Accurate decisions in an uncertain world: collective cognition increases true positives while decreasing false positives

Max Wolf¹,†, Ralf H. J. M. Kurvers¹,†, Ashley J. W. Ward¹,², Stefan Krause³ and Jens Krause¹

¹Department of Biology and Ecology of Fishes, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Mueggelseedamm 310, 12587 Berlin, Germany
²School of Biological Sciences, University of Sydney, Sydney, New South Wales 2006, Australia
³Department of Electrical Engineering and Computer Science, Lübeck University of Applied Sciences, 23562 Lübeck, Germany

In a wide range of contexts, including predator avoidance, medical decision-making and security screening, decision accuracy is fundamentally constrained by the trade-off between true and false positives. Increased true positives are possible only at the cost of increased false positives; conversely, decreased false positives are associated with decreased true positives. We use an integrated theoretical and experimental approach to show that a group of decision-makers can overcome this basic limitation. Using a mathematical model, we show that a simple quorum decision rule enables individuals in groups to simultaneously increase true positives and decrease false positives. The results from a predator-detection experiment that we performed with humans are in line with these predictions: (i) after observing the choices of the other group members, individuals both increase true positives and decrease false positives, (ii) this effect gets stronger as group size increases, (iii) individuals use a quorum threshold set between the average true- and false-positive rates of the other group members, and (iv) individuals adjust their quorum adaptively to the performance of the group. Our results have broad implications for our understanding of the ecology and evolution of group-living animals and lend themselves for applications in the human domain such as the design of improved screening methods in medical, forensic, security and business applications.

1. Introduction

Decision-makers in a wide range of contexts, including predator avoidance, medical decision-making, job candidate selection and security screening, face a fundamental dilemma [1–9]. The goal of the decision-maker is to take an action whenever a particular condition in its environment is fulfilled but not to take this action when the condition is not fulfilled. Animals, for example, aim to run away in the presence of predators but not in their absence, doctors aim to provide a treatment when a disease is present but not when it is absent. Whether or not the condition in question is fulfilled is unknown to the decision-maker at the point in time when the decision must be made and has to be inferred from cues. These cues, however, are seldom perfectly correlated with the condition: cues that are indicative for a certain condition (e.g. presence of predator/disease) may be present in the absence of that condition, conversely, cues may be absent in the presence of that condition. By increasing responsiveness to such cues, decision-makers thus increase not only their chance of correctly taking the action in the presence of the condition (e.g. run away in the presence of predators, provide treatment in the presence of the disease) but also that of erroneously taking this action in its absence (e.g. run away in the absence of predators, provide treatment in the absence of the disease). Increased true positives (a.k.a. hits in detection theory) are thus associated with increased false positives (a.k.a. false alarms in detection theory), giving rise to a fundamental limitation of decision accuracy.

*These authors contributed equally to this study.

under uncertainty [2,6,10,11]. Figure 1 illustrates this dilemma for a simple but generic decision-making context under uncertainty; table 1 summarizes five examples that fit this scenario. Decision accuracy under uncertainty is thus fundamentally constrained by the trade-off between true and false positives. A high rate of true positives is possible only at the cost of a
high rate of false positives; conversely, a low rate of false positives is possible only at the cost of a low rate of true positives. We here use an integrated theoretical and experimental approach to show that a group of decision-makers using a simple quorum decision rule can overcome this basic limitation: compared with a solitary decision-maker, a group of decision-makers can both increase true positives and decrease false positives simultaneously.

We proceed in two steps. First, we show mathematically that, compared with solitary decision-makers, a simple quorum decision rule allows decision-makers in groups to increase true positives and decrease false positives simultaneously. In particular, this effect is achieved whenever decision-makers use a quorum threshold that is set above the false-positive rate and below the true-positive rate of solitary decision-makers. Second, we present results from experiments that we conducted on humans that are consistent with our model predictions: (i) after observing the choices of the other group members, individuals both increase true positives and decrease false positives, (ii) this effect gets stronger as group size increases, (iii) individuals use a quorum threshold set between the average true and false-positive rates of the other group members, and (iv) individuals adjust their quorum adaptively to the performance of the group.

2. Mathematical analysis and predictions

Consider a group of \( N \) decision-makers that face a situation as depicted in figure 1 and assume that each of those decision-makers balances the trade-off between true and false positives such that he/she commits true and false positives with probabilities \( p_{\text{true}} \) and \( p_{\text{false}} \), respectively. Assume that each decision-maker uses the following two-step decision rule: ‘indicate to take action \( A \) whenever your personal information indicates that you should take action \( A \) (i.e. whenever the perceived cue intensity \( x \) exceeds the decision threshold \( t \), figure 1), take action \( A \) when at least a fraction \( q \) of the other group members indicate to take action \( A \).’ Figure 2 illustrates the consequences of this rule for a particular example with \( p_{\text{true}} = 0.6 \) (red hatched lines) and \( p_{\text{false}} = 0.3 \) (green hatched lines) for the three cases of a low (figure 2a), a high (figure 2b) and an intermediate (figure 2c) quorum threshold \( q \). When using a low or a high quorum threshold, decision-makers in groups cannot overcome the trade-off between true and false positives that solitary decision-makers face. For low thresholds (figure 2a, \( q = 0.1 \)), increasing group size gives rise to increased true positives (red dots), but these increases are associated with increased false positives (green dots). Similarly, for high thresholds (figure 2b, \( q = 0.8 \)), increasing group size gives rise to decreased false positives, but these decreases are associated with decreased true positives. A very different scenario occurs, however, for intermediate thresholds (figure 2c, \( q = 0.43 \)), where increasing group size is associated with increased true positives and decreased false positives simultaneously. Hence, when using intermediate thresholds (i.e. thresholds below the true-positive \( p_{\text{true}} \) and above the false positive \( p_{\text{false}} \), see below), decision-makers in groups can overcome the trade-off between true and false positives that solitary decision-makers face. In fact, even for relatively small group sizes (\( n=20 \)), this mechanism allows decision-makers in groups to achieve a very good match between their behaviour and the true state of the environment (e.g. run when a predator is present and stay when a predator is absent; provide treatment when disease present, do not provide treatment when disease absent).

How can this result be understood? Above, we assumed that, given the state of the environment (e.g. state 0 or state 1), the personal information (i.e. perceived cue intensity \( x \)) of each decision-maker is independent from that of the others (see §4). Consequently, in a group of size \( N \), the fraction of decision-makers that, based on their personal information, indicate to take action \( A \) when the environment is in state 1 is binomially distributed with mean \( p_{\text{true}} \) and variance \( p_{\text{true}}(1-p_{\text{true}})/N \). Analogously, when the environment is in state 0, the fraction of decision-makers that indicate to take action \( A \) is binomially distributed with mean \( p_{\text{false}} \) and variance \( p_{\text{false}}(1-p_{\text{false}})/N \). Thus, with increasing group size \( N \), the fraction of decision-makers that—based on their personal information—indicate to take action \( A \) approaches \( p_{\text{true}} \) (\( p_{\text{false}} \)) when the environment is in state 1 (state 0), because the variance of this fraction approaches zero. Therefore, whenever the probability \( p_{\text{true}} \) of committing true positives exceeds the probability \( p_{\text{false}} \) of false positives, any decision rule that sets the threshold \( q \) such that

\[
q < p_{\text{true}} < q < p_{\text{true}}
\]

can effectively deduce the true state of the environment (e.g. predator present or absent; disease present or absent). While perfect discrimination will typically be possible only for very large groups, figure 2c shows that substantial improvements for both true and false positives can be achieved already in relatively small groups of less than 20 individuals.

In sum, our analysis provides two key predictions. First, whenever the true positives \( p_{\text{true}} \) of solitary decision-makers exceed their false positives \( p_{\text{false}} \), a group of decision-makers can—compared with a solitary decision-maker—both increase true positives and decrease false positives simultaneously. Second, this collective intelligence effect is achieved whenever decision-makers use a quorum decision rule that sets its threshold \( q \) below the rate of true positives and above the rate of false positives of solitary decision-makers (equation (2.1)).

3. Experiment and results

To test these predictions, we recruited students from the University of Bielefeld, Germany, visitors to an open day of Wageningen University, the Netherlands and visitors to the natural history museum in Berlin, Germany. In total, we tested 436 participants divided over 24 groups (average group size: 18.2; range 9–25) in the following predator-detection test (see the electronic supplementary material).

Each group of individuals entered the test room, and individuals were asked to sit on a chair. For a fixed time period of 2 s, a school of 144 fish (aligned in \( 9 \times 16 \) grid) was projected onto a white screen. All fish in this school were identical, except one odd fish, which had either six or seven spines. The other 143 fish had no spines. We instructed our subjects to adopt the following decision rule: ‘If you see no odd fish or an odd fish with 6 spines then it is safe and you should stay. If you see an odd fish with 7 spines then it is dangerous and you should escape.’ After 2 s of observing the school, participants had 5 s to take a decision via an electronic keypad (‘polling 1’). We then projected for 5 s a bar chart showing
the number of individuals that decided to escape. Individuals were then asked to decide again (‘polling 2’) after which we presented the results of the second polling and the correct answer (stay or escape). There were two treatments: (i) one fish with six spines and 143 fish without spines, (ii) one fish with seven spines and 143 fish without spines. Each treatment was replicated 14 times resulting in a total of 28 rounds per group. The treatment order and the position of the odd fish were randomized. Prior to the 28 rounds, we performed two test rounds to instruct the participants about the procedure. The results of the test rounds were excluded from our analysis.

We calculated the average true and false positives for each individual before (polling 1) and after (polling 2) observing the decisions of the other group members. Note that a true positive is achieved whenever the seven-spined fish is present and a participant decides to escape; a false positive is committed whenever the six-spined fish is present and a participant decides to escape. As predicted from our analysis earlier, when comparing the second with the first polling (i.e. after observing the decisions taken by the others), individuals achieved both higher true positives (first polling: mean ± s.e. = 0.564 ± 0.011, second polling: 0.751 ± 0.009, est. ± s.e. = 0.8604 ± 0.148, z = 5.829, p < 0.001) and lower false positives (first polling: mean ± s.e. = 0.166 ± 0.006, second polling: 0.113 ± 0.005; est. ± s.e. = −0.448 ± 0.199, z = −2.257, p = 0.024). This pattern was very consistent between groups: when comparing polling 2 with polling 1, the average true positive of individuals increased in all of the 24 groups (figure 3a), the average false positive of individuals decreased in 20 of 24 groups (figure 3b). This pattern was also true at the individual level: a large proportion (40.2%) of our 436 participants both increased their true positives and decreased their false positives; and a substantial proportion (28.5%) either
increased true positives with no change in false positives or decreased false positives with no change in true positives (figure 3c). With increasing group size, there was a larger increase in true positives (est. ± s.e. = 0.004 ± 0.002, \(t = 2.002, p = 0.058\)) and a larger decrease in false positives (est. ± s.e. = 0.003 ± 0.001, \(t = 3.135, p = 0.005\)).

We investigated whether the participants used quorum responses in their decision to escape in polling 2 based on the social information provided after polling 1. In a quorum response, the probability that an individual decides for a particular option increases in a step-like way with the number of other individuals that have decided for that option. Such responses can be well described with equation (3.1) [12,13]:

\[
p = \frac{x^k}{x^k + T^k},
\]

where \(p\) is the probability that a focal individual chooses a particular option, \(x\) is the number of individuals that have already chosen this option, \(T\) is the threshold quorum at
which the response has the steepest increase and $k$ determines
the steepness of this increase. As a rule of thumb, a quorum
response occurs if $k/C_2 \geq 12,13$ and higher values of $k$
imply stronger quorum responses. (Note that our mathematical
analysis is based on the assumption that individuals use
strong quorum responses, that is large $k$-values. For large
$k$-values, the quorum threshold $T$ lies
below the average true-positive score and above the average false-positive score of individuals within that group. (b) There was a strong positive correlation between escape quorum and the average true-positive score during polling 1 (Spearman’s rho = 0.858, $p < 0.001$, $n = 24$) suggesting that individuals adaptively adjust their escape quorum in accordance to the performance of the group. For clarity, the groups in (a) are ranked according to the true-positive score of polling 1.

Figure 4. When taking the decisions of the other group members into account, individuals use a quorum threshold set between the average true- and false-positive rates of the other group members. (a) For each group, average true- (filled circle) and false- (filled triangle) positive scores during polling 1 and the estimated escape quorum $T$ (open circle) are shown. The escape quorum corresponds to the quorum threshold at which half of the individuals that decided to stay during polling 1 decided to escape during polling 2 (see the electronic supplementary material). As predicted, for 22 out of 24 of the groups, the escape quorum $T$ lies below the average true-positive score and above the average false-positive score of individuals within that group. (b) There was a strong positive correlation between escape quorum and the average true-positive score during polling 1 (Spearman’s rho = 0.858, $p < 0.001$, $n = 24$) suggesting that individuals adaptively adjust their escape quorum in accordance to the performance of the group. For clarity, the groups in (a) are ranked according to the true-positive score of polling 1.
4. Summary and discussion

(a) Summary
Using a mathematical model, we have shown that (i) by joining a group, individuals can both increase true positives and decrease false positives and (ii) this is achieved by a quorum decision rule with a threshold set between true- and false-positive rates of individual decision-makers. The results from a predator-detection test that we performed with humans are consistent with these predictions: (i) after observing the choices of the other group members, individuals both increase true positives and decrease false positives, (ii) this effect gets stronger as group size increases, (iii) individuals use a quorum threshold set between the average true- and false-positive rates of the other group members, and (iv) individuals adjust their quorum adaptively to the performance of the group.

(b) True and false positives of solitary decision-makers
A key assumption in our model is that the true-positive rate of a solitary decision-maker is higher than its false-positive rate. This is a realistic assumption [2]. Our analysis focuses on situations where the environment can be in either of two states, and the decision-maker aims at taking an action A in one of the states (state 1: e.g. predator/disease present) but not in the other (state 0: e.g. predator/disease absent). The true positive thus corresponds to the conditional probability \( p_A|\text{state} = 1 \) that the decision-maker takes the action, given that the environment is in state 1; the false positive corresponds to the conditional probability \( p_A|\text{state} = 0 \) that the decision-maker takes the action, given the environment is in state 0. In the extreme case, when the decision-maker has no information about the state of its environment, he/she can only randomize between taking the action or not, independent of the state of its environment, thus giving rise to equal true and false positive, i.e., \( p_A|\text{state} = 1 = p_A|\text{state} = 0 = p \). Consequently, any information about the true state of the environment will allow the decision-maker to achieve a rate of true positives that is higher than its rate of false positives.

(c) Independence of information
Another key assumption in our model is that, given the state of the environment, the information (i.e. perceived cue intensity \( x \), figure 1) held by different decision-makers is independent from each other. Full independence is not required and similar results will be obtained whenever the information of different decision-makers is correlated with each other. Some independence, however, is necessary. Put differently, when all decision-makers always perceive the same cue intensity \( x \) (figure 1), collective intelligence will not be possible.

In principle, independence of information arises via several potentially interacting mechanisms. First, when decisions have to be made fast (e.g. in an anti-predation context, during a security screen), a single decision-maker may not be able to fully evaluate the situation, and different decision-makers may focus on different aspects of the problem [14–16]. Second, when cues occur only for a brief moment in time and/or in one particular location in space (e.g. cues about a predator in an anti-predation context), the detection of these cues depends on the position of the decision-maker in space and the direction it faces. Third, when the decision problem is complex and there is no unique best way of evaluating the situation, different decision-makers (e.g. doctors/psychiatrist in a medical/clinical decision-making context) may evaluate identical pieces of information differently, dependent on their experience and cognitive style [17–19].

The fact that collective intelligence as reported here is possible only in the face of some degree of independent information has consequences for our understanding of animal groups and for the improvement of human decision-making processes. Group-living animals may have been shaped by natural selection in order to achieve some degree of independence—e.g. via favouring the synchronization and/or complementation of activities/positioning in space between different group members or via favouring particular mixtures of personality types [20] within a group. Similarly, decision accuracy in contexts such as medical decision-making, job candidate selection, clinical psychology or security screens (table 1) may be improved by group decisions in which decision-makers (i) make decisions without a prior exchange of opinions and (ii) differ in their experience, cognitive style and/or personality [21].

(d) Quorum decision-making
Quorum decision-making can be found in many animal species [13,22–25] ranging from insects to fish to humans. We have shown that quorum decision-making allows individuals in groups to both increase true positives and decrease false positives simultaneously, thereby overcoming a fundamental limitation to decision accuracy that solitary decision-makers face. Our analysis predicts that individuals should adjust their quorum threshold, dependent on the true- and false-positive rates of the other group members. In line with this prediction, we find a strong positive correlation between the escape quorum and the average true-positive score during polling 1 in our experiment (figure 4b), indicating that individuals adjust their escape quorum adaptively to the performance of the group. Non-human animals such as ants [26,27] and fish [28] are known to flexibly adjust their response threshold, dependent on environmental conditions. Whether flexible adjustment in response to the decision accuracy of other group members occurs in the way predicted from our model is an interesting open question.

(e) Condorcet’s jury theorem
Condorcet’s jury theorem [29,30] is a prominent example, illustrating how a group of decision-makers can improve decision accuracy compared with solitary decision-makers. Condorcet’s theorem applies to situations where (i) individuals face a binary choice, (ii) each individual has a probability \( p > 0.5 \) of making a correct decision in the absence of others, and (iii) different decision-makers make their choice independent of each other. Condorcet showed that in such situations, the probability that the majority of individuals make a correct decision increases with group size. While our results may seem to resemble Condorcet’s jury theorem, they differ in two key aspects. First, our modelling results are not based on the assumption that individuals have a probability \( p > 0.5 \) of making a correct decision in the absence of others. Analogously, participants in our experiments often had true positives below 50 per cent (figure 3a). Second, our results are not based on majority voting—indeed groups in our experiments often follow a minority of individuals (see the escape quorums in figure 4b).
(f) Mechanism of group-decision-making
For simplicity, we investigated situations where decision-makers in groups take a single and final decision after being presented with a summary measure of the decision of all other group members. This is a feasible mechanism in many applied contexts involving human decision-makers (table 1). In many natural situations involving non-human animals, however, both information transfer and decision-making will be more complex [31,32]. Individuals may, for example, not decide simultaneously thus giving rise to a much more dynamic and interdependent ‘voting process’; individuals may repeatedly switch between the different behavioural options; individuals in large groups may observe only the decisions of their local interaction partners. Future studies will investigate how such more complex decision-making mechanisms affect the ability of a group to achieve collective intelligence as reported here.

(g) Individual differences
Our modelling analysis was based on the assumption that all decision-makers are identical. However, in most natural situations—as for example in our experiments—group members may differ in a variety of relevant aspects, including their experience with the particular problem at hand, their cognitive abilities, their general tendency to rely on social information and their propensity to lead a group [33–35]. In our experiments, we have deliberately developed a set-up that minimizes individual differences, in particular, our set-up does not allow for the emergence of leaders and followers. Future studies will investigate how such differences affect the mechanisms of group-decision-making, the individual benefits associated with joining a group (i.e. improvement in decision accuracy) and, ultimately, the ability of a group to achieve collective intelligence as reported here.

(h) Predator avoidance
Research on collective behaviour and group-decision-making in animals has shown that individuals in groups can outperform solitary individuals [31,36–38]. Group-living has important consequences for the performance of individuals in anti-predator contexts [39,40]. Individuals in larger groups have, for example, a lower chance of being killed once detected (dilution and confusion effect) but the chance of being detected by a predator increases with group size. One important dimension of performance under predation risk is decision accuracy when detecting predators. Consequently, a key question is whether individuals in groups can achieve higher decision accuracy than solitary individuals, that is, combinations of true and false positives that are not feasible for solitary individuals.

It is well known that individuals in groups can detect predators earlier and/or with a higher probability than solitary individuals, i.e. achieve higher true positives. The basic intuition underlying this ‘many eyes effect’ [41,42] (also termed collective detection [4]) is simple: an individual in a group detects the predator not only when it detects the predator itself but also when another group member detects the predator and warns the others (e.g. via an alarm call). Pulliam [43] formalized this idea in an influential model, and several subsequent studies have found this effect empirically [41,44–48]. The same mechanism of social information use, however, is believed to give rise to a decreased performance in the absence of predators (i.e. a higher false positive) [4,7,9,49–51]. In a nutshell, an individual in a group not only commits a false positive when it erroneously detects a predator itself but also when another group member erroneously detects a predator and warns the others. As a consequence, individuals in groups achieve higher true positives than solitary individuals, but this increase comes at a cost of increased false positives, corresponding to the scenario depicted in figure 2a.

Our results suggest that groups of animals can do fundamentally better. In particular, when individuals in groups set intermediate response thresholds (i.e. thresholds below the true positive $p_{true}$ and above the false positive $p_{false}$, see above) they can—compared with solitary individuals—both increase true positives and decrease false positives simultaneously. This result is in line with previous findings that animals in groups can diminish the negative consequences of false positives by using a behavioural rule that does not respond to single but only to multiple other individuals [9,51–54].

(i) General applicability
While our basic arguments and our model apply to a diverse range of decision problems (table 1), in our experiments, we have focused on a predator avoidance context. Future studies will investigate the general applicability of our quorum rule. Multiple doctors may, for example, be presented with a series of cases (e.g. via diagnostic images such as mammography) where the correct diagnosis is known (e.g. cancer present or not). Part of the dataset may be used to estimate true and false positives of individual doctors. These estimates can be used to set a quorum threshold for this group of doctors, which would be used in the remaining part of the dataset. In this set-up, we could evaluate whether our quorum rule allows simultaneous increases in true positives and decreases in false positives in medical diagnostics; similar experiments are conceivable for all other decision contexts mentioned in table 1. As we have stressed repeatedly, our results may be of importance to decision-making in a wide variety of contexts and it will be exciting to further investigate the applicability of our quorum rule in these contexts.

The authors thank four anonymous referees for numerous helpful suggestions on previous versions of the manuscript; the participants of all experiments; Romain Clément and Sander van Doorn for stimulating discussions; Romain Clément, Chantal van Lin, Rudy Jonker, Vincent Comor, Kyle Tomlinson, Leif Engqvist and Julia Roehner for assisting in the experiment. R.H.J.M.K. was funded by an NWO Rubicon grant (no. 825.11.014).

References