## Research

Author for correspondence:
Alan Novaes Tump
e-mail: tump@mpib-berlin.mpg.de

# Individuals fail to reap the collective benefits of diversity because of over-reliance on personal information

Alan Novaes Tump[1], Max Wolf[2], Jens Krause[2,3] and Ralf H. J. M. Kurvers[1,2]

[1]Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
[2]Biology and Ecology of Fishes, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany
[3]Faculty of Life Science, Albrecht Daniel Thaer-Institut, Humboldt University, Invalidenstraße 42, 10115 Berlin, Germany

Collective intelligence refers to the ability of groups to outperform individuals in solving cognitive tasks. Although numerous studies have demonstrated this effect, the mechanisms underlying collective intelligence remain poorly understood. Here, we investigate diversity in cue beliefs as a mechanism potentially promoting collective intelligence. In our experimental study, human groups observed a sequence of cartoon characters, and classified each character as a cooperator or defector based on informative and uninformative cues. Participants first made an individual decision. They then received social information consisting of their group members' decisions before making a second decision. Additionally, individuals reported their beliefs about the cues. Our results showed that individuals made better decisions after observing the decisions of others. Interestingly, individuals developed different cue beliefs, including many wrong ones, despite receiving identical information. Diversity in cue beliefs, however, did not predict collective improvement. Using simulations, we found that diverse collectives did provide better social information, but that individuals failed to reap those benefits because they relied too much on personal information. Our results highlight the potential of belief diversity for promoting collective intelligence, but suggest that this potential often remains unexploited because of over-reliance on personal information.

## 1. Introduction

Decision makers are constantly confronted with uncertainty about the true state of the world (e.g. is this skin lesion malignant, is this person telling the truth) [1,2]. To improve decision accuracy, individuals need to obtain relevant information, often in the form of environmental cues (e.g. the shape, colour and symmetry of a skin lesion). Next to personal sampling, another powerful approach for reducing uncertainty is to use information provided by others, so-called social information [3–8]. By pooling the capacities of multiple individuals (either by combining independent judgements, or direct interaction mechanisms), groups of decision makers can reduce uncertainty—an effect known as collective intelligence, swarm intelligence or collective cognition [9–17]. In a variety of domains including predator detection [18–21], medical decision-making [1,22,23], geopolitical forecasting [24], prediction markets [25] and lie detection [26], it has been shown that groups can outperform the average individual and sometimes even the best individual. Understanding the conditions increasing (or decreasing) the performance of collectives is thus of key importance for several applied contexts. For example, in diagnostic decision-making, it is important to understand how to compose groups of doctors [1,22,23], and how to combine doctors with machine learning algorithms [27] to achieve collective intelligence.

The extent to which combining individual information allows collectives to outperform individuals is crucially mediated by the level of independence as highlighted by several theoretical studies [28–32]. In these studies, independence is usually treated as a statistical concept, meaning that the likelihood of an individual being correct does not change depending on whether other members are correct [33]. Decisions of group members can be negatively correlated, which generally is thought to enhance the group's potential to achieve collective intelligence [28,32], uncorrelated, or positively correlated, which is thought to be detrimental for collective intelligence. Despite the importance of the nature of such correlations for collective intelligence, knowledge of the mechanisms determining the level of independence—and the subsequent scope for collective intelligence—in the real world are largely lacking.

Many decision-making environments involve multiple cues that can be either informative or uninformative about the true state of the world [34–36]. For example, a jury member judging whether a defendant is guilty can use multiple cues including the defendant's testimony, previous interactions with the criminal justice system, witness statements and other cues. Such decisions involving multiple cues can be challenging for individuals because the validities of the cues are often only partly known. In such situations, we may expect the benefits of collective intelligence to depend on individual differences in cue beliefs. If all group members have similar beliefs about the cues present (e.g. by specializing on the same cues), they are likely to arrive at similar individual decisions, leaving little scope for collective intelligence. However, if individuals have different beliefs about the cues present (e.g. by specializing on different cues), combining their decisions is expected to improve collective accuracy. Former theoretical studies indeed identified variation in weights assigned to cues as an important source of independence [37,38]. Given that many decision-making contexts require the integration of multiple cues, diversity in cue beliefs is thus a potentially widespread mechanism underlying collective intelligence, but few experimental studies have investigated this. Sorkin *et al.* [39] showed that providing group members with more similar information reduced group performance, through an increased positive correlation among members' judgements [39]. The provisioning of information in this experiment, however, was highly controlled, providing each group member with unique information. This leaves unresolved the questions if and how individuals in groups naturally diverge in cue beliefs when confronted with the same information (e.g. the same hearing in court), which is the focus of the current study.

To address how diversity in cue beliefs affects collective intelligence, we performed an experiment in which individuals in groups were confronted with a binary classification task containing multiple cues, the meaning of which was learned through feedback. Cues could be either informative or uninformative. Participants first made an individual decision. They then received social information consisting of the decisions of the other group members before making a second decision. In addition to their decisions, we quantified participants' cue beliefs to determine how differences in cue beliefs affected collective intelligence. For the reasons given above, we predicted that individuals in high-diversity groups would show a higher collective improvement than individuals in low-diversity groups. This result was expected

for both informative cues (which would allow the collective to capitalize on different relevant information) and uninformative cues (in which case, different wrong beliefs can cancel each other out at the collective level).

## 2. Material and methods

### 2.1. Experimental procedures

Participants were 293 students from the University of Bielefeld, split into 14 groups (mean group size: 21, range: 17–24). We obtained informed consent from each participant prior to the experiment. Each group completed a sequence of binary classification tasks, in which they decided whether a cartoon character projected on a white screen was a cooperator or defector (figure 1*b*). The cues were various items worn by the character (beard, belt, briefcase, hat, shoes, glasses, tie and umbrella). In total, there were eight items: four uninformative (i.e. equally present for both cooperators and defectors) and four informative (i.e. more often present for cooperators than defectors, or vice versa). Of the four informative cues, two were more indicative of a cooperator (cooperator cue) and two of a defector (defector cue). There were 40 rounds in total and in each round the character was wearing four cues. All eight cues were shown equally often (i.e. 20 times). The weights of each of the informative cues were the same, meaning that, for example, a character wearing two cooperator cues and one defector cue was deemed a cooperator. Each character had a dominant majority of informative cues; thus, a participant with full knowledge of the cue meanings and weights would be able to classify all cases correctly.

After observing a character, each participant had 5 s to make an individual decision (poll 1) using a wireless electronic keypad (Key Point Interactive Audience Software for Power Point, version: 2.0.142 Standard Edition). After poll 1, individuals were shown social information—specifically, the number of group members voting cooperator/defector. They then made a second decision on the same character (without seeing it again) (poll 2). After poll 2, the number of group members voting cooperator/defector in poll 2 was shown along with the correct answer. This procedure was repeated 40 times, providing participants with the opportunity to learn the cue validities. Participants were instructed to maximize the accuracy of both their decisions (i.e. poll 1 & poll 2). Occasionally, participants missed the opportunity to respond. These missing values were excluded from all analyses ($\approx$ 2.4%). By filling out a survey after each 10 rounds, participants reported their beliefs about each cue on a seven-point confidence scale ranging from $-3$ (very confident that the cue is indicative of a defector), to 0 (uninformative cue) to $+3$ (very confident that the cue is indicative of a cooperator). We ran four versions of the experiment, counterbalancing the meaning of the cues across the four versions. Specifically, each cue was used once as a cooperator cue, once as a defector cue and twice as an uninformative cue. Each group experienced only one of the four versions (i.e. cue validity was stable across the experiment within each group).

### 2.2. Accuracy of decisions and beliefs

To analyse whether participants learned to discriminate between cooperators and defectors over time, we used a generalized linear mixed model (GLMM) [40] with 'individual decision correct' (yes/no) as a response variable and round as a fixed effect. Next, we compared the accuracy of individual decisions (i.e. 'poll 1') to that of decisions with social information (i.e. 'poll 2') and to the accuracy of a majority rule. For the majority rule, we determined for each round (within each group) whether there was a majority of individual votes for cooperator or
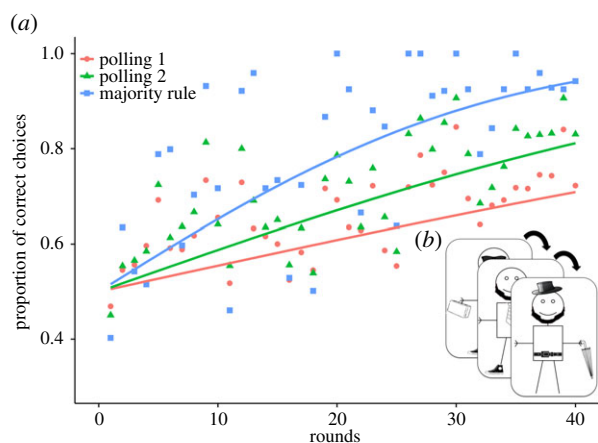
**Figure 1.** (a) Accuracy over the course of the experiment for individual decisions ('poll 1', red circles), decisions made after observing social information ('poll 2', green triangles) and when a simple majority rule was applied to the individual decisions (blue squares). All decision accuracies increased over time. Decisions made with social information were more accurate than individual decisions. However, applying a simple majority rule to the individual decisions resulted in even higher accuracy. (b) A sequence of three cartoon characters used in the experiment. Each character wore four of eight possible items (e.g. hat, beard, umbrella and belt). See Material and methods for more details.

defector. The majority rule took the option receiving most support as the final decision, and ties were broken randomly. The majority rule acts as a hypothetical benchmark, illustrating how much individuals could have improved if they had always followed the majority. We used a GLMM with 'decision correct' (yes/no) as a response variable and round in interaction with response type (poll 1, poll 2 and majority rule) as a fixed effect. For both GLMMs, we assumed that the decision accuracy at round 1 was 50% and therefore removed the intercept term. We included individual nested in group and experiment version (1–4) as random effects in both GLMMs, and used binomial errors and a probit-link function. Significance values were derived from the corresponding $z$-values and associated $p$-values, using the lme4 package in R [41].

To test whether participants became better at correctly classifying the cues, we used a GLMM (with probit-link function) with 'cue correctly classified' (yes/no) as a response variable, survey (1–4) as a fixed effect, and individual nested in group and experiment version (1–4) as random effects. We ran separate models for cooperator, defector and uninformative cues.

We used the weighted additive model (WADD) to verify that the reported cue beliefs correctly predicted the participants' responses [42]. WADD assumes that an individual sums the weights in favour of cooperator versus defector and chooses the option with the highest sum-score. Thereby, individuals are assumed to weigh their cues according to their reported subjective validity of the cue (i.e. the confidence in their cue beliefs). Each cue received a weight ranging from −3 (i.e. very confident that the cue indicates a defector) via 0 (i.e. uninformative cue) to +3 (i.e. very confident that the cue indicates a cooperator). Hence, a negative sum-score indicates that the participant has a net belief that the character is a defector, whereas a positive value indicates a net belief that the character is a cooperator. The higher the score, the stronger the belief. We tested whether these beliefs predicted participants' choices using a GLMM with 'decision' (cooperator/defector) as a response variable, 'sum-score' as a fixed effect, and individual nested in group and experiment version as random effects. Furthermore, we compared WADD with two other prominent decision strategies, namely the take-the-best (TTB) and the equal-weight (EQW)

strategy, using a maximum-likelihood approach following Pachur & Marinello [43]. Accordingly, we calculated for each strategy a $G^2$-value as a measure of the goodness of fit. The lower the $G^2$-value, the better the model fit. When applying TTB, individuals only rely on the subjectively most valid cue, whereas individuals using EQW count the cues in favour of each option (neglecting the weights) and take the option with the most cues favouring this option. If a strategy failed to make a prediction, we assumed a 50% probability of predicting cooperator or defector. Because cue validities were learned over time, this analysis considered only the cue beliefs reported after round 30 and decisions made from round 26 to 35.

## 2.3. Diversity

To analyse the effect of diversity in cue beliefs on collective improvement after observing social information, we calculated the intraclass correlation coefficient (ICC) for each group [44] as a measure of similarity in cue beliefs. We calculated the ICC separately for (i) informative and (ii) uninformative cues using the 'irr' package in R [45]. (i) For the informative cues, we treated the reported confidence scores as interval data ranging from −3 (high confidence in the wrong interpretation) via 0 (no cue usage) to +3 (high confidence in the correct interpretation). (ii) For uninformative cues, there is no correct interpretation (apart from not using them). A strong belief in an uninformative cue can drive the decision maker into making the correct decision if the belief coincides with the actual 'state of the world' (i.e. cooperator or defector) or, equally often, into making the wrong decision if the belief does not coincide with the 'state of the world'. Consequently, a strong belief in an uninformative cue is acting like a strong belief in two informative cues, with one being correct and one being wrong. Therefore, we treated each uninformative cue like two separate cues, each reflecting one of the two possible states (i.e. correct/wrong belief). More formally, we duplicated each cue and treated one duplicate as if the belief that the cue indicates a cooperator is correct and its counterpart as if this belief is wrong. The ICC ranges from negative (individuals with diverse beliefs), via 0, to positive (individuals with similar beliefs). Because cue validities needed to be learned over time, we limited all analyses on cue beliefs to decisions made in rounds 26 to 35 and to cue beliefs reported after round 30.

We used a linear mixed model to test the effect of collective cue diversity on improvement in individuals' decision accuracy after observing the social information. We used the absolute increase in decision accuracy at poll 2 as a response variable, including the similarity measures, average individual performance and group size as fixed effects. Significance values were derived from the corresponding $t$ values, using the lme4 package in R [41].

## 2.4. Model simulations

The extent to which individuals in groups exploit any potential benefits of collective diversity is expected to depend on (i) group size and (ii) their willingness to incorporate social information. We used model simulations to investigate these relationships further. From our data, we randomly sampled groups of $n$ (3, 7, 11 and 19) individuals exposed to the same version of the experiment, determined the diversity in their informative and uninformative cue beliefs (focusing again on rounds 26 to 35), and studied their collective improvement at different sensitivities to social information. To this end, we varied the voting quota [29], defined as the minimal proportion of individuals with an opposing decision needed to trigger a change in decision. If the voting quota is not reached, a voter will keep its initial decision. We varied the voting quota from 50% (i.e. majority rule) to 80%, with 10% increments. The higher the voting quota, the less sensitive individuals are to
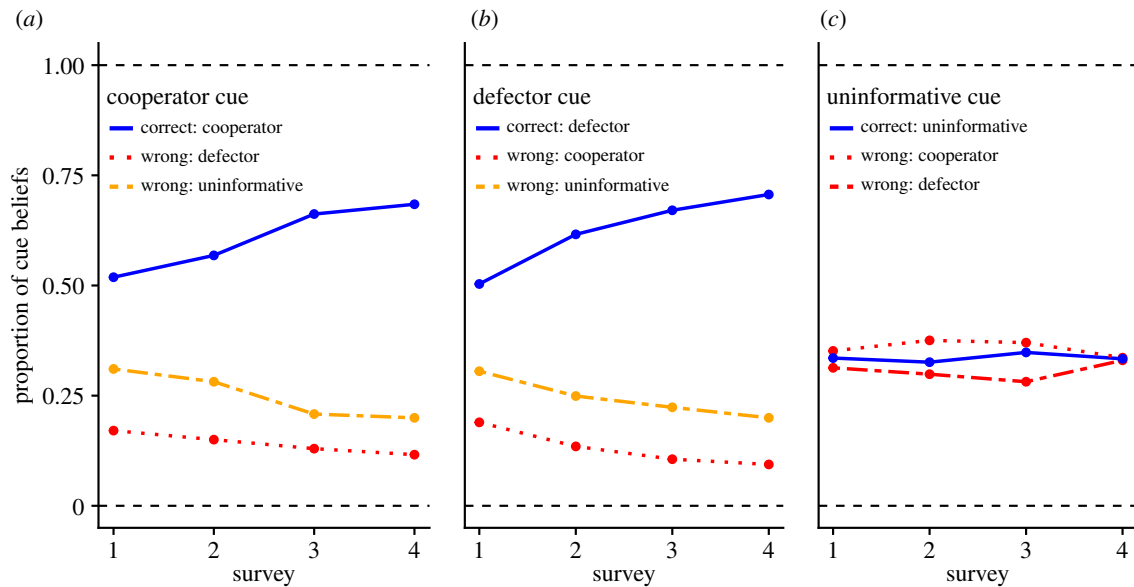
**Figure 2.** Proportion of correct and wrong classifications for (*a*) cooperator, (*b*) defector and (*c*) uninformative cues over the course of the experiment. Participants became better at correctly classifying the cooperator and defector cues, but not the uninformative cues.

social information. For each simulation run, we calculated a group's collective improvement as the difference between the collective performance and the average performance of individual group members. Importantly, these simulations were based on participants' actual cue beliefs and decisions in poll 1. We ran 1 000 000 simulation repetitions for each combination of group size and voting quota.

## 2.5. Empirical switching thresholds

Because we expected the effects of diversity on collective improvement to depend on the individual's willingness to incorporate social information (see results of model simulations below), we first investigated how individuals weighed personal and social information by studying their empirical switching thresholds over the course of the experiment. Specifically, we studied the likelihood that individuals changed their decision at poll 2 as a function of the proportion of group members making an opposing decision. We used GLMMs with 'decision switched' from poll 1 to poll 2 (yes/no) as a response variable and included the 'proportion of group members making the opposing decision from the focal individual' as a fixed effect. This fixed effect was included alone, in interaction with round, and in interaction with the past performance of the focal individual. We ran two separate GLMMs. In the first analysis, both round and performance were treated as dichotomous variables. For rounds, we compared the first with the second half of the experiment. For past performance, we compared individuals whose past performance was above the group average with individuals whose past performance was below the group average. In the second analysis, both round and past performance were included as continuous variables (with past performance for each individual being calculated as their average individual decision accuracy in all preceding rounds). From the model estimations, we determined the 'switching threshold', defined as the proportion of group members with an opposing decision required to induce a 50% likelihood of the focal individual switching his decision. The higher this threshold, the less sensitive individuals were to social information. Note that voting quota and switching thresholds are related in the sense that both parameters describe the sensitivity to social information. However, the first describes a strict decision rule, while the second describes probabilistic behaviour. In both models, individual nested in group and experiment version were fitted as random effects and a probit-link function was used. In the

second analysis, we additionally used the BOBYQA optimizer from the lme4 package in R because the default optimizer failed to converge.

# 3. Results

## 3.1. Accuracy of decisions and beliefs

The accuracy of individual decisions increased significantly over the course of the experiment from approx. 50% to 70% (estimate [est] $\pm$ s.e. $= 0.014 \pm 0.001$; $z = 13.18$; $p < 0.001$; figure 1*a*). Decisions made with social information (poll 2) were significantly more accurate than individual decisions (interaction: est $\pm$ s.e. $= -0.008 \pm 0.001$; $z = -10.62$; $p < 0.001$; figure 1*a*). However, decisions made under the majority rule were even more accurate than decisions made with social information (interaction: est $\pm$ s.e. $= 0.017 \pm 0.001$; $z = 19.2$; $p < 0.001$; figure 1*a*).

Participants became better at correctly classifying the informative cues over the course of the experiment (cooperator cues: est $\pm$ s.e. $= -0.194 \pm 0.026$; $z = 7.35$; $p < 0.001$; defector cues: est $\pm$ s.e. $= 0.222 \pm 0.027$; $z = 8.27$; $p < 0.001$; figure 2*a*,*b*) but not at correctly classifying the uninformative cues (est $\pm$ s.e. $= 0.018 \pm 0.001$; $z = 0.26$; $p = 0.799$; figure 2*c*). At each survey, participants identified only approximately 1/3 of the uninformative cues correctly; 1/3 were erroneously classified as cooperator cues and 1/3 as defector cues, indicating that participants incorporated many false beliefs. Individuals tended to classify certain uninformative cues more often as cooperator or defector. Glasses and beard, for example, were more often associated with a defector; tie and umbrella, in contrast, more often with a cooperator (electronic supplementary material, figure S1).

Cue beliefs, in turn, were good predictors of participants' decisions: the more strongly an individual believed that the cues present in a round were indicative of a cooperator/ defector, the more likely they were to choose cooperator/ defector (est $\pm$ s.e. $= 0.248 \pm 0.011$; $z = 20.65$; $p < 0.001$; electronic supplementary material, figure S2). In addition, the decisions were best described by WADD (G$^2$: 3,239) compared to TTB (G$^2$: 3,355) or EQW (G$^2$: 3,335).
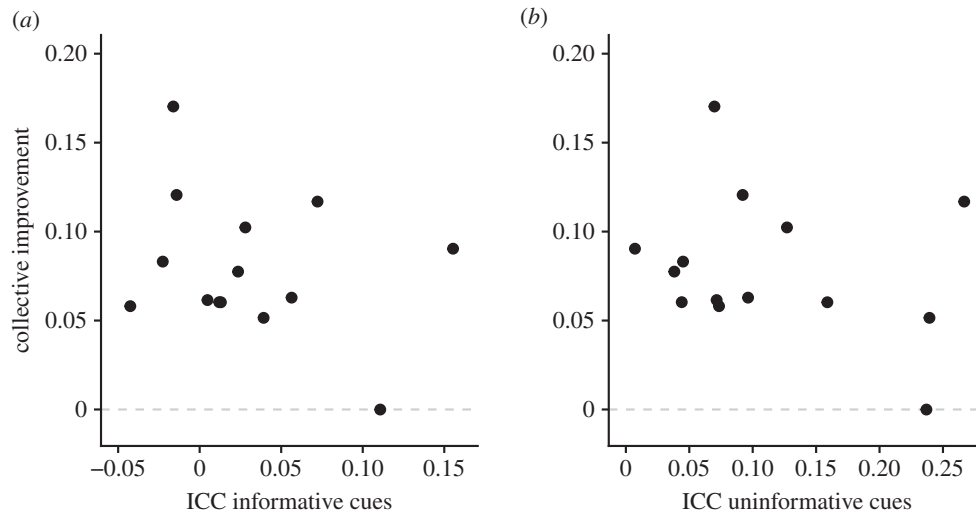
**Figure 3.** There was no significant effect of similarity in cue beliefs (i.e. the ICC) on collective improvement (i.e. the accuracy difference between poll 2 and poll 1) for (*a*) informative or (*b*) uninformative cues.

## 3.2. Diversity

Despite observing the same information, individuals developed different beliefs about the cues (electronic supplementary material, figures S3–S4). Contrary to our expectations, diversity in cue beliefs had no significant effect on collective improvement (figure 3). This was true for informative (est $\pm$ s.e. $= -0.511 \pm 0.23$; $t = -2.191$; $p = 0.562$) and uninformative (est $\pm$ s.e. $= 0.022 \pm 0.15$; $t = 0.15$.; $p = 0.884$) cues. There was also no significant effect of group size on collective improvement.

## 3.3. Model simulations

Figure 4 shows how group size and voting quota impacted the extent to which individuals benefited from diversity in cue beliefs. For both informative and uninformative cues, the effect of diversity on collective improvement depended on the voting quota. When individuals had a low voting quota of 50% (i.e. changed their decision as soon as they were in the minority), more diverse groups were able to improve more than less diverse groups. Likewise, larger groups were able to improve more than smaller groups. However, when the voting quota was increased, making individuals less sensitive to social information, three different effects emerged: (i) overall collective improvement decreased, (ii) the relative benefit of large groups decreased and (iii) the effect of diversity on collective improvement was substantially weakened. Importantly, these effects occurred for both informative and uninformative cues, despite uninformative cues being irrelevant for discriminating between cooperators and defectors. Electronic supplementary material, figure S6 shows how the importance of diversity interacted with average individual performance of group members. In short, the mediating role of diversity on collective improvement was highest when the average individual accuracy was between 60% and 90%.

## 3.4. Empirical switching thresholds

The likelihood of individuals to change their decision between poll 1 and poll 2 increased in a sigmoidal manner with an increasing 'proportion of opposing decisions' (figure 5a,b). There was a significant interaction effect of

rounds and 'proportion of opposing decisions' on switching threshold (est $\pm$ s.e. $= 0.351 \pm 0.073$; $z = 4.78$; $p < 0.001$; figure 5a,b): participants used lower switching thresholds in the second than in the first half of the experiment, indicating that individuals became more sensitive to social information. There was also a significant interaction effect of individuals' past performance and 'proportion of opposing decisions' on the switching threshold (est $\pm$ s.e. $= 0.381 \pm 0.088$; $z = 4.33$; $p < 0.001$; figure 5a,b): relatively poor performers used lower switching thresholds and were thus more sensitive to social information than relatively good performers. Figure 5c shows the model predictions when both rounds and individuals' past performance were treated as continuous variables, confirming that individuals lowered their switching thresholds over time (est $\pm$ s.e. $= 0.039 \pm 0.006$; $z = 6.43$; $p < 0.001$), and that low-performing individuals used lower switching thresholds (est $\pm$ s.e. $= -1.454 \pm 0.415$; $z = -3.5$; $p < 0.001$). Importantly, the estimated empirical switching threshold over the entire experiment was between 75% and 90%, indicating that a large majority with an opposing decision was needed before individuals changed their mind, and illustrating the relative reluctance to incorporate social information.

## 4. Discussion

Our results show that individuals experiencing the same environment can develop different beliefs about the cue validities. Given that individuals in many real-world situations do not experience the same environment, but are exposed to (partly) different sources of information, this 'diversification' of cue beliefs is likely to be even more pronounced in more realistic decision-making tasks. Although individuals generally profited from the decisions of others and diversity in cue beliefs directly affected the correlation among votes (electronic supplementary material, figure S5), the level of diversity did not predict the collective improvement in accuracy. Our model simulations provide a clue as to why this effect was absent: the mediating role of diversity on collective improvement depended on how sensitive individuals were to social information, with diversity being important when individuals used relatively low switching thresholds. However,
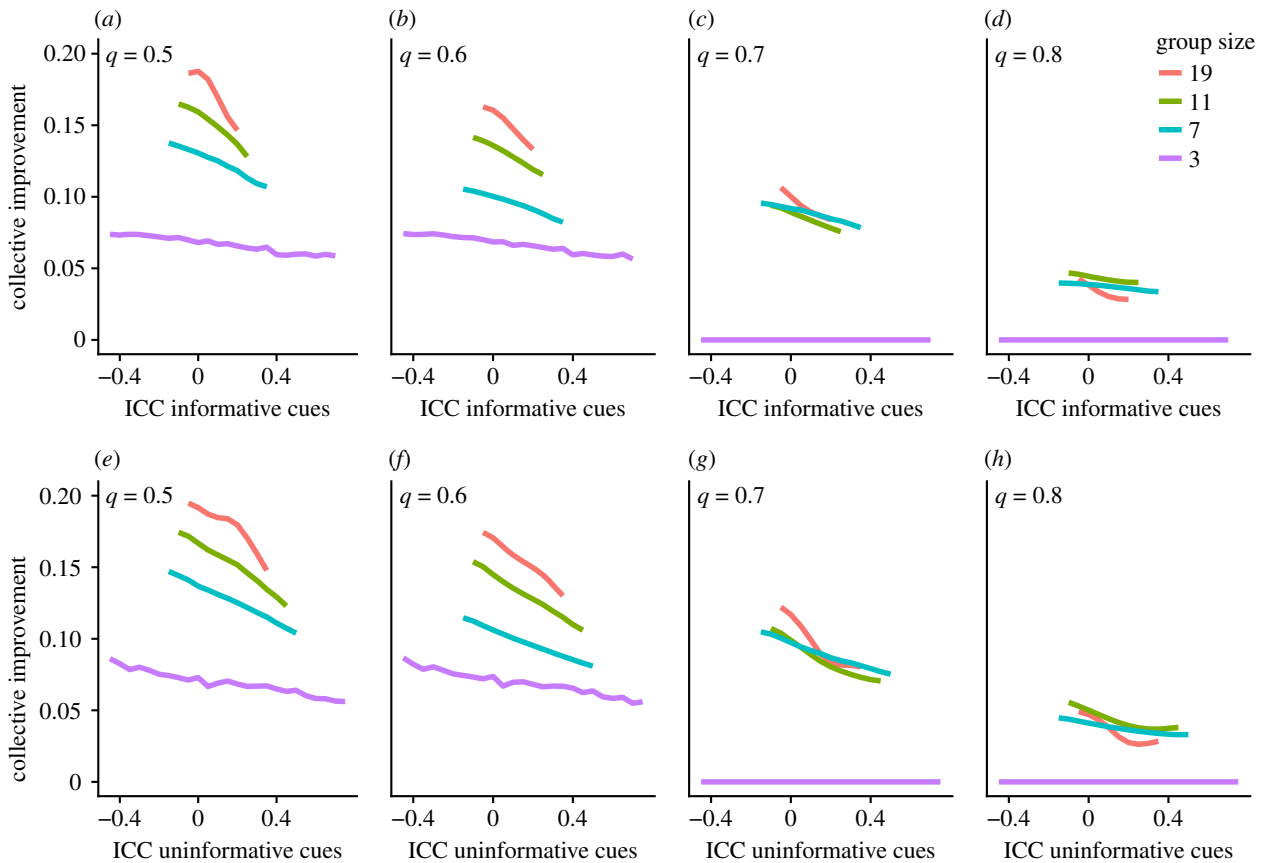
**Figure 4.** Collective improvement of resampled groups as a function of similarity in cue beliefs for (a–d) informative and (e–h) uninformative cues and different group sizes. Shown are the results with a voting quota (q) of (a,e) 50%, (b,f) 60%, (c,g) 70% and (d,h) 80%. At a low voting quota (a,e), individuals improved most when in groups with low ICC (i.e. high diversity in cue beliefs) and when in large groups. At a high voting quota (d,h), individuals generally improved less after observing social information (because they were unwilling to incorporate that information) and the effect of diversity disappeared (i.e. lines tended to become flat).

this effect disappeared when individuals became less sensitive to social information and required a larger majority before switching their initial decision. Interestingly, these effects applied to diversity in beliefs about informative and uninformative cues alike (figure 4).

Inspection of the empirical switching thresholds employed in the experiment revealed that individuals used relatively high switching thresholds. Although the switching threshold declined over the course of the experiment (showing that individuals became more willing to respond to social information), it remained, on average, above 75% even in the second half of the experiment (figure 5b). Individuals using such high thresholds are predicted to benefit only weakly, if at all, from diversity in cue beliefs. Although diversity is generally considered to be crucial in fostering collective intelligence [28–31,39,46], the general reluctance to using social information (see also below) might reduce its importance in natural decision environments. Interestingly, methods that algorithmically combine the decisions of different decision makers [22,23] will not suffer the same consequences, as the optimal thresholds can simply be set by an external decision maker.

Although individuals did not capitalize on the potential benefits of diversity, they did develop different beliefs while experiencing identical environments. The exact mechanisms underlying this diversity in our experiment are not yet clear. Individuals may have decided to focus on different cues and, as a result, generated different hypotheses and beliefs about them. Research on group memory has shown that group members making the same observation store

different information, making the information only partly redundant [47]. More mechanistic approaches, such as eye-tracking, could provide more insights into the development of (and individual differences in) cue beliefs [48]. Furthermore, a better understanding of how differences in cue sequences generate differences in beliefs could shed light on how individuals navigate such environments, and how different training regimes could be implemented to foster diversity between decision makers. Future studies could also investigate the importance of diversity in cue beliefs in interacting groups. For example, do individuals in interactive groups naturally specialize in different cues, thereby reaping increased collective benefits? Future work could also investigate and test our predictions in real-world decision-making contexts—for example, medical diagnostics [1], lie detection [26] or any context in which multiple cues need to be integrated to make a final decision.

Our model simulations suggest that diversity in uninformative cues can also be important for collective improvements. Intriguingly, individuals did not become better at correctly interpreting uninformative cues over time. Although each uninformative cue appeared equally often with a cooperator and defector, two-thirds of individuals considered them to be indicative of either a cooperator or a defector at any time in the experiment. Previous research has shown that individuals frequently observe patterns in random sequences of events [49–51]. Participants also showed propensities to associate certain uninformative items more often with cooperator or defector (electronic supplementary material, figure S1), which could increase the
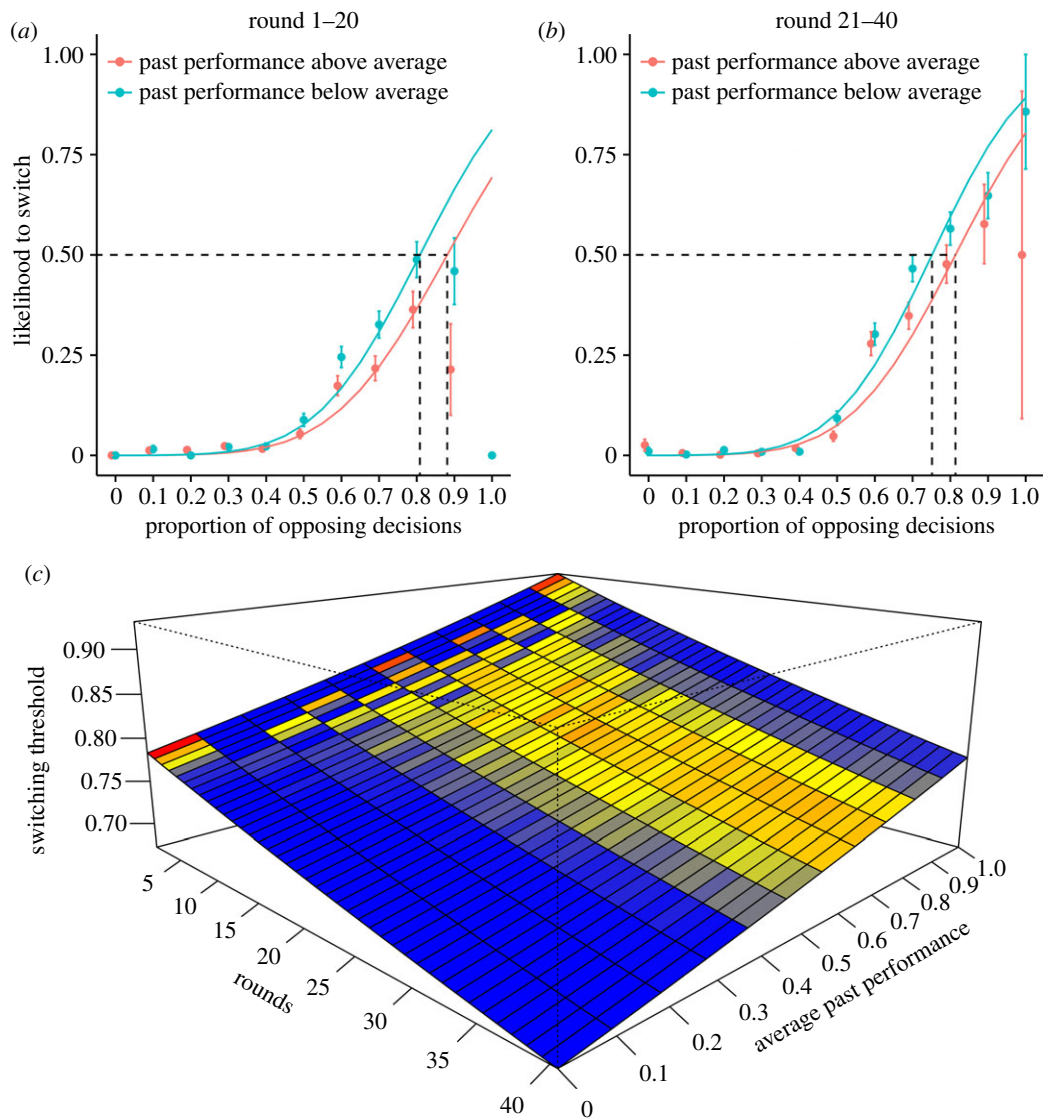
**Figure 5.** (*a,b*) Relationship between the proportion of opposing decisions in the group and the individual likelihood of switching between poll 1 and poll 2 (i.e. after observing the social information) for (*a*) the first 20 rounds and (*b*) the last 20 rounds. Red lines represent individuals performing above the group average; blue lines represent individuals performing below the group average. Individuals generally followed a sigmoidal curve in reacting to the decisions of others. Switching thresholds were estimated as the proportion of opposing votes needed to induce a 50% likelihood of an individual changing its decision. The dashed lines indicate the switching thresholds for high- and low-performing individuals. Higher-performing individuals used higher switching thresholds, and switching thresholds decreased over the course of the experiment. Error bars represent standard error. Note that the proportions are binned for visualization purposes. (*c*) Relationship between the switching threshold, rounds and average past performance, with rounds and average past performance expressed on a continuous scale. The switching threshold decreased over time and was higher for high-performing individuals. Colours indicate the frequency of individuals, with blue indicating low frequency, yellow medium frequency and red high frequency.

likelihood of retaining false beliefs about uninformative cues. Our simulations show that whether or not such false beliefs impair collective performance depends on whether individuals develop similar false beliefs, in which case collective improvements are hampered, or different false beliefs, which can be cancelled out at the collective level.

The simulation results further suggest that any positive effect of diversity on collective performance will depend on the average individual performance of a group: only when the average individual accuracy fell between approximately 60% and 90%, was there a positive effect of diversity on collective performance (electronic supplementary material, figure S6). When the average individual performance exceeded 90%, there was little room for collective improvement; when it fell below 60%, there was probably insufficient reliable information for groups to improve. Most groups in our experiment fell within this beneficial

range; the same is likely to apply in most real-world decision-making contexts. Moreover, we found a positive effect of combining diverse decisions only when the average individual performance was above 50%. This classical finding is consistent with Condorcet's jury theorem, which predicts that the accuracy of a majority vote increases with group size but only if the average individual accuracy is above 50% [52–54]. Increasing group size is predicted to worsen collective performance if the average individual accuracy is below 50%. Diversity is predicted to determine the strength of both effects: above 50% individual accuracy, the higher the diversity, the stronger the collective gains; below 50% individual accuracy, the higher the diversity, the stronger the collective losses [55].

One of the remaining questions is why individuals were relatively reluctant to incorporate social information and instead used suboptimal switching thresholds. For example,

applying a simple majority rule to the group members' first decisions resulted in decisions that were substantially better than participants' second decisions. Such suboptimal use of social information has been repeatedly reported [56–60] and various reasons for it have been suggested. One is that decision makers often mistrust others' opinions, a phenomenon called 'egocentric discounting' [61]. Another is that people underestimate the effectiveness of combining decisions [62]. Yet overvaluing personal information can be adaptive in some environments. In highly fluctuating environments, for example, social information can quickly become outdated, making personal information more reliable [63,64]. Moreover, over-reliance on social information can initiate false informational cascades [15,65].

Despite suboptimal use of social information, we found some hints of context-dependent threshold adjustments. First, individuals generally lowered their thresholds over time, suggesting that they first need to learn the validity of the social signal. This validity could be learned because correct feedback was provided after each round. Second, individuals adjusted their switching thresholds to their own performance: high performers used relatively high switching thresholds, whereas low performers used lower switching thresholds. This result corroborates the general finding that individuals with poorer personal information rely more on social information (copy when uncertain) [5,8,66–68]. Previous research has also shown that individuals adjust their switching thresholds to the quality of social information (i.e. the group performance) [69]. Most probably, individuals in groups monitor their own performance and the group performance and incorporate both when deciding how to respond to social information. In our study, individuals were provided with aggregated social information of the whole group (global information). Future work could further investigate how different social network structures and non-anonymous information lead to information use of (preferred) group members.

To conclude, our findings show that individuals developed different beliefs about cues despite observing the same information. Such diversity in beliefs about both informative and uninformative cues has the potential to promote collective intelligence. Individuals, however, did not benefit from this diversity because they over-relied on personal information. Because undervaluing social information is a widely observed phenomenon, we postulate that the potential collective benefits of diversity may frequently remain unexploited.

# References

1. Kurvers RHJM, Krause J, Argenziano G, Zalaudek I, Wolf M. 2015 Detection accuracy of collective intelligence assessments for skin cancer diagnosis. JAMA Dermatology 151, 1346–1353. (doi:10.1001/jamadermatol.2015.3149)

2. Ekman P, O'Sullivan M. 1991 Who can catch a liar? Am. Psychol. 46, 913–920. (doi:10.1037/0003-066X.46.9.913)

3. Danchin E, Giraldeau L-A, Valone TJ, Wagner RH. 2004 Public information: from nosy neighbors to cultural evolution. Science 305, 487–491. (doi:10.1126/science.1098254)

4. Galef BG, Laland KN. 2005 Social learning in animals: empirical studies and theoretical models. Bioscience 55, 489–499. (doi:10.1641/0006-3568(2005)055[0489:SLIAES]2.0.CO;2)

5. Morgan TJH, Rendell LE, Ehn M, Hoppitt W, Laland KN. 2012 The evolutionary basis of human social learning. Proc. R. Soc. B 279, 653–662. (doi:10.1098/rspb.2011.1172)

6. Seppänen JT, Forsman JT, Monkkönen M, Thomson RL. 2007 Social information use is a process across time, space, and ecology, reaching heterospecifics. Ecology 88, 1622–1633. (doi:10.1890/06-1757.1)

7. Valone TJ, Templeton JJ. 2002 Public information for the assessment of quality: a widespread social phenomenon. Phil. Trans. R. Soc. Lond. B 357, 1549–1557. (doi:10.1098/rstb.2002.1064)

8. Rendell L, Fogarty L, Hoppitt W, Morgan T. 2011 Cognitive culture: theoretical and empirical insights into social learning strategies. Trends Cogn. 15, 68–76. (doi:10.1016/j.tics.2010.12.002)

9. Krause J, Ruxton GD, Krause S. 2010 Swarm intelligence in animals and humans. Trends Ecol. Evol. 25, 28–34. (doi:10.1016/j.tree.2009.06.016)

10. Surowiecki J. 2004 The wisdom of crowds. New York, NY: Doubleday.

11. Couzin ID, Krause J, Franks NR, Levin SA. 2005 Effective leadership and decision-making in animal groups on the move. Nature 433, 513–516. (doi:10.1038/nature03236)

12. Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW. 2010 Evidence for a collective intelligence factor in the performance of human groups. Science 330, 686–688. (doi:10.1126/science.1193147)

13. Webb B. 2002 Swarm intelligence: from natural to artificial systems. Conn. Sci. 14, 163–164. (doi:10.1080/09540090210144948)

14. Conradt L, List C. 2009 Group decisions in humans and animals: a survey. Phil. Trans. R. Soc. B 364, 719–742. (doi:10.1098/rstb.2008.0276)

15. Sumpter DJT, Pratt SC. 2009 Quorum responses and consensus decision making. Phil. Trans. R. Soc. B 364, 743–753. (doi:10.1098/rstb.2008.0204)

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

17. Woolley AW, Aggarwal I, Malone TW. 2015 Collective intelligence and group performance. Curr. Dir. Psychol. Sci. 24, 420–424. (doi:10.1177/0963721415599543)

18. Wolf M, Kurvers RHJM, Ward AJW, Krause S, Krause J. 2013 Accurate decisions in an uncertain world: collective cognition increases true positives while decreasing false positives. Proc. R. Soc. B 280, 20122777. (doi:10.1098/rspb.2012.2777)

19. Clément RJG, Wolf M, Snijders L, Krause J, Kurvers RHJM. 2015 Information transmission via movement behaviour improves decision accuracy in human groups. Anim. Behav. 105, 85–93. (doi:10.1016/j.anbehav.2015.04.004)

20. Ward AJW, Herbert-Read JE, Sumpter DJT, Krause J. 2011 Fast and accurate decisions through collective vigilance in fish shoals. Proc. Natl Acad. Sci USA 108, 2312–2315. (doi:10.1073/pnas.1007102108)

21. Ward AJW, Sumpter DJT, Couzin ID, Hart PJB, Krause J. 2008 Quorum decision-making facilitates information transfer in fish shoals. Proc. Natl Acad. Sci. USA 105, 6948–6953. (doi:10.1073/pnas.0710344105)

22. Kurvers RHJM, Herzog SM, Hertwig R, Krause J, Carney PA, Bogart A, Argenziano G, Zalaudek I, Wolf M. 2016 Boosting medical diagnostics by

pooling independent judgments. *Proc. Natl Acad. Sci. USA* **113**, 201601827. (doi:10.1073/pnas. 1601827113)

23. Wolf M, Krause J, Carney PA, Bogart A, Kurvers RHJM. 2015 Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PLoS ONE* **10**, e0134269. (doi:10.1371/journal.pone.0134269)

24. Mellers B *et al.* 2014 Psychological strategies for winning a geopolitical forecasting tournament. *Psychol. Sci.* **25**, 1106–1115. (doi:10.1177/0956797614524255)

25. Arrow KJ *et al.* 2008 Economics: the promise of prediction markets. *Science* **320**, 877–878. (doi:10.1126/science.1157679)

26. Klein N, Epley N. 2015 Group discussion improves lie detection. *Proc. Natl Acad. Sci. USA* **112**, 7460–7465. (doi:10.1073/pnas.1504048112)

27. Tacchella A, Romano S, Ferraldeschi M, Salvetti M, Zaccaria A, Crisanti A, Grassi F. 2017 Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: a proof-of-principle study. *F1000Research* **6**, 2172. (doi:10.12688/f1000research.13114.1)

28. Kaniovski S. 2008 Aggregation of correlated votes and Condorcet's jury theorem. *Theory Decis.* **69**, 453–468. (doi:10.1007/s11238-008-9120-4)

29. Kaniovski S, Zaigraev A. 2011 Optimal jury design for homogeneous juries with correlated votes. *Theory Decis.* **71**, 439–459. (doi:10.1007/s11238-009-9170-2)

30. Ladha KK. 1992 The Condorcet jury theorem, free speech, and correlated votes. *Am. J. Pol. Sci.* **36**, 617–634. (doi:10.2307/2111584)

31. Ladha KK. 1995 Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *J. Econ. Behav. Organ.* **26**, 353–372. (doi:10.1016/0167-2681(94)00068-P)

32. Laan A, Madirolas G, de Polavieja GG. 2017 Rescuing collective wisdom when the average group opinion is wrong. *Front. Robot. AI* **4**, 56. (doi:10.3389/frobt.2017.00056)

33. Luan S, Katsikopoulos K V., Reimer T. 2012 The 'less-is-more' effect in group decision making. In *Simple heuristics in a social world* (eds R Hertwig, U Hoffrage, the ABC Research Group), pp. 293–318. New York, NY: Oxford University Press.

34. Dall SRX, Giraldeau L-A, Olsson O, McNamara JM, Stephens DW. 2005 Information and its use by animals in evolutionary ecology. *Trends Ecol. Evol.* **20**, 187–193. (doi:10.1016/j.tree.2005.01.010)

35. Candolin U. 2003 The use of multiple cues in mate choice. *Biol. Rev.* **78**, 575–595. (doi:10.1017/S1464793103006158)

36. Kao AB, Couzin ID. 2014 Decision accuracy in complex environments is often maximized by small group sizes. *Proc. R. Soc. B* **281**, 20133305. (doi:10.1098/rspb.2013.3305 or)

37. Keuschnigg M, Ganser C. 2016 Crowd wisdom relies on agents' ability in small groups with a voting aggregation rule. *Manage. Sci.* **1909**, 818–828. (doi:10.1287/mnsc.2015.2364)

38. Broomell SB, Budescu D V. 2009 Why are experts correlated? Decomposing correlations between judges. *Psychometrika* **74**, 531–553. (doi:10.1007/S11336-009-9118-Z)

39. Sorkin R, Hays C, West R. 2001 Signal-detection analysis of group decision making. *Psychol. Rev.* **108**, 183–203. (doi:30.1037/0033-295X)

40. Stroup WW. 2013 *Generalized linear mixed models: modern concepts, methods, and applications*. Boca Raton, FL: CRC Press.

41. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Grothendieck G, Green P. 2015 lme4: linear mixed-effects models using eigen and S4. *J. Stat. Softw.* **67**, 148. (doi:10.18637/jss.v067.i01)

42. Payne J, Bettman J, Johnson E. 1993 *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.

43. Pachur T, Marinello G. 2013 Expert intuitions: how to model the decision strategies of airport customs officers? *Acta Psychol.* **144**, 97–103. (doi:10.1016/J.ACTPSY.2013.05.003)

44. Shrout PE, Fleiss JL. 1979 Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, 420–428. (doi:10.1037/0033-2909.86.2.420)

45. Gamer M, Lemon J, Fellows I, Singh P. 2012 Various coefficients of interrater reliability and agreement. R package version 0.84. See https://cran.r-project.org/package=irr%0A

46. Luan S, Katsikopoulos K V., Reimer T. 2012 When does diversity trump ability (and vice versa) in group decision making? A simulation study. *PLoS ONE* **7**, e31043. (doi:10.1371/journal.pone.0031043)

47. Hinsz VB. 1990 Cognitive and consensus processes in group recognition memory performance. *J. Pers. Soc. Psychol.* **59**, 705–718. (doi:10.1037/0022-3514.59.4.705)

48. Holmqvist K, Nyström M, Andersson R, Dewhurst R. 2011 *Eye tracking: a comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

49. Ayton P, Fischer I. 2004 The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? *Mem. Cognit.* **32**, 1369–1378. (doi:10.3758/BF03206327)

50. Bressan P. 2002 The connection between random sequences, everyday coincidences, and belief in the paranormal. *Appl. Cogn. Psychol.* **16**, 17–34. (doi:10.1002/acp.754)

51. Falk R, Konold C. 1997 Making sense of randomness: implicit encoding as a basis for judgment. *Psychol. Rev.* **104**, 301–318. (doi:10.1037/0033-295X.104.2.301)

52. Boland PJP. 1989 Majority systems and the Condorcet jury theorem. *J. R. Stat. Soc. Ser. D* **38**, 181–189. (doi:10.2307/2348873)

53. List C. 2004 Democracy in animal groups: a political science perspective. *Trends Ecol. Evol.* **19**, 168–169. (doi:10.1016/j.tree.2004.02.004)

54. Condorcet MJAN. 1785 *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris, France: Imprimerie Royale.

55. King AJ, Cowlishaw G. 2007 When to use social information: the advantage of large group size in individual decision making. *Biol. Lett* **3**, 137–139. (doi:10.1098/rsbl.2007.0017)

56. Yaniv I, Milyavsky M. 2007 Using advice from multiple sources to revise and improve judgments. *Organ. Behav. Hum. Decis. Process.* **103**, 104–120. (doi:10.1016/j.obhdp.2006.05.006)

57. Mannes AE. 2009 Are we wise about the wisdom of crowds? The use of group judgments in belief revision. *Manage. Sci.* **55**, 1267–1279. (doi:10.1287/mnsc.1090.1031)

58. 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)

59. Mesoudi A. 2011 An experimental comparison of human social learning strategies: payoff-biased social learning is adaptive but underused. *Evol. Hum. Behav.* **32**, 334–342. (doi:10.1016/j.evolhumbehav.2010.12.001)

60. Jayles B, Kim H-R, 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**, 12 620–12 625. (doi:10.1073/pnas.1703695114)

61. Yaniv I, Kleinberger E. 2000 Advice taking in decision making: egocentric discounting and reputation formation. *Organ. Behav. Hum. Decis. Process.* **83**, 260–281. (doi:10.1006/obhd.2000.2909)

62. Larrick RP, Soll JB. 2006 Intuitions about combining opinions: misappreciation of the averaging principle. *Manage. Sci.* **52**, 111–127. (doi:10.1287/mnsc.l060.0518)

63. Feldman MW, Aoki K, Kumm J. 1996 Individual versus social learning: evolutionary analysis in a fluctuating environment. *Anthropol. Sci.* **104**, 209–231. (doi:10.1537/ase.104.209)

64. Henrich J, Boyd R. 1998 The evolution of conformist transmission and the emergence of between-group differences. *Evol. Hum. Behav.* **19**, 215–241. (doi:10.1016/S1090-5138(98)00018-X)

65. Giraldeau LA, Valone TJ, Templeton JJ. 2002 Potential disadvantages of using socially acquired information. *Phil. Trans. R. Soc. Lond. B* **357**, 1559–1566. (doi:10.1098/rstb.2002.1065)

66. Laland KN. 2004 Social learning strategies. *Anim. Learn. Behav.* **32**, 4–14. (doi:10.3758/BF03196002)

67. van Bergen Y, Coolen I, Laland K. 2004 Nine-spined sticklebacks exploit the most reliable source when public and private information conflict. *Proc. R. Soc. Lond. B* **271**, 957–962. (doi:10.1098/rspb.2004.2684)

68. Galef BG. 2009 Strategies for social learning: testing predictions from formal theory. *Adv. Study Behav.* **39**, 117–151. (doi:10.1016/S0065-3454(09)39004-X)

69. Kurvers RHJM, Wolf M, Krause J. 2014 Humans use social information to adjust their quorum thresholds adaptively in a simulated predator detection experiment. *Behav. Ecol. Sociobiol.* **68**, 449–456. (doi:10.1007/s00265-013-1659-6)