Impact of correlated information on pioneering decisions

Megan Stickler,[1,*] William Ott,[1,†] Zachary P. Kilpatrick,[2,‡] Krešimir Josić,[1,3,§] and Bhargav R. Karamched[4,5,6,‖]

[1]*Department of Mathematics, University of Houston, Houston, Texas 77004, USA*
[2]*Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado 80309, USA*
[3]*Department of Biology and Biochemistry, University of Houston, Houston, Texas 77004, USA*
[4]*Department of Mathematics, Florida State University, Tallahassee, Florida 32306, USA*
[5]*Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA*
[6]*Program in Neuroscience, Florida State University, Tallahassee, Florida 32306, USA*

(Received 28 March 2023; accepted 9 June 2023; published 10 July 2023)

Normative models are often used to describe how humans and animals make decisions. These models treat deliberation as the accumulation of uncertain evidence that terminates with a commitment to a choice. When extended to social groups, such models often assume that individuals make independent observations. However, individuals typically gather evidence from common sources, and their observations are rarely independent. Here we ask: For a group of ideal observers who do not exchange information, what is the impact of correlated evidence on decision accuracy? We show that even when agents are identical, correlated evidence causes decision accuracy to depend on temporal decision order. The first decider is less accurate than a lone observer, and early deciders are less accurate than late deciders. These phenomena occur despite the fact that the rational observers use the same decision criterion, so they are equally confident in their decisions. We analyze discrete and macroscopic evidence-gathering models to explain why the first decider is less accurate than a lone observer when evidence is correlated. Pooling the decisions of early deciders using a majority rule does not rescue accuracy results in only a modest accuracy gain. Although we analyze an idealized model, we believe that our analysis offers insights that do not depend on exactly how groups integrate evidence and form decisions.

DOI: 10.1103/PhysRevResearch.5.033020

I. INTRODUCTION

Most organisms and many computational algorithms make decisions based on a sequence of noisy observations of the environment [1]. Normative models that describe how evidence should be integrated to make the best choice are central to our understanding of such decisions [2]. When an observer needs to choose between alternatives, accumulating evidence refines their perceived probability of the truth of each alternative. Decision policies often prescribe a threshold on the accumulated evidence in order to balance the speed and accuracy of decisions [3,4]. These theories have been developed and validated over decades in experiments with humans and other animals [5–9]. However, most previous work was focused on individual decision makers, and less is known about groups of observers who make choices based on streams of evidence [10,11].

Each member of a social group often needs to choose between the same alternatives based on a combination of correlated and independent observations [12]. For instance, when deciding whom to vote for, two individuals may see some of the same media coverage, but each may also read opinion pieces that the other does not [13]. Conspecifics deciding where to forage are likely to rely on some of the same cues but can also learn from distinct experiences [14]. Traders may have access to private information but often track the same aggregate market indices and reports to decide what stocks to buy and sell, and the processes governing the valuation of distinct commodities are known to be correlated [15]. Thus, even in the absence of direct communication, the measurements individuals in a group use to make decisions are generally *imperfectly* correlated.

Here we assess the impact of such correlated measurements on the accuracy of individual decisions within groups of agents who do not share information (see Fig. 1). When identical, rational, unbiased agents make *independent* observations the probability of a correct decision is independent of the order or the time at which the decision is made [16,17]. However, when such agents makes correlated measurements, early deciders tend to make decisions based on misleading observations, and their choices are less accurate than those of later deciders by as much as 20%. The order of a decision can therefore determine its accuracy, despite each agent subjectively believing their decision is based on the same amount of evidence, and thus as accurate as that of anyone else. Yet an outsider who

*Corresponding author: mstickler@carthage.edu; present address: Carthage College, Kenosha, Wisconsin 53140, USA.
†Corresponding author: william.ott.math@gmail.com
‡Corresponding author: zpkilpat@colorado.edu
§Corresponding author: kresimir.josic@gmail.com
‖Corresponding author: bkaramched@fsu.edu

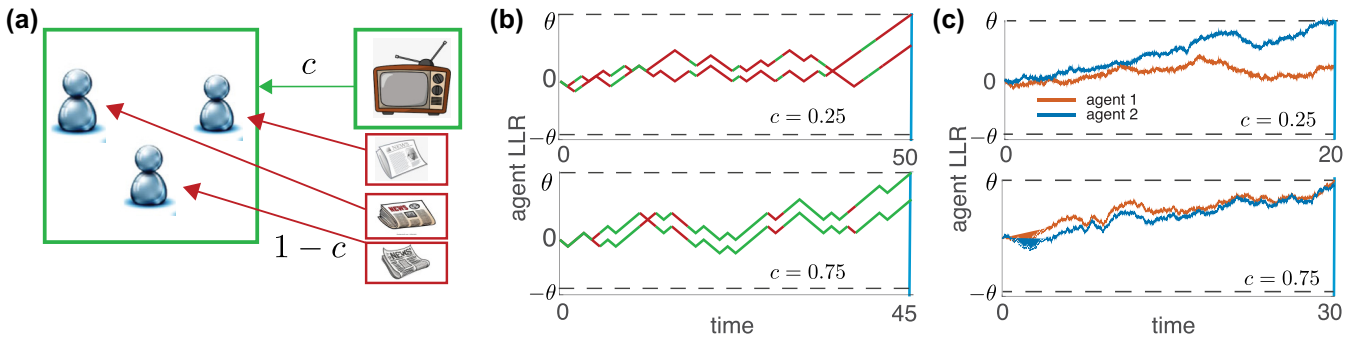

FIG. 1. Agents receiving partially correlated evidence. (a) Agents make a sequence of measurements to decide between alternatives. At each time step agents all make the same observation with probability c and independent observations with probability $1-c$. (b) Representative trajectories of the log-likelihood ratios (LLRs) computed at discrete times [Eq. (2)] for $c = 0.25$ and $c = 0.75$ and two agents. Green segments correspond to increments due to common observations, and red segments arise from independent observations. An agent commits to a decision ($\pm$) when the computed LLR crosses $\pm\theta$. (c) Analogous trajectories generated from simulations of the limiting drift-diffusion model [Eq. (3)].

observes the order in which decisions are made knows that early decisions are less likely to be correct than later ones. We analytically show why this is the case in tractable examples and provide an intuitive argument explaining why the same holds more generally. Our analysis demonstrates why this difference in accuracy depends on how strongly evidence is correlated and on the size of the population. We also show that pooling early decisions does not always help, but weighting individual decisions according to their order can produce better results.

II. MODEL

We consider a community of N agents who accumulate evidence to decide between two states, or hypotheses, H^+ or H^-. Each agent accumulates evidence (observations) to decide between the two hypotheses. Agents are rational (Bayesian) and compute the probability that either hypothesis holds based on all evidence they accrue. Each makes a decision once the log-likelihood ratio (LLR) of the conditional probabilities between the two hypotheses, given all the accumulated observations, crosses one of two predetermined decision thresholds [2,18]. For simplicity, we assume that the observations the agents make are statistically identical and that they use the same decision policy.

A. Independent evidence accumulation

The problem of a single agent integrating evidence to decide between two options has been thoroughly studied [2,10,18–21]. In the simplest setting, an agent makes a sequence of noisy observations (measurements), $\xi_{1:t}$, with $\xi_i \in \Xi$ for $i \in \{1, \ldots, t\}$, where $\Xi \subset \mathbb{R}$. The observations, ξ_i, are independent and identically distributed, conditioned on the true state, $H \in \{H^+, H^-\}$,

$$P(\xi_{1:t}|H^\pm) = \prod_{i=1}^{t} P(\xi_i|H^\pm) = \prod_{i=1}^{t} f_\pm(\xi_i).$$

Here the conditional probability of each measurement is given by the probability mass functions, $f_\pm(\xi) := P(\xi|H^\pm)$, when the conditional probability distributions are discrete, or by density functions when they are absolutely continuous. Observations, ξ_i, are drawn from the same set, Ξ, in either state $H^\pm$, and the two states are distinguished by the differences in the conditional probabilities of making certain measurements. See Appendix A for details on how the restriction $\Xi = \{\xi^+, \xi^-\}$ can confine beliefs to evolving on the integer lattice.

To compute the probability of the two choices, given all observations, $P(H^\pm|\xi_{1:t})$, an ideal observer uses Bayes' rule. For simplicity, we assume that the agent knows the measurement distributions, $f_\pm(\xi)$, and knows that both environmental states are equally likely, and hence uses a flat prior, $P(H^+) = P(H^-) = 1/2$. The log-likelihood ratio (LLR) of the two states at time t is then

$$y_t := \log\left(\frac{P(H^+|\xi_{1:t})}{P(H^-|\xi_{1:t})}\right) = \sum_{s=1}^{t} \mathrm{LLR}(\xi_s) = y_{t-1} + \mathrm{LLR}(\xi_t), \tag{1}$$

where $\mathrm{LLR}(\cdot) \equiv \log\frac{P(\cdot|H^+)}{P(\cdot|H^-)}$. We also refer to y_t as the *belief* of the agent at time t. The magnitude of the LLR can be viewed as the information an agent has gathered in support of a hypothesis, while its sign describes the choice preference (H^+ or H^-) of the agent. The flat prior implies $y_0 = \log\frac{P(H^+)}{P(H^-)} = \log\frac{1/2}{1/2} = 0$.

The optimality of the sequential probability ratio test [18] implies that an individual agent best manages speed and accuracy by waiting to decide until their belief reaches or crosses above (below) an upper (lower) threshold $\theta_+ > 0$ ($\theta_- < 0$). Thus, an ideal agent continues making observations while $\theta_- < y_t < \theta_+$ and makes a decision after acquiring sufficient evidence, choosing H^+ (H^-) once $y_t \geqslant \theta_+$ ($y_t \leqslant \theta_-$). We have analyzed a generalization of this model to social networks both small [22] and large [17], where each agent accrues independent information according to Eq. (1) and shares their decision state with some or all other agents in the group. These models of normative information exchange based on neighbors' decisions build on previous work on normative confidence weighting for majority rules [23–26], locally optimal Bayesian integration on sparse graphs [27],

the impact of common observations [16], and nonnormative decision sharing [28].

B. Accumulation of correlated measurements

We analyze the impact of correlated information on the accuracy of the decisions of a community of N independent and isolated agents. At each time step, t, every agent, i, makes an observation (measurement), $\xi_t^i \in \Xi$, and updates their private belief, y_t^i, according to Eq. (1). However, an individual agent does not know whether others have made decisions or what those decisions were, in contrast to social network models studied in the past [17,22–32]. This could be a model of a sample of voters, each of whom does not know the others, or traders deciding to buy or sell a stock without tipping their hand.

To model correlated measurements, we assume that on each time step all agents make an identical observation with probability c. An identical observation means that $\xi_t^i = \xi_t$ for all agents, $i = 1, \ldots, N$, where ξ_t is a single sample from the measurement distribution, $f_\pm(\xi)$. With probability $1 - c$ agents make independent observations during a time step, and the N measurements, ξ_t^i, are sampled independently from the distribution $f_\pm(\xi)$. This is equivalent to having N *private*, independent sources of evidence, each accessible to a single agent, and one *common* evidence source accessible to all agents (see the Discussion for less restrictive assumptions). Therefore, the belief of each agent evolves according to

$$y_t^i = y_{t-1}^i + (1 - \chi_t)\mathrm{LLR}\left(\xi_t^i\right) + \chi_t \mathrm{LLR}(\xi_t), \qquad (2)$$

where χ_t are i.i.d. Bernoulli random variables each with parameter c. When $c = 1$ agents make only common observations, and when $c = 0$ agents make only independent observations. As c increases from zero, each observation is more likely to be common, and the overall evidence becomes more correlated.

Each agent makes observations until their belief, y_t^i, reaches one of the thresholds, $\theta_\pm$, at which point they make the corresponding decision, $H^\pm$. For simplicity we henceforth assume the thresholds are symmetric about zero, i.e., $\theta_\pm = \pm\theta$, with $\theta > 0$. We denote the decision time of agent i by T_i, and assume that decisions are immutable. Thus, decision times are uniquely defined, and only undecided agents continue to make observations.

Importantly, each agent is isolated and *does not* observe others' decisions or their decision state (decided or undecided), in contrast with [17,22]. Agents do not know whether an observation is common or private, and each uses the evidence they have collected to make the best possible decision based on their belief (LLR) given by Eq. (2).

We ask how the accuracy of an agent's decision depends on the order in which the decision is made. In particular, how accurate is the first decider? If multiple agents make a decision at first-decision time, the "first" decider is chosen randomly with equal probability from that group. The probability of a correct first decision then equals the probability that this first decider makes the correct choice, i.e., that the belief of the first decider reaches the threshold, $\pm\theta$, whose sign agrees with that of the true environmental state, $H^\pm$. We briefly discuss other ways of defining a first decision in Appendix B.

C. Scaling limit of correlated evidence accumulation

Computing decision accuracy and the distribution of decision times reduces to a first-passage problem [2]. Often it is easier to solve such problems in the scaling limit, thus avoiding the combinatorial challenges common in discrete problems [33]. By invoking the Donsker Invariance Principle [34], in the limit of infinitely many infinitesimally informative measurements we obtain the macroscopic version of Eq. (2), often referred to as a *drift-diffusion equation*:

$$dy_i = \pm\mu\, dt + \sqrt{2(1-c)\mu}\, dW_i + \sqrt{2c\mu}\, dW_c. \qquad (3)$$

Here $y_i(t)$ is the limit of the LLR of agent i and μ scales both the drift and diffusion terms. See Appendix C for a derivation of Eq. (3), verification of agreement with the discrete model, and definition of μ, which is proportional to the square of the signal-to-noise ratio of the sample distribution. The sign of the drift agrees with the sign of the environmental state, $H^\pm$. The Wiener processes, $W_i(t)$ and $W_c(t)$, capture the variability of belief increments due to independent and common observations, respectively. Thus, the belief of each observer, $y_i(t)$, evolves according to a drift-diffusion model [2] with a combination of independent and correlated noise sources. This model has been analyzed previously [16,35], but we are not aware of a previous derivation from the normative model (see Discussion).

III. RESULTS

We first asked how correlated evidence impacts the accuracy of decisions within a group of rational, identical agents. The probability that a randomly selected agent in the group makes a correct choice does not depend on the number of other agents or on how strongly the evidence is correlated. However, for all $0 < c < 1$, the probability that the *first* decider in the group is correct is *smaller* than the probability that a lone observer is correct, reaching an internal minimum [Fig. 2(a)].

Since every individual agent's perception of the correct hypothesis and decision process are described by the same stochastic process, each agent has the same subjective estimate that their choice is correct, $(1/(1 + e^{-\theta})$ [2,36]. Indeed, this is the probability that a randomly chosen agent makes a correct decision. However, the first agent to make a decision is less likely to make a correct choice than all other agents in a group, and this probability decreases with the number of agents in the community [Fig. 2(b)]. Furthermore, decider accuracy increases almost monotonically with the order of the decision [Fig. 2(c)]. Thus, someone observing the order in which decisions are made should trust later decisions more than early ones. Decision times of distinct agents get closer as common observations become more probable [Fig. 2(d)], since the observers' beliefs evolve more synchronously.

The decreased accuracy of the first decider for $0 < c < 1$ relative to single-decider accuracy is not a trivial consequence of early deciders spending less time accumulating evidence. If this were the case, the first decider would be less accurate than later ones when $c = 0$. But when all observations are independent, the probability of a correct decision is independent of the order in which the decision is made and is determined

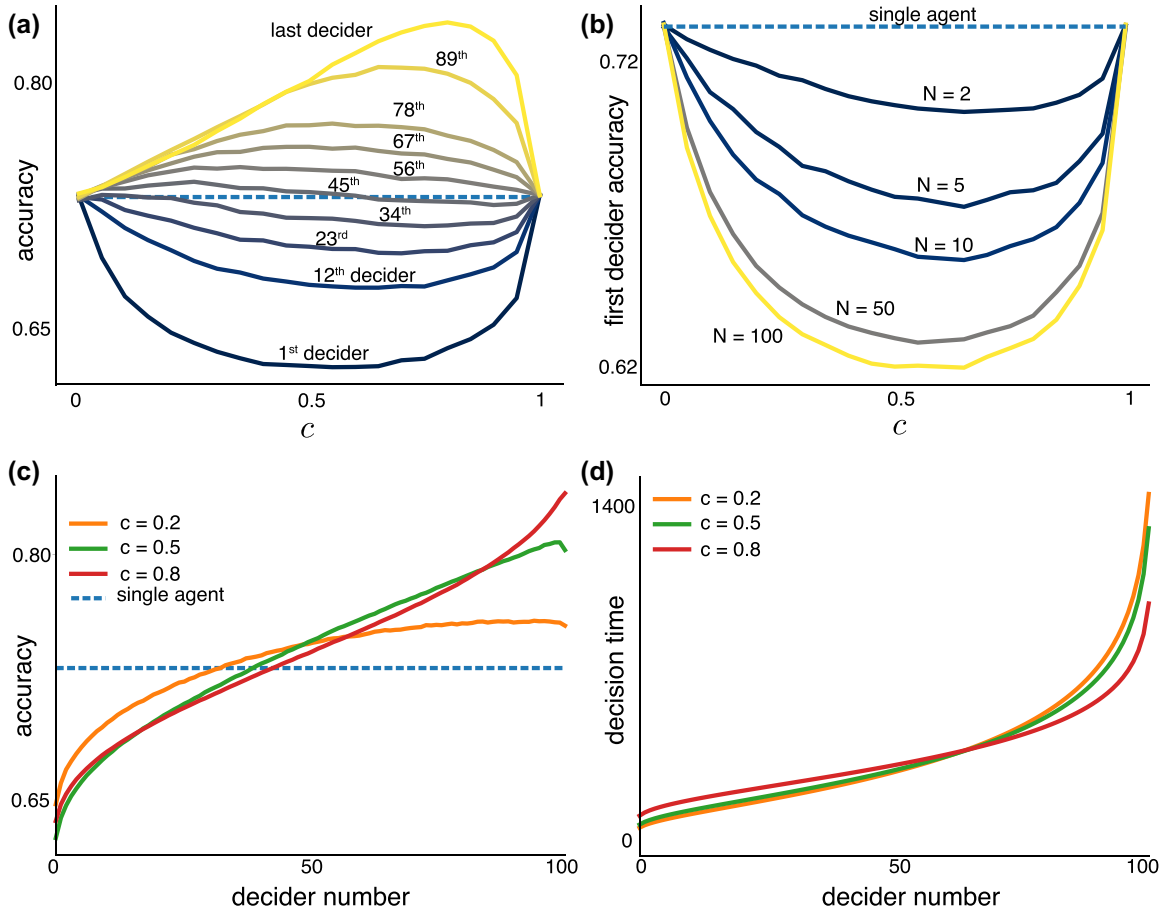

FIG. 2. Impact of the probability of making a common observation, c, on decider accuracy and timing. (a) The probability of a correct decision increases with the order in which the decision is made (solid lines) and varies nonmonotonically with c. The average accuracy computed over all deciders (dashed line) equals the accuracy of a randomly chosen agent and is constant with c. $N = 100$. (b) The accuracy of the first decider varies nonmonotonically with c, possessing an internal minimum $0 < c_{\min} < 1$. As N increases, the lowest accuracy decreases. (c) The accuracy of each of $N = 100$ deciders increases with decision order almost monotonically, so the first (last) decider is less (more) accurate than a lone decider for $c \neq 0, 1$. (d) The time of the decision of $N = 100$ agents as a function of order is approximately invariant to changes in the probability of common evidence, c. We used the discrete LLR model Eq. (2) with $\theta = 10$ and binary likelihood functions $f_{\pm}$ as described in Appendix A. Specifically we chose $f_{\pm}$ so that the update size is ± 0.05.

by the decision threshold. However, a common initial bias can also lead to accuracies that depend on decision order, even when measurements are independent (see Discussion as well as [22,37]). Moreover, as c is increased from 0 to approximately 0.5, the average time to the first decision increases, but the average accuracy of this decision decreases. We next provide an explanation of this observation.

A. An intuitive explanation for the decrease in first decision accuracy

Why do common observations lead to less accurate first decisions? At the time of the first decision, the remaining undecided agents have likely made independent observations that counter the common observations that often contribute to the first decider's choice. Indeed, if these independent observations aligned with this choice and the common evidence, the other agents would likely have already made a decision as well. For small c, little information is gained from common evidence, and not much independently gathered evidence is needed to counter it. As c increases, common evidence more often drives the first decision, so we expect a substantial fraction of the independent evidence collected by an undecided agent will often counter the common evidence. However, when c is large, most evidence is common, and fewer observations are independent, leaving less time for strong, contrary independent observations. Thus, at a critical value of c, the average total independent evidence obtained by undecided agents countering common observations reaches a maximum. The probability of a correct first decision is smallest at this critical value. In the next subsection, we sharpen this argument by showing independent observations made by undecided agents that favor the correct decision are stronger when the first decider makes an incorrect choice than in the opposite case.

B. Reduction of the log-likelihood ratio of the first decider

We next show mathematically why the first decider's choice is less accurate than that of a randomly chosen agent

selected with equal probability from all agents in the group prior to evidence accumulation. To do so, we write the log-likelihood ratio (LLR) associated with the probability the first decider makes the correct choice as a sum of two terms: One term is the LLR of a randomly selected agent at decision time, while the second incorporates the condition that this agent is the first decider. We show that the first term's magnitude equals θ, while the second term is negative for $0 < c < 1$. Thus, the information obtained by undecided agents reduces the probability of a correct first decision. We begin by considering a pair of agents and obtain expressions for the sum of LLR terms in the case of beliefs evolving on a lattice. We then extend this calculation to an arbitrary number of agents.

1. Pair of agents in discrete time

We randomly number the agents using indices $j = 1, 2$, and let FD be the index of the first decider. Let T_j be the time of the decision of agent j, and denote the decision of agent j by $d_j \in \{H^+, H^-\}$, so that $y_j(T_j) = \pm\theta$ and $|y_j(t)| < \theta$ when $0 \leqslant t < T_j$. Let $T = \min(T_1, T_2)$ denote the time of the first decision. We assume the first decider chooses H^+ ($d_{FD} = H^+$) without loss of generality (WLOG), and write the conditional probability $P^\pm(d_{FD} = H^+) := P(d_{FD} = H^+|H^\pm)$ as

$$\begin{aligned} P^\pm(d_{FD} = H^+) &= \sum_{j=1}^{2} P^\pm(d_j = H^+, FD = j) \\ &= \sum_{j=1}^{2} P^\pm(FD = j|d_j = H^+)P^\pm(d_j = H^+) \\ &= 2P^\pm(d_1 = H^+)P^\pm(FD = 1|d_1 = H^+), \end{aligned} \tag{4}$$

where the final line follows from the exchange symmetry between the two agents. The first term in Eq. (4) is the $P^\pm$ probability that a randomly chosen agent (here agent 1, WLOG) selects H^+, depending only on agent 1's observations. The second term is the $P^\pm$ probability that, conditioned on choosing H^+, agent 1 is also the first to decide, which depends on information gathered by agent 2.

The second term on the right side of Eq. (4) can be rewritten as a sum over $T_1, T_2 \in \mathbb{N}$ and then simplified by noting that $FD = 1$ with certainty if $T_1 < T_2$ and with probability $1/2$ if $T_1 = T_2$:

$$\begin{aligned} P^\pm(d_{FD} = H^+) &= 2P^\pm(d_1 = H^+) \sum_{t_1\in\mathbb{N}} \sum_{t_2\in\mathbb{N}} P^\pm(FD = 1|T_1 = t_1, T_2 = t_2, d_1 = H^+)P^\pm(T_1 = t_1, T_2 = t_2|d_1 = H^+), \\ &= 2P^\pm(d_1 = H^+) \sum_{t_1\in\mathbb{N}} \left[\frac{1}{2}P^\pm(t_1 = T_1 = T_2|d_1 = H^+) + P^\pm(t_1 = T_1 < T_2|d_1 = H^+)\right]. \end{aligned}$$

Using Eq. (4) we can thus write the corresponding LLR of the first decider at the time of their decision as

$$\mathrm{LLR}(d_{FD} = H^+) = \log \frac{P^+(d_{FD} = H^+)}{P^-(d_{FD} = H^+)} = \mathrm{LLR}(d_1 = H^+) + \mathrm{LLR}(FD = 1|d_1 = H^+).$$

The first term in this sum is the LLR of the decision a randomly chosen agent (taken here to be agent 1 WLOG), $\mathrm{LLR}(d_1 = H^+) = \theta$. The second term is given by

$$\mathrm{LLR}(FD = 1|d_1 = H^+) = \log \frac{\sum_{t_1\in\mathbb{N}} \left[\frac{1}{2}P^+(t_1 = T_1 = T_2|d_1 = H^+) + P^+(t_1 = T_1 < T_2|d_1 = H^+)\right]}{\sum_{t_1\in\mathbb{N}} \left[\frac{1}{2}P^-(t_1 = T_1 = T_2|d_1 = H^+) + P^-(t_1 = T_1 < T_2|d_1 = H^+)\right]}. \tag{5}$$

Now, if agent 1 has made an incorrect decision, one inconsistent with the true hypothesis, both this agent's common and independent observations are likely to support the incorrect decision. But, by assumption, any randomly sampled observation is more likely to be consistent with the true than the wrong hypothesis. Thus, the independent observations of agent 2 are likely to point to the correct hypothesis, conflicting with the common observations supporting the incorrect decision of agent 1. As a result, agent 2 more likely decides after T_1 when agent 1's choice is wrong than when it is correct. This argument shows that we expect

$$\begin{aligned} &\sum_{t_1\in\mathbb{N}} \left[\frac{1}{2}P^+(t_1 = T_1 = T_2|d_1 = H^+) + P^+(t_1 = T_1 < T_2|d_1 = H^+)\right] \\ &\quad < \sum_{t_1\in\mathbb{N}} \left[\frac{1}{2}P^-(t_1 = T_1 = T_2|d_1 = H^+) + P^-(t_1 = T_1 < T_2|d_1 = H^+)\right] \end{aligned} \tag{6}$$

for $0 < c < 1$, so that Eq. (5) implies $\mathrm{LLR}(FD = 1|d_1 = H^+) < 0$ for such values of c. As a result, $\mathrm{LLR}(d_{FD} = H^+) < \mathrm{LLR}(d_i = H^+) = \theta$ for $i = 1, 2$ and $0 < c < 1$, so the first decider makes a correct choice less often than an agent chosen at random. This argument can be extended to $N > 2$ agents in most cases, demonstrating an increased probability and volume of contrary evidence in more remaining undecided agents, causing a larger drop in the first decider's accuracy (see Appendix C).

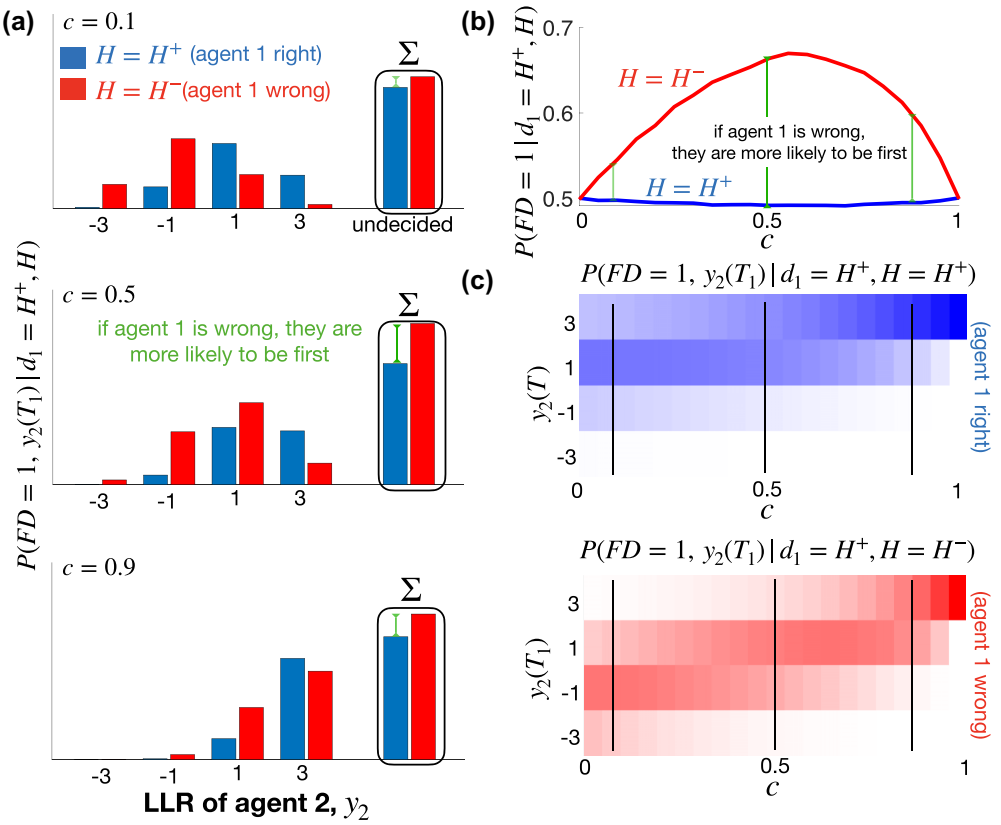

FIG. 3. When evidence is partially correlated, a randomly selected agent is more likely to be the first decider if they are wrong. (a) Joint distribution of the probability that agent 1 decides first ($FD = 1$) and the belief of agent 2 at that decision time, $y_2 = y_2(T_1)$, conditioned on agent 1 being right ($d_1 = H^+ = H$, blue) or wrong ($d_1 = H^+ \neq H$, red). When $0 < c < 1$, the accuracy of the first decider is strictly below that of a randomly selected agent (here agent 1, WLOG) because of inequality (6). When c is small, $P^+(FD = 1|d_1 = H^+)$ nearly equals $P^-(FD = 1|d_1 = H^+)$ (difference indicated by green line), since the joint distributions are approximately reflections of one another, i.e., $P^+(FD = 1, y_2(T_1)|d_1 = H^+) \approx P^-(FD = 1, -y_2(T_1)|d_1 = H^+)$, with equality holding when $c = 0$. As c increases, the difference $P^-(FD = 1|d_1 = H^+) - P^+(FD = 1|d_1 = H^+)$ first grows ($c = 0.5$) and then shrinks ($c = 0.9$), as both terms converge to $1/2$ as $c \to 1$. As discussed in Appendix A, each observation changes an agent's belief, y_j, by ± 1; e.g., when $FD = 1$ and $y_1(T_1) = \pm 3$, then $y_2(T_1)$ is also an odd integer. (b) The probability that agent 1 decides first (conditioned on $d_1 = H^+$ and $H = H^-$) as a function of c peaks around $c = 0.5$. (c) Colormap of the joint distributions from (a) as functions of c. Here we used the discrete LLR model (2) with $\theta = 3$ and binary likelihood functions $f_\pm$ as described in Appendix A. $f_\pm$ are chosen so the update size is ± 1.

Moreover, as c increases, so does the fraction of wrong common observations that can be countered by correct independent observations of agent 2. This initially increases the likelihood that agent 2 remains undecided following incorrect decisions by agent 1. But if c is high, most observations are common, and agent 2 makes few independent observations. Thus, as c approaches 1 the agents' beliefs tend to evolve more synchronously, and the difference between the left and right sides of inequality (6) decreases. This tension between the increase, with c, in the fraction of wrong common observations that are likely to be counteracted, and the decrease in the fraction of correct independent observations that can counteract them causes Eq. (5) to achieve an internal minimum, $0 < c_{\min} < 1$.

Numerical experiments support this explanation. Figure 3 illustrates the case of two agents, each with decision threshold magnitude $\theta = 3$. As our argument predicts, $P^+(FD = 1|d_1 = H^+) < P^-(FD = 1|d_1 = H^+)$ for all $0 < c < 1$ [Fig. 3(b)]. Further, the difference $P^-(FD = 1|d_1 = H^+) - P^+(FD = 1|d_1 = H^+)$ first grows and then shrinks as c increases, due mainly to the unimodalilty of the conditional probability that agent 1 decides first when their choice is wrong, $P^-(FD = 1|d_1 = H^+)$. Looking at the joint conditional probabilities of $FD = 1$ and the belief of agent 2 at the time of the first decision, $P^+(FD = 1, y_2(T)|d_1 = H^+)$ and $P^-(FD = 1, y_2(T)|d_1 = H^+)$ helps illuminate the situation. Figure 3(a) shows these joint distributions for representative values of c with $\theta = 3$. The distribution of beliefs, y_2, concentrates more on values $y_2(T) = \pm 1$, away from the thresholds, when $H = H^-$ than when $H = H^+$ for intermediate values of c [Fig. 3(c)].

2. Two agents with decision threshold magnitude $\theta = 2$

We now discuss the case where $\theta = 2$, allowing us to compute exact expressions for Eq. (5), since two measurements are sufficient for belief magnitude to reach the bound. As in Appendix A, we assume there can only be two measurement values ($\xi^\pm$), and $f_\pm(\xi^\pm) = p_+ = e/(e+1)$ and $f_\pm(\xi^\mp) = 1 - p_+ \equiv p_-$, so beliefs are restricted to the integer lattice.

Setting thresholds to $\pm\theta = \pm 2$, the belief of any undecided agent, i, must equal $y_t^i = \pm 1$, at any odd time, and $y_t^i = 0$ at any even time. Thus, the stochastic process governing the evidence accumulation of undecided agents resets to 0 (renews) every two time steps. If T is the time of the first decision, then

$$P(d_{FD} = H^{\pm}, T = t | T > t-2) = P(d_{FD} = H^{\pm}, T = 2)$$

for all even $t > 0$, since if $T > t-2$ then at time $t-2$ both agents must have been undecided with beliefs $y_{t-2}^i = 0$.

We now enumerate and sum the probabilities of all cases in which agent 1 (not necessarily the first decider) makes decision $d_1 = H^+$ under either condition, $H = H^{\pm}$. There are four ways for the two agents to make a simultaneous decision: If $d_1 = H^+$, agent 2 can make the same decision ($d_2 = H^+$) given zero, one, or two independent measurements, or the opposite choice ($d_2 = H^-$), if they made two independent measurements. Therefore,

$$\begin{aligned} &P^{\pm}(T_1 = T_2 = t | d_1 = H^+, T > t-2) \\ &\quad = c^2 + 2c(1-c)p_{\pm} + (1-c)^2(p_+^2 + p_-^2). \end{aligned}$$

The second agent may remain undecided at the time of the first agent's decision if they made one independent measurement that conflicts with the first agent's decision, or two independent measurements that conflict with each other:

$$\begin{aligned} &P^{\pm}(T_1 = t < T_2 | d_1 = H^+, T > t-2) \\ &\quad = 2c(1-c)p_{\mp} + 2(1-c)^2 p_+ p_-. \end{aligned} \tag{7}$$

Now let $m \in \mathbb{N}$ and $t_1 = 2m$. Referring to the sums in Eq. (5), we have

$$\begin{aligned} &P^{\pm}(t_1 = T_1 = T_2 | d_1 = H^+) \\ &\quad = P^{\pm}(t_1 = T_1 = T_2 | d_1 = H^+, T > t_1 - 2) \\ &\qquad \times P^{\pm}(T > t_1 - 2 | d_1 = H^+) \\ &\quad = P^{\pm}(t_1 = T_1 = T_2 | d_1 = H^+, T > t_1 - 2) P(T > t_1 - 2) \\ &\quad = P^{\pm}(t_1 = T_1 = T_2 | d_1 = H^+, T > t_1 - 2)[P(T > 2)]^{m-1}. \end{aligned}$$

A similar calculation gives

$$\begin{aligned} &P^{\pm}(t_1 = T_1 < T_2 | d_1 = H^+) = P^{\pm}(t_1 = T_1 < T_2 | d_1 \\ &\quad = H^+, T > t_1 - 2)[P(T > 2)]^{m-1}. \end{aligned}$$

We factor common terms out of the sums in Eq. (5) and cancel sums over factors of $P(T > 2)^{m-1}$ in the numerator and denominator to obtain an explicit form of Eq. (5),

$$\begin{aligned} &\mathrm{LLR}(FD = 1 | d_1 = H^+) \\ &\quad = \log \frac{[c^2 + 2c(1-c)(1+p_-) + (1-c)^2(1+2p_+p_-)]}{[c^2 + 2c(1-c)(1+p_+) + (1-c)^2(1+2p_+p_-)]}. \end{aligned} \tag{8}$$

The numerator and the denominator in this expression differ only in the middle terms, $2c(1-c)(1+p_-) < 2c(1-c)(1+p_+)$ for $0 < c < 1$, which is the probability that agent 2 makes an independent observation that counters the agents' common observation, in agreement with our general explanation. As discussed previously, this is more likely when the decision of the first agent (and the common measurement) is wrong.

3. Macroscopic case

Our results for the discrete model extend to agents with continuously evolving beliefs, obtained in the limit of many weak observations (see Appendix C). Agents' beliefs, $y_j(t)$, each evolve according to Eq. (3) until crossing a threshold $\pm\theta$, determining the choice $d_j = H^{\pm}$ and decision time $T_j \in (0, \infty)$ for $j = 1, \ldots, N$. Define $\mathbf{T} = (T_1, \ldots, T_N) \in (0, \infty)^N \equiv \mathbb{R}_+^N$. For finite N and $c < 1$, the probability that two agents decide at the same time is zero, so we need not account for simultaneous decisions. By marginalizing over all agents and decision times, we obtain

$$\begin{aligned} &P^{\pm}(d_{FD} = H^+) = N P^{\pm}(d_1 = H^+) \\ &\quad \times \int_{\mathbb{R}_+^N} P^{\pm}(FD = 1 | \mathbf{T} = \boldsymbol{t}, d_1 = H^+) g^{\pm}(\boldsymbol{t} | d_1 = H^+)\, d\boldsymbol{t}. \end{aligned}$$

Here $g^{\pm}(\cdot | d_1 = H^+)$ is the conditional probability density function for $\mathbf{T}$, conditioned on the state, $H = H^{\pm}$, and on the decision $d_1 = H^+$. We have $P^{\pm}(FD = 1 | \mathbf{T} = \boldsymbol{t}, d_1 = H^+) = 1$ if $t_1 = \min_{1 \leqslant i \leqslant N} t_i$ and otherwise this quantity is zero, simplifying the multi-dimensional integral in the preceding expression to an integral over the t_1 axis and allowing us to write

$$\mathrm{LLR}(d_{FD} = H^+) = \theta + \log \frac{\int_{\mathbb{R}_+} G^+(t_1 | d_1 = H^+)\, dt_1}{\int_{\mathbb{R}_+} G^-(t_1 | d_1 = H^+)\, dt_1},$$

where the second term is the log of the ratio of the probabilities that all other agents are undecided at the time at which agent 1 chooses H^+. Terms in the ratio are obtained by integrating the probability density

$$\begin{aligned} &G^{\pm}(t_1 | d_1 = H^+) \\ &\quad = \int_{(t_1, \infty)^{N-1}} g^{\pm}(t_1, t_2, \ldots, t_N | d_1 = H^+)\, dt_2 \cdots dt_N, \end{aligned}$$

across all possible times of the decision of agent 1, given that agent 1 chooses H^+.

When $N = 2$, the nonmonotonicity of the first decider's accuracy in c is due to the tension between opportunity for contradiction in agent 2's observations and the decreasing prevalence of independent observations, as c increases. The densities $\frac{d}{dz} P(FD = 1, y_2(T) \leqslant z | d_1 = H^+, H)$ are nearly reflections of one another for small c [Fig. 4(a), top left]. Integrating over z, the difference $P^-(FD = 1 | d_1 = H^+) - P^+(FD = 1 | d_1 = H^+)$ is small when c is small [red bar minus blue bar, Fig. 4(a), top center]. For intermediate values of c, the distribution of beliefs of agent 2 is pulled away from the correct threshold when agent 1 decides incorrectly, due to common observations, causing $P^-(FD = 1 | d_1 = H^+) - P^+(FD = 1 | d_1 = H^+)$ to reach a maximum within the intermediate c range. When c is close to 1, both $P^-(FD = 1 | d_1 = H^+)$ and $P^+(FD = 1 | d_1 = H^+)$ converge to 1/2, so the difference converges to zero. Figure 4(b) shows that the unimodal response of first-decider accuracy as c increases occurs because the probability $P^-(FD = 1 | d_1 = H^+)$ of an incorrect agent deciding first increases for small c and then decreases in c (red curve), while $P^+(FD = 1 | d_1 = H^+)$ is approximately insensitive to c (blue curve).

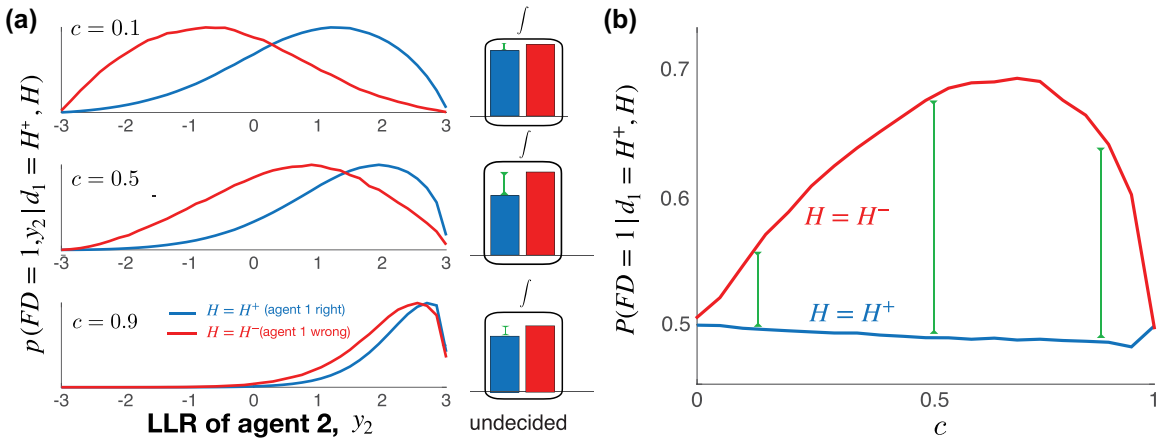

FIG. 4. When beliefs evolve continuously and evidence is correlated, a randomly selected agent is again more likely to decide first if they are wrong. (a) As in the discrete model, the densities $p(FD = 1, y_2|d_1 = H^+, H)$ are nearly reflections of one another for small c. By marginalizing over the distribution of beliefs, y_2, we can obtain the difference $P^-(FD = 1|d_1 = H^+) - P^+(FD = 1|d_1 = H^+)$. This difference is small when c is small (red bar minus blue bar). As c increases, this difference first increases and then decreases, the latter because each term in the difference converges to $1/2$ as $c \to 1$. (b) The unimodal response of first-decider accuracy as c increases is again due to $P^-(FD = 1|d_1 = H^+)$ obtaining a maximum around $c = 0.5$. We used the macroscopic model (3) with unit drift and variance and threshold $\theta = 3$.

C. Pooling over early deciders does not rescue accuracy

The "wisdom of crowds" is the idea that collective decision by a group of people is more likely to be correct than the decision of any single member of the group [23,38]. A group's decision accuracy can be improved when individuals exchange information preceding their final decisions or when the group decision is determined by the majority of individual choices [17,23,27,39,40]. However, this improvement can be diminished, and individuals can even outperform crowds when biases in individual decisions are not accounted for when forming the group decision [41,42]. Applying a majority rule to an initial pool of early deciders, we show that even modest correlations in information can cause this pool to make less accurate choices than a randomly selected agent and only slightly improves on the accuracy of the first decider [Figs. 5(a) and 5(b)]. The additional time required to obtain these additional opinions is appreciable and roughly independent of the population size, N [Fig. 5(c)]. Hence, even weak correlations in evidence impact the accuracy of collective decisions.

IV. DISCUSSION

Humans and other animals integrate evidence to make decisions. Often members of a group or community are faced with the same choices and will use evidence that is available to all of them to decide between a common set of options [43–45]. We have shown that when some observations are made in common, even when no social information is exchanged, the first individual to decide makes the least accurate decision. The accuracy of subsequent decisions increases in the order in which they are made, with few exceptions.

We have focused on agents deciding between two options, so that response accuracy can be computed as exit probabilities of populations of univariate stochastic processes

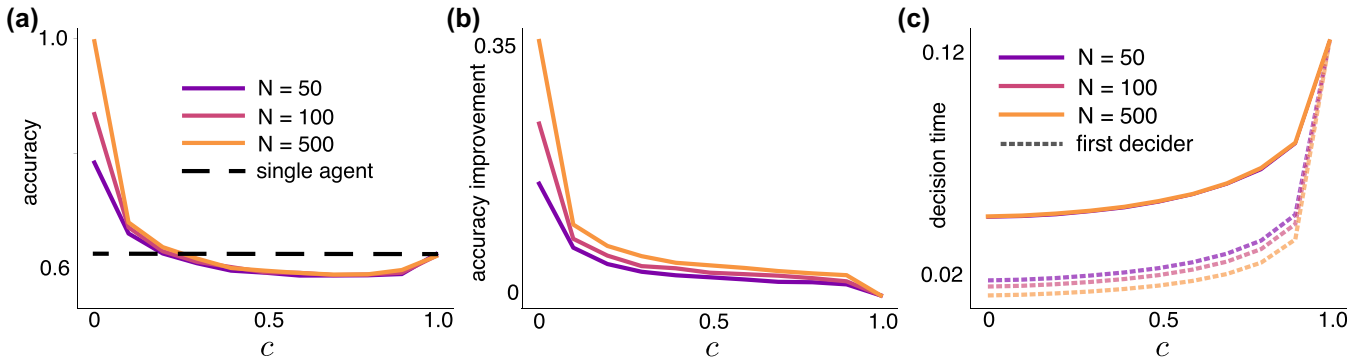

FIG. 5. Pooling choices of early deciders using a majority rule mildly improves accuracy compared to the first decision when evidence is correlated. (a) The group's decision is determined by the majority of the first N_{pool} deciders. For different population sizes, N, the accuracy of the group decision at first decreases as c is increased and can be lower than the accuracy of a single decider in isolation (dashed line). $N_{\text{pool}} = 0.2N$. (b) Improvement in the accuracy obtained by pooling the first $N_{\text{pool}} = 0.2N$ decisions compared with the accuracy of the first decision drops substantially even for small values of c and is nearly independent of N. (c) The mean time at which the last decider in the pool makes a decision increases with c (solid curves). Dashed curves give the mean time of the first decision. We used the macroscopic LLR (3) with threshold $\theta = 10$.

driven by common and independent noise [46]. The accuracy of the first agent to make a decision depends nonmonotonically on the probability c of making a common measurement. When the accuracy of the first decision is at a minimum, roughly half the observations are common. The remaining independent observations allow the agent's beliefs to diverge, leading the first agent to often choose differently than later deciders.

A similar result holds for groups of observers who have a common initial bias and integrate independent evidence [22]. If there are many such agents, the first decision will almost always correspond to the decision boundary closest to the initial belief [37] and is thus wholly determined by the initial common bias.

We made the simplifying assumption that all agents either jointly make a common observation or all make private observations on each time step. This requires a coordinated measurement process, which is counter to our assumption that agents do not share social information. We could relax this assumption and allow agents to each independently make measurements from two sources, one common to the group and one available only to the agent. With two agents this model is equivalent to the model we analyzed. More generally, different subsets of agents could have access to separate sources of shared information, rather than a single common source available to the entire community. The analysis of these cases becomes more cumbersome, but we expect that our general conclusions will hold.

Agents could also share their decisions, in which case the fact that no decision has been announced up to a time t can be informative. The first agent to reach threshold will know that no other decisions have been made yet. This reveals that other agents have gathered independent evidence that disagrees with the first agent's choice. The first agent can take this information into account reducing their belief that what they thought is the correct choice holds true. Similar reasoning can lead to intricate social information exchanges [22]. However, humans frequently exhibit correlation neglect [47]. If observers assume that information is uncorrelated, then the model we described here may be applicable even when they observe each other's decisions.

We have assumed that the agents in the population are identical. If agents have different decision thresholds, early decisions tend to be driven by less evidence [2], generating a decrease in accuracy unrelated to the effect of common observations. Correlated evidence could exacerbate this decrease in accuracy. However, if agents have access to information of different quality, early deciders tend to be those with access to the best information [27]. In this case early decisions can be more accurate than later ones. We expect correlated evidence to still impact the accuracy of the first decision, but the specifics would depend on the quality of common and private evidence.

Except for limiting cases, we found it quite cumbersome to obtain analytical expressions for the accuracy of the first decider and other statistics of the agents' decisions. However, prior work has shown that the correlated drift diffusion model generated in the macroscopic limit can be solved explicitly using method of images solutions for specific threshold values [35]. In our case thresholds always form a square domain encompassing both agents' beliefs for $N = 2$ or cubes or hypercubes for $N > 2$, but method of images approach may still be applicable.

Like other mathematical models of cognition, our model only roughly approximates decision-making processes used by humans and animals. Despite its limitations, we believe that our analysis offers important insights independent of the exact way evidence is integrated and decisions formed in groups: Common observations drive the beliefs of individuals in the community in the same direction. If those common observations are misleading, it takes time for private evidence to counter their effect. When deciders use a substantial fraction of common observations to make their decisions, early decisions are most likely consistent with common observations. Thus, if common observations are right (wrong), the first decision tends to be as well. First decisions thus tend to be based predominantly on common evidence, which offers less information than what is implied by the decision threshold. We expect that the resulting asymmetric weight of common evidence in determining the first decision leads to similar effects more generally, e.g., when the population is heterogeneous, faced with more than two choices, or when observations are made asynchronously. Social information exchange would lead to more subtle effects, modulating the impact of common measurements. We have thus described a general mechanism that can affect group decision-making, with implications that transcend specific scenarios. The insights we provided can describe decision-making processes across a range of contexts and could be used to organize and guide more effective individual and group choices.

ACKNOWLEDGMENTS

Z.P.K. and K.J. were supported by NSF DMS-2207700. Z.P.K. was also supported by NIH BRAIN 1R01EB029847-01. K.J. was also supported by NSF NeuroNex Grant No. DBI-1707400. W.O. was supported by NSF DMS-1816315.

APPENDIX A: BELIEFS EVOLVING ON THE INTEGER LATTICE

In the simplest case we can assume that an observer can make only two measurements, ξ^+ and ξ^-. We let $P(\xi^\pm|H^\pm) = p_+$ and $P(\xi^\pm|H^\mp) = p_-$ with $p_+ + p_- = 1$ and $p_- < p_+$. Assuming $p_- = p_+/e$, gives $p_+ + p_+/e = 1$ so $p_+ = e/(1+e)$, $p_- = 1/(1+e)$, and hence, $\mathrm{LLR}(\xi^\pm) = \pm 1$. Binarized evidence samples $\xi^\pm$ then increment or decrement each agent's belief y_t^i by one, so the sum of an even (odd) number of odd numbers, ± 1, will be even (odd). In particular, when $p_- = p_+/a$ the information provided by observation $\xi^\pm$ equals $\pm \ln a$. As a result, the belief of each agent, y_i, lies on a lattice defined by $\{n \ln a\}_{n\in\mathbb{Z}}$, and we can use the mapping $n \to n \ln a$ or a logarithm in base a to place beliefs on an integer lattice. In the double limit of infinitesimal evidence, $\lim_{a\to 1^+} \ln a$, and infinitesimal time between observations, we can recover a continuous-time model.

APPENDIX B: ALTERNATIVE DEFINITIONS OF THE FIRST DECIDER

In the main text we defined the "first decider" as an agent chosen with equal probability from the set of all agents who

reach threshold at the same time. Alternatively, we could pool all first deciders across trials and compute the probability an agent in this entire pool makes a correct choice. In the scaling limit, the probability that multiple agents reach the threshold at the same time converges to zero, and the two definitions are equivalent. However, when evidence increments are finite, multiple agents can decide first simultaneously. In that case choosing the first decider within a trial and pooling across trials gives different results.

APPENDIX C: DERIVATION OF THE SCALING LIMIT

Let $f_+(\xi)$ be a probability distribution of observations, ξ, over an arbitrary set Ξ obtained in state H^+, and $f_-(\xi)$ the probability distribution of observations over that same set Ξ in state H^-. Note if the sets of observations for either state differ, then there will be infinitely informative observations which, when observed, would immediately make an agent certain of the state. However, these occurrences could be rare, in which case an accumulation process would still be needed. As previously, we use y to denote accumulated LLR so that in the discrete case we have

$$y(t) = \sum_{s \leqslant t} \mathrm{LLR}(\xi_s), \tag{C1}$$

where ξ_s is the observation obtained at time $s \leqslant t$. Similarly, in the continuous case

$$y(t) = \int_0^t \frac{dy(s)}{ds}\, ds, \tag{C2}$$

where $\frac{dy}{ds}$ is given by the stochastic drift-diffusion equation described previously.

We again assume that in a group of N observers each observer at each time step t makes an independent private observation with probability $1-c$, and all observers make a common observation with probability c. Private and common observations have the same conditional distributions, $f_\pm(\xi)$ given the state $H^\pm$.

For observations drawn from such general likelihood functions, we can determine the statistics of the limiting stochastic accumulation process by averaging the impact of multiple "subobservations" on short intervals which we shrink to be infinitesimal. Focusing on a single observer i, define a family of stochastic processes parameterized by k, the number of subobservations made in an interval of length Δt. Thus, we expect the LLR increment obtained each Δt is given

$$\begin{aligned}\Delta y_t &= \sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)} \\ &= \mathrm{E}_\xi\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)}\middle| H\right] \\ &\quad + \left(\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)} - \mathrm{E}_\xi\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)}\middle| H\right]\right).\end{aligned}$$

We can split the sum not contained in an expectation into those observations drawn from the common pool and those not,

$$\Delta y_t = \mathrm{E}_\xi\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)}\middle| H\right] + \left(\sum_{l=1}^{k_c} \log \frac{f_+\left(\xi_{i,t}^{l,c}\right)}{f_-\left(\xi_{i,t}^{l,c}\right)} + \sum_{l=1}^{k-k_c} \log \frac{f_+\left(\xi_{i,t}^{l,n}\right)}{f_-\left(\xi_{i,t}^{l,n}\right)} - \mathrm{E}_\xi\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)}\middle| H\right]\right),$$

where $\xi_{i,t}^{l,c}$ are samples the ith agent sees from the common pool and $\xi_{i,t}^{l,n}$ are those they see from the independent pool. For large k while keeping Δt fixed, we know the number of common observations will scale as $k_c \approx ck$, so assigning

$$\pm\mu\Delta t \equiv \mathrm{E}_\xi\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)}\middle| H = H^\pm\right],$$

assuming $f_\pm(\xi)$ are scaled appropriately as $\Delta t \to 0$. We then estimate the variability in the incremental process as $k \to \infty$ by computing

$$\begin{aligned}&\left\langle\left(\sum_{l=1}^{k_c} \log \frac{f_+\left(\xi_{i,t}^{l,c}\right)}{f_-\left(\xi_{i,t}^{l,c}\right)} + \sum_{l=1}^{k-k_c} \log \frac{f_+\left(\xi_{i,t}^{l,n}\right)}{f_-\left(\xi_{i,t}^{l,n}\right)} - \mathrm{E}_\xi\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^l\right)}{f_-\left(\xi_{i,t}^l\right)}\middle| H\right]\right)^2\right\rangle \\ &= \left\langle\sum_{l=1}^{k_c}\left[\log \frac{f_+\left(\xi_{i,t}^{l,c}\right)}{f_-\left(\xi_{i,t}^{l,c}\right)}\right]^2\right\rangle - c\mu^2\Delta t^2 + \left\langle\sum_{l=1}^{k-k_c}\left[\log \frac{f_+\left(\xi_{i,t}^{l,n}\right)}{f_-\left(\xi_{i,t}^{l,n}\right)}\right]^2\right\rangle - (1-c)\mu^2\cdot\Delta t^2 \\ &= c\mathrm{Var}\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^{l,c}\right)}{f_-\left(\xi_{i,t}^{l,c}\right)}\right] + (1-c)\mathrm{Var}\left[\sum_{l=1}^{k} \log \frac{f_+\left(\xi_{i,t}^{l,n}\right)}{f_-\left(\xi_{i,t}^{l,n}\right)}\right].\end{aligned}$$

We can thus approximate the update in the limit of rapid and infinitesimally weak observations using the Donsker Invariance Principle [34]

$$\Delta y_{i,t} \approx \pm\mu\Delta t + \sqrt{\Delta t}(\rho_{1-c,\Delta t}(t)\eta_{1-c} + \rho_{c,\Delta t}(t)\eta_c),$$

where η_c and η_{1-c} are random variables with standard normal distributions, and

$$\pm\mu = \frac{1}{\Delta t}\mathbb{E}_\xi\left[\log\frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})}\middle|H^\pm\right],$$
$$\rho^2_{1-c,\Delta t}(t) = \frac{(1-c)}{\Delta t}\mathrm{Var}_\xi\left[\log\frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})}\middle|H^\pm\right],$$
$$\rho^2_{c,\Delta t}(t) = \frac{c}{\Delta t}\mathrm{Var}_\xi\left[\log\frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})}\middle|H^\pm\right]. \tag{C3}$$

The drift $h_{\Delta t}$ and the variances $\rho^2_{c,\Delta t}$, $\rho^2_{1-c,\Delta t}$ will diverge unless $f_\pm(\xi)$ are properly scaled in the $\Delta t \to 0$ limit.

Taking $\Delta t \to 0$ gives

$$dy = \pm\mu dt + \rho_{1-c}\, dW_i + \rho_c\, dW_c, \tag{C4}$$

where

$$\pm\mu = \lim_{\Delta t\to 0} h_{\Delta t}(t) = \mathbb{E}_\xi\left[\log\frac{f_+(\xi)}{f_-(\xi)}\middle|H^\pm\right],$$
$$\rho^2_c(t) = \lim_{\Delta t\to 0}\rho^2_{c,\Delta t}(t) = c\,\mathrm{Var}_\xi\left[\log\frac{f_+(\xi)}{f_-(\xi)}\middle|H^\pm\right],$$
$$\rho^2_{1-c}(t) = \lim_{\Delta t\to 0}\rho^2_{c,\Delta t}(t) = (1-c)\,\mathrm{Var}_\xi\left[\log\frac{f_+(\xi)}{f_-(\xi)}\middle|H^\pm\right].$$

We note that dW_i corresponds to private noise, which is generated independently for each agent. The term dW_c is common to all agents.

APPENDIX D: EXTENDING THE ANALYSIS OF THE DISCRETE MODEL TO MORE THAN TWO AGENTS

Accuracy of the first decider dips even lower when considering more than two agents $N > 2$ [see Fig. 2(b)]. To explain this more general observation, we extend our two-agent analysis. We denote the decision of agent $j \in \{1,\ldots,N\}$ by d_j and the corresponding decision time by T_j. The probability that the first decider chooses H^+ conditioned on the true state is given by

$$P^\pm(d_{FD} = H^+) = \sum_{j=1}^{N} P^\pm(FD = j|d_j = H^+)P^\pm(d_j = H^+).$$

Leveraging exchange symmetry of distinct agents and defining $\mathbf{T} = (T_1,\ldots,T_N) \in \mathbb{N}^N$ (the vector of decision times) and $T = \min_j T_j$ (the time of the first decision), then

$$P^\pm(d_{FD} = H^+) = NP^\pm(d_1 = H^+)\sum_{\boldsymbol{t}\in\mathbb{N}^N} P^\pm(FD = 1|\mathbf{T} = \boldsymbol{t},$$
$$d_1 = H^+)P^\pm(\mathbf{T} = \boldsymbol{t}|d_1 = H^+), \tag{D1}$$

where the first term in the sum vanishes if $t_1 > \min_{1\leqslant j\leqslant N} t_j$. On the other hand, if $t_1 = \min_{1\leqslant j\leqslant N} t_j$, the conditional probability that agent 1 is chosen as the first decider depends on the number of indices j for which $t_j = t_1$, i.e., the number of agents who simultaneously decide at the time of the first decision. Let $n_{FD}(\boldsymbol{t})$ denote the number of these first deciders. Overall, we have

$$P^\pm(FD = 1|\mathbf{T} = \boldsymbol{t}, d_1 = H^+) = \begin{cases} 0, & t_1 > \min_{1\leqslant j\leqslant N} t_j, \\ 1/n_{FD}(\boldsymbol{t}), & t_1 = \min_{1\leqslant j\leqslant N} t_j. \end{cases}$$

Thus, we can turn the second term within the sum from Eq. (D1) into an additional sum over the count of agents deciding at the first decision time:

$$P^\pm(d_{FD} = H^+) = NP^\pm(d_1 = H^+)\sum_{t_1\in\mathbb{N}}\sum_{k=1}^{N}\frac{1}{k}P^\pm$$
$$\times(t_1 = T_1 = T, n_{FD}(\mathbf{T}) = k|d_1 = H^+).$$

As before, we write the LLR as a sum of two terms, one given by the LLR of a randomly selected agent (agent 1) choosing H^+, $\mathrm{LLR}(d_1 = H^+) = \log[P^+(d_1 = H^+)]/P^-(d_1 = H^+) = \theta$, and a second term involving conditional probabilities that the randomly selected agent is the first decider,

$$\mathrm{LLR}(d_{FD} = H^+) = \mathrm{LLR}(d_1 = H^+) + \mathrm{LLR}(FD = 1|d_1 = H^+),$$

where

$$\mathrm{LLR}(FD = 1|d_1 = H^+) = \log\frac{\sum_{t_1\in\mathbb{N}}\sum_{k=1}^{N}\frac{1}{k}P^+(t_1{=}T_1 = T, n_{FD}(\mathbf{T}){=}k|d_1{=}H^+)}{\sum_{t_1\in\mathbb{N}}\sum_{k=1}^{N}\frac{1}{k}P^-(t_1 = T_1{=}T, n_{FD}(\mathbf{T}){=}k|d_1{=}H^+)}.$$

This term has the same form as in the case of two agents and is negative for $0 < c < 1$ for the same reason: Common observations are likely to be in agreement with the decision of the first decider. However, when the first decider is wrong, independent observations of the other observers are more likely to point away from the first decision threshold than when the first decision is correct. Thus, the first decider is less likely to be correct than a randomly selected agent when $0 < c < 1$, in agreement with simulation results. Moreover, the difference between the numerator and denominator grows with the number of agents, reflecting the additional information provided by having even more undecided agents [Fig. 2(b)]. Other agents will make observations countering the first decision when it is incorrect, and consistent with it when it is correct.

[1] J. I. Gold and M. N. Shadlen, Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward, Neuron **36**, 299 (2002).

[2] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks, Psych. Rev. **113**, 700 (2006).

[3] L. Chittka, P. Skorupski, and N. E. Raine, Speed–accuracy tradeoffs in animal decision making, Trends Ecol. Evol. **24**, 400 (2009).

[4] R. Bogacz, E.-J. Wagenmakers, B. U. Forstmann, and S. Nieuwenhuis, The neural basis of the speed–accuracy tradeoff, Trends Neurosci. **33**, 10 (2010).

[5] R. Ratcliff, A theory of memory retrieval, Psych. Rev. **85**, 59 (1978).

[6] L. Chittka, A. G. Dyer, F. Bock, and A. Dornhaus, Bees trade off foraging speed for accuracy, Nature (London) **424**, 388 (2003).

[7] W. T. Newsome, K. H. Britten, and J. A. Movshon, Neuronal correlates of a perceptual decision, Nature (London) **341**, 52 (1989).

[8] N. Uchida and Z. F. Mainen, Speed and accuracy of olfactory discrimination in the rat, Nat. Neurosci. **6**, 1224 (2003).

[9] J. A. Swets, W. P. Tanner Jr., and T. G. Birdsall, Decision processes in perception, Psych. Rev. **68**, 301 (1961).

[10] J. I. Gold and M. N. Shadlen, The neural basis of decision making, Annu. Rev. Neurosci. **30**, 535 (2007).

[11] T. Bose, A. Reina, and J. A. R. Marshall, Collective decision-making, Curr. Opin. Behav. Sci. **16**, 30 (2017).

[12] A. B. Kao and I. D. Couzin, Decision accuracy in complex environments is often maximized by small group sizes, Proc. R. Soc. B: Biol. Sci. **281**, 20133305 (2014).

[13] A. S. Gerber, D. Karlan, and D. Bergan, Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions, Am. Econ. J. Appl. Econ. **1**, 35 (2009).

[14] T. J. Valone, Group foraging, public information, and patch estimation, Oikos **56**, 357 (1989).

[15] W. Mensi, M. Beljid, A. Boubaker, and S. Managi, Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold, Econ. Model. **32**, 15 (2013).

[16] R. Moreno-Bote, Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators, Neural Comput. **22**, 1786 (2010).

[17] B. Karamched, M. Stickler, W. Ott, B. Lindner, Z. P. Kilpatrick, and K. Josić, Heterogeneity Improves Speed and Accuracy in Social Networks, Phys. Rev. Lett. **125**, 218302 (2020).

[18] A. Wald and J. Wolfowitz, Optimum character of the sequential probability ratio test, Ann. Math. Stat. **19**, 326 (1948).

[19] R. Ratcliff and G. McKoon, The diffusion decision model: Theory and data for two-choice decision tasks, Neural Comput. **20**, 873 (2008).

[20] M. Usher and J. L. McClelland, The time course of perceptual choice: The leaky, competing accumulator model, Psych. Rev. **108**, 550 (2001).

[21] A. Veliz-Cuba, Z. P. Kilpatrick, and K. Josić, Stochastic models of evidence accumulation in changing environments, SIAM Rev. **58**, 264 (2016).

[22] B. Karamched, S. Stolarczyk, Z. P. Kilpatrick, and K. Josić, Bayesian evidence accumulation on social networks, SIAM J. Appl. Dyn. Syst. **19**, 1884 (2020).

[23] N. De Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (Cambridge University Press, Cambridge, 2014).

[24] S. Nitzan and J. Paroush, Optimal decision rules in uncertain dichotomous choice situations, Int. Econ. Rev. **23**, 289 (1982).

[25] P. J. Boland, Majority systems and the Condorcet jury theorem, J. R. Stat. Soc. D (The Statistician) **38**, 181 (1989).

[26] J. A. R. Marshall, G. Brown, and A. N. Radford, Individual confidence-weighting and group decision-making, Trends Ecol. Evol. **32**, 636 (2017).

[27] A. Reina, T. Bose, V. Srivastava, and J. A. R. Marshall, Asynchrony rescues statistically optimal group decisions from information cascades through emergent leaders, R. Soc. Open Sci. **10**, 230175 (2023).

[28] R. J. Caginalp and B. Doiron, Decision dynamics in groups with interacting members, SIAM J. Appl. Dyn. Syst. **16**, 1543 (2017).

[29] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma, Belief consensus and distributed hypothesis testing in sensor networks, in *Networked Embedded Sensing and Control: Workshop NESC'05* (Springer, Berlin, Heidelberg, 2006), pp. 169–182.

[30] A. V. Banerjee, A simple model of herd behavior, Q. J. Econ. **107**, 797 (1992).

[31] E. Mossel and O. Tamuz, Opinion exchange dynamics, Prob. Surveys **14**, 155 (2017).

[32] B. Çelen and S. Kariv, Observational learning under imperfect information, Games Econ. Behav. **47**, 72 (2004).

[33] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, 2001).

[34] P. Billingsley, *Probability and Measure* (John Wiley & Sons, Hoboken, New Jersey, 2008).

[35] H. Shan, R. Moreno-Bote, and J. Drugowitsch, Family of closed-form solutions for two-dimensional correlated diffusion processes, Phys. Rev. E **100**, 032132 (2019).

[36] A. Wald, Sequential tests of statistical hypotheses, Ann. Math. Stat. **16**, 117 (1945).

[37] S. Linn and S. D. Lawley, Extreme hitting probabilities for diffusion, J. Phys. A: Math. Theor. **55**, 345002 (2022).

[38] F. Galton, Vox populi, Nature **75**, 450 (1907).

[39] L. Conradt and T. J. Roper, Consensus decision making in animals, Trends Ecol. Evol. **20**, 449 (2005).

[40] J. Surowiecki, *The Wisdom of Crowds* (Anchor Books, 2005).

[41] J. P. Simmons, L. D. Nelson, J. Galak, and S. Frederick, Intuitive biases in choice versus estimation: Implications for the wisdom of crowds, J. Consumer Res. **38**, 1 (2011).

[42] A. W. Lo and R. Zhang, The wisdom of crowds versus the madness of mobs: An evolutionary model of bias, polarization, and other challenges to collective intelligence, Collect. Intell. **1**, 26339137221104785 (2022).

[43] T. D. Seeley, S. Camazine, and J. Sneyd, Collective decision-making in honey bees: How colonies choose among nectar sources, Behav. Ecol. Sociobiol. **28**, 277 (1991).

[44] S. C. Pratt, E. B. Mallon, D. J. Sumpter, and N. R. Franks, Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant *Leptothorax albipennis*, Behav. Ecol. Sociobiol. **52**, 117 (2002).

[45] I. D. Couzin, Collective cognition in animal groups, Trends Cogn. Sci. **13**, 36 (2009).

[46] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, Berlin, 2009), Vol. 4.

[47] B. Enke and F. Zimmermann, Correlation neglect in belief formation, Rev. Econ. Stud. **86**, 313 (2019).