and half the groups estimated its numeric complement (eg, “what percent of people in this crowd are not wearing hats”). Example prompts from the count-based tasks and the full set of prompts for percentage-based tasks are shown in Figs S5 and S6 below.

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 was used for each question set. 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

Question 1a - 1f. Health estimation: how many calories are in this meal?

Question 2a - 2f. Economic estimation: what is the combined cost of these items (in dollars)?

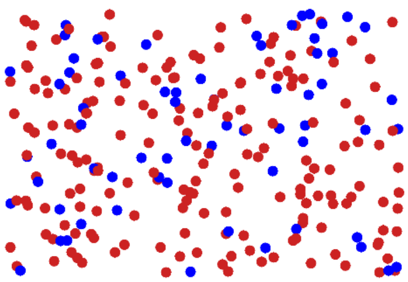
Question 3a - 3f. Arithmetic estimation 1: how many candies are in this container?

Question 4a - 4f. Arithmetic estimation 2: how many pennies are in this photo?

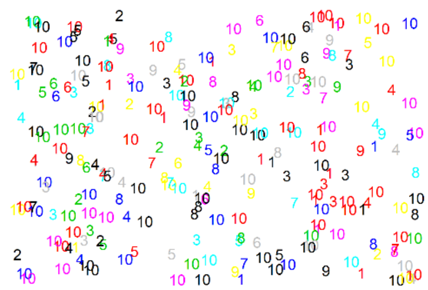
Fig. S5. Each panel shows a sample count-based estimation task with the accompanying text. We provided participants with six unique sets (a through f) of four questions, each of which varied slightly in the details (eg, a different food item). In total, we collected data for 24 unique count-based estimation tasks.

Question 5a: What % of the rectangle is filled by the dark purple?
Question 5b: What % of the rectangle is filled by the light purple?

Question 6a: What % of these people are wearing hats?
Question 6b: What % of the people are NOT wearing hats?



Question 7a: What % of the dots in this image are blue?
Question 7b: What % of the dots in this image are red?



Question 8a: What percent of the characters in this figure are the number 10?
Question 8b: What percent of the characters in this figure are NOT the number 10?

Question 9a: What % of the umbrellas in this image are NOT solid black?
Question 9b: What % of the umbrellas in this image are solid black?

Fig. S6. Each panel shows one of the five prompts used for percentage-based estimations. For each prompt, participants were asked to estimate one of two possible complementary parameters. In total, we collected data for 10 unique percentage-based estimation tasks.

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.

The answers to each of the questions were verified by the nutrition label for the health estimation, the manufacturer websites for the economic estimation, and independent counts for the arithmetic and percentage based estimation questions. In the case of the economic estimation task, the pictures contained images of three items, the prices of which were independently researched and aggregated. Participants in the study had one minute for each round, which prevented them from discovering the retail value of all three items in the time allotted. The wide standard deviation in participants’ responses for each question (as shown in the Supplementary Data tables) confirmed that subjects did not have independent access to the correct answers.

Network Metrics. While asymptotic results (21) identify eigenvector centrality as the determining factor in collective beliefs, we study collective beliefs after only two belief revisions. In diffusion processes, eigenvector centrality represents the limiting measure of influence as time increases, where degree centrality reflects a node’s influence after only one time period (42). Similarly in the DeGroot model, an individual’s weight in the collective belief is equal to their degree centrality after a single revision, while eigenvector centrality represents the long term limit

(21). In the random networks used here, degree centrality is highly correlated with eigenvector centrality, and the most central node is generally the same with both metrics. Our experimental design utilizes a network that is highly centralized as measured by both metrics of centralization (28). In the figures showing simulated results, we measure networks with degree centralization.

Degree centrality is a network centrality metric assigned to each node in a network that is measured by simply counting the number of edges that are connected to a given node (40). In our experimental design, we consider only undirected networks, and therefore make no distinction between in-degree and out-degree. As a result, degree centrality for a given node is simply its row (or column) sum in the network adjacency matrix.

Centralization is a generic measure that describes the equality of distribution for any centrality metric in a network (28). A network with highly unequal centrality (i.e., a small number of central nodes and many peripheral nodes) has a high degree centralization, while a network where all nodes are equally central has a low degree centralization. Centralization scores are normalized by the maximum possible, where *C*=0 indicates no centralization (all nodes are equally central) and *C*=1 indicates the maximum possible centralization. Centralization is defined as the following:

$$C = \frac{\sum_{i=1}^{N} c^* - c_i}{C_{\max}}$$

where c_i indicates the centrality score for node *i*, and c^* indicates the centrality score for the most central node, and C_{max} is a normalizing constant equal to the maximum possible value for the numerator for a graph with N nodes (28).

Experimental and Simulation Network Conditions. Our experimental test employed two network conditions. The decentralized condition employed a random network where every node had the same number of connections, yielding degree centralization C=0. We constructed a network with 4 edges per node, and employed the same network across all decentralized trials to minimize variance. The centralized condition employed a star network, which consists of a single central node with one connection to all peripheral nodes, yielding a network with degree centralization C=1. Robustness checks using our theoretical model show a continuous effect of degree centralization on the wisdom of crowds in the range C=0 to C=1 (Fig. S7).

All networks used in simulations were generated using the igraph package v0.7.1 in R v3.2.0. Decentralized networks were generated using the degree.sequence.game function with the "vl" method, which samples uniformly from the set of possible graphs. In our case, a network is fully determined by two parameters: N, which determines the number of nodes; and k, which determines the degree of each node. In order to generate a set of networks that vary continuously in their centralization score, centralized networks were generated using the barabasi.game function, which runs a variant of the preferential attachment algorithm described by Barabasi and Albert (41). These networks are fully determined by three parameters: N, which determines the number of nodes; m, which determines the minimum degree of any node; and p, which determines the strength of preferential attachment. By varying the strength of preferential attachment ("power" parameter) this algorithm can generate networks with a range of centralization. When power is sufficiently high, this algorithm generates a star network, allowing us to test a continuous range of centralized networks. All centralized networks in the simulations were generated with m=1, and power (p) ranged from 0 to 3. We generated decentralized networks with k = 4, which by construction have an average degree Z = 4 and density D = 0.1. In the centralized networks,

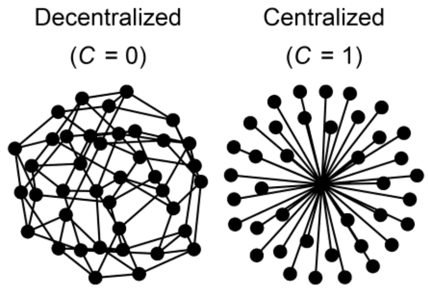


Fig. S7. The two network conditions used in the experimental trials were a decentralized network and a maximally centralized star network.

average degree is fully determined by *N* and *m*. In our case, *m*=1, which yielded an average degree Z = 1.95 and density D = 0.05. Note that this is fixed by *N* and *m*, and thus does not vary with *p*. As a result, in the simulations where centralization varied (Fig. S12) average degree and density remained constant.

Supporting Data Analysis

For each estimation task by each experimental or control group, we measure collective belief in terms of the central tendency of the estimate distribution, i.e., its mean or median. To assess collective error, we used the absolute value of the difference between the collective belief and the true answer. In order to facilitate comparisons across questions with varying response distributions, all outcomes are normalized by the standard deviation. This is equivalent to normalizing all responses prior to analysis. For example, the error of a group estimate as measured by the mean is defined as:

$$error_{j,q,t} = \frac{|\,mean_{j,q,t} - truth_q\,|}{\sigma_{q,t=1}}$$

where the numerator indicates the absolute value of the difference from truth for group *j* on the *q*th question at time *t*, and the denominator indicates the pooled standard deviation of independent (baseline) responses for all individuals who answered that question in both of the network conditions and the control condition. I.e., the standard deviation in the denominator is based on all independent responses to a given question. All outcomes, including change in error, are therefore measured in terms of standard deviations.

To assess individual error, we measured the absolute value of the difference between an individual's response and the true answer, and normalized this value by the standard deviation of baseline responses for that question as follows:

$$error_{i,q,t} = \frac{|\,response_{i,q,t} - truth_q\,|}{\sigma_{q,t=1}}$$

where individual *i*'s error is the absolute distance between their response and the truth for that question, normalized by the standard deviation of initial responses for that question at Round One.

To assess whether crowds were wise in their baseline judgments, we compared the error of the group mean, $error_{j,t=1}$, to each individual response, $response_{i,q,t=1}$ and calculated the proportion of individuals for group *j* on question *q* that were more accurate, less accurate, or equally accurate as compared to the group belief. When reporting this percentage across multiple group estimates, we report the average value of this percentage as calculated for each group estimate for each question.

To assess the properties of estimate distributions, we measured the skew for each group estimation as well as the relative location of the mean, median, and true answer. In 21 of the 24 group estimations for count-based estimation tasks by decentralized networks, the estimate distribution at Round One had a mean that fell between the median and truth, or a mean that fell on the opposite side of truth from the median (Fig S8). In 18 of the 35 estimations for percentage-based estimation tasks by decentralized networks, the estimate distribution at Round One had a mean that fell between the median and truth, or a mean that fell on the opposite side of truth from the median. In these cases, social influence is expected to pull the median towards the mean, improving the accuracy of the mean, as described in our theoretical overview below.

All statistical tests presented here and in the main text used two-tailed tests. Because each network and control group completed multiple estimation tasks, all of our analyses control for correlations between estimates made by the same group. The main analyses are conducted as follows. For each task within a given experimental trial, we first measured the error of the mean for that task as the group error for the 40 subjects in a network. We then took the average of these group errors over all tasks that were completed within the trial. This average group-level error was measured both at Round One and at Round Three for every experimental trial. This yields a single, network-level value for the initial group error and the final group error for each condition in the trial. We repeated this process for each of the 13 experimental trials. This produces 13 independent observations for each network condition. We then calculated the average error for each network condition by taking the mean of these 13 observations, both at Round One and Round Three, for decentralized and centralized networks. Complete data for all tests are provided in the Supporting Data File.

Wisdom of the Crowds. We test the wisdom of the crowds by calculating, for each group estimation, the percentage of individuals who hold an independent response that is more accurate than the mean or median of independent responses. Across all centralized group estimations, decentralized group estimations, and control group estimations, an average of 35% of individuals have an independent response that is more accurate than the mean, and 34% are more accurate than the median. To generate this estimate, we first calculate the percentage of initial responses in each group estimation task that are more accurate than the initial mean/median of individual responses. We average that value across all estimation tasks completed by each of the 13 decentralized trials, 13 centralized trials, and 8 control groups. Finally, we average that value across all trials.

Analysis for Decentralized Networks. The statistical analysis for the decentralized networks is based on the analysis of 13 experimental trials, comparing the estimates before social influence (Round One) to the third estimate after two revisions (Round Three). Because individual data are not independent, we conduct all analysis at the group level. Because individual groups completed multiple estimation tasks, we analyze the average effect of the 13 independent experimental trials. Table S1 shows the cumulative change from Round One to Round Three for the average revision coefficient, error of the mean, and standard deviation for each of the 13 experimental trials.

Across all 13 experimental trials, the average standard deviation within each group was 0.96 s.d. at Round One and 0.55 s.d. at Round Three. To calculate these values, we first measure the standard deviation for each estimation task at Round One and Round Three. We then measure, for each experimental trial, the average standard deviation across all tasks. To test whether this decrease is significant, we compare the 13 measurements of average standard deviation at Round One with the 13 measurements of average standard deviation at Round Three with a Wilcoxon signed rank test, and find that mean of the standard deviation was significantly lower at Round Three than Round One ($P<0.001$).

Group ID	Tasks Completed	Avg. Revision Coefficient	Cumulative Change, Round One to Round Three: Individual Error	Error of Mean	Standard Deviation
2	4	0.31	-0.18	0.03	-0.39
3	4	0.47	-0.25	-0.13	-0.65
5	4	0.29	-0.16	-0.01	-0.25
7	4	0.06	-0.10	-0.01	-0.33
9	4	0.30	-0.25	0.00	-0.58
12	4	0.34	-0.26	-0.09	-0.44
13	5	0.15	-0.18	-0.06	-0.34
16	5	0.30	-0.24	-0.09	-0.40
18	5	0.22	-0.25	-0.12	-0.42
19	5	0.24	-0.28	-0.09	-0.37
22	5	0.46	-0.33	-0.16	-0.45
23	5	0.21	-0.26	-0.10	-0.37
24	5	0.23	-0.24	-0.09	-0.34

Table S1. Average revision coefficient and cumulative change in average individual error, average error of the mean, and average standard deviation for estimation tasks completed by 13 experimental trials with decentralized networks. In 12 out of 13 trials, the error of the mean decreased on average across all completed estimation tasks.

The error of the mean, averaged across all 13 trials with decentralized groups, was 0.70 s.d. at Round One and 0.62 s.d. at Round Three. To calculate these values, we first measure the error of the mean for each estimation task at Round One and Round Three. We then measure, for each experimental trial, the average error of the mean across all tasks. In 12 of the 13 trials, average error decreased after social influence (Table S1). To test whether this decrease is significant, we compare the 13 measurements for average error of the mean at Round One with the measurements for average error of the mean at Round Three with a Wilcoxon signed rank test, finding that the average error of the mean was significantly lower at Round Three than Round One (P<0.01). Using a similar test, we also find that the error of the median significantly decreased after social influence (P<0.001).

In addition to measuring change in group error, we also examine changes in individual error. Because network data are not independent, we estimate this value by first measuring the average individual error for each independent group estimation task, and then average this value across all estimation tasks completed by each experimental trial. For all 13 experimental trials with decentralized networks, mean individual error at at Round One was 0.99 s.d. and 0.76 s.d. at Round Three. To test whether this decrease is significant, we compare the 13 measurements for average individual error at Round One with the 13 measurements at Round Three using a Wilcoxon signed rank test, and found that the average individual error was lower at Round Three than Round One ($P<0.001$).

To explain the decrease in error by decentralized networks, we measure the correlation between the revision coefficient (see Analysis of Individual Behavior, below) for each estimation task by each network and the change in the error of the mean from Round One to Round Three. Because each network completed multiple estimation tasks, we measure this as the partial correlation after controlling for correlation between estimation tasks completed by the same group. To do this, we first regress each outcome variable (revision coefficient or change in error) for each of the 59 estimation tasks on a categorical variable indicating membership in one of the 13 groups. We then measured the correlation between the residuals of these two regressions. This is equivalent to directly measuring the correlation between revision coefficient and change in error after subtracting the mean value for each experimental group from each outcome variable. We describe the process of measuring partial correlation in more detail below, in Analysis of Individual Behavior. To generate confidence intervals on this estimate, we use the percentile bootstrap method (44) drawing each bootstrap sample at the cluster level, following standard methods for clustered data analysis (45).

Comparison of Decentralized Networks with Control Group. As a robustness test, we compare the behavior of decentralized networks with the behavior by individuals in control groups. We

Group ID	Tasks Completed	Cumulative Change, Round One to Round Three: Individual Error	Error of Mean	Standard Deviation
27	4	0.07	0.10	0.57
28	5	-0.04	-0.05	-0.04
29	4	-0.03	0.04	-0.10
30	5	-0.09	-0.07	-0.08
31	4	-0.13	-0.06	-0.44
32	4	0.03	-0.03	0.04
33	4	-0.04	-0.02	-0.05
34	4	-0.03	0.06	-0.09

Table S2. Cumulative change in average individual error, average error of the mean, and average standard deviation for estimation tasks completed by 8 control groups.

first consider whether standard deviation significantly decreased in control groups. For the 8 experimental trials with control groups, average standard deviation was 0.97 at Round One at 0.95 at Round Three. To test whether this decrease is significant, we compare the 8 values for average standard deviation at Round One with the 8 values for standard deviation at Round Three using a Wilcoxon signed rank test, and find that standard deviation was not significantly lower at Round Three than Round One (P>0.25). Because individual responses within control groups are independent, we also compare variance at Round One with the variance at Round Three for each group estimate separately. Using an F-test for change in variance, we generate a P-value for each of the 34 estimates by the 8 control groups. We adjust the P-values for multiple testing using the Holm method, which is less likely than Bonferroni correction to produce Type II error (43). In one out of the 34 groups, standard deviation decreased significantly (adjusted P<0.001).

To test whether overall decrease in standard deviation is significantly smaller than the decrease in standard deviation for centralized and decentralized networks, we first calculate the average change in standard deviation for each of the 8 control trials (Table S2), 13 decentralized trials (Table S1), and 13 centralized trials. We compare these using a Wilcoxon rank sum test, and conclude that the change in standard deviation is significantly greater in decentralized networks than control groups (*P*<0.001) and is significantly greater in centralized networks than control groups (*P*<0.001).

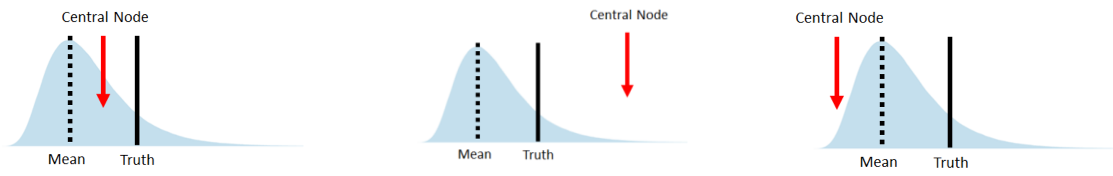
We next consider whether individual error decreased after revisions by isolated individuals. Because individual responses are independent, we can conduct this analysis at the individual level. For the 320 individual assigned to a control group, we first calculate the average error at Round One and Round Three across all the estimation tasks completed by each individual. We then compare the average individual error for Round One (1.0 s.d.) to Round Three (0.97 s.d.) using a Wilcoxon signed-rank test, and conclude that average individual error decreased after revisions in control groups (p<0.01).

Notably, the magnitude of the reduction in individual error was only 0.03 s.d. in the control condition conditions, while average individual error decreased by 0.23 s.d. in decentralized networks. To test whether this difference is significant, we calculate the difference between the average individual error at Round Three and the average individual error at Round One for each of the 8 control groups using a similar process to the group analyses described above. We then compare the 8 values for the control condition (Table S2) with the 13 values for the decentralized condition (Table S1) using a Wilcoxon rank sum test, and conclude that individual improvement was significantly greater for individuals in decentralized networks than individuals in the control condition (*P*<0.001).

Importantly, a reduction in individual error does not guarantee collective improvement if it does not lead to a shift in the overall distribution. To test for change in group accuracy in the control case, we first conduct an analysis comparable to our main analysis on networked groups, analyzing the change the error of the mean estimate for each of the 8 control trials. The average error of the group mean was 0.69 s.d. at Round One and 0.68 s.d. at Round Three. To test if this decrease is significant, we compare the 8 values for average error of the mean at Round One with the 8 values at Round Three with a Wilcoxon signed rank test, and find that error does not significantly change (P>0.87).

Because individual estimates are statistically independent, we can also directly examine the shift in group distributions. To test whether individual revisions had any effect at all on the mean of estimate distributions, we test for a change in the group mean for each individual estimation task, each of which was completed by 40 independent individuals. For each of the 34 unique estimation tasks, we compare the distribution of the Round One estimates with the distribution of the Round Three estimations using a Wilcoxon signed rank test. This generates 34 P-values for the change in mean, which we then correct for multiple tests using the Holm method, which is less likely to produce Type II error than the Bonferroni correction (43). None of these are significant, with a range of 0.32 to 1.0 and a mean value of 0.90. As a result, we conclude that the group mean for estimates by control groups did not significantly change after revisions.

Analysis for Centralized Networks. The statistical analysis for the centralized networks is based on the analysis of 59 group estimation tasks completed in 13 experimental trials, comparing the independent estimates (Round One) to the final socially influenced estimates (Round Three). In two cases, the central node failed to provide a response at either Round One or Round Two, and thus the 39 other nodes in this network were not exposed to any social information. Thus, to determine the effect of social influence on change in group estimates, this analysis is conducted on the remaining 57 group estimates.



Case 1 **Case 2** **Case 3**

Fig. S8. Each of these panels illustrates a different possible relationship between the estimate of the central node and the group belief. In Case 1, the central node is more accurate than the group mean, and pulls the group mean towards truth. In Case 2, the central node is less accurate than the group mean, but still pulls the group mean towards truth. In Case 3, the central node is less accurate than the group mean, and falls on the opposite side of the group mean from truth, and thus pulls the group away from the true value.

To test the effects of social influence in centralized networks, we conducted the same test for the standard deviation that we conducted for decentralized networks as described above. In this test, we compared the 13 values for the average standard deviation of estimates at Round One with the 13 values for average standard deviation of estimates at Round Three using a Wilcoxon signed rank test, and found that the standard deviation significantly decreased in centralized networks ($P<0.001$).

To explain the dynamics of centralized networks, we study change in error conditioned on whether the central member held a belief that was in the direction of the ruth relative to collective judgment or in the opposite direction of truth (see Fig. S8). For each experimental trial, we measure the change in error for those estimations where the central node was in the direction of the truth separately from the change in error for those estimations where the central node was away from the truth. In 12 out of 13 trials, at least one estimation fell in each category (see Table S3). In one trial, the central node was away from the truth in all four estimation tasks.

To calculate the change in error for these two categories, we first measure, for a given experimental trial, the error of the mean (or median) at Round One and Round Three across all estimation tasks that fall into a given category (eg, central node toward truth). We then measure

Group ID	Central Node Belief In Right Direction?	Number of Tasks	Cumulative Change Round One to Round Three: Individual Error	Cumulative Change Round One to Round Three: Error of Mean
1	No	2	0.06	0.12
	Yes	2	0.48	-0.26
4	No	1	-0.14	0.29
	Yes	2	0.25	-0.08
6	No	2	0.13	0.29
	Yes	2	0.20	0.04
8	No	4	-0.06	0.22
	Yes	0	--	--
10	No	2	-0.03	0.07
	Yes	2	0.31	-0.14
11	No	1	-0.01	-0.12
	Yes	3	0.15	-0.01
14	No	1	-0.07	0.44
	Yes	4	0.48	-0.35
15	No	4	-0.10	0.24
	Yes	1	0.62	-0.60
17	No	3	-0.05	0.19
	Yes	2	0.27	-0.10
20	No	2	0.21	0.14
	Yes	2	0.41	-0.14
21	No	2	0.12	0.07
	Yes	3	0.37	-0.31
25	No	1	-0.04	0.19
	Yes	4	0.45	-0.37
26	No	2	0.16	0.04
	Yes	3	0.50	-0.54

Total for All Trials: Central Node Belief In Right Direction	Total for All Trials: Number of Trials	Cumulative Change Round One to Round Three: Individual Error	Cumulative Change Round One to Round Three: Error of Mean
No	13	0.17	0.29
Yes	12	-0.24	-0.34

Table S3. Summary data for centralized networks, conditioned on the relative location of the central node. When the central node held a belief that pulled the group toward truth, error of the mean and median decreased on average across all estimation tasks. When the central node held a belief that pulled the group away from truth, error of the mean and median increase on average across all estimation tasks. The values for each row indicate the average across all group estimates in that category. For change in the error of the mean and change in standard deviation, results are conditioned on the relative location of the mean and the central node. For change in the error of the median, results are conditioned on the relative location of the median and the central node.

the average change in error for each category for each experimental trial. For estimations in which the central node was toward truth, the average error of the mean was 0.56 s.d. at Round One and 0.32 s.d at Round Three. We compare the 12 values for average error at Round One to the 12 values at Round Three with a Wilcoxon signed rank test, and conclude that average error of the mean decreased after social influence when the central node was toward truth ($P<0.001$). Across all 13 trials, the average error of the mean estimate increased when the central node held an estimate that was in the direction away from truth (Table S3). For this category, average error at Round One was 0.84 s.d. and was 1.01 s.d at Round Three. We test whether this increase is significant with a Wilcoxon signed rank test, finding that average error at Round Three was significantly greater than average error at Round One ($P<0.001$). We conduct a similar test for the change in the median. Using the same analysis, we found that when the central node was in the direction of truth relative to the group median, the error of the median decreased (N=12, $P<0.001$, Wilcoxon signed rank test). When the central node was away from truth relative to the group median, the error of the median significantly increased (N=13, $P<0.001$, Wilcoxon signed rank test).

One important possibility is whether a central node that was accurate on one question is also likely to be accurate on another question. To examine if individual accuracy on one question was predictive of accuracy on another, we examined the correlation among individual errors across the four categories of count-based questions and the five categories of percentage-based questions. For each set of questions, we measured the pairwise correlation between the complete set of independent responses for each set of questions (eg, we measured the correlation between individual error for Arithmetic Estimation 1 and individual error for Health Estimation, and repeated this measurement for all possible pairs). We then averaged the correlation for all pairwise comparisons to yield an estimate of the overall correlation between questions. The average numerical correlation was extremely small, with $\bar{\rho}$=0.042 for count-based questions and $\bar{\rho}$=0.029 for percentage-based questions. We therefore conclude that individuals who are accurate on one question are not likely to be accurate on another question. Complete data for this test is provided as a Supporting Data file.

We also examine the effect of social influence on the accuracy of the 13 subjects placed in the central position in a centralized network. Average error for these subjects was 0.98 s.d. before social influence and 0.78 s.d. after social influence. To test if this decrease is significant, we compare the 13 values for the average error of the central node at Round One with the 13 values at Round Three using a Wilcoxon signed rank test, and find that the average error of the central node at Round Three was significantly lower than Round One (P<0.01).

Analysis of Individual Behavior. To calculate the revision coefficient for each estimation task by each group, we measure the correaltion between the magnitude of individual revisions ΔR_i = $|R_{1,i} - R_{3,i}|$ and individual error $E_{1,i} = |R_{1,i} - T|$, after controlling for the magnitude of the social signal $S_{1,i} = |\ R_{1,i} - \bar{R}_{1,j\in N_i}\ |$ where $R_{t,i}$ indicates the estimate for subject i at time t, $\bar{R}_{t,j\in N_i}$ indicates the average estimate of subject i's network neighbors at time t, and T is the true answer. Individual revision magnitude for subject i is measured as the absolute value of the difference between their estimate at Round Three and their estimate at Round One. The magnitude of the social signal is measured as the absolute value of the difference between an individual's independent estimate at Round One and the average independent estimate of their network neighbors at Round One. To measure the partial correlation, we first regress $\Delta R_\cdot$ on $S_{0,\cdot}$ and measure the residuals from this regression for each individual i as $e_{1,i}$. We then regress $E_{0,\cdot}$ on $S_{0,\cdot}$ and obtain the residuals from this regression for each individual *i* as $e_{2,i}$. Finally, the partial correlation is measured as the correlation between $e_{1,\cdot}$ and $e_{2,\cdot}$, which we refer to as "revision coefficient."

Because each individual completed multiple estimation tasks, our correlation analyses for the entire population (Fig. 2, main text) include the addition of a categorical control variable indicating which individual provided each estimate. This is equivalent to correlating each individual's error and revision magnitude after subtracting their individual average for each value, and follows the same procedure described above for our correlation analysis of decentralized networks. The procures a partial correlation ρ=0.25 that is equivalent to the coefficient in a linear regression, scaled to fall between zero and one. Analysis of covariance indicates the relationship is significant at $P<0.001$, and therefore we conclude that error is correlated with the magnitude of an individual's revision when the magnitude of their social signal is held constant.

In the main text, all analyses compare estimates at Round One with estimates at Round Three. We also analyze individual behavior separately for the first revision between Round One and Round Two, and the second revision between Round Two and Round Three. Using the same method as described above, we find that error is significantly correlated with the magnitude of revision after controlling for social information at both stages of revisions (ρ=0.13, $P<0.001$ for the first revision; ρ=0.37, $P<0.001$ for the second revision).

Our theoretical model assumes that an individual's revised answer falls somewhere between (or is equal to) their initial answer and the average of their peers. The amount of distance that is closed is defined in the model by the parameter α_i assigned to each individual. As we describe in the model definition below, we can identify this parameter for each revision by each subject as a function of that subject's estimate before social influence, the average estimates of that subject's network neighbors before social influence, and the subject's estimate after social

		Proportion of Estimations	Mean
Revision 1	$\alpha < 0$	9.0%	-1.52
	$\alpha > 0$	7.5%	2.15
	$0 \leq \alpha \leq 1$	83.5%	0.63
Revision 2	$\alpha < 0$	8.8%	-1.53
	$\alpha > 0$	10.4%	2.22
	$0 \leq \alpha \leq 1$	80.8%	0.74

Table S4. Our theoretical model assumes that individuals have a coefficient between zero and one inclusive. This table shows the distribution of α_i as measured for both revisions by N=1040 individuals completing 4340 estimation tasks in network conditions. Most individuals displayed behavior consistent with the theoretical model, with an estimated α_i between zero and one inclusive. Of these, 19.1% of users did not change their answer at all in the first revision (ie, $\alpha_i = 1$) and 34.7% made no change in the second revision. When $\alpha_i < 0$ an individual moved away from the average of peer estimates, and when $\alpha_i > 0$ the revision overshot the average of peer estimates.

influence. Our theoretical model assumes that this value falls between zero and one. However, some individuals revise their answer such that it moves in the opposite direction of social information, yielding a negative value when this parameter is calculated, or they adjust their answer so that it moves past the average of their network neighbors, yielding a value greater than one. Table S4 shows the distribution of each class of outcomes, measuring α_i for both the first and second revisions by all individuals in a network condition. In both revision processes, over 80% of people displayed behavior consistent with our theoretical model, providing estimates that yielded a parameter value $0 \leq \alpha \leq 1$. Considering only those individuals with a valid α in both revisions (ie, $0 \leq \alpha \leq 1$), we compare the mean α for Revision 1 with the mean α for Revision 2 using a Wilcoxon signed rank test, and find that α increased significantly from the first revision to the second revision ($P<0.01$, Wilcox rank-sum test). This increase in α indicates that the magnitude of revisions decreased slightly between rounds.

Table S4 omits revisions for which α_i is not defined, including 107 Round One responses which were exactly equal to the social signal at time 1, and 168 Round Two responses that exactly equaled the second social signal. These totaled 1.9% and 2.9% of estimates, respectively. Complete experimental data is provided as a Supporting Data file.

Attrition. The experiment lasted 15 minutes, which resulted in very few cases of attrition. Across all trials, only 6.5% of initial estimates and 5.3% of final estimates were non-responses. There were no significant differences in attrition between centralized and decentralized network conditions, for either initial estimations (P>0.69, Chi-squared test) or final estimations (P>0.99, Chi-squared test). For tests, figures, and tables showing change in error between Round One and Round Three, we use only those responses by individuals who provided a response at both Round One and Round Three. An analysis on the full data set provided qualitatively similar results to those presented here.

Ensuring Data Quality. We took precautions to ensure that the subjects did not violate the design of the experiment. Such precautions can be more difficult in online experiments because researchers may have less control over the behavior of the subjects than in traditional laboratory settings. We took several steps to ensure that the data collection was sound. In order to prevent individuals from playing the “Intelligence Game” multiple times, we designed the system so that if a user tried to use a second browser tab to simultaneously access the game, the system would produce an error, and allow only one active browser tab to communicate with our servers. As a result, users were prohibited from playing simultaneously on the same computer. For users recruited from the web at large, we required users to enter their email address before playing the game, and all payments were sent to these addresses, which made it more difficult for users to gain access to the system multiple times. To do so, a user would have had to enroll with multiple email addresses. The interface was very simple and was explained with a set of instruction pictures as users waited for the game to start, so there was very little reason to believe that there was any skill or learning that could occur from having played the game before. For those trials where users were recruited at-large from the web (count-based questions) we used unique sets of questions for each trial (where a trial includes one centralized, one decentralized, and control group) so that repeat users would not have any advantage over new players. For percentage-based estimates, we recruited participants from Amazon Mechanical Turk platform limiting our sample to U.S. participants, which provides strong safeguards ensuring that each registered user was unique. Because of these safeguards and the short time period for this phase of the data collection, we repeated percentage-based questions across multiple trials.

Simulation Analysis

This study builds on theoretical models of opinion formation (20,21) in which agents who revise their beliefs indefinitely will eventually reach consensus, so that every agent in a population shares the same belief (20). DeMarzo et al. (21) analyzed the asymptotic properties of DeGroot's model of consensus (20), showing that after infinite revisions, group opinions converged to a weighted mean of the initial, independent opinions. Each individual's weight in that final collective estimate, or "social influence weight," is determined by that node's eigenvector centrality in the weighted social influence network (21). After only a single revision, however, each node's social influence weight is determined by the sum of incoming tie weights. In decentralized networks (where agents all have the same centrality) this means simply that the group will converge toward the mean of independent estimates.

In populations where the initial beliefs are distributed asymmetrically around their mean, this process leads the median to converge toward the mean in decentralized networks. In count-based questions, we observed initial belief distributions that were heavily right-skewed so that the median was less than the mean. Additionally, these opinion distributions were negatively biased, so that participants tended to underestimate the true answer. These properties produced estimate distributions where the median was less than the mean, and in turn the mean was less than the true answer. These properties of belief distributions, along with the properties of estimation revision, suggest that under minimal satisfying conditions, social influence will improve the median of group estimates in decentralized networks (Fig. S10).

Empirically, however, we observed a reduction in accuracy not only for the median belief, but also the mean belief, and we explain this result by analyzing individual behavior. We observed that individuals who were more accurate made smaller revisions to their estimate, even after controlling for the distance between their estimate and their neighborhood signal. One explanation for this observation emerges from Bayesian decision theory, which forms the basis for the social learning model we study here (21). Demarzo, Vayanoz, and Zweibel (21) note that a rational actor should place more weight on estimates they consider to be more reliable. Thus, if agents have information about the accuracy of their own estimates (e.g., if self-confidence is correlated with accuracy) then it follows that self-weight is correlated with accuracy. Another possibility is that subjects who are more analytically focused on the estimation task might (due to limited cognitive resources) give less attention to social information and also generate more accurate answers. Both of these possibilities offer interesting directions for future research.

To model this effect, 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, however, 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.

As shown in Table S4, we observe a slight decrease in the magnitude of individual revisions between Round One and Round Two. However, DeGroot (20) and DeMarzo (21) assume that individual responsiveness to social influence remains constant over time. Our simulations show that a decrease in revision magnitude over time leads to a minor decrease in effect sizes, but does not otherwise change the network dynamics of social influence on the wisdom of crowds.

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 (20) 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, $\bar{R}_{t,j\in N_i}$. 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:

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

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- α_i) indicates the weight they place on the average estimate of their network neighbors; and $\bar{R}_{t,j\in N_i}$ 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 - α_i).

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

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

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 $(1-\alpha_i)$, 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 (20).

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 (Table S4).

In the main text, we directly measured the relationship between error and revision magnitude, which was statistically significant without any value re-scaling. For count-based estimations, which generate a highly skewed distribution, we find that the partial correlation between error and the scaled parameter α is not significant, even after controlling for question and distance from the social signal ($P>0.49$, Analysis of Covariance). However, when error is calculated on the log-transformed estimate, this correlation is significant ($P<0.05$, Analysis of Covariance). This is due to the fact that the domain for α is bounded by zero and one, while the domain for count-based estimates is bounded on the left side by zero and is unbounded on the right side, generating log-normal distributions (14). For percentage-based estimations, which are bounded by zero and one hundred, α is significantly correlated with the untransformed estimate (P<0.001, Analysis of Covariance).

As in the main text, all responses are normalized by standard deviation prior to analysis. Using the transform described above and standard OLS regression, we estimate the following relationship between error and α:

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

where ε_i indicates the absolute value of the error for an estimate by agent *i*. For count-based response distributions, $\varepsilon_i = |ln(R_i) - ln(truth)|$ where ln indicates the natural log function, while for symmetric distributions $\varepsilon_i = | R_i - truth |$ 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.

As shown in Table S4, mean α_i increases slightly between the first revision (Round One to Round Two) and the second revision (Round Two to Round Three), indicating that responsiveness to social influence decreases. To test the effect of increasing α (decreasing responsiveness to social influence) we introduce a decay parameter, $0 \leq \delta \leq 1$, that controls how much an agent's responsiveness to social influence decreases each round. After each revision, each agent's α_i is modified according to the rule:

$$\alpha_i = 1 - (1 - \alpha_i)(1 - \delta) \ . \tag{6}$$

When δ=0, this reduces to $\alpha_i = \alpha_i$, and each agent's self-weight (α_i) and responsiveness to social influence ($1 - \alpha_i$) is constant over time and the model is identical to that developed by DeGroot (20). When δ=1, this reduces to α_i=1. Thus, when δ=1, each agent makes only one revision, after which α_i=1 and the agent ignores all social information. When 0< δ<1, α_i gradually approaches 1 with each revision, and responsiveness to social information gradually decreases.

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 skew 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

[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.

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 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. S9 and S10 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 (Fig S10, Panels A-C), the group median is drawn toward the mean. When more accurate individuals have a higher value for α (correlation > 0) the group mean becomes more accurate (Fig. S9). When inaccurate individuals make smaller revisions than accurate individuals the group mean becomes less accurate (Fig. S9). 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. Fig. S11 shows the effect of social influence on the group mean when individual responsiveness to social information decreases over time. We find that while effect sizes are slightly reduced, the overall dynamics are identical.

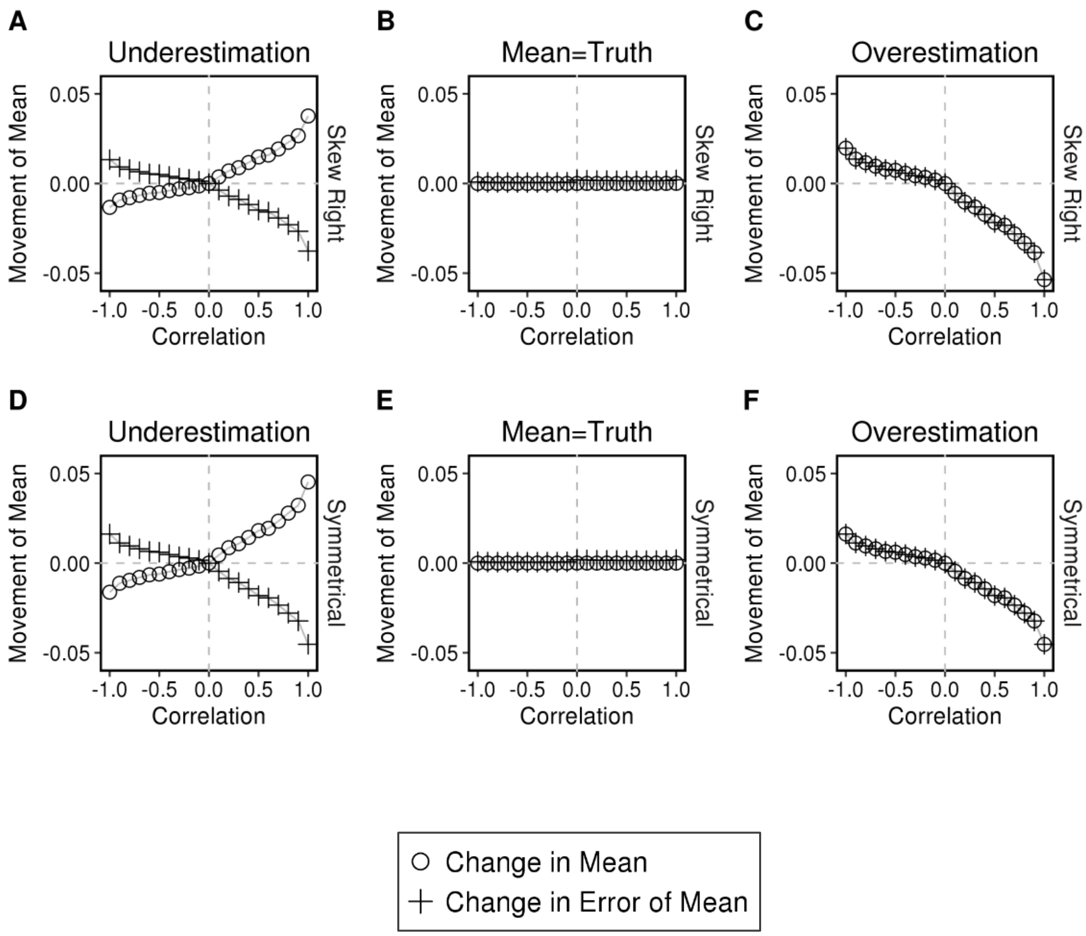


Fig. S9. This figure shows the change in the mean and change in the error of the mean for simulated trials. The x-axis of each panel indicates the correlation between accuracy and α_i. The y-axis for each panel indicates the change in the mean and the change in the error of the mean, as measured in units 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. 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.

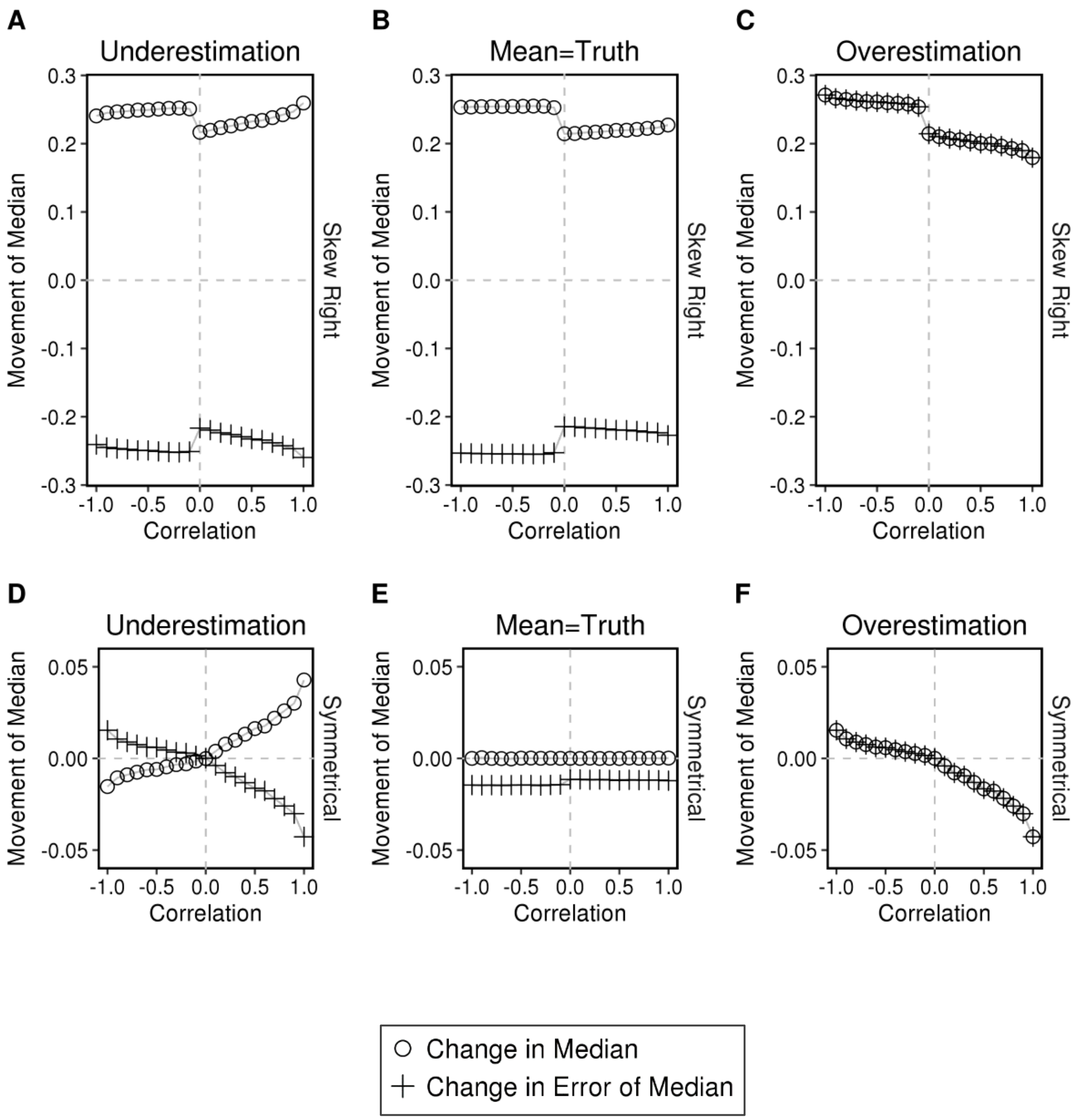


Fig. S10. This figure shows the same model parameters as Fig. S9, but displays results for the median instead of the mean. The x-axis of each panel indicates the correlation between accuracy and α_i. The y-axis for each panel indicates the change in the mean and the change in the error of the mean, as 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.

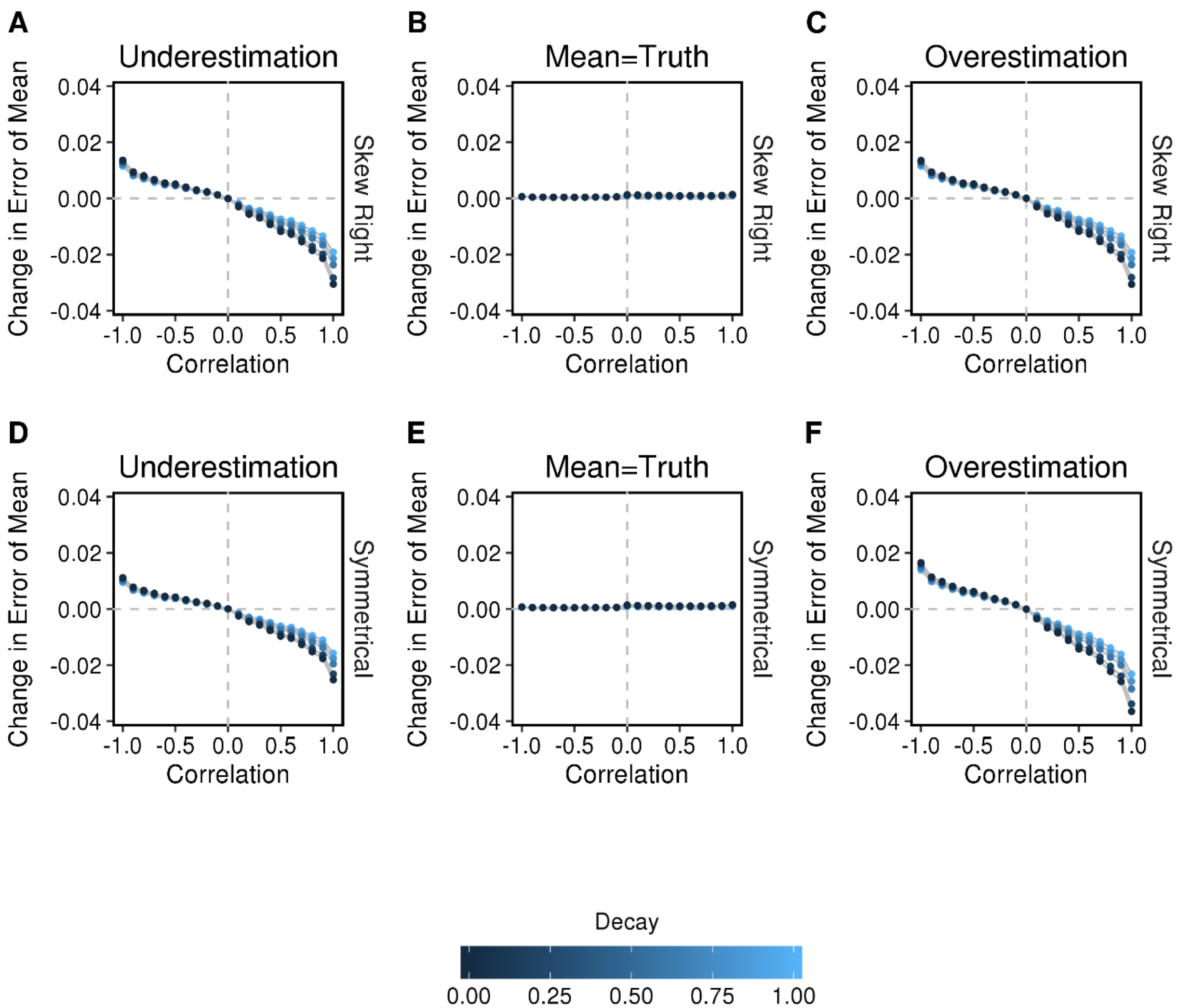


Fig. S11. This figure shows the same model parameters as Fig. S9, but with variation in the decay (δ) of responsiveness to social influence (1-α) over time. When revision magnitude decrease over time, effect sizes are decreased, but overall effects remain the same. Light blue indicates a rapid decay, while dark blue indicates a slow decay. When δ=0, the model is equivalent to that shown in Fig. S9 and S10. The value for decay indicates the percent decrease of responsiveness to social information (1-α) at each round. For example, when δ=0.5, (1-α) decreases by 50% each round. When the decay=0, α and (1-α) both remain constant. N=1,000 nodes per network, 10,000 simulations per point

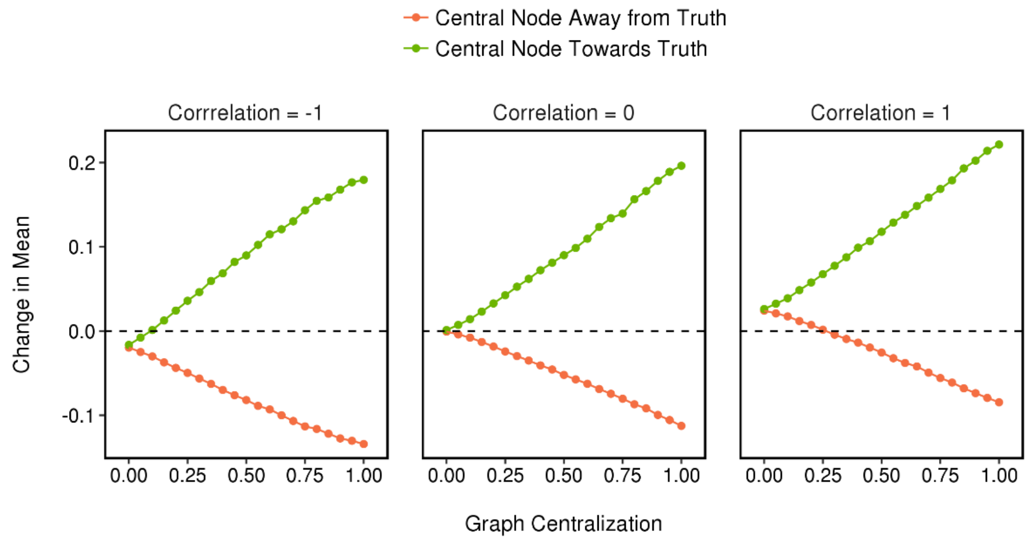


Fig. S12. The effect of centralization in networks on the change in the group mean. These simulations reflect a group that underestimates the true value, and therefore an increase in group mean indicates a decrease in error. The influence of prominent nodes has a stronger effect on group beliefs than 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.

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 (Fig. S8). When the most central member held a belief that represented a movement away from truth (Fig. S12, bottom set of points), the group mean after social influence decreased with centralization, leading

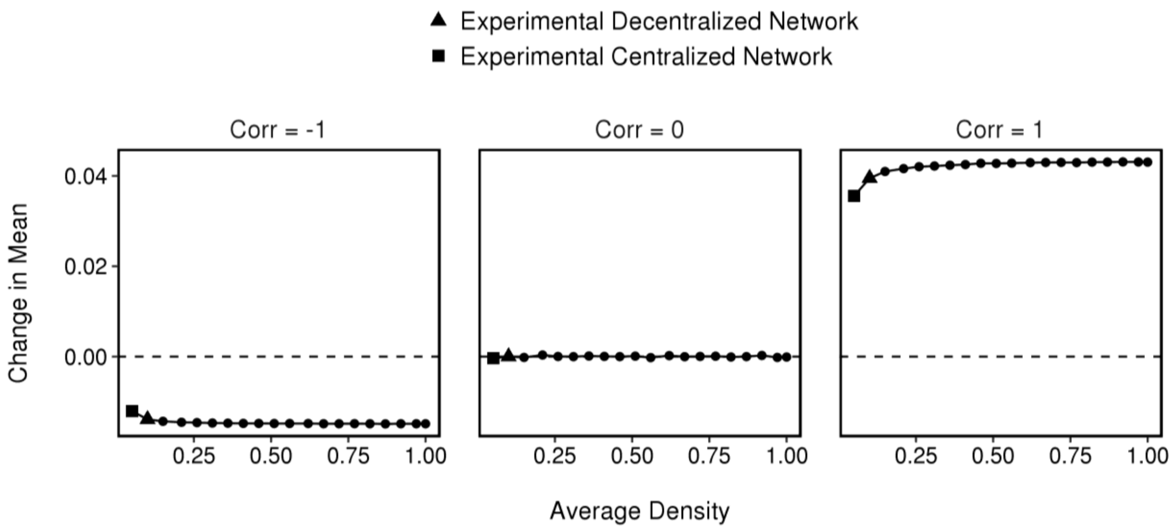


Fig. S13. The effect of density in networks on the change in the group mean. These simulations reflect a group that underestimates the true value, and therefore an increase in group mean indicates a decrease in error. These networks are generated using the degree.sequence.game algorithm (see Experimental and Simulation Network Conditions) and vary only the degree parameter, 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

to an increase in error. When the most central member held a belief that fell in the direction of truth (Fig. S12, 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, other network properties such as density (Fig. S13) and average degree (Fig. S14) have negligible effects on social information and collective 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. S13). In the long run dynamics, density has no effect asymptotically (as it is determined only by the eigenvector of the adjacency matrix for the weighted influence network) and this minimal effect is not enough to account for the difference between networks in our empirical observations. In our experimental trials, the effect of network centralization more than overcomes the small effect of density. 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 densities of the two networks used in our study are noted in both panels: the centralized (box) and decentralized (triangle) networks.

The effects of average degree are similar to the effects of density in direction and magnitude for the change in the group mean (Fig. S14). 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.

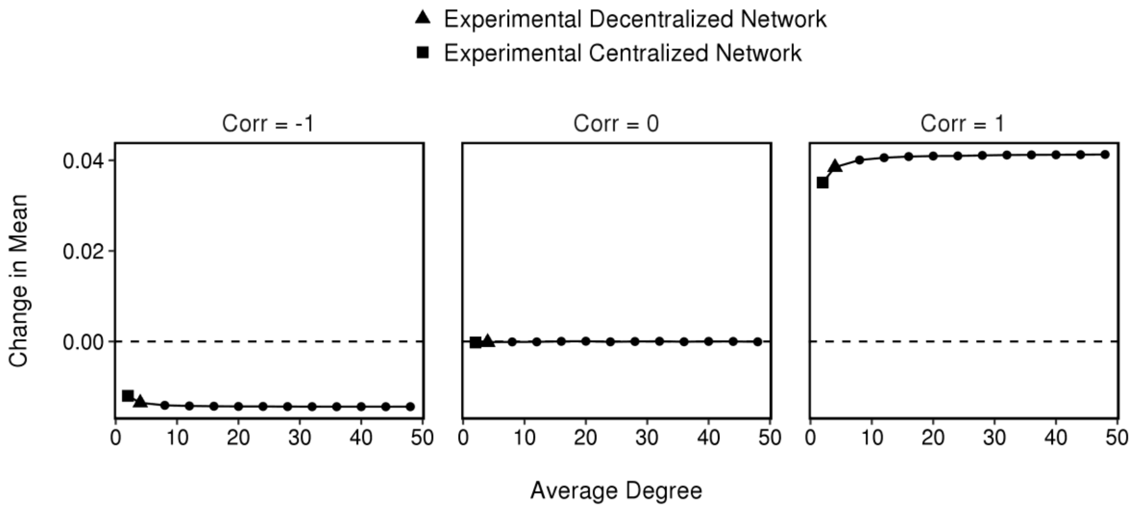


Fig. S14. The effect of average degree in networks on the change in mean. These simulations reflect a group that underestimates the true value, and therefore an increase in group mean indicates a decrease in error. These networks are random networks where every node has the same degree varying N and 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.

Relationship to Previous Work. Some research has examined the effect of social influence on individual accuracy, but has largely focused on individual outcomes and psychological factors (29,30). In one experiment, teams were found to produce better estimates than isolated individuals, but this design was unable to distinguish the mechanisms generating the improvement (9). In one study that examines group accuracy when individuals are exposed to the beliefs of others, Lorenz et al (14) present the primary result that social influence "substantially reduces the diversity of the group without improving its accuracy." They clearly show decrease in diversity as measured by standard deviation, an outcome consistent with our own results, and necessary if individuals are to improve at all. While their interpretation rests on the claim that collective error does not decrease,

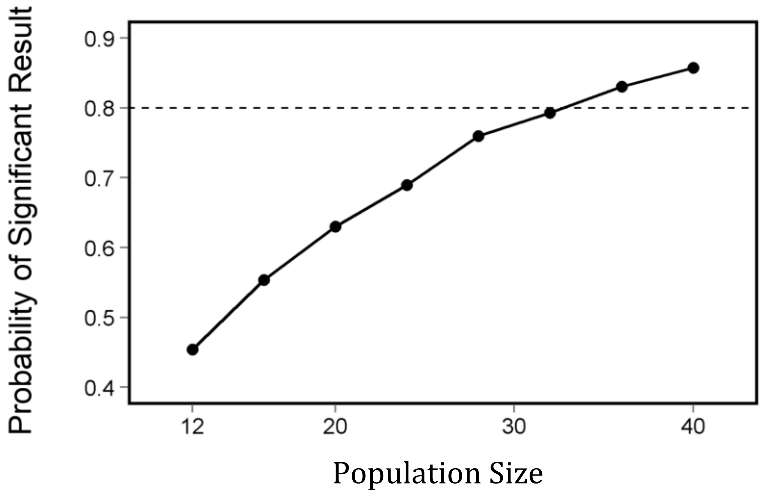


Fig. S15. Points show the expected experimental power as a function of population size, with a dashed line indicating standard design with 80% probability of significant results. Because the wisdom of crowds emerges only in large groups, low population sizes are less likely to reliably demonstrate improvement 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.

their absence of evidence is not evidence of absence. Their null result is most likely due to the small group size. Using simulations seeded with their data, we estimate power tests at various population sizes. Fig. S15 suggests that their experimental design did not have sufficient statistical power to detect the effects described in our study. Our simulations show that with the population size used in their study (N=12), there is more than a 50% chance of a Type II error (power < 50%). Only at larger population sizes like the one used in our study (N=40) does the probability of detecting a significant effect meet conventional levels of statistical power (i.e., power > 80%).

To generate Fig. S15, we conducted a power analysis based on their experimental design and empirical data. In their study, groups of 12 individuals answered 6 questions and were given 5 rounds to revise their answers. The social information in Lorenz et al. (14) was drawn from a decentralized network (fully connected). The theory and results from our study indicate that both individuals and groups should improve when embedded in decentralized networks. While, Lorenz et al. (14) conclude that their population exhibited no change in collective accuracy, their own analysis is highly suggestive of a positive but not statistically significant reduction in accuracy. While they demonstrate clear evidence that populations show herding dynamics as a decrease in standard deviation, and that participants increase in confidence, they do not show that groups do not improve.

These simulations used their experimental design with our theoretical model assuming no particular relationship between α_i and accuracy. Based on individual response data, we estimated α_i as described above. We simulated 5 responses for 4 trials of 6 questions each, producing a total of 24 unique trials with 12 users per trial. Simulated user responses were drawn from empirical distributions associated with each question. We then conducted simulated tests with varied population sizes, generating boot-strapped estimates for the probability of type II error as a function of population size.

The Lorenz et al. (14) data are available at:

http://www.pnas.org/content/suppl/2011/05/10/1008636108.DCSupplemental/sd01.xls.

Table S5. Summary Data for All Estimation Tasks by Decentralized Networks.

This table shows the revision coefficient, the change in the error of the mean, the change in the error of the median, and the change in standard deviation for each estimation task. All changes are measured by subtracting the value at Round One from the value at Round Three. The revision coefficient is measured as the partial correlation between individual error at Round One and the magnitude of the total revision for each individual between Round One and Round Three, after controlling for the distance between the estimate at Round One and the average estimate of network neighbors at Round One.

Group ID	Task	Revision Coefficient Round One to Round Three	Cumulative Change in Error of Mean Round One to Round Three	Cumulative Change in Error of Median Round One to Round Three	Cumulative Change in Standard Deviation Round One to Round Three
2	1e	0.02	0.11	0.00	-0.29
2	2e	0.15	-0.06	-0.14	-0.37
2	3e	0.15	0.00	0.01	-0.29
2	4e	0.92	0.07	-0.04	-0.62
3	1d	1.00	-0.27	0.00	-1.74
3	2d	0.32	-0.13	-0.12	-0.03
3	3d	-0.06	-0.09	-0.22	-0.36
3	4d	0.61	-0.03	-0.24	-0.47
5	1c	0.28	0.05	-0.07	-0.25
5	2c	0.33	-0.11	-0.18	-0.11
5	3c	0.40	-0.01	-0.03	-0.08
5	4c	0.13	0.04	-0.18	-0.54
7	1b	0.15	-0.11	-0.09	-0.27
7	2b	-0.27	0.06	0.04	-0.39
7	3b	0.30	0.06	-0.09	-0.50
7	4b	0.07	-0.05	-0.08	-0.17
9	1a	0.11	0.04	0.00	-0.47
9	2a	-0.08	-0.04	0.01	-0.26
9	3a	0.91	-0.02	-0.11	-1.05
9	4a	0.24	0.01	-0.07	-0.55
12	1f	0.00	0.04	-0.13	-0.46
12	2f	0.46	-0.10	-0.09	-0.35
12	3f	0.28	-0.13	0.03	-0.29
12	4f	0.64	-0.15	-0.08	-0.67
13	5b	0.08	-0.01	0.03	-0.31
13	6b	0.28	-0.16	-0.27	-0.35
13	7b	0.38	-0.13	-0.17	-0.40
13	8b	-0.23	0.08	0.10	-0.30
13	9b	0.24	-0.08	-0.06	-0.34
16	5a	0.64	-0.30	0.00	-0.54
16	6a	0.12	-0.06	-0.14	-0.37
16	7a	0.44	-0.16	-0.30	-0.45

16	8a	0.24	0.03	0.00	-0.34
16	9a	0.07	0.04	0.00	-0.29
18	5a	0.77	-0.29	-0.25	-0.55
18	6a	0.33	-0.17	-0.23	-0.41
18	7a	0.43	-0.26	-0.15	-0.45
18	8a	0.01	-0.06	-0.24	-0.36
18	9a	-0.46	0.17	0.23	-0.31
19	5b	0.32	-0.19	-0.12	-0.22
19	6b	0.07	-0.04	-0.14	-0.30
19	7b	0.69	-0.26	-0.17	-0.38
19	8b	0.11	0.08	0.00	-0.47
19	9b	0.01	-0.04	0.06	-0.50
22	5a	0.70	-0.41	-0.19	-0.65
22	6a	0.56	-0.21	-0.35	-0.49
22	7a	0.63	-0.23	-0.20	-0.30
22	8a	0.05	0.17	-0.09	-0.49
22	9a	0.36	-0.12	-0.33	-0.34
23	5b	0.17	0.02	0.00	-0.24
23	6b	0.12	-0.18	-0.18	-0.36
23	7b	0.81	-0.33	-0.21	-0.54
23	8b	-0.08	0.10	0.21	-0.35
23	9b	0.06	-0.10	-0.19	-0.39
24	5a	0.71	-0.38	-0.32	-0.58
24	6a	0.33	-0.07	-0.14	-0.32
24	7a	0.32	-0.23	0.00	-0.25
24	8a	0.16	0.06	0.06	-0.28
24	9a	-0.40	0.16	0.00	-0.26

Table S6. Summary Data for All Estimation Tasks by Centralized Networks.

This table shows the change in the error of the mean, the change in the error of the median, and the change in standard deviation for each estimation task. All changes are measured by subtracting the value at Round One from the value at Round Three. For each estimation task, the third column ("Central Node Belief in Right Direction?") indicates whether the central node was in the direction of truth relative to the group mean, as shown in Fig. S8. Complete experimental data is available as a Supporting Data file.

Group ID	Task	Central Node Belief in Right Direction Relative to Mean?	Cumulative Change in Error of Mean Round One to Round Three	Cumulative Change in Error of Median Round One to Round Three	Cumulative Change in Standard Deviation Round One to Round Three
1	1e	No	0.22	0.00	-0.48
1	2e	Yes	-0.58	-0.64	-0.19
1	3e	Yes	0.06	0.20	-0.87
1	4e	No	0.01	0.00	-0.05
4	2d	No	0.29	0.26	-0.56
4	3d	Yes	-0.24	-0.43	-0.41
4	4d	Yes	0.07	0.05	-0.16
6	1c	Yes	0.07	-0.08	-0.94
6	2c	No	0.24	-0.12	-1.08
6	3c	Yes	0.01	-0.04	-0.11
6	4c	No	0.35	0.63	-0.17
8	1b	No	0.49	0.69	-0.41
8	2b	No	0.14	0.14	-0.20
8	3b	No	0.04	-0.04	-0.23
8	4b	No	0.21	0.07	-0.88
10	1a	No	-0.04	0.05	0.07
10	2a	Yes	-0.15	0.05	-0.41
10	3a	No	0.18	0.22	-0.18
10	4a	Yes	-0.12	-0.34	-0.64
11	1f	No	-0.12	0.05	0.11
11	2f	Yes	-0.08	0.41	-0.25
11	3f	Yes	0.11	0.20	-0.31
11	4f	Yes	-0.07	-0.18	-0.32
14	5b	No	0.44	0.54	-0.47
14	6b	Yes	-0.19	-0.23	-0.53
14	7b	Yes	-0.28	0.00	-0.52
14	8b	Yes	-0.27	-0.34	-0.53
14	9b	Yes	-0.64	-0.82	-0.15
15	5a	No	0.04	0.25	-0.34
15	6a	No	0.25	0.23	-0.12
15	7a	Yes	-0.60	-0.50	-0.53

15	8a	No	0.40	0.60	-0.54
15	9a	No	0.26	0.36	-0.34
17	5a	Yes	-0.13	0.00	-0.20
17	6a	Yes	-0.06	-0.02	-0.35
17	7a	No	-0.07	0.00	-0.27
17	8a	No	0.57	0.78	-0.40
17	9a	No	0.06	0.00	-0.29
20	5b	No	0.22	-0.15	-0.46
20	6b	Yes	0.29	0.50	-0.57
20	7b	No	0.06	0.42	-0.54
20	8b	Yes	-0.57	-0.69	-0.25
21	5a	Yes	-0.50	-0.13	-0.63
21	6a	No	-0.11	-0.33	-0.50
21	7a	Yes	-0.52	-0.80	-0.53
21	8a	Yes	0.09	-0.06	-0.35
21	9a	No	0.24	0.33	-0.36
25	5a	Yes	-0.51	-0.32	-0.63
25	6a	Yes	-0.58	-0.94	-0.52
25	7a	Yes	-0.19	-0.25	-0.33
25	8a	Yes	-0.20	-0.60	-0.24
25	9a	No	0.19	0.33	-0.24
26	5b	Yes	-0.63	-0.48	-0.38
26	6b	Yes	-0.18	-0.14	-0.32
26	7b	No	-0.29	0.33	-0.69
26	8b	No	0.36	0.34	-0.46
26	9b	Yes	-0.80	-0.95	-0.11

Table S7. Summary Data for All Estimation Tasks by Control Subjects

This table shows the change in the error of the mean, the change in the error of the median, and the change in standard deviation for each estimation task. All changes are measured by subtracting the value at Round One from the value at Round Three. Complete experimental data is available as a Supporting Data file.

Group ID	Task	Cumulative Change in Error of Mean Round One to Round Three	Cumulative Change in Error of Median Round One to Round Three	Cumulative Change in Standard Deviation Round One to Round Three
27	1a	0.06	0.00	-0.21
27	2a	0.44	0.04	2.24
27	3a	-0.07	-0.02	0.12
27	4a	-0.06	0.07	0.15
28	5a	-0.24	-0.19	0.10
28	6a	-0.25	0.00	-0.17
28	7a	0.01	0.00	0.00
28	8a	0.14	0.00	-0.08
28	9a	0.10	0.00	-0.08
29	1b	0.12	0.19	-0.09
29	2b	-0.01	-0.10	-0.07
29	3b	0.05	-0.18	-0.24
29	4b	-0.01	0.08	0.00
30	5b	-0.06	-0.06	-0.21
30	6b	-0.09	0.00	-0.13
30	7b	-0.06	0.00	-0.08
30	8b	0.10	0.14	0.08
30	9b	-0.22	-0.19	-0.03
31	1c	0.11	-0.07	-0.47
31	2c	-0.25	-0.09	0.47
31	3c	-0.14	-0.01	-1.62
31	4c	0.05	-0.11	-0.14
32	1d	0.00	0.00	0.00
32	2d	0.03	0.05	-0.03
32	3d	-0.10	-0.03	0.13
32	4d	-0.05	0.08	0.06
33	1e	-0.16	-0.17	-0.04
33	2e	0.06	0.02	-0.02
33	3e	0.00	-0.09	-0.08
33	4e	0.02	0.02	-0.05
34	1f	0.02	0.12	0.00
34	2f	0.16	0.08	0.11
34	3f	0.02	0.17	-0.86
34	4f	0.06	0.01	0.38

Supporting References

40. Wasserman, S, Faust, K (1994). *Social Network Analysis: Methods and applications.* (Cambridge University Press).

41. Barabási, A, Albert, R (1999). Emergence of scaling in random networks. *Science*. 286 (5439): 509–512.

42. Banerjee, A, Chandrasekhar, A G, Duflo, E, & Jackson, M O (2013). The diffusion of microfinance. *Science*, 341(6144): 1236498.

43. Holm, S (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2): 65-70.

44. Wilcox, R. R. (1996). Statistics for the social sciences. (Academic Press).

45. Cameron, A. C., & Miller, D. L. (2015). *A practitioner's guide to cluster-robust inference.* J Hum Resour, 50(2): 317-372.