Exchanging small amounts of opinions outperforms sharing aggregated opinions of large crowds

Bertrand Jayles\textsuperscript{1,*}, Ralf H.J.M Kurvers\textsuperscript{1}

\textsuperscript{1}Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin

Abstract

The digital revolution has fundamentally changed social information exchange and vastly increased exposure to the opinions of others. However, it is unclear whether exchanging such large amounts of information benefits decision making. Exchanging a moderate amount or aggregated forms of social information may indeed avoid information overload and foster better decisions. We performed experiments in which participants were asked to estimate quantities twice, before and after receiving either all of their peers’ estimates or the geometric mean thereof. We find that second estimates were more accurate when participants observed all estimates than when they saw their geometric mean. Using a model, we predict that accuracy improves most when about twelve estimates are exchanged, independent of group size. Taken together, our results thus suggest that to optimize collective decisions, individuals should receive all decisions from a moderate number of group members, rather than aggregated opinions of large crowds.

\textsuperscript{*}Corresponding author – jayles@mpib-berlin.mpg.de
Introduction

Social information is a decisive component of human decision making. Most of people’s everyday choices, whether picking a movie, finding the best school for one’s children, or gathering information before voting in an election, are influenced by the experiences and intuitions of others. On a broader perspective, social learning strategies, which consist in using social information selectively, continue to play a central role in the emergence and evolution of cultures and their startling diversity [1, 2]. Understanding the impact of social information on human decision making is thus crucial for comprehending human behaviour.

Information technology has altered how people relate to information and how individuals interact with and influence each other. People are more connected to each other than ever before: social networks, blogs and websites, and the massive diffusion of smartphones have made information and virtual others instantaneously available, anywhere and at any time [3]. Moreover, pervasive online recommender systems and social networks have considerably extended people’s exposure to others’ opinions and recommendations [4, 5, 6]. For instance, when selecting a restaurant, a travel destination, or a hotel, the first thing one often does is look at others’ ratings and reviews. This permanent exchange of social information, generally mediated by digital interfaces, is likely to amplify in the coming years, with new generations being born and raised with smartphones and the Internet. This brings about new challenges, such as how to process so much information and make efficient decisions, especially given people’s limited time and cognitive resources [7, 8, 9]. One issue of particular importance is how to best exchange social information in human groups in a way that improves individual and collective decisions. On the one hand, providing individuals access to all the available information gives them more ground to make proper judgments, but at the risk of cognitive overload [10, 11, 12]. On the other hand, aggregated information is easier to process, but lacks potentially important cues about the underlying distribution of information (e.g., the variance in a sample). Between these extremes, exchanging a moderate number of pieces of social information may be valuable for enhancing the quality of decisions.

Here we address this important issue through the prism of estimation tasks, a highly suitable paradigm for quantitative studies on social influenceability [13, 14, 15, 16, 17, 18]. We performed experiments in which subjects were asked to estimate a series of quantities both before and after receiving social information from other group members. Social information consisted of a varying number of estimates \( \tau \) from other group members (\( \tau = 1, 3, 5, 7, 9, \) or 11), and two main conditions were tested: either (i) all \( \tau \) estimates (“full information” condition), or (ii) their geometric mean (“aggregated information” condition) were presented to the subjects. Crucially, in the aggregated information condition, and contrary to previous studies [19, 20], subjects were aware of the number of estimates used to calculate the geometric mean.

Previous studies have analyzed the patterns of social influenceability and the conditions under which
social information exchange can improve estimation accuracy [13, 14, 15, 16, 17, 19, 20, 21, 22]. However, to 
the best of our knowledge, a direct and systematic comparison of the effects of full versus aggregated social 
information, in interaction with the number of estimates exchanged, on collective decisions in estimation 
tasks has been lacking.

We proceed in four steps. First, we present the results of an experiment comparing the effects of aggregated 
versus full information exchange on individual and collective accuracy in estimation tasks. We show that 
exchanging full information leads to improved collective accuracy, whereas exchanging aggregated information 
does not. We did not find a difference in individual improvements between both conditions. Second, we 
investigate the mechanisms underlying these results. We show that the collective improvement in the full 
condition results from subjects’ tendency to favor estimates that are higher than their own over those that 
are lower, thus counteracting the well-known underestimation bias [19, 23, 24, 25]. This effect is not observed 
in the aggregate condition, explaining the lack of collective improvement. Moreover, subjects relied more on 
social information in the aggregate condition than in the full condition, and increasingly so as they knew 
more estimates were involved in the computation of the aggregate (i.e., geometric mean). Third, we present 
a computational model that reproduces the empirical results well, showing that the mechanisms observed are 
key to explaining collective and individual improvements. Finally, we show the model’s predictions for larger 
groups. We find that improvements in collective and individual accuracy are predicted to be optimal when 
about 12 estimates are exchanged, independent of the actual group size.

**Experimental design**

A total of 216 subjects (sex: 138 females, 70 males, 8 unreported; mean age ± standard deviation: 26 ± 
4, 17 unreported), distributed over 18 groups of 12 individuals, took part in the experiment. They signed 
an informed consent form prior to participating. The experiment was approved by the Institutional Review 
Board of the Max Planck Institute for Human Development (ARC 2018/08). Each individual was confronted 
with 42 estimation questions (see the list in the Supplementary Information) on a tactile tablet. Each question 
was asked twice: First, subjects provided their personal estimate $E_p$. Next, they saw the estimate(s) of one 
or several group member(s) as social information and provided a second estimate $E_s$ (see Supplementary 
Figure S1). The social information never contained a participant’s own estimate.

Social information was displayed to subjects in three conditions: (i) the “sorted full information condition,” where $\tau$ estimates ($\tau = 1, 3, 5, 7, 9, 11$) were presented to the subjects, sorted by increasing values; (ii) the “unsorted full information condition,” where the $\tau$ estimates were presented in unsorted order to the 
subjects; and (iii) the “aggregated information condition,” where the geometric mean of the $\tau$ estimates was
presented. In all conditions the exchanged estimates were selected randomly, and in the aggregated information condition subjects were informed about the number of estimates used to compute the geometric mean. Participants in each group only experienced one of the three display conditions (i.e., between-subject design). This was done to avoid the “leakage” of strategies and/or information across treatments. For instance, being exposed to the dispersion of estimates in the full information condition may impact a person’s subsequent decisions on the integration of social information in the aggregated information condition. The number $\tau$ of estimates exchanged did, however, vary within groups.

The 42 questions were randomly assigned to seven blocks of six questions. Across groups, the order of the blocks and the questions within a block were randomized. A block always contained each number of estimates to be exchanged (1, 3, 5, 7, 9, and 11) once. All subjects thus experienced each level of $\tau$ the same amount of times. The randomization was constrained so that across all of the 18 groups, each unique question was asked once at each unique combination of display (three levels) and number of estimates exchanged (six levels). All tablets were controlled by a central server, and participants could only proceed to the next question once all individuals had provided their second estimate. A 30-second countdown was shown on the screen to motivate subjects to answer within this time window, although they were allowed to take more time. Subjects received a flat fee of €15 for participation and a bonus payment of €1 to €5 depending on their performance (see Supplementary Information for detailed payment information).

Since both full information conditions gave relatively similar results (see Figures S2 and S3), we focus here on comparing the sorted full information condition and the aggregated information condition. We refer to them as “full condition” and “aggregate condition” for simplicity.

**Experimental results**

Because of the human logarithmic internal representation of numbers [26], it is more appropriate to consider the logarithm of estimates than the estimates themselves in estimation tasks. Moreover, to make estimates of different quantities comparable, it is necessary to normalize them by the true value of their respective quantities. We therefore use the quantity $X = \log \left( \frac{E}{T} \right)$ as our variable of interest, where $E$ is the actual estimate and $T$ the corresponding true value. $X$ represents a deviation from the truth in terms of orders of magnitude. For simplicity, we will refer to the log-normalized estimates $X$ as “estimates”, with $X_p$ being personal estimates and $X_s$ being second estimates (i.e., after social information exchange). We compared the performance of groups when subjects received the full information (i.e., all pieces of social information) and when they received aggregated information (i.e., the geometric mean of the pieces of social information). Following [19], we define (i) collective accuracy as $\left| \text{Median}_{i,q}(X_{i,q}) \right|$, where $i$ runs over individuals and $q$ over
quantities, and (ii) individual accuracy as $\text{Median}_{i,q}(|(X_{i,q})|)$. Collective accuracy measures how close the median estimate $X$ of all group members is to the truth (the log-normalized transform of the truth is 0), and individual accuracy measures how close, on average, individual estimates are to the truth.

Figure 1: Relative improvement in (a) collective accuracy and (b) individual accuracy, against the number of estimates exchanged, in the full (black) and aggregate (red) conditions. Positive/negative values indicate that accuracy improved/declined. Dots and error bars (computed using a bootstrap procedure described in the Supplementary Information) show empirical data. Dashed lines and empty dots are model simulations.

Figure 1 shows the relative improvements in collective and individual accuracy after social information exchange, as a function of the number of estimates exchanged, for the full and aggregate conditions. A value of 0 implies no improvement (i.e., collective and individual accuracy do not change) and 1 implies maximum improvement (i.e., collective and individual accuracy correspond to the truth after social information exchange). An improvement in collective accuracy (Figure 1a) amounts to a shift of the median estimate toward the truth, which is perforce accompanied by an improvement in individual accuracy (Figure 1b), as individual estimates also get, on average, closer to the truth. However, there can be individual improvement without collective improvement if estimates converge after social information exchange, but without a shift of the median of the $X$ (as shown in [19]). This is the case, for example, when a single estimate is exchanged in both conditions (see Figure 1). Figure S4 shows collective and individual accuracy both before and after social information exchange, to supplement the improvement thereof presented in Figure 1.

In the full condition, increasing the number of estimates resulted in increased collective improvement (Figure 1a). In the aggregate condition, we did not observe such an increase. Moreover, in both conditions, increasing the number of estimates exchanged led to higher individual improvements (Figure 1b). Note that individual improvement in the aggregate condition at $\tau = 11$ was unexpectedly low, which is likely a statistical artefact due to limited samples (i.e., noise). Indeed, collective improvement at $\tau = 11$ in the aggregate condition was also lower than expected (it should be close to 0, as also predicted by the model), and thus negatively affected individual accuracy at $\tau = 11$. Moreover, the error bar for individual accuracy
at \( \tau = 11 \) after social information exchange (Figure S4f) points strongly downwards (i.e., toward higher accuracy), suggesting that higher improvement could have been expected.

We next investigated the mechanisms underlying these results by studying the level of social information use across conditions. We define the value assigned by subjects to the social information as the weight \( S \) they give to the arithmetic mean \( M \) of the social information. Note that the arithmetic mean of log-transformed estimates \( X \) is equivalent to the log of the geometric mean of the actual estimates \( E \). We define a subject’s second estimate \( X_s \) as the weighted arithmetic mean of their personal estimate \( X_p \) and the social information \( M \):

\[
X_s = (1 - S) X_p + S M.
\]

\( S \) can thus be expressed as \( S = \frac{X_p - X_s}{M - X_p} \). \( S = 0 \) implies that subjects keep their personal estimate \( (X_s = X_p) \)—that is, they disregard social information—and \( S = 1 \) implies that their second estimate equals the geometric mean \( (X_s = M) \)—that is, they adopt the central tendency of the social information. Figure 2 shows the average value of \( S \) across all conditions.

![Figure 2: Average weight given to the social information, against the number of estimates exchanged, in the (a) full and (b) aggregate conditions. Shown are the values when (i) all data are combined (dots), (ii) social information is higher than the personal estimate (squares), and (iii) social information is lower than the personal estimate (triangles). Filled symbols and error bars (computed using a bootstrap procedure described in the Supplementary Information) show empirical data. Empty symbols and dashed lines indicate the values obtained from the model simulations.](image-url)

Figure 2a shows that, in the full condition, subjects weighted social information more when it was higher (squares) than their personal estimate than when it was lower (triangles). This mechanism is known as the “asymmetry effect” and has been observed before [21]. Because individuals favor values that are higher than their personal estimates over those that are lower, second estimates tended to shift toward higher values. Subjects thus partly compensated for the underestimation bias, thereby improving collective and individual accuracy. The asymmetry effect increased with the number of estimates exchanged \( \tau \), explaining the increase in collective and individual improvements with \( \tau \) in the full condition (Figure 1a).

In the aggregate condition, subjects weighted social information more the higher the number of exchanged estimates (i.e., the number of estimates used to compute the geometric mean). We call this the average size effect (Figure 2b). Prior research has shown that a partial weighting of social information on average (i.e.,
$0 < \langle S \rangle < 1$ entails individual improvement, a mechanism known as the “herding effect” [21]. Moreover, the authors showed that individual improvement increases as $\langle S \rangle$ increases, peaking at a value of 0.5, after which individual improvement decreases again. The increased weighting of the social information up to $\langle S \rangle \approx 0.5$ in Figure 2b thus explains the increased individual improvement in Figure 1b. Subjects did, however, weigh social information that was higher or lower than their own estimate equally (no asymmetry effect). The absence of the asymmetry effect explains the lack of collective improvement in the aggregate condition (Figure 1a). Subjects, on average, followed social information more in the aggregate than in the full condition. One may thus expect individual improvement to be higher in the aggregate condition than in the full condition. However, this expected difference was compensated by the asymmetry effect, which was only present in the full condition.

In the next section, we describe the agent-based model we built to test whether the proposed mechanisms of information integration can indeed explain the observed patterns of collective and individual improvement.

**Models of social information integration**

The models for the full and aggregate conditions are both adaptations of a model developed in [19], which explains how individuals integrate a single piece of social information (the average of an unknown number of estimates from other group members). It consists of three key components which are also at the heart of the models for full and aggregate conditions presented below. Figure 3 presents these three components, described in more detail below. Parameter values are provided at first mention of each parameter.

First, personal estimates $X_p$ are drawn from Laplace distributions (Figure 3a) for each question, the center and width of which are the median $m_p$ and dispersion $\sigma_p = \langle |X_p - m_p| \rangle$ of the experimental personal estimates, respectively. Because they are made before social information is exchanged, personal estimates are by definition independent of the condition and value of $\tau$.

Second, each agent receives, as social information, the average $M$ of $\tau$ personal estimates from other group members. Note that this happens naturally in the aggregate condition, while for the full condition, we assume that participants also adjust their estimates to the average social information $M$, following previous findings [21, 27]. $S$ is thus defined in the exact same way in both models. After receiving the social information $M$, each agent then either keeps its personal estimate ($S = 0$) with probability $P_0$ or draws an $S$ in a Gaussian distribution of mean $m_g = 0.5$ and standard deviation $\sigma_g = 0.3$ with probability $P_g$. The Gaussian distribution encompasses the probabilities to contradict the social information ($S < 0$), to compromise with it ($0 < S < 1$), to adopt it ($S = 1$), or to overreact to it ($S > 1$). The overall distribution of $S$ is thus composed of a Gaussian distribution and a Dirac peak at $S = 0$ (Figure 3b). Since the distribution
of \( S \) depends on both the condition and value of \( \tau \), data from different cases cannot be combined, contrary to the distribution of personal estimates. Figs. 3b and 3c each show one specific case: the distribution of the \( S \) in the aggregate condition for \( \tau = 1 \) (to remain close to [19]).

Third, the average weight \( S \) given to social information increases linearly with the distance \( D = M - X_p \) between the personal estimate \( X_p \) and the average social information \( M \): \( \langle S \rangle = P_g m_g = \alpha + \beta |D| \) (Figure 3c), where \( \alpha \) is the intercept (\( \alpha = 0.12 \) in the full condition, and \( \alpha = 0.2 \) in the aggregate condition) and \( \beta \) the slope of the linear cusp relationship (\( \beta = 0.1 \) for both conditions). This is the distance effect, described in [19].

On top of these three common components, one additional effect was included in each condition. In the full condition, we introduced the asymmetry effect, built as a linear dependence of \( \langle S \rangle \) on \( \tau \), with a positive slope \( \gamma_+ = 0.03 \) when \( D > 0 \) and a negative slope \( \gamma_- = -0.01 \) when \( D < 0 \), as suggested by Figure 2a: \( \langle S \rangle = P_g m_g = \alpha + \beta |D| + \gamma \pm \tau \). In the aggregate condition, we introduced the average size effect as another linear dependence of \( \langle S \rangle \) on \( \tau \), as observed in Figure 2b: \( \langle S \rangle = P_g m_g = \alpha + \beta |D| + \gamma \tau \), with \( \gamma = 0.022 \).

The probability \( P_g \) for an agent to draw an \( S \) in the Gaussian part of the distribution is given by \( P_g = \langle S \rangle / m_g \). Then \( P_0 \) is given by \( P_0 = 1 - P_g \). Finally, an agent’s second estimate \( X_s \) is defined as \( X_s = (1 - S) X_p + S M \).

Each simulation of the model mimicked our experiment, and the model predictions (shown in all figures presented here) were averaged over 10,000 simulations.
As can be seen from the model simulation results in Figures 1 and 2, both models quantitatively reproduced the empirical results, suggesting that they capture the key mechanisms at play in the integration of social information in both conditions. In the next section, we use the models to make comparative predictions about the full and aggregate conditions for larger group sizes.

Model predictions

Figure 4 shows how the number of estimates exchanged shapes collective and individual improvements, for groups of 50 individuals in both conditions.

Figure 4: Predicted improvement in collective and individual accuracy for groups of 50 individuals. Model simulations showing the predicted relative improvement in (a) collective accuracy and (b) individual accuracy, against the number of estimates exchanged, in the full (black) and aggregate (red) conditions. In both conditions, collective and individual improvements are predicted to saturate when approximately 12 estimates are exchanged.

The model predicts that collective and individual improvements saturate in both conditions when about 12 estimates are exchanged. In the full condition, individual improvement is even predicted to slightly decline when more than 12 estimates are exchanged. The saturation follows from the constraint that probabilities ($P_g$ and $P_b$) must remain between 0 and 1, which imposes limits on the maximum and minimum average weight $S$ given to the social information (see Figure 5), according to the equation $\langle S \rangle = P_g m_g$. The asymmetry (Figure 5a) and average size (Figure 5b) effects are thus bounded, leading to the saturation in collective and individual accuracy, respectively. In the aggregate condition, the collective improvement slightly increases with the number of estimates exchanged, but remains negligible in comparison to the collective improvement in the full condition. However, when more than 12 estimates are exchanged, individual improvement is predicted to be slightly higher in the aggregate condition than in the full condition. Supplementary Figure S5 and S6 show that we obtain qualitatively similar results for different group sizes.
Figure 5: Saturation of the weight given to the social information. Model simulations showing the average weight given to the social information ($S$), against the number of estimates exchanged, in the (a) full and (b) aggregate conditions, for groups of 50 individuals. Shown are the values when (i) all data are combined (circles), (ii) social information is higher than the personal estimate (squares), and (iii) social information is lower than the personal estimate (triangles). $S$ saturates in both conditions when more than approximately 12 estimates are exchanged. Note that all symbols overlap in (b) because there is no asymmetry effect in the aggregate condition.

**Discussion**

The ever-increasing amount of information available online raises two questions: would exchanging aggregated forms of social information improve decision making, as compared to making the complete information available, and is there a limit to the number of exchanged estimates that improve the quality of decisions? We compared the performance of groups in estimation tasks, when subjects received either all the available social information or an aggregate version of it (the geometric mean of the estimates of other group members).

Collective and individual improvements in the full condition were primarily driven by subjects’ tendency to favor estimates that were higher than their personal estimate over those that were lower (asymmetry effect), which shifted second estimates towards higher values. Since humans have a tendency to underestimate quantities [28, 29, 30, 31], this shift towards higher values is a shift towards the truth. This effect, originally unveiled and discussed in [21], has a valuable effect on collective performance in human groups. The authors argued that because people have “difficulties to reason about magnitudes outside of human perception” [32], they may assess the reliability of relatively low numbers more easily than they would very high numbers, making them more likely to discard low estimates compared to high estimates. A concomitant explanation is that people usually know that the quantities they were asked to estimate are supposed to be large, even if they have a poor idea of the actual value. It is therefore conceivable that people are more likely to assume they have underestimated quantities than they are to assume they have overestimated them; as a consequence, they are more likely to follow high estimates than low estimates.

Surprisingly, the asymmetry effect was absent in the aggregate condition, which explains the lack of collective improvement in this condition. A possible explanation is that because averaging smooths out
extreme values, the aggregated social information would usually seem reasonable, independent of whether it
was lower or higher than the personal estimates. Subjects thus could not differentially assess the reliability
of the social information they received.

Moreover, subjects followed social information more (i.e., they “herded” more) when it was aggregated
than when it was fully displayed, and increasingly so as more estimates were exchanged. Previous studies
demonstrated that people are sensitive to the central tendency and dispersion of estimates, and weigh social
information more when the dispersion is low; this mechanism is called the “similarity effect” [21, 27]. Since
in the aggregate condition the dispersion is zero, one may expect maximum herding. Note that subjects
herded more in the aggregate condition even when a single estimate was exchanged, although in this case we
expected both conditions to be equivalent. Since subjects experienced every level of $\tau$ during an experimental
session in the full condition, it is conceivable that the similarity effect negatively affected the overall weight
subjects gave to the social information, including when a single estimate was exchanged. In other words,
experiencing high levels of dispersion at higher values of $\tau$ could, in turn, also reduce social information use
when a single estimate was exchanged. Moreover, the higher weight given to the social information as more
estimates underlie the aggregate reflects people’s statistical intuition that the reliability of averages generally
increases with the number of samples they are computed from [33, 34]. The herding that increased with
the number of estimates exchanged (in the aggregate condition) led to increased individual improvement (as
discussed in [21]), but was not accompanied by collective improvement.

To investigate the generalizability of our results to larger group sizes, we built and calibrated a model
of social information integration based on the results presented in Figure 2. We found that, independent
of group size, collective improvement in the full condition is predicted to improve sharply with the num-
ber of estimates exchanged, up to about 12 estimates, after which it is predicted to quickly saturate. In
the aggregate condition, however, there was barely any improvement in collective accuracy. Individual im-
provement is predicted to saturate in both conditions when about 12 estimates are exchanged, with slightly
higher improvements in the aggregate condition. These results combined suggest that about 12 estimates
should be exchanged and presented to individuals in large groups in order to maximize both collective and
individual improvements. Interestingly, this number, 12, is of the same order as the “magical number seven,
plus or minus two,” a limit in people’s capacity to process information suggested by Miller in his seminal
1956 paper [35]. Although these numbers should be taken with a grain of salt, it makes intuitive sense that
processing more than a few pieces of social information is a difficult task. It is important to understand that
this limit (12) is related the way people use social information (distribution of $S$ and strength of the effects).
One may thus reasonably wonder if the way social information use is distributed in human groups may have
evolved in order to optimize collective decision making, given humans’ inherent cognitive limitations.
It is possible that the parameter values describing social information use in the model would be different at large group sizes when more than 12 estimates are exchanged, such that the observed saturation values may be different from what we observed for a group size of 12. Indeed, if subjects know that the aggregated social information they receive was computed from a large number of estimates, they might weigh the social information more on average, such that \( \langle S \rangle \) might go beyond 0.5. However, even if this was the case, individual improvement would degrade, since it reaches its highest value for \( \langle S \rangle = 0.5 \) (as discussed above), the value obtained when 12 estimates are exchanged. This argument, therefore, does not change our conclusion that about 12 estimates should be exchanged in groups for maximizing collective and individual accuracy. One may also argue that the model ignores potential cognitive overload phenomena [10, 11, 12] when a large number of estimates are exchanged, such that it may be inaccurate for large group sizes. However, the number of estimates does not matter in the aggregate condition. Moreover, in the full condition when estimates are sorted, people are expected to use the same strategy as when a few estimates are exchanged—that is, to focus mostly on the central tendency of the social information, as shown in previous studies [21, 27, 36]. The model predictions should therefore be quite similar to what would be observed in experimental conditions. The argument is more apt when the social information is not sorted, but in this case, collective and individual accuracy are expected to degrade rather than improve with increasing number of estimates exchanged, as people would be less and less able to process social information properly. Therefore, even when social information is not sorted, no more than about 12 estimates should be exchanged in order to maximize improvements in collective and individual accuracy. Finally, if one imagines a situation in which all available pieces of social information must be exchanged (and a substantial amount of them), then it is likely to be preferable to present them in an aggregated manner. Indeed, the main advantage of aggregates is that they free individuals from cognitive overload, independent of the number of estimates exchanged. One can also presume that a critical number of estimates exchanged exists, beyond which displaying aggregates yields higher collective and individual improvement than does displaying the full information. This could constitute an interesting direction for future research.

Overall, we have demonstrated that to optimize collective decisions in human groups, providing individuals with all the decisions or opinions of a moderate number of peers is better than providing aggregates, as people are able to base their decision on more than a generalized tendency.
References


Acknowledgments

We are grateful to Felix Lappe for programming the experiment.

Author contributions

B.J. and R.K. designed research; B.J. and R.K. performed research; B.J. analysed data; B.J. and R.K. wrote the paper.

Competing interests

The authors declare no competing interests.
Supplementary Information. Belonging to:

Exchanging small amounts of opinions outperforms sharing aggregated opinions of large crowds

Bertrand Jayles¹,* Ralf H.J.M Kurvers¹

¹Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

March 9, 2020

1 List of questions

Below is the list of questions used in the experiment and the corresponding true values $T$. In the original experiment, the questions were asked in German. Questions were a mix of general knowledge questions and estimating the number of objects (e.g., marbles, matches, animals) in an image. Images were shown for 6 seconds.

1. Marbles 1: How many marbles are in the jar in the following image? $T = 100$

2. Marbles 2: How many marbles are in the jar in the following image? $T = 450$

3. Matches 1: How many matches are in the following image? $T = 240$

*Corresponding author – jayles@mpib-berlin.mpg.de
4. Matches 2: How many matches are in the following image? $T = 480$

5. Matches 3: How many matches are in the following image? $T = 720$

6. How many gnus are in the following image? $T = 483$

7. Ants 1: How many ants are in the following image? $T = 587$

8. Ants 2: How many ants are in the following image? $T = 672$

9. How many herrings are in the following image? $T = 939$

10. How many ladybugs are in the following image? $T = 812$

11. How many flamingos are in the following image? $T = 528$

12. How many bees are in the following image? $T = 976$

13. What is the population of Tokyo and its conurbation? $T = 38,000,000$

14. What is the population of Shanghai and its conurbation? $T = 25,000,000$

15. What is the population of Seoul and its conurbation? $T = 26,000,000$

16. What is the population of New York City and its conurbation? $T = 21,000,000$

17. What is the population of Madrid and its conurbation? $T = 6,500,000$

18. What is the population of Melbourne and its conurbation? $T = 4,500,000$

19. What is the population of Amsterdam and its conurbation? $T = 1,600,000$
20. How many e-books were sold in Germany in 2016? $T = 28,100,000$
21. How many books does the American Library of Congress hold? $T = 16,000,000$
22. How many people died from cancer worldwide in 2015? $T = 8,800,000$
23. How many smartphones were sold in Germany in 2017? $T = 24,100,000$
24. What was the total distance of the 2016 Tour de France (in kilometers)? $T = 3,529$
25. How many insured cars were stolen in Germany in 2016? $T = 18,227$
26. How many cars were registered in Germany in 2016? $T = 45,071,000$
27. What is the average annual salary of a player in the Bundesliga (in euros)? $T = 1,456,565$
28. What is the average annual salary for players at Bayern Munich (in euros)? $T = 5,460,000$
29. What is the distance from Berlin to New York (in kilometers)? $T = 6,188$
30. How many international tourists were recorded in France in 2016? $T = 82,600,000$
31. How many bicycles are there in Germany? $T = 62,000,000$
32. How many people identify as indigenous in Mexico? $T = 6,000,000$
33. What is the total length of the metal threads used in the braided cables of the Golden Gate Bridge (in kilometers)? $T = 129,000$
34. What is the mass of the Pyramid of Cheops (in tons)? $T = 5,000,000$
35. How much did the Burj Khalifa tower in Dubai cost to build (in dollars)? $T = 1,500,000,000$
36. How many UFO sightings have been reported to the National UFO Reporting Center since its inception? $T = 90,000$
37. What is the diameter of the Sun (in kilometers)? $T = 1,391,400$
38. What is the distance from Mercury to the Sun (in kilometers)? $T = 58,000,000$
39. What is the distance between Earth and the Moon (in kilometers)? $T = 384,400$
40. How many stars are in the Milky Way? $T = 235,000,000,000$
41. How many kilometers are in one light year (in billion kilometers)? $T = 9,460$
42. How many cells are there in the human body (in billion cells)? $T = 100,000$
2 Computation of the error bars

The error bars indicate the variability of our results depending on the $N_Q = 42$ questions presented to the subjects. We call $x_0$ the actual measurement of a quantity appearing in the figures by considering all $N_Q$ questions asked. We then generate the results of $N_{\text{exp}} = 1,000$ new effective experiments. For each effective experiment indexed by $n = 1, ..., N_{\text{exp}}$, we randomly draw $N'_Q = N_Q$ questions among the $N_Q$ questions asked (so that some questions can appear several times, and others may not appear) and recompute the quantity of interest which now takes the value $x_n$. The upper error bar $b_+$ for $x_0$ is defined so that $C = 68.3\%$ (by analogy with the usual standard deviation for a normal distribution) of the $x_n$ greater than $x_0$ are between $x_0$ and $x_0 + b_+$. Similarly, the lower error bar $b_-$ is defined so that $C = 68.3\%$ of the $x_n$ lower than $x_0$ are between $x_0 - b_-$ and $x_0$. The introduction of these upper and lower confidence intervals is adapted to the case that the distribution of the $x_n$ is unknown and potentially not symmetric.

3 Incentive structure

The performance $P$ of an individual $i$ was defined as:

$$P_i = \frac{1}{2} \left( \text{Median}_q \left| \log \left( \frac{E_{p,i,q}}{T_q} \right) \right| + \text{Median}_q \left| \log \left( \frac{E_{s,i,q}}{T_q} \right) \right| \right),$$

where $i$ and $q$ are, respectively, indexes for individuals and questions; $E_p$ and $E_s$ are, respectively, estimates before (personal) and after (second) social information exchange; and $T$ is the correct answer to the question.

This performance criterion measures the median distance to the correct answer—in terms of orders of magnitude—over all questions, averaged over the two estimates (before and after social information exchange).

The payments were defined according to the following distribution of performances:

- $P < 0.3$: €20
- $0.3 \leq P < 0.35$: €19
- $0.35 \leq P < 0.4$: €18
- $0.4 \leq P < 0.45$: €17
- $0.45 \leq P$: €16
Figure S1: Experimental procedure for an example question. The left panel shows the first screen, where subjects provided their personal estimate. The question was asked on the first line, and subjects could type their answer on the second line, using a keyboard that appeared when they clicked on Ihre Antwort (Your answer in German). Subjects submitted their estimates by pushing the OK button. A timer was displayed in the top right corner of the screen to remind subjects to answer within 30 seconds. The right panel shows the second screen, where subjects could revise their estimate after observing answers from other group members. In this example five answers were shown (full information condition). As a reminder, the original question and the subject’s personal estimate were shown. Subjects provided their second estimate in the same way as the first one, with the countdown again set to 30 seconds.

Figure S2: Relative improvement in (a) collective accuracy and (b) individual accuracy, against the number of estimates exchanged, in the sorted (black) and unsorted (blue) conditions. Positive/negative values indicate that accuracy improved/declined.
Figure S3: Average weight given to social information, against the number of estimates exchanged, in the (a) sorted and (b) unsorted conditions. Shown are the values when (i) all data are combined (dots), (ii) social information is higher than the personal estimate (squares), and (iii) social information is lower than the personal estimate (triangles). The asymmetry effect (see main text) is present in both the sorted and unsorted full information conditions.

Figure S4: (a, b, c) Collective accuracy and (d, e, f) individual accuracy, against the number of estimates exchanged, before and after social information exchange, in the sorted (a, d), unsorted (b, e), and aggregate (c, f) conditions.
Figure S5: Model simulations of collective improvement for different group sizes. Relative improvement in collective accuracy against the number of estimates exchanged $\tau$, in the sorted (black) and aggregate (red) conditions. Independent of group size, collective improvement barely increases with $\tau$ in the aggregate condition, and saturates when more than about 12 estimates (blue dashed vertical line) are exchanged in the full condition.

Figure S6: Model simulations of individual improvement for different group sizes. Relative improvement in individual accuracy against the number of estimates exchanged, in the sorted (black) and aggregate (red) conditions. Independent of group size, individual improvement saturates when more than about 12 estimates (blue dashed vertical line) are exchanged in both conditions. Moreover, the saturation value is slightly higher in the aggregate condition.