Research

Author for correspondence:
Bertrand Jayles
e-mail: jayles@mpib-berlin.mpg.de

The impact of incorrect social information on collective wisdom in human groups

Bertrand Jayles[1,2,3], Ramón Escobedo[2], Stéphane Cezera[4], Adrien Blanchet[4,5], Tatsuya Kameda[6], Clément Sire[1] and Guy Theraulaz[2,5]

[1]Laboratoire de Physique Théorique, and [2]Centre de Recherches sur la Cognition Animal-Centre de Biologie Intégrative, Centre National de la Recherche Scientifique (CNRS), Université de Toulouse - Paul Sabatier (UPS), Toulouse, France
[3]Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
[4]Toulouse School of Economics, INRA, Université de Toulouse (Capitole), 31000 Toulouse, France
[5]Institute for Advanced Study in Toulouse, 31015 Toulouse, France
[6]Department of Social Psychology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

BJ, 0000-0002-4550-7919; RE, 0000-0002-9945-0460; CS, 0000-0003-4089-4013; GT, 0000-0002-3062-4215

A major problem resulting from the massive use of social media is the potential spread of incorrect information. Yet, very few studies have investigated the impact of incorrect information on individual and collective decisions. We performed experiments in which participants had to estimate a series of quantities, before and after receiving social information. Unbeknownst to them, we controlled the degree of inaccuracy of the social information through 'virtual influencers', who provided some incorrect information. We find that a large proportion of individuals only partially follow the social information, thus resisting incorrect information. Moreover, incorrect information can help improve group performance more than correct information, when going against a human underestimation bias. We then design a computational model whose predictions are in good agreement with the empirical data, and sheds light on the mechanisms underlying our results. Besides these main findings, we demonstrate that the dispersion of estimates varies a lot between quantities, and must thus be considered when normalizing and aggregating estimates of quantities that are very different in nature. Overall, our results suggest that incorrect information does not necessarily impair the collective wisdom of groups, and can even be used to dampen the negative effects of known cognitive biases.

1. Introduction

The digital revolution has changed the way people access and share information. In particular, the past few decades have seen an exponential increase of media sources and amount of available information [1]. Moreover, a growing distrust in traditional media has given an increasing share of news consumption to social networks and other pathways to relay information. This facilitated and more diverse access to information may arguably enhance people's ability to make informed decisions, but at the same time such an information overload dramatically increases the difficulty to verify information, understand an issue, or make efficient decisions [2,3]. In certain cases, it has also disrupted the relationship between citizens and the truth [4,5], leading to polarized communities unable to listen to each other [6]. Recently, the effects of large scale diffusion of incorrect information and fake news on the behaviour of crowds have gained increasing interest, because of their major social and political impact [7]. The propagation of false information is also reinforced by the use of social bots simulating the behaviour of Internet users [8]. In particular, there has been recent evidence that fake news can propagate faster and affect people deeper than true information on Twitter, especially when they carry political content [9]. In this context, there is a strong need to understand how the

diffusion of incorrect information among group members affects individual and collective decisions.

To address this issue, we use the experimental framework of estimation tasks, which is highly suitable for quantitative studies on social influenceability [10–16]. We performed experiments in which subjects had to estimate a series of quantities with varying levels of demonstrability, before and after having received social information. The demonstrability of a quantity can be interpreted as the amount of prior information a group has about it. To put it in simple terms, it represents the 'difficulty' to determining the actual value of a quantity. We first show that this quantity is closely related to the dispersion of estimates, and *must* be taken into account when normalizing estimates, which has hitherto largely been neglected [11,12,16–19]. We provide an adequate normalization procedure, and discuss its implications in terms of distributions of estimates.

We then investigate how incorrect social information affects estimation accuracy. Participants estimated each quantity sequentially, i.e. one after another. By introducing *virtual influencers*, providing either the true value or some incorrect information (without the subjects being aware of it) in the sequence of estimates, we controlled the *quality* of the information provided to the subjects, allowing us to quantify its resulting impact on individual and collective accuracy. We demonstrate that providing incorrect information that overestimates the truth compensates the underestimation bias—a pervasive bias in estimations of large quantities [20–23]—and thereby improves individual and collective accuracy, the so-called wisdom of crowds [24,25].

Finally, we use a modified version of an agent-based model developed in [18] to better understand the present results, and to analyse the collective response of human groups to information of which levels of inaccuracy go beyond the values tested in our experiments. The model quantitatively reproduces the experimental results, and confirms the counterintuitive observation that incorrect information can improve a group's performance more than correct information, in particular when the group underestimates the true value and the social information overestimates it.

2. Experimental design

One-hundred and eighty subjects participated in our experiment. Twenty sessions were organized, in each of which nine subjects were asked to estimate 32 quantities. Each quantity was estimated twice: subjects first provided their personal/prior estimate E_P. Next, they received as social information the geometric mean G of the τ previous estimate(s) in the sequence ($\tau = 1$ or 3), and were then asked to provide a second/final estimate E_s. The choice of the geometric mean is consistent with humans perceiving numbers roughly as their order of magnitude [26–28]. The value of τ was unknown to the subjects and so was the exact nature of the mean provided. Moreover, this second estimate E_s was used to update the social information for the corresponding subject in the next session. Hence, our experiment produced $9 \times 32 \times 20 = 5760$ personal and second estimates, adding up to a total of 11 520 estimates.

We controlled the quality of the social information provided to the subjects, without them being aware of it. To that end, we inserted in the sequence of 20 final estimates

given by the subjects—unbeknownst to them—$n = 0$, 5 or 15 artificial estimates. These additional estimates correspond to a fraction $\rho = n/(20 + n) = 0\%$, 20% or 43% of *virtual influencers*. Each sequence thus consisted of $N = 20 + n = 20$, 25 or 35 estimates overall, among which 20 were estimates given by 20 actual participants, one per session. The influencers' estimates were introduced at random locations in the sequences. The value T_I of the influencers' estimates provided to the participants was controlled through a parameter α, which represents a normalized distance to the true value quantifying the (in)correctness of the influencers' estimates ($\alpha = -2, -1, -0.5, 0.5, 1, 1.5, 2, 3$), and which will be defined in the Results section. In each session and for each question, a subject was thus assigned a value of ρ, τ and α, and his/her second estimate was a single step in a sequence of 20, 25 or 35 estimates. Figure 1 below provides a graphic representation of the protocol. Note that the estimates of the *virtual influencers* are also used to update the social information which is then provided to the next subjects in the sequence.

The quantities to estimate were grouped into four categories: visual perception (number or length of objects in an image); population of large cities in the world; daily life facts; extreme astronomical, biological or geological events. As we will see, the separation into these loosely defined categories is reflected in the collected data. Three additional questions were asked, which cannot be assigned to any of these categories (see the list of questions in the electronic supplementary material). All experimental details are given in the electronic supplementary material, Material and methods.

3. Results

3.1. Comparing quantities of very different nature

Because humans perceive numbers roughly as their order of magnitude [26–28], the logarithm of estimates is the natural quantity to consider in estimation tasks, especially for large quantities, rather than the actual estimates themselves. Distributions of estimates have indeed often been found highly right-skewed, while the distribution of their common logarithm is generally much more symmetric [11,13,17,28]. An important issue in estimation tasks is to find a proper way to normalize and aggregate estimates arising from questions with very different quantitative answers. Within studies, how can one aggregate estimates of quantities that differ by several orders of magnitude? Between studies, how can one compare findings coming from different sets of quantities?

In line with other works [29,30], we find that the median log-estimate scales linearly with the logarithm $\log(T)$ of the true value (figure 2a), which leads to the natural normalization: $X_P = \log(E_P/T)$. X_P represents the deviation of an estimate from the true value in orders of magnitude, and is often used as the quantity of interest in estimation tasks [11,12,17,18]. However, this normalization does not take into account the dispersion of the log-estimates $\langle |\log(E_P) - \text{Median}(\log(E_P))| \rangle$ (where $\langle x \rangle$ refers to the mean of x) which can vary considerably for different questions (figure 2b). In the following, we simply refer to X as the 'estimates', dropping the 'log-' prefix.

Figure 3a presents the median m_P and figure 3b the dispersion $\sigma_P = \langle |X_P - m_P| \rangle$ of the personal estimates X_P, for
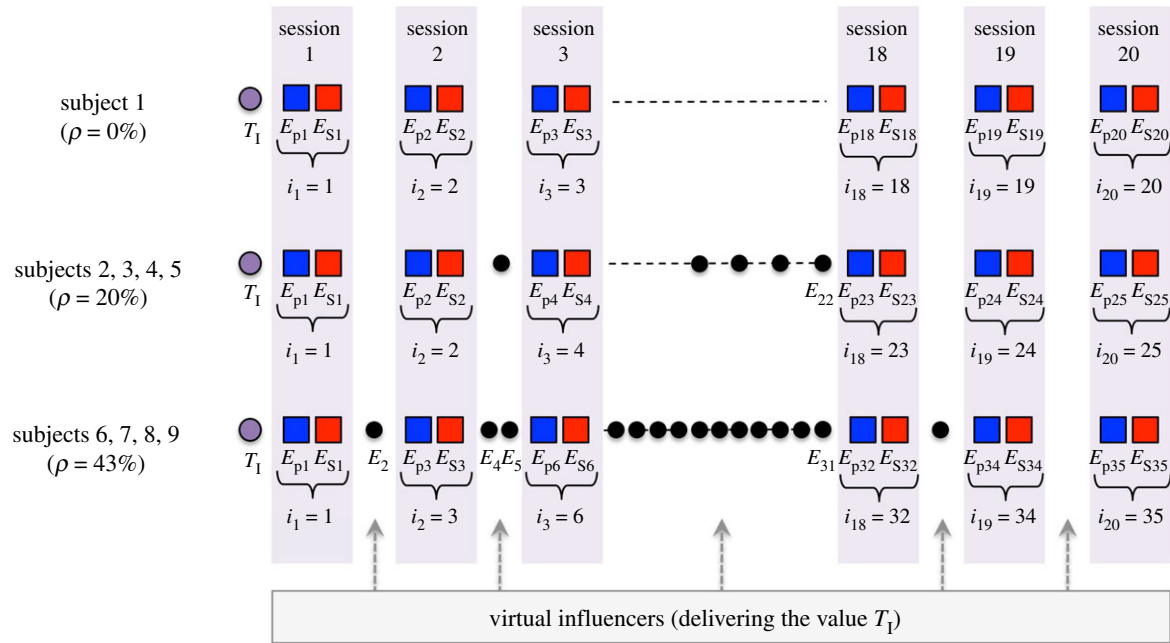
Figure 1. Illustration of the experimental set-up: 20 sessions were organized. Each session is a single step in a sequence of $N = n + 20$ estimates, where $n = 0$, 5 or 15 is the number of *virtual influencers* providing artificial estimates which value was controlled. In each session, nine individuals answer 32 questions twice, once providing their personal estimate E_p (blue squares), and then providing a second estimate E_s (red squares), possibly revised after they received social information. Social information was the geometric mean of the last τ second estimates in the sequence ($\tau = 1$ or 3), and could include artificial estimates from *virtual influencers* (black dots). Every new second estimate was thus used to update the social information provided to the next participant in the next session. The value T_I provided by the *virtual influencers* varied between questions (see the electronic supplementary material, table S1), and was used as initial condition (purple dots).
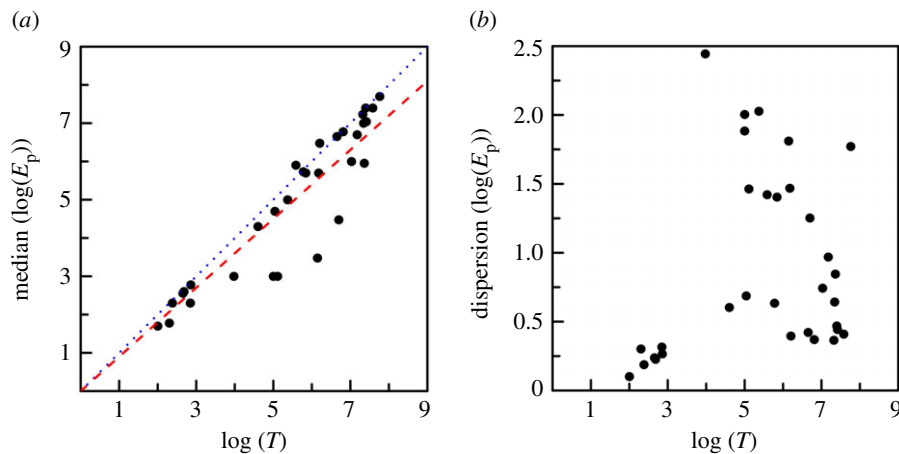

Figure 2. (*a*) Median and (*b*) dispersion $\langle |\log(E_p) - \mathrm{median}(\log(E_p))| \rangle$ of the logarithms of the personal estimates E_p, for the 32 questions asked in the experiment. Median($\log(E_p)$) scales linearly with the log of the true value T. The red dashed line is the linear regression, which slope is lower than 1 (blue dotted line), revealing the human tendency to underestimate quantities.

all questions asked in this experiment (sorted by category of questions). One can note the extreme variation of both quantities depending on the question, suggesting that including m_p and σ_p in the normalization process is crucial to compare quantities of a different nature. Figure 3*b* shows that the category of a question is clearly identifiable by the dispersion of estimates σ_p (but not by the median m_p, see figure 3*a*). The natural classification that we have chosen a priori is thus reflected in the experimental data. Moreover, we see that the less demonstrable a question is, the higher the dispersion of estimates. This is further supported by the three unclassified questions (30–32): one could have predicted that they had a low demonstrability (i.e. that people have little prior information about them), and that they would therefore be

closer to the 'extreme events' category than to the other categories, as observed.

3.2. Full normalization of estimates

In a previous study, we found and justified that the estimates X_p for low demonstrability questions have a probability distribution function (PDF) close to the Cauchy distribution [18]. This property can be explained by a simple probabilistic argument: if two people provide estimates X_1 and X_2 of a quantity about which they have *no information at all*, then the average $(X_1 + X_2)/2$ of both estimates cannot be a statistically better estimation of the correct answer T. Hence, this average has necessarily the same probability distribution as
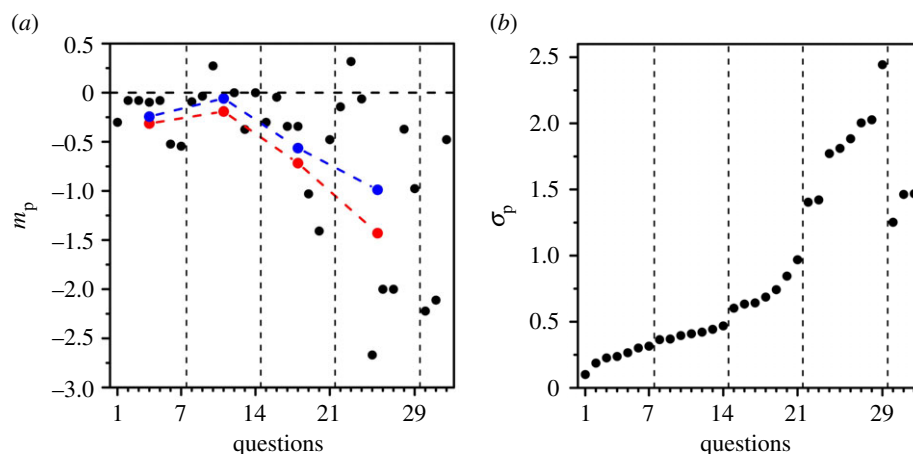
Figure 3. (a) Median m_p and (b) dispersion $\sigma_p = \langle |X_p - m_p| \rangle$ of estimates $X_p = \log(E_p/T)$, for the 32 questions asked in the experiment, whose identity (ID) are ranked according to their σ_p which also reflects their demonstrability. The four categories of questions (from left to right: visual perception, population of large cities in the world, daily life facts, extreme events), plus the three additional questions, are separated by dashed lines. The categories are well distinct in (b), indicating that σ_p is characteristic of the type of quantity to estimate, and more precisely of a question's demonstrability. In (a), the correlation with demonstrability is much less clear, although m_p tends to grow on average when the demonstrability decreases (i.e. when the question ID and σ_p increase). The blue and red dashed lines in (a) are respectively the average value of m_p, and the quantity $-\sqrt{\langle m_p^2 \rangle}$, in each category.
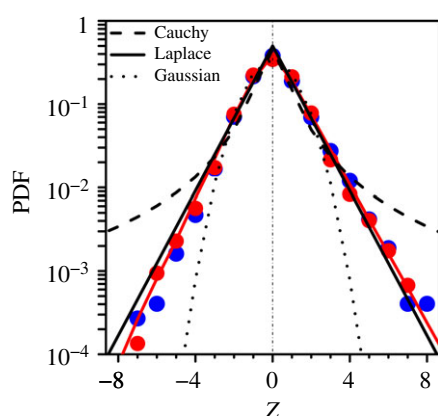

Figure 4. Distribution of fully normalized estimates $Z = (X - m)/\sigma$, before (blue) and after (red) social influence. m and σ are, respectively, the median of the estimates $X = \log(E/T)$ and their dispersion, for each corresponding question. E are the actual estimates and T the true value for each corresponding quantity. The black lines are the standard (centre 0 and width 1) Laplace distribution (full line), the Cauchy distribution (dashed line) and the Gaussian distribution (dotted line) of same width. The Laplace distribution fits the experimental data the best. Red lines (overlapping blue lines) are model simulations.

X_1 and X_2, and the only distribution that satisfies such a property is the Cauchy distribution (see also electronic supplementary material, Material and methods). Our model based on Cauchy distributions convincingly reproduced the experimental data, and in particular, the experimental distribution of personal estimates X_p [18].

However, as we pointed out above, *both m_p and σ_p* have to be considered to compare estimates for questions with answers spanning several orders of magnitude. Hence, for each question characterized by its intrinsic median m_p and dispersion σ_p, we normalize the estimate as $Z_p = (X_p - m_p)/\sigma_p$. Figure 4 shows that the normalized estimates Z_p follow the standard Laplace distribution (i.e. with centre 0 and width 1), $f(Z) = \exp(-|Z|)/2$, implying that the X_p are also Laplace distributed for *individual questions*. It is only when different questions with arbitrary dispersions σ_p are aggregated without

our normalization that an overall Cauchy-like distribution for the X_p emerges. Similarly, note that after social influence (red dots), the $Z_s = (X_s - m_s)/\sigma_s$, with $m_s = \text{median}(X_s)$ and $\sigma_s = \langle |X_s - m_s| \rangle$ also follow the standard Laplace distribution, implying that the X_s also follow a Laplace distribution for each question. We will, therefore, slightly modify the model developed previously [18], to replace Cauchy distributions by Laplace distributions (see Model section).

By measuring m_p and σ_p and using them in the normalization process, we fix the quantity $\langle |Z_p| \rangle = 1$, and therefore have *some* information about the distribution, instead of *none* for the Cauchy distributions argument presented above. As shown in the electronic supplementary material, Material and methods, by exploiting the principle of maximum entropy, the most likely distribution satisfying such a constraint is indeed the Laplace distribution.

This constraint on the dispersion of estimates can be understood as an intrinsic property of the system {group of individuals, question}: the dispersion is characteristic of a given group of individuals estimating a given quantity, and gives the typical range of answers that would seem reasonable to most people in the group for that question. The lower the demonstrability of a question (i.e. the lower the amount of prior information held by individuals in a group about that question), the larger this range. This is intuitive when considering the following example: an estimate three orders of magnitude away from the true value would seem absurd if one considers the age of death of a celebrity, while it would seem perfectly plausible if one considers the number of stars in the universe. While the normalization by m_p is somewhat trivial (it simply shifts the centre of the distribution of X to 0 for every question), the normalization by σ_p is therefore crucial in order to be able to properly compare and aggregate estimates from different questions (and possibly, from different studies). We wish to insist on the fact that this prescription is not a mere methodological detail and that it should be adopted by future works in the field.

In the electronic supplementary material, figure S1, we show the distribution of Z for the four categories of questions. One can note that for very large quantities (electronic supplementary material, figures S1c and S1d), the left side of
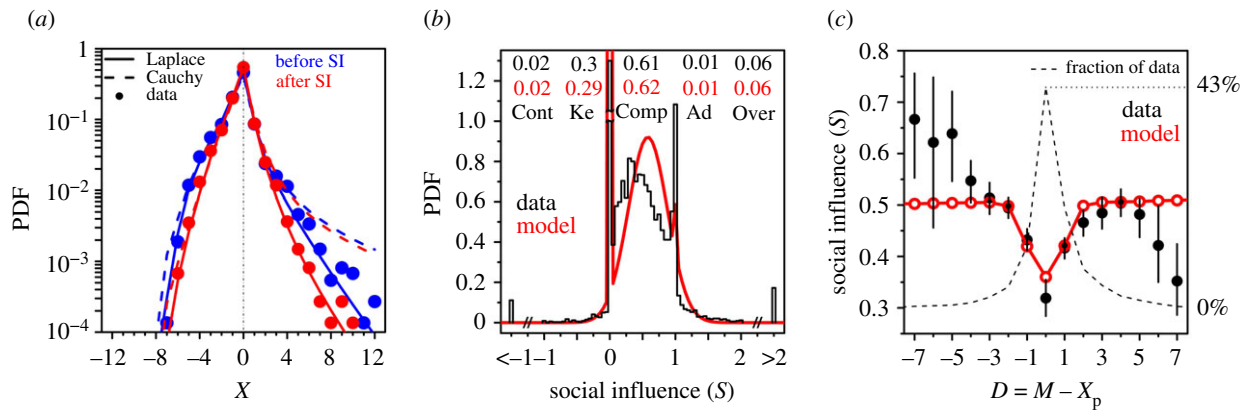
Figure 5. (a) Probability density function (PDF) of individual estimates, before (X_p, blue) and after social influence (X_s, red). Dots show experimental data, dashed lines are model simulations based on Cauchy distributions and full lines model simulations based on Laplace distributions. Note the sharp decay on the left side of the distribution, well reproduced by the model. The questions asked in our experiment imposed answers higher than one, which translates into $X > -\log(T)$. (b) PDF of the sensitivity to social influence S. The fractions of the five behavioural categories are shown, from left to right: contradicters ('Cont', $S < 0$), keepers ('Ke', $S = 0$), compromisers ('Comp', $0 < S < 1$), adopters ('Ad', $S = 1$) and overreacters ('Over', $S > 1$). Experimental data are shown in black, and model simulations in red. (c) Average sensitivity to social influence S against the distance $D = M - X_p$ between the social information M and the personal estimate X_p. Because the average is sensitive to extreme values, we excluded the values such that $|S| > 100$, which represent less than 1% of the data. Black dots correspond to the experimental data, and red empty circles to the model simulations. The dashed line shows the fraction of data for each dot.

the distribution collapses faster than the right side, suggesting that people have an intuition that such quantities must be large, even though they know little about them, such that very small estimates are less frequent. Such asymmetric Laplace distributions can also be derived from the principle of maximum entropy, by adding a constraint that penalizes small or large estimates (see electronic supplementary material, Material and methods).

3.3. Model

In [18], we have introduced an agent-based model to better understand the effects of individual sensitivity to social influence, and of the quantity of information delivered to the individuals, on collective performance and accuracy observed at the group level in estimation tasks. The model uses as basic variables the log-transformed estimates $X = \log(E/T)$, called 'estimates' for simplicity.

Personal estimates X_p are drawn from Laplace distributions, the centre and width of which are, respectively, the median m_p and dispersion $\sigma_p = \langle | X_p - m_p | \rangle$ of the experimental personal estimates X_p for each question. Figure 5a presents the distribution of estimates X for all questions combined, before (blue) and after social influence (red), as well as the corresponding distributions generated by our model, when the X_p are generated from Cauchy distributions (as in our previous research [18], dashed lines) and Laplace distributions (full lines).

The Laplace distribution is able to capture the estimates far from the truth ($X_{p,s} > 5$) better than the Cauchy distribution. It is important to mention that in our previous study [18], the range of possible answers were limited to plus or minus 3, 5 or 7 orders of magnitude from the true value, depending on the question. By not allowing extreme answers, we probably increased artificially the probability of estimates in the interval [5,7], making the distribution even closer to a Cauchy distribution.

After providing its personal estimate X_p, each agent receives as social information the *arithmetic* mean M of the τ previous final estimates in the sequence, among which

some information V (provided by the *virtual influencers*) is introduced with probability ρ. Note that the actual participants were provided the geometric mean G of the τ previous estimates. In terms of log-estimates, the social information $M = \log(G)$ indeed transforms into the standard arithmetic mean. The agent then provides a second estimate X_s, defined as the weighted average of its personal estimate X_p and the social information M: $X_s = (1 - S) X_p + S M$, where S is the weight given to the social information, that we call *sensitivity to social influence*. S can thus be expressed as $S = \frac{X_s - X_p}{M - X_p}$. In figure 5b, we show the distribution of S from which five natural behavioural categories can be identified, in accordance with our previous findings [18]: subjects keep their opinion ('keepers', $S = 0$), compromise with the social information ('compromisers', $0 < S < 1$), adopt the social information ('adopters', $S = 1$), contradict it ('contradicters', $S < 0$), or overreact to it ('overreacters', $S > 1$). In the model, after receiving the social information, an agent keeps its personal estimate ($S = 0$) with probability P_0, adopts the social information ($S = 1$) with probability P_1, or draws an S in a Gaussian distribution of centre m_g and width σ_g with probability $P_g = 1 - P_0 - P_1$.

Figure 5c shows that the average sensitivity to social influence S increases linearly with the distance $D = M - X_p$ between the average social information M and the personal estimate X_p. This is implemented in the model by making the probability P_g increase linearly with D, according to the equation: $\langle S \rangle = P_1 + P_g m_g = a + b \, |D|$, where the intercept a and the slope b characterize the linear cusp observed in figure 5c. More details can be found in the electronic supplementary material, Material and methods section. Note the subjects' tendency to give more weight to social information that is much lower than their personal estimate ($D < -3$), than to social information that is much higher ($D > 3$). Because this concerns only about 7.6% of the data, we neglect this effect in the model.

Note that the distribution of X narrows after social influence (red dots and lines in figure 5a), implying that estimates have overall got closer to the truth, all conditions mixed. This may seem counterintuitive, because in most conditions,
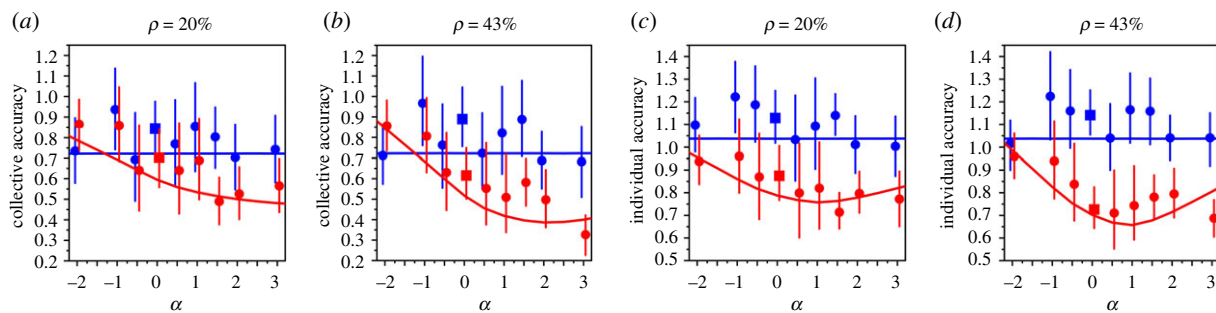
Figure 6. Collective (*a* and *b*) and individual (*c* and *d*) accuracy, as a function of the quantifier of information quality α, before (blue) and after (red) social influence, for $\rho = 20\%$ (*a* and *c*) and $\rho = 43\%$ (*b* and *d*) of influencers in the sequence of estimates. Dots are experimental data from the experiment presented here, while squares at $\alpha = 0$ are experimental data from a previous study, in which the same percentage of *virtual influencers* provided some perfectly accurate information [18]. Full lines are model simulations. Surprisingly, incorrect information can be beneficial to collective and individual accuracy, which optima are reached for positive values of α, i.e. for incorrect information that overestimates the truth. The computation of the error bars is explained in the electronic supplementary material, Material and methods.

incorrect information was provided into the sequence of estimates. To understand this result, we next investigate the impact of incorrect information on estimation accuracy for each condition separately.

3.4. Impact of incorrect information on estimation accuracy

As explained above, we controlled the quality of the social information received by the individuals, by introducing $n = 0$, 5 or 15 *virtual influencers* providing artificial estimates of value T_I randomly inserted in the sequences of 20 estimates provided by the participants, and hence affecting the social information delivered to them. Because we are looking for an information parameter that is independent of the questions, we define, consistently with the previous discussion on the normalization procedure, the normalized (log) deviation from the truth $\alpha = \log(T_I/T)/\sigma_{P_{\exp}} = V/\sigma_{P_{\exp}}$ as an indicator of information quality, where $\sigma_{P_{\exp}}$ is an expected value of the dispersion of personal estimates X_P (the values of the $\sigma_{P_{\exp}}$ are given in the electronic supplementary material, table S1), and V the (log) deviation from the truth of the *virtual influencers* estimates T_I. We obviously did not know the dispersion of estimates before running the experiment. Yet, because the questions were similar to others used in a previous study [18], we could formulate reasonable expectations. Indeed, electronic supplementary material, figure S2 shows that $\sigma_{P_{\exp}}$ scales linearly with the actual dispersion of estimates σ_P, although it tends to underestimate it. α thus represents the deviation of T_I from the truth T in the (expected) natural scale of each question. The value T_I introduced in the sequence of estimates is hence $T_I = T \times 10^{\alpha \sigma_{P_{\exp}}}$, and equals the true value T when $\alpha = 0$. Subsequently, to study the impact of information quality on the group performance, we introduce the variable $Y = X/\sigma_P$, where σ_P is the dispersion of X_P for a given question, and define:

(i) **individual accuracy** as the median of the absolute values of the Y of all individuals i, averaged over all questions q: $\langle \mathrm{median}_i(|Y_{i,q}|)\rangle_q$, and

(ii) **collective accuracy** as the absolute value of the median of the Y of all individuals i, averaged over all questions q: $\langle |\mathrm{median}_i(Y_{i,q})|\rangle_q$.

Individual accuracy measures how close individual estimates are to the truth (i.e. close to 0 in terms of log variables X) on average, while collective accuracy measures how close the median estimate is to the truth. Both measures are distinct, although related. Indeed, an improvement in collective accuracy amounts to a shift of the median estimate towards the truth, which is perforce accompanied by an improvement in individual accuracy, 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 influence, but without a shift of the median (as shown and discussed in [18]).

Figure 6 shows that both measures improve after social influence (i.e. red dots are closer to 0 than blue dots), *over almost the whole range of the considered values of* α, suggesting that incorrect information can, counterintuitively, be beneficial to the performance of groups. Our results also suggest that individual accuracy slightly improves after social influence when $\rho = 0\%$ (i.e. no *virtual influencers*), but not collective accuracy, confirming previous findings [18].

Moreover, the optimum value α_{opt} of α at which collective or individual accuracy improves the most is strictly positive, confirming the model prediction in [18] that such improvement is maximized not by providing perfectly accurate information to individuals, but information that overestimates the true value. Such incorrect information partly compensates the underestimation bias, thus bringing second estimates closer to the truth.

Collective accuracy before social influence (blue dots and lines) represents the absolute value of the collective bias of the group, i.e. the distance between the median estimate and the truth, averaged over all questions. If the value of the collective bias is $\alpha_0 \approx -0.72$, one may naively expect that $\alpha_{\mathrm{opt}} = -\alpha_0$ in order to compensate the collective bias and thus optimize collective accuracy. However, because not all subjects follow the social information fully, one should rather expect $\alpha_{\mathrm{opt}} > -\alpha_0$, as supported by the data and model.

The fraction ρ of *virtual influencers* has no significant effect on collective accuracy in the data in figure 6. However, the simulations of the model predict that collective accuracy degrades after social influence either when $\alpha < \alpha_{\min} \approx -1.2$ (for both $\rho = 20\%$ and $\rho = 43\%$) or when $\alpha > \alpha_{\max} \approx 13.4$ for $\rho = 20\%$ and $\alpha_{\max} \approx 7.2$ for $\rho = 43\%$, which corresponds, respectively, to $\langle \sigma_{P_{\exp}q}\rangle_q \times \alpha_{\max} \approx 0.46 \times 13.4 \approx 6.2$ and
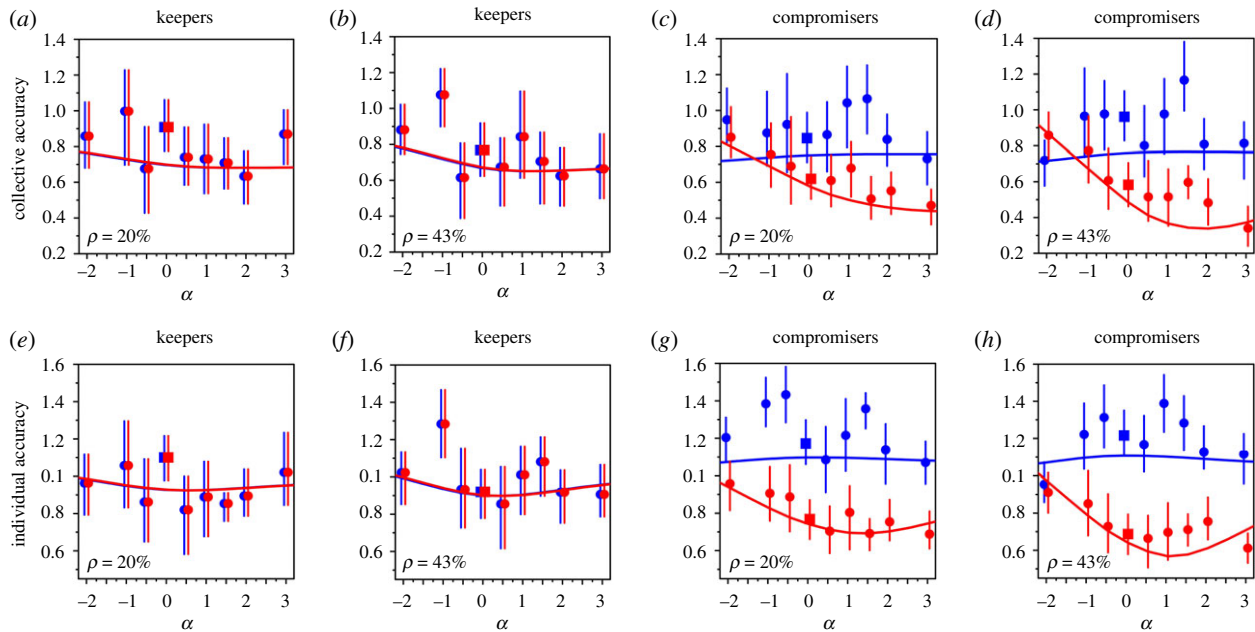
Figure 7. Collective (*a–d*) and individual accuracy (*e–h*) as a function of the quantifier of information quality α, before (blue) and after (red) social influence, for $\rho = 20\%$ (*a,c,e,g*) and $\rho = 43\%$ (*b,d,f,h*) of influencers in the sequence of estimates. Keepers ($S = 0$) are shown in (*a,b,e,f*) and compromisers ($0 < S < 1$) in (*c,d, g,h*). Dots are experimental data from the experiment presented here, while squares at $\alpha = 0$ are experimental data from a previous study, in which the same percentage of influencers provided some perfectly accurate information [18]. Full lines are model simulations.

$\langle \sigma_{P_{\exp q}} \rangle_q \times \alpha_{max} \approx 0.46 \times 7.2 \approx 3.3$ orders of magnitude beyond the true value (see the electronic supplementary material, figure S3). The impact of α is therefore not symmetric with respect to its optimum α_{opt}: incorrect information that largely overestimates the truth can still be beneficial to collective accuracy, while incorrect information that only moderately underestimates the true value is enough to damage collective accuracy. The same analysis remains true for individual accuracy, only with different values of α_{opt}, α_{min}, α_{max} and α_0.

Electronic supplementary material, figure S4 shows individual and collective accuracy for the four categories of questions. Unexpectedly, subjects are found less accurate for questions involving visual perception (electronic supplementary material, figure S4 left column), compared to the other categories which show similar levels of accuracy to one another as well as to when all questions are combined (figure 6). This suggests that estimations involving visual perception are not strictly identical to estimations based on personal knowledge/memory only, although they both are based on similar cognitive processes [31,32]. Indeed, to estimate populations of large cities, daily life facts or extreme events, subjects can only rely on their prior, personal information. However, when actually seeing objects in an image, subjects can attempt to directly measure the lengths/areas/volumes or number of objects to ultimately answer the questions of the first category. In the electronic supplementary material, figure S5, we show the *absolute* collective and individual accuracy for the four categories of questions, i.e. the accuracy before normalizing estimates X by their intrinsic dispersion σ_P for each question. Before the normalization is done, individuals are more accurate for visual perception questions and populations of large cities than for daily life facts and extreme events, as expected from their respective demonstrability (figure 3b). It is, therefore, only when measuring subjects performance relative to the intrinsic dispersion of questions that the difference between visual perception questions and other questions is revealed.

3.5. Incorrect information and sensitivity to social influence

It has been shown that estimation accuracy strongly depends on the sensitivity to social influence of individuals in groups [18]. Analysing the above results in the light of the five behavioural categories of sensitivity to social influence (figure 5b) helps us to understand the mechanisms underlying them. They cannot be explained by contradicters ($S < 0$), adopters ($S = 1$) or overreacters ($S > 1$), who only represent a small percentage of the population. Figure 7 shows collective and individual accuracy as a function of α, for the keepers and compromisers, which together represent a substantial fraction of the population (approx. 91%). Note that the effects are clearer when this separation into behavioural categories is made (compare to figure 6).

Because keepers disregard social information, we observe no improvement in individual or collective accuracy after social influence (figure 7a,b,e,f). However, compromisers (figure 7c,d,g,h), who partly follow the social information, significantly improve their performance over the whole range of incorrect information tested here (except for $\alpha = -2$ and $\rho = 43\%$ of *virtual influencers*). Indeed, because subjects in general, and compromisers in particular, tend to substantially underestimate quantities, they can improve their estimates by following incorrect social information that is closer to the true value than their own personal estimate. Moreover, partially following social information that overestimates the truth allows their second estimates to reach more accurate values, even when the overestimation is quite pronounced. Conversely, individual and collective accuracy degrade quickly when compromisers are given incorrect social information which reinforces their natural cognitive bias by underestimating the true value. Compromising thus allows group members to take advantage of incorrect information, as long as this information goes against their cognitive bias.
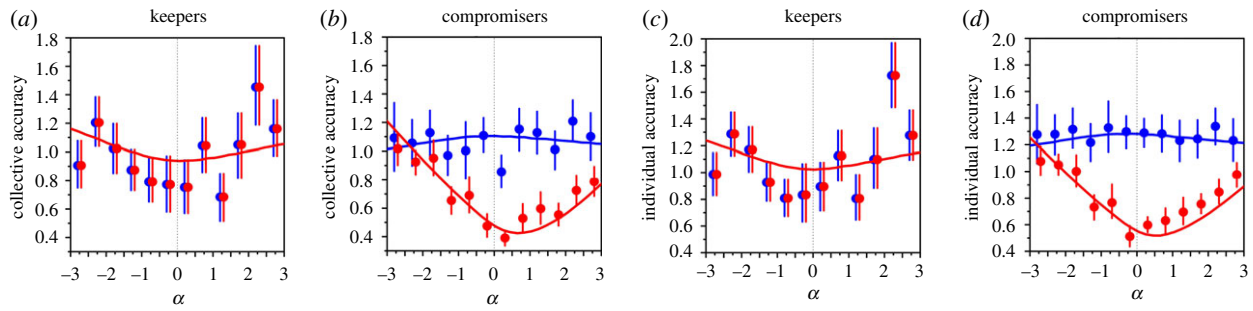
Figure 8. Collective (*a,b*) and individual (*c,d*) accuracy against the quantifier of information quality α, before (blue) and after (red) social influence, in the separated experiment with the isolated subjects. (*a,c*) Keepers ($S = 0$); (*b,d*) compromisers ($0 < S < 1$). Dots are experimental data, and full lines are model simulations.
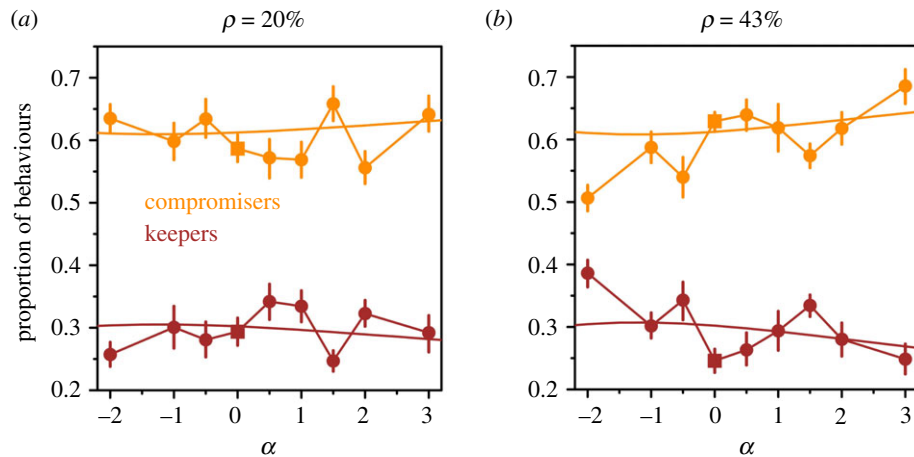

Figure 9. Proportion of compromisers (orange) and keepers (brown) as a function of the quantifier of social information quality α, for $\rho = 20\%$ (*a*), and $\rho = 43\%$ (*b*), of influencers in the sequence of estimates. Dots are experimental data from the present experiment, while squares at $\alpha = 0$ are taken from [18], in which the same percentage of *virtual influencers* provided some perfectly accurate information. Full lines are the model predictions. When $\rho = 43\%$, the fraction of compromisers increases with α at the expense of the fraction of keepers.

Figure 8 shows the equivalent graphs for the 'isolated' subjects of our experiment (see the electronic supplementary material, Material and methods). Isolated subjects received as social information for each question, an estimate T_I generated from a *random* value of α uniformly distributed in the interval [−3, 3]. Figure 8 confirms the above conclusions, but displays sharper patterns, owing to a discretization effect: social information in the main experiment was generated from a discrete set of values of α, whereas for isolated subjects, it was drawn from a continuous distribution.

Before social influence (blue), we find that keepers are slightly more accurate than compromisers (average collective accuracy: 0.98 versus 1.07; average individual accuracy: 1.08 versus 1.28). This was already observed in [18], and justified by the fact that a higher tendency to disregard social information is usually associated with a higher average confidence of the subjects in their own estimates, which often comes with a higher prior knowledge about the quantity to estimate.

Note the slight U-shaped curve for keepers in figure 8*a,c*. This effect is a direct consequence of people's tendency to stick to their personal estimate more when the social information is closer to it (figure 5*c*): when participants receive inaccurate information and retain their opinion, it is often because they were close to it and therefore relatively inaccurate too. Conversely, when participants receive accurate information and keep their opinion, it is often because they

were close to it and therefore quite accurate too. Both effects can be observed in figure 7, but are less pronounced there.

3.6. Influence of the fraction of virtual influencers on individual behaviour

We have seen that compromisers, by partially following social information, were able to improve their accuracy over a wide range of incorrect social information. Figure 9 shows the fraction of keepers and compromisers as a function of α, when $\rho = 20\%$ (figure 9*a*) and $\rho = 43\%$ (figure 9*b*) of *virtual influencers* are introduced in the sequence of estimates.

When $\rho = 20\%$, both fractions of compromisers and keepers remain more or less independent of α (figure 9*a*). However, when the proportion of *virtual influencers* providing incorrect information is doubled ($\rho = 43\%$, figure 9*b*), the fraction of compromisers gradually increases (from 0.5 to 0.68, orange line) with α (from −2 to 3), at the expense of the fraction of keepers which decreases (from 0.38 to 0.25, brown line). For this to happen, this increasing transition from keeping to compromising behaviour as α increases thus necessitates a significant proportion of subjects to be provided with incorrect social information. Moreover, this result also suggests that subjects not only adapt their behaviour to the degree of incorrectness of the social information they receive but also tend to compromise more with some social information that overestimates the truth, than with

some social information that underestimates it. The model predicts this increased behavioural transfer with α, even when $\rho = 20\%$. This is a direct consequence of the cusp relationship between the sensitivity to social influence S and the distance to the social information D (figure 5c): people tend to compromise with the social information more as it gets farther from their personal estimates (and from the truth) on average. However, this effect is significantly stronger in the data, suggesting that other mechanisms exist that are not implemented in the model. Electronic supplementary material, figure S6 demonstrates that this increasing fraction of compromisers with α when $\rho = 43\%$ leads to an increased improvement in individual and collective accuracy for the whole group after social influence (but not when $\rho = 20\%$).

4. Discussion

Understanding the effects of incorrect information on individual and collective decisions is crucial in modern digital societies, where social networks and other vectors of information allow a fast and deep flow of information, the accuracy of which is increasingly hard to verify [33]. Here, we rigorously controlled the quality of the information delivered to subjects in estimation tasks, by means of *virtual influencers*, i.e. virtual agents inserted into the sequence of estimations—unbeknownst to the subjects—and providing a value whose level of inaccuracy was controlled. We were thus able to precisely quantify the impact of information quality on individual and collective accuracy in those tasks.

We demonstrated that a proper normalization of estimates must take into account their dispersion, which gives the natural range of 'reasonable' estimates of a given quantity for a given group. This normalization process led to the conclusion that estimates follow a Laplace distribution when subjects have little prior information about a quantity to estimate. Early research showed that in many datasets, estimates X (i.e. deviations from the truth) were often close either to Gaussian distributed or to Laplace distributed [34,35]. Later work have encompassed Laplace and Gaussian distributions into a broader family of exponential distributions, the generalized normal distributions (GND) family [36,37], described by their centre m (often called *location* parameter), width σ (often called *scale* parameter) and tailedness η (often called *shape* parameter), which controls the thickness of the tails. The fatter the tails of a distribution, the higher the probability to find outliers (i.e. estimates that are very far from the distribution centre). More recent work has studied various datasets of estimates and forecasts in the light of GND, and showed that the tailedness of distributions ranged from $\eta = 1$ (Laplace distribution) to $\eta = 1.6$, η being equal to 2 for Gaussian distributions [38]. They concluded that most distributions of estimates for usual quantities are actually closer to Laplace distributions than to Gaussian distributions. This discussion can be related to the amount of prior information held by a group about a certain quantity. We found that when a quantity is 'hard' to estimate (i.e. low demonstrability, corresponding to a low amount of prior information about the quantity in the group), the expected distribution of estimates is very close to a Laplace distribution. When a quantity is 'easy' to estimate (i.e. high demonstrability, corresponding to a high amount of prior information about the quantity in

the group), few outliers are expected, such that the distribution of estimates could be expected to be closer to a Gaussian distribution. However, our results show that regardless of the questions' demonstrability, distributions of estimates are significantly closer to Laplace distributions than to Gaussian distributions when properly normalized, in agreement with [38]. In any case, we consider that future studies involving estimation tasks should apply the normalization procedure presented here when comparing and aggregating the estimates of different quantities, for which the width σ_{p} should be used to quantify their demonstrability.

We then studied the impact of incorrect information on individual and collective accuracy, and found that providing incorrect information that overestimates the true value can help a group perform better than providing the correct value itself, by partly compensating for the human underestimation bias. Moreover, collective and individual accuracy can improve after social influence over a surprisingly wide range of incorrect information. This counterintuitive result is a consequence of a large proportion of individuals compromising with the social information, i.e. partially following it. By doing so, subjects are able to benefit not only from relatively accurate social information but also from incorrect information that goes against their cognitive bias. Indeed, because of the human tendency to underestimate quantities, partially following an overestimation of the truth—even a large one—can bring second estimates closer to the truth, thus improving accuracy. However, incorrect social information can also harm accuracy if it amplifies the bias. This may be related to some deleterious effects of social information observed at times, for instance, how the spread of misinformation can deeply affect the behaviour of crowds as well as public opinion [39,40].

In a former study [18], we showed that adopting the social information was the best strategy in order to improve accuracy, if *virtual influencers* provide perfectly accurate information in the sequence of estimates. However, while adopting can lead to higher performance than compromising in this particular case, our results show that compromising offers more resilience when the information provided is potentially less accurate.

We also found that subjects were sensitive to the degree of incorrectness of the social information they received. They adapted their behaviour to the social information, by compromising more with the social information as it overestimated the truth more, and compromising less as it underestimated the truth more. This asymmetric strategy is surprisingly well adapted to counter the human underestimation bias. Indeed, as explained above, following (even partly) social information that underestimates the truth may increase the bias, while following social information that overestimates the truth may decrease it. Following less in the former case, and more in the latter is thus bound to increase the performance of groups. Former studies have already observed this subjects' tendency to rely more on social information that is higher than their personal estimate, than on social information that is lower, and showed that it had valuable consequences for collective performance in estimation tasks [29,30]. In [30], it is suggested that people can more easily assess the validity of small numbers compared to large numbers, because they have no direct experience with events related to those large numbers [41], and as a

consequence reject more often small numbers provided by the social information.

We then used a modified version of a model of collective estimation developed in [18]. The predictions of the model are in good agreement with the experimental data, and confirm that to optimize collective accuracy, social information must overestimate the truth further than merely compensating the initial collective bias, as most individuals only partly follow social information. In addition, the impact of the quality of information is not symmetric with respect to its optimum: collective accuracy can be improved by delivering incorrect information which overestimates the true value by up to several orders of magnitude, whereas it decays fast if the information delivered only slightly underestimates it. In other words, social information reinforcing the bias of the group has a strong negative impact on its accuracy.

Overall, we found that incorrect social information does not necessarily impair the collective wisdom of groups, and can even be used to counter some deleterious effects of cognitive biases. Individuals demonstrated an ability to discriminate the validity of the social information, depending on its distance from their personal estimates, and thus to benefit from accurate social information, while at the same time resisting inaccurate social information. These results suggest that groups of people may be more resilient to malicious information than is often thought, and at the same time that the negative effects of identified biases can be dampened by exchanging relevant social information, thus improving collective decisions.

References

1. Castells M. 2009 *The rise of the network society*. Oxford, UK: Wiley-Blackwell.

2. Schick AG, Gordon LA. 1990 Information overload: a temporal approach. *Account. Organ. Soc.* **15**, 199–220. (doi:10.1016/0361-3682(90)90005-F)

3. Klingberg T. 2009 *The overflowing brain: information overload and the limits of working memory*. Oxford, UK: Oxford University Press.

4. Viner K. 2016 How technology disrupted the truth. *Guardian* (London), July 12. Accessed 26 December 2016. See https://www.theguardian.com/media/2016/jul/12/how-technology-disrupted-the-truth.

5. Lewandowsky S, Ecker UK, Cook J. 2017 Beyond misinformation: understanding and coping with the 'post-truth' era. *J. Appl. Res. Mem. Cogn.* **6**, 353–369. (doi:10.1016/j.jarmac.2017.07.008)

6. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W. 2015 Science vs conspiracy: collective narratives in the age of misinformation. *PLoS ONE* **10**, e0118093. (doi:10.1371/journal.pone.0118093)

7. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W. 2016 The spreading of misinformation online. *Proc. Natl Acad. Sci. USA* **113**, 554–559. (doi:10.1073/pnas.1517441113)

8. Mønsted B, Sapiezynski P, Ferrara E, Lehmann S. 2017 Evidence of complex contagion of information in social media: an experiment using Twitter bots. *PLoS ONE* **12**, e0184148.

9. Vosoughi S, Roy D, Aral S. 2018 The spread of true and false news online. *Science* **359**, 1146–1151. (doi:10.1126/science.aap9559)

10. Moussaïd M, Kämmer JE, Analytis PP, Neth H. 2013 Social influence and the collective dynamics of opinion formation. *PLoS ONE* **8**, e78433. (doi:10.1371/journal.pone.0078433)

11. Mavrodiev P, Tessone CJ, Schweitzer F. 2013 Quantifying the effects of social influence. *Sci. Rep.* **3**, 1360. (doi:10.1038/srep01360)

12. Madirolas G, de Polavieja GG. 2015 Improving collective estimations using resistance to social influence. *PLoS Comput. Biol.* **11**, e1004594. (doi:10.1371/journal.pcbi.1004594)

13. Chacoma A, Zanette DH. 2015 Opinion formation by social influence: from experiments to modeling. *PLoS ONE* **10**, e0140406. (doi:10.1371/journal.pone.0140406)

14. Yaniv I. 2004 Receiving other people's advice: influence and benefit. *Organ. Behav. Hum. Decis. Process.* **93**, 1–13. (doi:10.1016/j.obhdp.2003.08.002)

15. Soll JB, Larrick RP. 2009 Strategies for revising judgement: how (and how well) people use others' opinions. *J. Exp. Psychol.: Learn. Mem. Cogn.* **35**, 780–805. (doi:10.1037/a0015145)

16. Luo Y, Iyengar G, Venkatasubramanian V. 2018 Social influence makes self-interested crowds smarter: an optimal control perspective. *IEEE Trans. Comput. Soc. Syst.* **5**, 200–209. (doi:10.1109/TCSS.2017.2780270)

17. Lorenz J, Rauhut H, Schweitzer F, Helbing D. 2011 How social influence can undermine the wisdom of crowd effect. *Proc. Natl Acad. Sci. USA* **108**, 9020–9025. (doi:10.1073/pnas.1008636108)

18. Jayles B, Kim HR, Escobedo R, Cezera S, Blanchet A, Kameda T, Sire C, Theraulaz G. 2017 How social information can improve estimation accuracy in human groups. *Proc. Natl Acad. Sci. USA* **114**, 12620–12625. (doi:10.1073/pnas.1703695114)

19. Vande Kerckhove C, Martin S, Gend P, Rentfrow PJ, Hendrickx JM, Blondel VD. 2016 Modelling influence and opinion evolution in online collective behaviour. *PLoS ONE* **11**, e0157685. (doi:10.1371/journal.pone.0157685)

20. Krueger LE. 1982 Single judgements of numerosity. *Percept. Psychophys.* **31**, 175–182. (doi:10.3758/BF03206218)

21. Krueger LE. 1989 Reconciling Fechner and Stevens: toward a unified psychophysical law. *Behav. Brain Sci.* **12**, 251–320. (doi:10.1017/S0140525X0004855X)

22. Izard V, Dehaene S. 2008 Calibrating the mental number line. *Cognition* **106**, 1221–1247. (doi:10.1016/j.cognition.2007.06.004)

23. Scheibehenne B. 2019 The psychophysics of number integration: evidence from the lab and from the field. *Decision* **6**, 61–76. (doi:10.1037/dec0000089)

24. Galton F. 1907 Vox populi. *Nature* **75**, 450–451. (doi:10.1038/075450a0)

25. Surowiecki J. 2005 *The Wisdom of Crowds*. New York, NY: Anchor Books.

26. Dehaene S. 2003 The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* **7**, 145–147. (doi:10.1016/S1364-6613(03)00055-X)

27. Dehaene S, Izard V, Spelke E, Pica P. 2008 Log or linear? Distinct intuitions of the number scale in western and Amazonian indigene cultures. *Science* **320**, 1217–1220. (doi:10.1126/science.1156540)

28. Ioannou CC, Madirolas G, Brammer FS, Rapley HA, de Polavieja GG. 2018 Adolescents show collective intelligence which can be driven by a geometric mean rule of thumb. *PLoS ONE* **13**, e0204462.

29. Kao AB, Berdahl AM, Hartnett AT, Lutz MJ, Bak-Coleman JB, Ioannou CC, Giam X, Couzin ID. 2018 Counteracting estimation bias and social influence to improve the wisdom of crowds. *J. R. Soc. Interface* **15**, 20180130. (doi:10.1098/rsif.2018.0130)

30. Jayles B, Kurvers RHJM. 2020 Debiasing the crowd: how to select social information for improving collective judgments? Preprint at https://doi.org/10.31234/osf.io/hn8rz.

31. Algom D, Wolf Y, Bergman B. 1985 Integration of stimulus dimensions in perception and memory: composition rules and psychophysical relations. *J. Exp. Psychol.: Gen.* **114**, 451–471. (doi:10.1037/0096-3445.114.4.451)

32. Chong SC, Evans KK. 2011 Distributed versus focused attention (count vs estimate). *WIREs Cogn. Sci.* **2**, 634–638. (doi:10.1002/wcs.136)

33. Centola D. 2018 *How behavior spreads: the science of complex contagions*. Princeton, NJ: Princeton University Press.

34. Laplace PS. 1774 Méemoire sur la probabilité des causes par les évènements. *Mémoires de l'Academie Royale des Sciences Presentés par Divers Savan* **6**, 621–656.

35. Wilson EB. 1923 First and second laws of error. *J. Am. Stat. Assoc.* **18**, 841–851. (doi:10.1080/01621459.1923.10502116)

36. Rider PR. 1924 A generalized law of error. *J. Am. Stat. Assoc.* **19**, 217–220. (doi:10.1080/01621459.1924.10502883)

37. Nadarajah S. 2005 A generalized normal distribution. *J. Appl. Stat.* **32**, 685–694. (doi:10.1080/02664760500079464)

38. Lobo MS, Yao D. 2010 Human judgement is heavy tailed: empirical evidence and implications for the aggregation of estimates and forecasts. INSEAD working paper No. 2010/48/DS.

39. Karlova NA, Fisher KE. 2013 A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Inf. Res.* **18**, 1–12.

40. Mocanu D, Rossi L, Zhang Q, Karsai M, Quattrociocchi W. 2015 Collective attention in the age of (mis)information. *Comp. for Hum. Learn. Behav. Collab. Social Mobile Netw. Era.* **51(Part B)**, 1198–1204.

41. Resnick I, Newcombe NS, Shipley TF. 2017 Dealing with big numbers: representation and understanding of magnitudes outside of human experience. *Cogn. Sci.* **41**, 1020–1041. (doi:10.1111/cogs.12388)