Fast decisions reflect biases, slow decisions do not

Samantha Linn,[1, *] Sean D. Lawley,[1, †] Bhargav R.
Karamched,[2, 3, 4, ‡] Zachary P. Kilpatrick,[5, §] and Krešimir Josić[6, 7, ¶]

[1]*Department of Mathematics, University of Utah, Salt Lake City, Utah, USA*

[2]*Department of Mathematics, Florida State University, Tallahassee, Florida 32306, USA*

[3]*Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, USA*

[4]*Program in Neuroscience, Florida State University, Tallahassee, Florida 32306, USA*

[5]*Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado 80309, USA*

[6]*Department of Mathematics, University of Houston, Houston, Texas 77004, USA*

[7]*Department of Biology and Biochemistry, University of Houston, Houston, Texas 77004, USA*

(Dated: January 3, 2024)

Abstract

Decisions are often made by heterogeneous groups of individuals, each with distinct initial biases and access to information of different quality. We show that in large groups of independent agents who accumulate evidence the first to decide are those with the strongest initial biases. Their decisions align with their initial bias, regardless of the underlying truth. In contrast, agents who decide last make decisions as if they were initially unbiased, and hence make better choices. We obtain asymptotic expressions in the large population limit that quantify how agents' initial inclinations shape early decisions. Our analysis shows how bias, information quality, and decision order interact in non-trivial ways to determine the reliability of decisions in a group.

1

Evidence accumulation models are used widely to describe how different organisms integrate information to make choices [3]. Experimental evidence shows that these models capture the dynamics of the decision making process of humans and other animals, including the tradeoff between speed and accuracy [5, 25, 30, 34, 37]. Such models can also be used to understand how decisions are made in social groups, both when individuals observe each other's choices [10, 19, 31, 36] and when they act independently [33].

The accumulation of evidence is often modeled using biased Brownian motion with the quality of evidence determining the magnitude of drift and diffusion. An agent is assumed to commit to a decision when the process crosses a threshold. Most previous evidence accumulation models describe a single agent. However, questions remain about how the order of choices in a group is related to their accuracy [40]. In a group of initially unbiased individuals accumulating evidence of different quality, the fastest and most accurate decisions are made by those accessing the highest quality information [31]. Here we ask how the initial biases of individuals in a group impact the order and accuracy of their choices. When is a decision driven mainly by an agent's initial bias as opposed to accumulated evidence?

We show that in large groups of agents starting with different initial biases, early decisions tend to be made by agents with the most extreme predispositions. The choices of these agents agree with their initial bias, regardless of the quality of the evidence they have access to. On the other hand, decisions of late deciders do not depend on their initial bias. Thus, in large groups early decisions reflect only initial inclinations, regardless of which choice is right. Late decisions reflect only accumulated evidence and are more likely to be correct. These effects hold generically, but not in the special case of initially unbiased agents [31].

Model description. We first assume that each individual in a population of N agents has to decide between two choices (hypotheses), H^+ and H^-. They do so by accumulating evidence and computing the conditional probabilities, $P(H^\pm|\text{evidence})$, that one of the two hypotheses is correct. When observations are independent and each provides weak evidence, the log likelihood ratio, or *belief*, of agent i in the group, $X_i = \log(P(H^+|\text{evidence}_i)/P(H^-|\text{evidence}_i))$, evolves approximately as a biased Brownian motion [3, 7] (See Fig. 1A),

$$\mathrm{d}X_i = \mu_i\,\mathrm{d}t + \sqrt{2D_i}\,\mathrm{d}W_i, \tag{1}$$

where the drift, μ_i, and diffusion coefficient, D_i, capture the strength and noisiness of the evidence, respectively [26]. For all agents the correct choice ($H \in H^\pm$) is given by the sign

of the drift ($\text{sign}[\mu_i] = \pm 1$). Eq. (1) is widely used and accurately captures the dynamics of decisions in humans and animals, including variability in response time and the impact of evidence quality and biases on choice [8, 23, 30].

Agents start with an initial bias, $X_i(0)$, reflecting information or assumptions they have about the prior probability of either hypothesis [23]. We denote by y the initial data for a generic agent. Each agent then accumulates evidence, and its beliefs evolve according to Eq. (1). Agent i makes a decision when its belief reaches one of two thresholds, $-\theta < 0 < \theta$, at *decision time* $\tau_i := \inf\{t > 0 : X_i(t) \notin (-\theta, \theta)\}$. This decision, denoted by $d_i = H^\pm$, is determined by the sign of the threshold reached, $\text{sign}[X_i(\tau_i)]$. If decision criteria differ between agents an appropriate rescaling of $X_i(0)$, μ_i, and D_i allows us to assume that all agents use the same thresholds [3].

Agents with the most extreme initial biases decide first. We show that in large groups agents whose initial biases are closest to one of the thresholds make the earliest decisions. We first assume observers are identical except for their initial biases, so that $\mu_i = \mu$ and $D_i = D$ in Eq. (1). We denote by T_i the ith decision time so that $T_1 \leq T_2 \leq \cdots \leq T_N$, where $T_i = \tau_{n(i)}$ and $n(i)$ is the index of the ith agent to decide. Hence, the index of the first decider is $n(1)$.

For simplicity, we assume that each agent starts with one of finitely many initial beliefs, $\{x_0, x_1, \ldots, x_{I-1}\}$, sampled with probability $q_i = P(X_j(0) = x_i)$ for $i = 0, ..., I-1$. The distance of the initial belief x_i to the closest threshold is $L_i = \min\{\theta - x_i, x_i + \theta\}$. Let $i = 0$ be the index of the unique most extreme initial belief held by an agent, so $L_0 < L_i$ for $i \neq 0$.

For a fixed number of initial beliefs, I, the first agent to decide in a large group is the one with the largest initial bias (Fig. 1A), in the sense that

$$P(X_{n(1)}(0) = x_0) \to 1 \quad \text{as } N \to \infty. \tag{2}$$

More precisely, in the Supplemental Material (SM) we show that

$$P(X_{n(1)}(0) = x_i) \sim \eta_i (\ln N)^{(\beta_i - 1)/2} N^{1-\beta_i} \tag{3}$$

as $N \to \infty$ for each $i \neq 0$, where

$$\beta_i = (L_i/L_0)^2 > 1,$$
$$\eta_i = \frac{q_i}{q_0^{\beta_i}} \exp\left(\frac{\sqrt{\beta_i}}{2D}(\mu_i L_0 - \mu_0 L_i)\right)\sqrt{\beta_i \pi^{\beta_i - 1}}\Gamma(\beta_i) > 0,$$
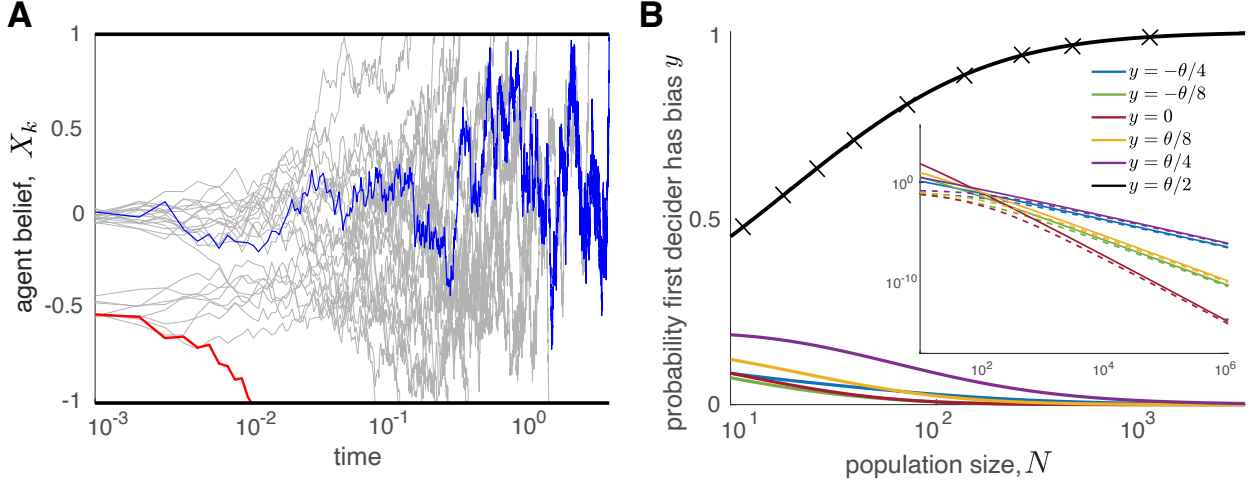
3

FIG. 1. Initial bias determines the choice of early deciders. (A) Evolution of beliefs of $N = 10^4$ agents who each have even odds of initially being unbiased or biased ($P(X_j(0) = x_i) = 0.5$, $x_i = 0, -0.5$). The first agent (red) decides according to their initial bias, and makes the wrong decision at $T_1 \approx 0.01$. The last agent (blue) decides correctly at $T_{10000} \approx 10$. (B) Probability that the agent with the largest initial bias decides first as a function of population size, N. Solid curves were determined by numerical quadrature (Eq. (S4)) with initial biases assigned with uniform probability from values listed in the legend; black crosses denote results of a stochastic simulation averaged over 10^6 trials. Inset: Log-log plot of the same results with dashed curves showing the asymptotic results in Eq. (3). Throughout, agents use identical thresholds $\pm\theta = \pm 1$, drift $\mu = 1$, and diffusivity $D = 1$.

and $\mu_i = \pm\mu$ if $x_i \gtrless 0$. The same statement holds if $n(1)$ is replaced by $n(j)$ in Eq. (3), but with a change in the prefactor, η_i (See SM). Thus, the probability that the first decision is *not* made by the agent with the most extreme initial belief decreases as a negative power of the population size N (Fig. 1B). The approximation given by Eq. (3) is in excellent agreement with the true probabilities when $N \gtrsim 10^3$ (See Fig 1B inset). Moreover, the probability that the agents with the most extreme initial beliefs make the first decision is close to unity already for $N \approx 100$ when initial beliefs are well separated and drift is not too strong.

The choice of the fastest decider agrees with their initial bias: e.g., if θ is the threshold closest to the most extreme initial belief, x_0, then $P(X_{n(1)}(T_1) = \theta) \to 1$ as $N \to \infty$ (See Fig. 2A,B). Similar results hold when initial beliefs are drawn from a continuous distribution (See SM and next section). Thus, although all agents behave rationally, early decisions of
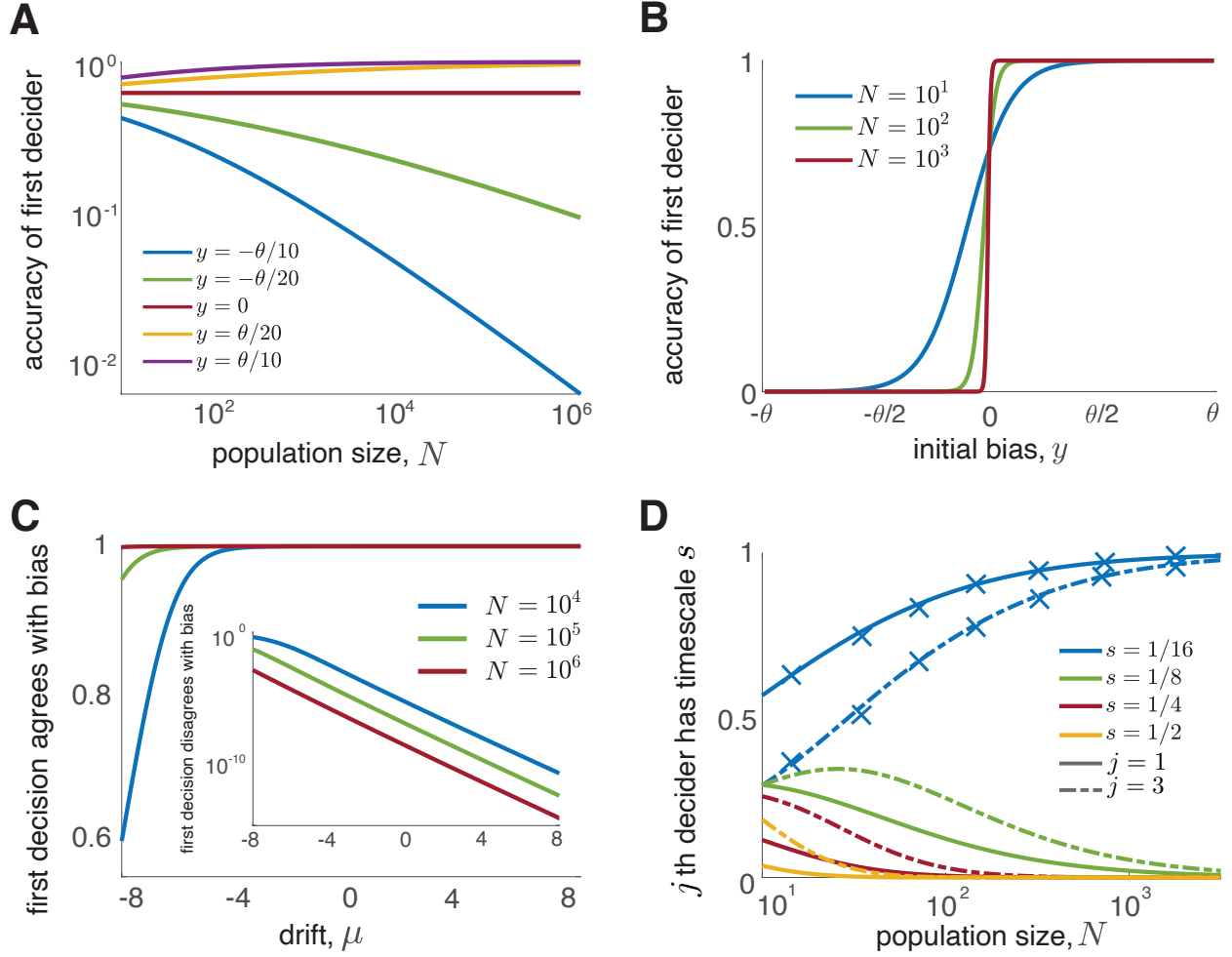
FIG. 2. First decider accuracy is determined by its initial bias. (A) The accuracy of the first decider as a function of population size, N, for different initial biases, y, obtained by quadrature. Curves are ordered by the proximity of the initial bias y of the first decider to the correct threshold $+\theta$. The drift, and hence the correct decision, are positive. (B) Under the same assumptions a small deviation from an unbiased initial belief strongly affects the probability of a correct first decision when N is large. (C) Drift weakly affects the first decision in populations with biased agents ($y = \theta/4$ here) when N is large. See SM for decision polarity formulas. (D) In large populations in which all agents have the same initial bias, $y = \theta/2$, but different diffusivities, early deciders (here first and third) have the shortest diffusive timescale. X's represent averages of stochastic simulations over 10^6 trials.

biased agents tend to be less accurate [11, 33].

In contrast, the probability that a single agent - or one chosen randomly without regard

5

to decision order - decides incorrectly can be made arbitrarily small by increasing the drift or threshold [3]. In large populations with biased agents, drift and diffusion impact the probability of the first decision only through the prefactor in Eq. (3), η_i, and thus decrease in importance as population size diverges. If even a small proportion of a large population holds an initial bias, early decisions are determined by the most extreme bias (Fig. 2B) regardless of the drift (Fig. 2C). On the other hand, if all deciders are initially unbiased ($X_i(0) = 0$ for all i), the probability the first decider makes a correct choice is $(1 + \exp(-\mu\theta/D))^{-1}$ [3].

Heterogeneous population and continuous distribution of initial biases. While we can obtain the most precise asymptotic results in the homogeneous case, our conclusions extend to populations of agents with heterogeneous distributions of initial biases, drifts, diffusivities, and thresholds. We again assume that each agent again starts with one of finitely many initial beliefs, $X_i(0) \in \{x_0, x_1, \ldots, x_{I-1}\}$ with drift and diffusivity sampled from a finite set of fixed size. For each agent we define the diffusive timescale,

$$S_i = \frac{L_i^2}{4D_i} > 0. \tag{4}$$

By assumption, the timescales S_i follow a discrete distribution $\mathbb{P}(S = s_i) > 0$ with support on a finite set $0 < s_0 \leq s_1 \leq s_2 \leq s_3 \cdots \leq s_J$, and $S_{n(j)}$ refers to the timescale of the jth agent to decide (See Fig. 2D). We denote by s the diffusive timescale of a generic agent.

In large populations, early deciders are those with the shortest diffusive timescales. In particular, we show in the SM that for every $\varepsilon > 0$ and *fixed $j \geq 1$*,

$$N^{1-s_1/s_0-\varepsilon} \ll \mathbb{P}(S_{n(j)} > s_0) \ll N^{1-s_1/s_0+\varepsilon} \quad \text{as } N \to \infty, \tag{5}$$

where we use the notation $f \ll g$ to mean $\lim_{N\to\infty} f/g = 0$. We can thus conclude that $N^{1-s_1/s_0-\varepsilon} = o\big(\mathbb{P}(S_{n(j)} > s_0)\big)$ and $\mathbb{P}(S_{n(j)} > s_0) = o\big(N^{1-s_1/s_0+\varepsilon}\big)$ as $N \to \infty$.

These results agree with our earlier conclusion: If all agents share the same diffusivity, then the fastest deciders are the agents who start closest to their decision thresholds. This is true regardless of the quality of the evidence they receive. Diffusivity can reduce the effective distance to the threshold according to Eq. (4). Thus, the fastest deciders are either those with the most extreme initial biases or those with the noisiest integration process, regardless of the drift, μ_i. Indeed, how we model drift does not impact these conclusions, and they hold even if we model the evolution of beliefs as an Ornstein-Uhlenbeck process, as is frequently done in the psychophysics literature [4, 38].
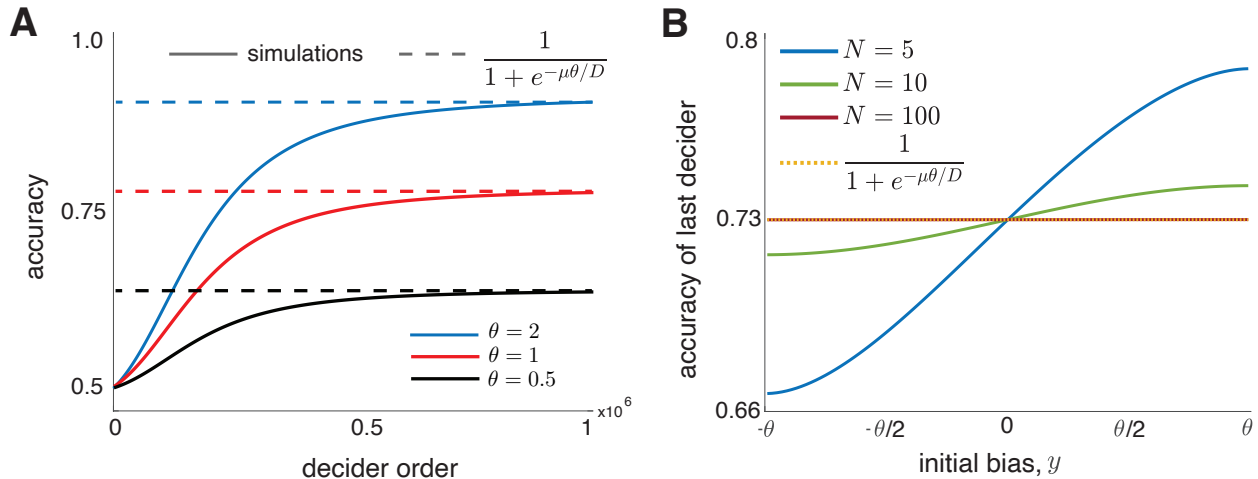
6

FIG. 3. Late deciders make choices as if they held no initial bias. (A) For large N, decision accuracy monotonically increases with decision order. The accuracy of late deciders approaches the accuracy of a single, initially unbiased agent. Here, all agents have initial bias $\theta/3$, and on each trial, $\mathbb{P}(H = H^+) = 0.5$. (B) In large groups even a large initial bias has no impact on the decision of later agents. Here, initial biases are sampled with uniform probability from $(-\theta, \theta)$.

Late deciders make decisions as if initially unbiased. We expect in large populations the inaccuracy of early deciders to be balanced by higher accuracy of late deciders [33]. Thus, we next determine the probability that the last agent to decide makes a correct decision. In the SM we show that this probability has an intuitive form,

$$\mathbb{P}(X_{n(N)}(T_N) = \theta) \to \int_{-\theta}^{\theta} p_\theta(x) q(x) \, dx \quad \text{as } N \to \infty. \tag{6}$$

Here $p_\theta(x)$ is the probability that a single agent with initial bias $X(0) = x$ makes a correct decision, and $q(x)$ is the quasi-steady state distribution [27] of beliefs evolving according to Eq. (1). Thus the decision of the last decider is made as if they forget their actual initial bias and instead sample an initial belief from the quasi-stationary distribution, $q(x)$.

Eq. (6) is general and can be extended to arbitrary domains. When applied to the drift-diffusion process with decision boundaries at $\pm\theta$ we show in the SM that $\mathbb{P}(X_{n(N)}(T_N) = \theta) \to (1+\exp(-\mu\theta/D))^{-1}$ as $N \to \infty$, which is the probability that a single, initially unbiased decider makes a correct decision (See Fig. 3A) [3]. Thus, the last decider forgets their initial bias and makes decisions based only on the accumulated evidence. The probability that an agent with a large initial bias makes a late decision is small. But should this happen, the
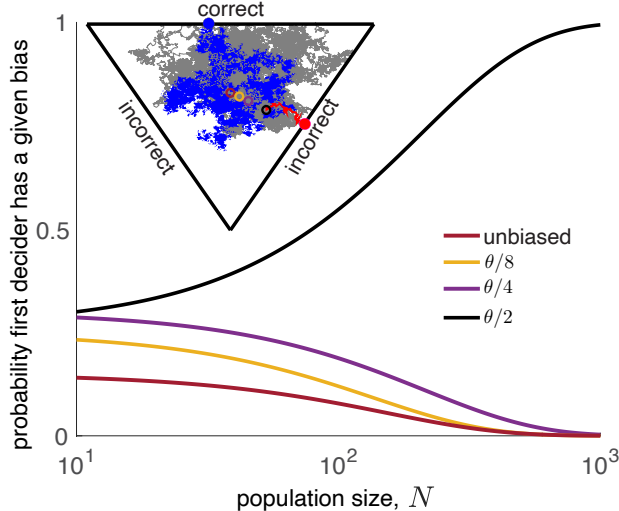
FIG. 4. Bias impacts multi-alternative and two-alternative decisions similarly in large groups. Beliefs about three options evolve on an equilateral triangle. Here, θ is the closest distance from the center of the triangle (burgundy ring) to the boundary. The initial bias is the distance from the triangle center to the initial belief, $X_i(0)$. As N increases, the probability that the most biased agent chooses first grows. Curves are computed by averaging 10^6 stochastic simulations. Inset: Sample trajectories from a trial with biases sampled with equal probability from $\{\theta/2, \theta/4, \theta/8\}$. The first agent to decide (red) has the largest initial bias. The belief of the last decider (blue) explores the space before reaching a threshold.

initial bias will have little impact on their decision (See Fig. 3B).

Extension to multiple alternatives. We can extend these results to decisions between k alternatives. Eq. (1) again describes the evolution of beliefs, but now $X_i(t), \mu_i \in \mathbb{R}^{k-1}$ and W_i is a vector of independent Wiener processes [28]. Each belief evolves on a domain, $\Omega \subset \mathbb{R}^{k-1}$, with k boundaries [21], each associated with one of the alternatives. Agent i chooses alternative j if its belief, $X_i(t)$, crosses the associated boundary first. The boundaries that lead to the best decisions are difficult to find analytically [35], but their exact shape is immaterial for our result.

In the SM we show that Eq. (3) holds for general domains in arbitrary dimensions (See Fig. 4). We therefore reach our earlier conclusions: In large homogeneous populations, the agents holding the most extreme initial beliefs make the first decisions, and their choices are consistent with their initial biases. Our conclusions about the late decisions also carry

over to agents facing multiple choices: The natural extension of Eq. (6) holds with $q(x)$ the quasi-stationary distribution on Ω. The last decider makes a choice as if it sampled its initial belief from this quasi-stationary distribution.

Discussion Our decisions are often influenced by information we obtained previously and predilections we develop. In drift-diffusion models, prior evidence and initial inclinations are often represented by a shift in the initial state. We have shown that initial biases determine early decisions and have a diminishing impact on later decisions.

An agent unaware of the order of their decision would believe this decision was made according to the evidence the agent accumulated and that the accuracy of their choice is determined only by the decision threshold [3]. Though early decisions are not always necessarily less accurate [10], our work identifies a clear case in which hasty choices tend to be the most unreliable. Our findings also suggest a means of weighting choices of biased agents according to decision order in a large group when formulating collective decisions [20]. However, in social groups the exchange of social information between agents [1, 22] or correlations in the evidence [33] will affect these results.

Ramping activity of individual neurons during decision making has been observed across the brain [12, 32] (although see [14]). Such dynamics may reflect the underlying evidence accumulation process preceding a decision and is often modeled by a drift-diffusion process. Decisions are thought to be triggered by the elevated activity of sufficiently many choice-related neurons [39]. Our results suggest that in large neural populations decisions reflect the most extreme initial neural states, rather than the accumulated evidence, if the activity is uncorrelated. Since neural activity is often correlated [6], the effect of such biases could be tempered.

While we have interpreted our results in the context of social decision theory, they apply more generally to independently evolving drift-diffusion processes on bounded domains [17]: In large populations early threshold crossings reflect only the initial states, while late crossings are independent of initial states and reflect the quasi-stationary distribution. Hence, early crossings reflect initial biases providing fast reactions needed for deadlined biophysical processes [9]. If time allows, quorum sensing processes that weight passages by order could be used [13]. Thus, our theory shows how initial biases can be used to implement population level tradeoffs between speed and accuracy.

ACKNOWLEDGEMENTS

Supplemental Material

MATHEMATICAL PRELIMINARIES

Suppose $\{(\tau_n, Z_n)\}_{n \geq 1}$ is an independent and identically distributed (iid) sequence of realizations of the pair of (possibly correlated) random variables (τ, Z). We have in mind that τ is the decision time (or first passage time (FPT)) of some decider whose stochastic evolution of beliefs is denoted by $\{X(t)\}_{t \geq 0}$ and Z is a vector containing information about this decider, such as their random initial position, drift, diffusivity, and decision made. Define the cumulative distribution function (CDF) of τ,

$$F(t) := \mathbb{P}(\tau \leq t).$$

Further, for any event E that is in the σ-algebra generated by Z, define

$$F_E(t) := \mathbb{P}(\tau \leq t \cap E).$$

In words, E is any event for which we can know whether or not it occurred by knowing Z. For example, we are interested in events E like $E = \{X(0) = \theta/2\}$, $E = \{X(0) \leq 0\}$, $E = \{X(\tau) = \theta\}$, etc.

For a given $N \geq 1$, let $n(j) \in \{1, \ldots, N\}$ denote the (random) index of the jth fastest decider out of the first N deciders to make a decision. That is, suppose we order the first N FPTs (or first decision times),

$$T_{1,N} \leq T_{2,N} \leq \cdots \leq T_{N-1,N} \leq T_{N,N},$$

where $T_{j,N}$ denotes the jth fastest FPT,

$$T_{j,N} := \min \left\{ \{\tau_1, \ldots, \tau_N\} \setminus \cup_{i=1}^{j-1} \{T_{i,N}\} \right\}, \quad j \in \{1, \ldots, N\}. \tag{7}$$

Then $n(j)$ is such that

$$\tau_{n(j)} = T_{j,N}. \tag{8}$$

In the examples of interest, the FPTs, τ, have continuous probability distributions (i.e. $F(t)$ is a continuous function) so that the event $\tau_{n^*} = \tau_{n'} < \infty$ for $n^* \neq n'$ has probability zero so there is no ambiguity in Eq. (8).

11

Since we have the sequence $\{(\tau_n, Z_n)\}_{n \geq 1}$, we denote E_n the event E as it pertains to the nth element in the sequence $\{(\tau_n, Z_n)\}_{n \geq 1}$. For example, if $E = \{X(0) = \theta/2\}$, then $E_n = \{X_n(0) = \theta/2\}$. Similarly, $E_{n(j)}$ is the event E as it pertains to $Z_{n(j)}$.

Throughout the Supplemental Material, we use the notation $\int f(t) \, dg(t)$ to denote the Riemann-Stieltjes integral of a function f with respect to a function g.

Proposition 1. *For any $j \in \{1, 2, \ldots, N\}$ (denoting an agent by the order j of their decision), we have that*

$$\mathbb{P}(E_{n(j)}) = j \binom{N}{j} \int_0^\infty [F(t)]^{j-1} [1 - F(t)]^{N-j} \, dF_E(t). \tag{9}$$

In the case $j = 1$ (i.e. the fastest decider), Proposition 1 implies

$$\mathbb{P}(E_{n(1)}) = N \int_0^\infty [1 - F(t)]^{N-1} \, dF_E(t). \tag{10}$$

Since $1 - F$ is a decreasing function, Eq. (10) implies that the short-time behavior of F and F_E determine the large N behavior of $\mathbb{P}(E_{n(1)})$. More generally, Proposition 1 implies that the short-time behavior of F and F_E determine the large N behavior of $\mathbb{P}(E_{n(j)})$ for $1 \leq j \ll N$.

In the case $j = N$ (i.e. the slowest decider), Proposition 1 implies

$$\mathbb{P}(E_{n(N)}) = N \int_0^\infty [F(t)]^{N-1} \, dF_E(t). \tag{11}$$

Since F is an increasing function, Eq. (11) implies that the large-time behavior of F and F_E determine the large N behavior of $\mathbb{P}(E_{n(N)})$. More generally, Proposition 1 implies that the large-time behavior of F and F_E determine the large N behavior of $\mathbb{P}(E_{n(N-j)})$ for $1 \ll N-j$.

SOME INTEGRAL ASYMPTOTICS

The following proposition is useful for estimating the large N behavior of some integrals of the form in Eq. (9) and was proved in [17] (See Proposition 2 in [17]). Throughout the Supplemental Material, "$f \sim g$" denotes $f/g \to 1$ (e.g., as $N \to \infty$ or as $t \to 0$).

Proposition 2. *Assume $C_+ > C > 0$, $A > 0$, and $p, q \in \mathbb{R}$. Then there exists a $\delta_0 > 0$ so that for all $\delta \in (0, \delta_0]$, we have*

$$\int_0^\delta t^{q-2} e^{-C_+/t} \left(1 - At^p e^{-C/t}\right)^{N-1} dt \sim \eta (\ln N)^{p\beta - q} N^{-\beta} \quad \text{as } N \to \infty,$$

where

$$\beta = C_+/C > 1, \quad \eta = C^{q-1}(AC^p)^{-\beta}\Gamma(\beta) > 0,$$

and $\Gamma(\beta) := \int_0^\infty z^{\beta-1} e^{-z} \, \mathrm{d}z$ *denotes the gamma function.*

The following result estimates integrals of the form in Eq. (9) for $1 \leq j \ll N$ assuming that $F(t)$ and $F_+(t)$ have short-time t behavior that is characteristic of diffusion.

Theorem 3. *Assume $F(t)$ and $F_+(t)$ are bounded, nondecreasing, continuous from the right, and satisfy*

$$F(t) \sim At^p e^{-C_0/t} \quad \text{as } t \to 0^+, \tag{12}$$

$$F_+(t) \sim Bt^q e^{-C_+/t} \quad \text{as } t \to 0^+, \tag{13}$$

where $C_+ > C_0 > 0$, $A > 0$, $B > 0$, and $p, q \in \mathbb{R}$. Then for any fixed integer $j \geq 1$, we have

$$j\binom{N}{j} \int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j} \, \mathrm{d}F_+(t) \sim \eta(j)(\ln N)^{p\beta-q} N^{1-\beta} \quad \text{as } N \to \infty,$$

where

$$\beta := C_+/C_0 > 1, \quad \eta(j) := B(C_0)^{q-p\beta} A^{-\beta} \Gamma(j)\Gamma(\beta + j) > 0, \tag{14}$$

and $\Gamma(x) := \int_0^\infty z^{x-1} e^{-z} \, \mathrm{d}z$ denotes the gamma function.

Notice that the asymptotic behavior found in Theorem 3 as $N \to \infty$ is independent of $j \geq 1$, except for the constant prefactor $\eta(j)$. Further, this prefactor is an increasing function of j and satisfies

$$\eta(j) = \frac{(j-1)!\Gamma(\beta + j)}{\Gamma(\beta + 1)}\eta(1), \quad j \geq 1.$$

The asymptotic behavior in Eq. (12)-(13) is typical for diffusion, but computing the prefactors A and B and the powers p and q can be challenging [15]. Indeed, these constants depend on the details of the system (e.g., drift, space dimension, geometry of the domain, etc.). However, the constants in the exponents C_0 and C_+ are more universal and can be obtained in a very general mathematical setting [16]. The following result yields estimates on the fastest deciders when we only know these constants, which is equivalent to knowing the short-time behavior of $F_+(t)$ and $F(t)$ on a logarithmic scale.

Theorem 4. *Assume $F(t)$ and $F_+(t)$ are bounded, nondecreasing, continuous from the right, and satisfy*

$$\lim_{t \to 0^+} t \ln F(t) = -C_0 < 0, \quad \lim_{t \to 0^+} t \ln F_+(t) \leq -C_+ < 0, \tag{15}$$

where $C_+ > C_0 > 0$. Then for every $\varepsilon > 0$,

$$j\binom{N}{j} \int_0^\infty [F(t)]^{j-1} [1 - F(t)]^{N-j} \, dF_+(t) = o(N^{1-\beta+\varepsilon}) \quad \text{as } N \to \infty, \tag{16}$$

where

$$\beta := C_+/C_0 > 1.$$

If, in addition, we assume that

$$\lim_{t \to 0^+} t \ln F_+(t) = -C_+ < 0, \tag{17}$$

then for every $\varepsilon > 0$,

$$N^{1-\beta-\varepsilon} = o\left(j\binom{N}{j} \int_0^\infty [F(t)]^{j-1} [1 - F(t)]^{N-j} \, dF_+(t) \right) \quad \text{as } N \to \infty.$$

The following result estimates integrals of the form in Eq. (9) for $1 \ll N - j \leq N$ assuming that $F(t)$ and $f_i(t) = F_i'(t)$ have large-time t behavior that is characteristic of diffusion in a bounded domain.

Theorem 5. *Assume $F(t) \in [0, 1)$ is continuous and nondecreasing and $f_i(t)$ is continuous and bounded and*

$$F(t) = 1 - ce^{-\lambda t} + \text{h.o.t.} \quad \text{as } t \to \infty,$$

$$f_i(t) = \lambda c_i e^{-\lambda t} + \text{h.o.t.} \quad \text{as } t \to \infty,$$

where $\lambda > 0$, $c > 0$, $c_i > 0$. Then for any fixed $j \geq 0$, we have that

$$(N - j)\binom{N}{N - j} \int_0^\infty [F(t)]^{N-j-1} [1 - F(t)]^j f_i(t) \, dt \to \frac{c_i}{c} \quad \text{as } N \to \infty.$$

PROOF OF EQ. (3) IN MAIN TEXT

We now apply Theorem 3 to obtain Eq. (3) in the main text. Suppose the belief of each agent evolves independently according to the following stochastic differential equation (SDE),

$$dX = \mu \, dt + \sqrt{2D} \, dW, \tag{18}$$

where $\mu \in \mathbb{R}$ is a constant drift, $D > 0$ is a constant diffusivity, and $W = \{W(t)\}_{t \geq 0}$ is a standard Brownian motion. Define the FPT,

$$\tau := \inf\{t > 0 : X(t) \notin (-\theta, \theta)\},$$

for some threshold $\theta > 0$. Assume that the initial distribution $\mathbb{P}(X(0) = x_i)$ of each agent is a sum of Dirac masses at a finite set of points $\{x_0, x_1, \ldots, x_{I-1}\}$,

$$\mathbb{P}(X(0) = x) = \begin{cases} q_i & \text{if } x = x_i \text{ for some } i \in \{0, 1, \ldots, I-1\}, \\ 0 & \text{if } x \notin \cup_{i=0}^{I-1} x_i. \end{cases},$$

Letting $F_i(t) \equiv F_{X(0)=x_i}(t) = \mathbb{P}(\tau \leq t \cap X(0) = x_i)$, we have that [17]

$$F_i(t) \sim q_i A_i t^{1/2} e^{-C_i/t} \quad \text{as } t \to 0^+, \tag{19}$$

where

$$C_i = \frac{(L_i)^2}{4D},$$

and

$$A_i = \begin{cases} \exp\left(\frac{-\mu L_i}{2D}\right)\sqrt{\frac{4D}{\pi(L_i)^2}} & \text{if } x_i < 0 \\ \exp\left(\frac{\mu L_i}{2D}\right)\sqrt{\frac{4D}{\pi(L_i)^2}} & \text{if } x_i > 0 \\ \left[\exp\left(\frac{-\mu L_i}{2D}\right) + \exp\left(\frac{\mu L_i}{2D}\right)\right]\sqrt{\frac{4D}{\pi(L_i)^2}} & \text{if } x_i = 0, \end{cases}$$

where L_i is the distance to the closest threshold from x_i,

$$L_i = \min\{\theta - x_i, x_i + \theta\}.$$

Further, we assume $0 \in \{0, 1, \ldots, I-1\}$ is the index of the unique starting location closest to a threshold

$$L_0 = \min\{L_0, L_1, \ldots, L_{I-1}\} < L_i \quad \text{if } i \neq 0,$$

then

$$F(t) \sim F_0(t) \quad \text{as } t \to 0^+.$$

We claim that

$$\mathbb{P}(X_{n(1)}(0) = x_0) \to 1 \quad \text{as } N \to \infty, \tag{20}$$

15

Thus, when N is large the first decider out of many deciders is always the one with the most extreme initial bias. Using the integral representation in Proposition 1 and applying Theorem 3 yields

$$\mathbb{P}(X_{n(1)}(0) = x_i) \sim \eta_i(1)(\ln N)^{(\beta_i-1)/2}N^{1-\beta_i} \quad \text{as } N \to \infty \quad \text{for each } i \neq 0,$$

where

$$\beta_i = (L_i/L_0)^2 > 1,$$

and

$$\eta_i(1) = \begin{cases} \dfrac{q_i}{q_0^{\beta_i}}\sqrt{\dfrac{\pi^{\beta_i-1}}{\beta_i}}\Gamma(\beta_i+1)\exp\left(\dfrac{\sqrt{\beta_i}}{2D}\left(\mu_i L_0 - \mu_0 L_i\right)\right) & \text{if } x_i \neq 0, \\[4ex] \dfrac{q_i}{q_0^{\beta_i}}\sqrt{\dfrac{\pi^{\beta_i-1}}{\beta_i}}\Gamma(\beta_i+1)\left[\exp\left(\dfrac{\sqrt{\beta_i}}{2D}\left(\mu_i L_0 - \mu_0 L_i\right)\right) + \exp\left(\dfrac{\sqrt{\beta_i}}{2D}\left(-\mu_i L_0 - \mu_0 L_i\right)\right)\right] & \text{if } x_i = 0, \end{cases}$$

where $\mu_i = \pm\mu$ if $x_i \gtrless 0$.

FIRST DECISION AGREES WITH INITIAL BIAS

The analysis above shows that the first agent to decide in a large group has the most extreme initial bias. We now show the intuitive result that this first decider's decision agrees with their initial bias. Without loss of generality, assume that the most extreme initial bias is negative, $x_0 < 0$. Letting $F_+(t) = \mathbb{P}(\tau \leq t \cap X(\tau) = +\theta)$, we have

$$\begin{aligned} F_+(t) &= \sum_i \mathbb{P}(\tau \leq t \cap X(\tau) = +\theta \mid X(0) = x_i)q_i \\ &\sim \mathbb{P}(\tau \leq t \cap X(\tau) = +\theta \mid X(0) = x_{i^+})q_{i^+} \\ &\sim q_{i^+}A_{i^+}t_{i^+}^p e^{-C_{i^+}/t} \quad \text{as } t \to 0^+, \end{aligned}$$

where $i^+ \in \{1, \ldots, I\}$ is the index of the starting location closest to $+\theta$. Using the integral representation in Proposition 1 and applying Theorem 3 yields

$$\mathbb{P}(X_{n(1)}(\tau) = +\theta) \sim \eta_{i^+}^{(1)}(\ln N)^{(\beta_{i^+}-1)/2}N^{1-\beta_{i^+}} \quad \text{as } N \to \infty.$$

CONTINUOUS INITIAL BELIEF DISTRIBUTION

In Section , we showed that the first of many deciders have the most extreme initial beliefs in the case that the population has a discrete initial belief distribution. We now generalize

this calculation to the case that the deciders have a continuous initial belief distribution. In particular, suppose that the decider's initial belief (position) has a smooth probability density $\nu(x)$ with support (a, b) with $-\theta < a < b < \theta$. Suppose that

$$\nu(x) \sim (x - a)^{\alpha_a}\nu_a \quad \text{as } x \to a^+,$$
$$\nu(x) \sim (b - x)^{\alpha_b}\nu_b \quad \text{as } x \to b^-,$$

where the coefficients are positive, $\nu_a > 0$, $\nu_b > 0$, and the powers ensure that ν is integrable, $\alpha_a > -1$, $\alpha_b > -1$. In light of (19), suppose that

$$\mathbb{P}(\tau \leq t \mid X(0) = x) \sim A(x)t^p e^{-C(x)/t} \quad \text{as } t \to 0^+, \quad \text{uniformly for all } x \in [a, b],$$

where

$$C(x) = (L(x))^2/(4D) > 0, \quad L(x) = \min\{\theta - x, \theta + x\},$$

and $A(x) > 0$ for all $x \in [a, b]$.

It follows that

$$F(t) = \mathbb{P}(\tau \leq t) = \int_a^b \mathbb{P}(\tau \leq t \mid X(0) = x)\nu(x)\,dx$$
$$\sim t^p \int_a^b A(x)\nu(x)e^{-C(x)/t}\,dx \quad \text{as } t \to 0^+.$$

We thus need to estimate the small time t asymptotics of the integral

$$I := \int_a^b A(x)\nu(x)e^{-C(x)/t}\,dx,$$

which is an exercise in Laplace's method [2]. If $b > 0$, then for any $\varepsilon \in (0, b)$, we have

$$\int_0^b A(x)\nu(x)e^{-C(x)/t}\,dx \sim \int_{b-\varepsilon}^b A(x)\nu(x)e^{-C(x)/t}\,dx$$
$$\sim A(b)e^{-C(b)/t}\nu_b\Gamma(\alpha_b + 1)t^{\alpha_b+1} \quad \text{as } t \to 0^+.$$

Similarly, if $a < 0$, then for any $\varepsilon \in (0, |a|)$, we have

$$\int_a^0 A(x)\nu(x)e^{-C(x)/t}\,dx \sim \int_a^{a+\varepsilon} A(x)\nu(x)e^{-C(x)/t}\,dx$$
$$\sim A(a)e^{-C(a)/t}\nu_a\Gamma(\alpha_a + 1)t^{\alpha_a+1} \quad \text{as } t \to 0^+.$$

Putting this together, we have that if $b > |a|$, then

$$F(t) \sim A(b)\nu_b\Gamma(\alpha_b + 1)t^{p+\alpha_b+1}e^{-C(b)/t} \quad \text{as } t \to 0^+,$$

and similarly if $|a| > b$ or $|a| = b$.

With these estimates, we can apply Theorem 3 to obtain estimates that the fastest decider(s) have extreme initial beliefs. In particular, suppose we want to estimate

$$\mathbb{P}(a + \varepsilon < X_{n(1)}(0) < b - \varepsilon) \quad \text{for some small } 0 < \varepsilon \ll 1,$$

which is the probability that the fastest decider does not have extreme initial beliefs. If we define the event

$$E = \{a + \varepsilon < X(0) < b - \varepsilon\},$$

then using the notation of Section , we have that

$$\begin{aligned}
F_E(t) := \mathbb{P}(\tau \le t \cap E) &= \int_{a+\varepsilon}^{b-\varepsilon} \mathbb{P}(\tau \le t \,|\, X(0) = x)\nu(x)\,\mathrm{d}x \\
&\sim t^p \int_{a+\varepsilon}^{b-\varepsilon} A(x)\nu(x)e^{-C(x)/t}\,\mathrm{d}x \quad \text{as } t \to 0^+,
\end{aligned}$$

which can be estimated as above using Laplace's method [2]. In particular, if $b > |a|$, then

$$F_E(t) \sim A(b - \varepsilon)\nu(b - \varepsilon)t^{p+1}e^{-C(b-\varepsilon)/t} \quad \text{as } t \to 0^+,$$

assuming $\nu(b - \varepsilon) > 0$, and similarly if $|a| > b$ or $|a| = b$. With this short-time behavior of $F_E(t)$, we can then plug this into Theorem 3 to show that the first deciders have the most extreme initial beliefs.

HETEROGENEOUS POPULATION WITH MULTIPLE ALTERNATIVES

We next consider the generalized case where the beliefs of the agents in the population evolve as processes with (possibly space-dependent) drift, diffusion coefficient, initial position, and even domain (in their own arbitrary space dimension $d \ge 1$). Suppose the belief of the ith decider evolves according to the following d-dimensional SDE,

$$\mathrm{d}X_i = \mu_i(X_i)\,\mathrm{d}t + \sqrt{2D_i}\,\mathrm{d}W_i, \tag{21}$$

where $\mu_i : \mathbb{R}^d \to \mathbb{R}^d$ is a possibly space-dependent drift, $D_i > 0$ is the diffusion coefficient, and $W(t) \in \mathbb{R}^d$ is a standard Brownian motion in d-dimensional space.

Let $L > 0$ denote an agent's (random) shortest distance they must travel to hit the closest target and let $D > 0$ denote the agent's diffusion coefficient. Define the random timescale

$$S = \frac{L^2}{4D} > 0.$$

Suppose that S has a discrete distribution on a finite set

$$0 < s_0 < s_1 < s_2 < s_3 \cdots < s_I,$$

where

$$\mathbb{P}(S = s_i) = q_i > 0, \quad \sum_{i=0}^{I} q_i = 1.$$

Since we have $N \geq 1$ iid agents indexed from $n = 1$ to $n = N$, we let S_n denote the value of S for the nth agent and $S_{n(j)}$ the value of S for the jth fastest to decide.

We have that [16]

$$\lim_{t \to 0^+} t \ln \mathbb{P}(\tau \leq t) = -s_0 < 0, \quad \lim_{t \to 0^+} t \ln \mathbb{P}(\tau \leq t \cap S = s_i) = -s_i < 0.$$

Hence, Proposition 1 and Theorem 4 imply that for any fixed $j \geq 1$ and $i \in \{1, \ldots, I\}$ and any $\varepsilon > 0$,

$$N^{1-s_i/s_0-\varepsilon} \ll \mathbb{P}(S_{n(j)} = s_i) \ll N^{1-s_i/s_0^\varepsilon} \quad \text{as } N \to \infty, \tag{22}$$

where we use the notation $f \ll g$ to mean $\lim f/g = 0$. That is, in more traditional notation,

$$N^{1-s_i/s_0-\varepsilon} = o\big(\mathbb{P}(S_{n(j)} = s_i)\big) \quad \text{as } N \to \infty,$$
$$\mathbb{P}(S_{n(j)} = s_i) = o\big(N^{1-s_i/s_0+\varepsilon}\big) \quad \text{as } N \to \infty.$$

In the special case that the agents all move in one space dimension and the drifts are spatially constant (but may differ between agents), we can get the constant and logarithmic prefactors on the decay of $\mathbb{P}(S_{n(j)} = s_i)$ as $N \to \infty$.

The result in Eq. (22) says that in a large population if all the agents have the same diffusion coefficient, then the fastest deciders started closest to their decision thresholds (targets). If we allow the diffusion coefficients to vary between agents, then (22) implies that the fastest deciders started close to their decision thresholds and/or they had big diffusion coefficients.

SLOWEST DECIDERS

Suppose the beliefs of the iid agents diffuse in some d-dimensional spatial domain $U \subset \mathbb{R}^d$ and can be absorbed at one of $m \geq 2$ targets $V_0, \ldots, V_{m-1}$ and let $\kappa \in \{0, \ldots, m-1\}$ indicate which target the decider eventually hits. Here, we will think of the m targets as parts of the $d-1$ dimensional boundary of the domain, and assume that hitting one of the targets triggers a decision. Following [18, 29], suppose the beliefs of the deciders evolve as stochastic process $\{X(t)\}_{t \geq 0}$ that diffuse according to the SDE

$$\mathrm{d}X(t) = -\nabla V(X(t))\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W(t), \tag{23}$$

with reflecting boundary conditions. In Eq. (23), the drift term is the gradient of a given potential, $V(x)$, and the noise term depends on the diffusion coefficient $D > 0$ and a standard d-dimensional Brownian motion (Wiener process) $\{W(t)\}_{t \geq 0}$. The survival probability conditioned on the initial position,

$$\mathbf{S}(x, t) := \mathbb{P}(\tau > t \mid X(0) = x),$$

satisfies the backward Kolmogorov (also called backward Fokker-Planck) equation,

$$
\begin{aligned}
\tfrac{\partial}{\partial t}\mathbf{S} &= \mathcal{L}\mathbf{S}, && x \in U, \\
\mathbf{S} &= 0, && x \in \text{targets}, \\
\tfrac{\partial}{\partial \mathbf{n}}\mathbf{S} &= 0, && x \in \text{reflecting boundary (if there is one)}, \\
S &= 1, && t = 0.
\end{aligned}
\tag{24}
$$

In Eq. (24), the differential operator $\mathcal{L}$ is the generator (i.e. the backward operator) of Eq. (23),

$$\mathcal{L} = -\nabla V(x) \cdot \nabla + D\Delta,$$

and $\frac{\partial}{\partial \mathbf{n}}$ is the derivative with respect to the inward unit normal $\mathbf{n} : \partial U \to \mathbb{R}^d$.

Using the following weight function of Boltzmann form ,

$$\rho(x) := \frac{e^{-V(x)/D}}{\int_U e^{-V(y)/D}\,\mathrm{d}y}, \tag{25}$$

one can check that the differential operator $\mathcal{L}$ is formally self-adjoint on the weighted space of square integrable functions (see, for example, Ref. [29]),

$$L^2_\rho(U) := \left\{ f : \int_U |f(x)|^2 \rho(x)\,\mathrm{d}x < \infty \right\},$$

20

using the boundary conditions in (24) and the following weighted inner product,

$$(f, g)_\rho := (f, g\rho) = \int_U f(x)g(x)\rho(x)\,\mathrm{d}x,$$

where $(f, g) = \int_U f(x)g(x)\,\mathrm{d}x$ denotes the standard L^2-inner product (i.e. with no weight function). Expanding the solution to (24) yields,

$$\mathbf{S}(x, t) = \sum_{n \geq 1}(u_n, 1)_\rho e^{-\lambda_n t}u_n(x) = \sum_{n \geq 1}(u_n, \rho)e^{-\lambda_n t}u_n(x), \tag{26}$$

where

$$0 < \lambda_1 < \lambda_2 \leq \dots, \tag{27}$$

denote the (necessarily positive) eigenvalues of $-\mathcal{L}$. The corresponding with eigenfunctions $\{u_n(x)\}_{n \geq 1}$ satisfy the following time-independent equation,

$$-\mathcal{L}u_n = \lambda_n u_n, \quad x \in U, \tag{28}$$

and identical boundary conditions as $\mathbf{S}$. Further, the eigenfunctions are orthogonal and are taken to be orthonormal, which means that

$$(u_n, u_m)_\rho = \delta_{nm} \in \{0, 1\}, \tag{29}$$

where δ_{nm} denotes the Kronecker delta function (i.e. $\delta_{nn} = 1$ and $\delta_{mn} = 0$ if $n \neq m$).

If the initial distribution of an agent has probability measure μ_0,

$$\mathbb{P}(X(0) \in B) = \mu_0(B) = \int_B 1\,\mathrm{d}\mu_0(x), \quad B \subset U, \tag{30}$$

then the FPT τ has survival probability given by

$$S(t) := \mathbb{P}(\tau > t \mid X(0) =_\mathrm{d} \mu_0) = \int_U \mathbf{S}(x, t)\,\mathrm{d}\mu_0(x),$$

where the condition $X(0) =_\mathrm{d} \mu_0$ in the conditional probability merely denotes that $X(0)$ has initial distribution given by μ_0. Hence, we obtain the following representation for the survival probability,

$$S(t) = \sum_{n \geq 1} A_n e^{-\lambda_n t} = \sum_{n \geq 1}(u_n, \rho)(u_n, \mathrm{d}\mu_0)e^{-\lambda_n t}, \tag{31}$$

21

where the coefficients are given by the following integrals,

$$A_n := (u_n, 1)_\rho \int_U u_n(x) \, d\mu_0(x), \quad n \geq 1. \tag{32}$$

We have that the FPT τ to one of the targets has CDF

$$\begin{aligned} F(t) = \mathbb{P}(\tau \leq t) &= 1 - \mathbb{P}(\tau > t) \\ &= 1 - \sum_{k \geq 1} (u_k, \rho)(u_k, d\mu_0) e^{-\lambda_k t}, \end{aligned}$$

If

$$p_i(x) = \mathbb{P}(\kappa = i | X(0) = x),$$

then

$$\begin{aligned} F_i(t) := \mathbb{P}(\tau \leq t \cap \kappa = i) &= \mathbb{P}(\kappa = i) - \mathbb{P}(\tau > t \cap \kappa = i) \\ &= \mathbb{P}(\kappa = i) - \sum_{k \geq 1} (u_k, p_i\rho)(u_k, d\mu_0) e^{-\lambda_k t}, \end{aligned}$$

and therefore

$$f_i(t) := F_i'(t) = \sum_{k \geq 1} \lambda_k (u_k, p_i\rho)(u_k, d\mu_0) e^{-\lambda_k t}$$

Applying Proposition 1 and Theorem 5 yields

$$\mathbb{P}(\kappa_{n(N-j)} = i) \to \frac{(u_1, p_i\rho)}{(u_1, \rho)} = \frac{(u_1\rho, p_i)}{(u_1, \rho)} \quad \text{as } N \to \infty.$$

Now, the solution to the forward Fokker-Planck equation is given by

$$p(x, t) = \mathbb{P}(X(t) = dx \,|\, \tau > t) = \sum_{k \geq 1} e^{-\lambda_k t} (u_k, d\mu_0) \rho(x) u_k(x).$$

Hence, $u_1(x)\rho(x)/(u_1, \rho)$ is the quasi-stationary distribution (QSD), $q(x)$, defined by

$$\begin{aligned} q(x) = \lim_{t \to \infty} \mathbb{P}(X(t) = dx \,|\, \tau > t) &= \lim_{t \to \infty} \frac{\mathbb{P}(X(t) = dx \cap \tau > t)}{\mathbb{P}(\tau > t)} \\ &= \lim_{t \to \infty} \frac{\sum_{k \geq 1} e^{-\lambda_k t} (u_k, d\mu_0) \rho(x) u_k(x)}{\sum_{k \geq 1} (u_k, \rho)(u_k, d\mu_0) e^{-\lambda_k t}} \\ &= \lim_{t \to \infty} \frac{e^{-\lambda_1 t} (u_1, d\mu_0) \rho(x) u_1(x)}{(u_1, \rho)(u_1, d\mu_0) e^{-\lambda_1 t}} \\ &= \frac{\rho(x) u_1(x)}{(u_1, \rho)}. \end{aligned}$$

Summarizing, we have shown that

$$\mathbb{P}(\kappa_{n(N-j)} = i) \to \int_U p_i(x) q(x) \, dx \quad \text{as } N \to \infty. \tag{33}$$

The case of drift-diffusion processes in one dimension. For the one-dimensional example in which all the beliefs of all the agents evolve according to (18), we can compute the QSD, and find that

$$q(x) = \frac{(\pi^2 D^2 + \theta^2 \mu^2) \cos(\frac{\pi x}{2\theta}) e^{\frac{\mu(\theta+x)}{2D}}}{2\pi D^2 \theta (e^{\frac{\theta \mu}{D}} + 1)}.$$

Further, it is straightforward to show that the probability that a decider reaches $+\theta$ before $-\theta$ conditioned on the initial belief $x \in [-\theta, \theta]$ is

$$p_1(x) := \mathbb{P}(X(\tau) = +\theta) = \frac{1}{2}\left(\coth\left(\frac{\theta\mu}{D}\right) - 1\right) e^{\frac{\mu(\theta-x)}{D}} \left(e^{\frac{\mu(\theta+x)}{D}} - 1\right)$$

Therefore, applying (33) and explicitly computing the integral yields

$$\mathbb{P}(\kappa_{n(N-j)} = 1) \to \int_{-\theta}^{\theta} p_1(x) q(x) \, \mathrm{d}x = \frac{1}{1 + e^{-\frac{\theta\mu}{D}}} = p_1(0) \quad \text{as } N \to \infty.$$

Hence, the slowest deciders out of $N \gg 1$ deciders make a decision as if they were initially unbiased (i.e. as if $X(0) = 0$).

PROOFS

Proof of Proposition 1. Since $\{(\tau_n, Z_n)\}_{n \geq 1}$ are identically distributed, we have that

$$\mathbb{P}(A_{n(j)}) = \sum_{\substack{\text{distinct indices} \\ n_1, \ldots, n_N \in \{1, \ldots, N\}}} \mathbb{P}(\max\{\tau_{n_1}, \ldots, \tau_{n_{j-1}}\} < \tau_{n_j} < \min\{\tau_{n_{j+1}}, \ldots, \tau_{n_N}\} \cap A_{n_j})$$

$$= j\binom{N}{j} \mathbb{P}(\max\{\tau_1, \ldots, \tau_{j-1}\} < \tau_j < \min\{\tau_{j+1}, \ldots, \tau_N\} \cap A_j), \tag{34}$$

where the coefficient comes from noting that the number of terms in the sum is obtained by choosing the j fastest FPTs out of N and then choosing which of those j will be the jth fastest. Define

$$\tau_j^{(A_j)} = \begin{cases} \tau_j & \text{if } A_j \text{ occurs,} \\ +\infty & \text{if } A_j \text{ does not occur,} \end{cases}$$

so that if $j < N$,

$$\mathbb{P}(\max\{\tau_1, \ldots, \tau_{j-1}\} < \tau_j < \min\{\tau_{j+1}, \ldots, \tau_N\} \cap A_j)$$

$$= \mathbb{P}(\max\{\tau_1, \ldots, \tau_{j-1}\} < \tau_j^{(A_j)} < \min\{\tau_{j+1}, \ldots, \tau_N\})$$

To handle the case $j = N$, we can simply replace $+\infty$ by $-\infty$ in the definition of $\tau_j^{(A_j)}$.

Since $\{\tau_n\}_{n\geq 1}$ are iid, we have that

$$\mathbb{P}(\max\{\tau_1,\dots,\tau_{j-1}\} < t) = \mathbb{P}(\max\{\tau_1,\dots,\tau_{j-1}\} \leq t) = [F(t)]^{j-1},$$

where we have used that $F(t)$ is continuous. Similarly,

$$\mathbb{P}(\min\{\tau_{j+1},\dots,\tau_N\} > t) = [1 - F(t)]^{N-j},$$

Using that $\{\tau_n\}_{n\geq 1}$ are independent, we have

$$
\begin{aligned}
G(t) :&= \mathbb{P}(\max\{\tau_1,\dots,\tau_{j-1}\} < t < \min\{\tau_{j+1},\dots,\tau_N\}) \\
&= \mathbb{P}(\max\{\tau_1,\dots,\tau_{j-1}\} < t)\mathbb{P}(t < \min\{\tau_{j+1},\dots,\tau_N\}) \\
&= [F(t)]^{j-1}[1 - F(t)]^{N-j}.
\end{aligned}
$$

Combining the above finally yields

$$
\begin{aligned}
\mathbb{P}(A_{n(j)}) &= j\binom{N}{j}\mathbb{P}(\max\{\tau_1,\dots,\tau_{j-1}\} < \tau_j^{(A_j)} < \min\{\tau_{j+1},\dots,\tau_N\}) \\
&= j\binom{N}{j}\mathbb{E}[G(\tau_j^{(A_j)})] \\
&= j\binom{N}{j}\int_0^\infty [F(t)]^{j-1}[1 - F(t)]^{N-j}\,\mathrm{d}F_E(t),
\end{aligned}
$$

which completes the proof. $\qquad\square$

The proof of Theorem 3 is similar to the proof of Theorem 3 in [17].

Proof of Theorem 3. Define the integral from $t = a$ to $t = b$,

$$I_{a,b} := \int_a^b [F(t)]^{j-1}[1 - F(t)]^{N-j}\,\mathrm{d}F_+(t).$$

Let $\varepsilon \in (0,1)$. By the assumptions in Eq. (12)-(13), there exists a $\delta > 0$ so that

$$A_{-\varepsilon}t^p e^{-C_0/t} \leq F(t) \leq A_{+\varepsilon}t^p e^{-C_0/t} \quad \text{for all } t \in (0,\delta), \tag{35}$$

$$B_{-\varepsilon}t^q e^{-C_+/t} \leq F_+(t) \leq B_{+\varepsilon}t^q e^{-C_+/t} \quad \text{for all } t \in (0,\delta), \tag{36}$$

where $A_{\pm\varepsilon} := A(1 \pm \varepsilon)$ and $B_{\pm\varepsilon} := B(1 \pm \varepsilon)$. Using Eq. (35) and integrating by parts yields

$$
\begin{aligned}
I_{0,\delta} &\leq \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (1 - A_{-\varepsilon} t^p e^{-C_0/t})^{N-j} \, \mathrm{d}F_+(t) \\
&= (A_{+\varepsilon} \delta^p e^{-C_0/\delta})^{j-1} (1 - A_{-\varepsilon} \delta^p e^{-C_0/\delta})^{N-j} F_+(\delta) \\
&\quad + (N-j) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (pt^{-1} + C_0 t^{-2}) A_{-\varepsilon} t^p e^{-C_0/t} \left(1 - A_{-\varepsilon} t^p e^{-C_0/t}\right)^{N-j-1} F_+(t) \, \mathrm{d}t \\
&\quad - (j-1) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (pt^{-1} + C_0 t^{-2}) \left(1 - A_{-\varepsilon} t^p e^{-C_0/t}\right)^{N-j} F_+(t) \, \mathrm{d}t \\
&\leq (A_{+\varepsilon} \delta^p e^{-C_0/\delta})^{j-1} (1 - A_{-\varepsilon} \delta^p e^{-C_0/\delta})^{N-j} F_+(\delta) \\
&\quad + (N-j) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^j (pt^{-1} + C_0 t^{-2}) \left(1 - A_{-\varepsilon} t^p e^{-C_0/t}\right)^{N-j-1} B_{+\varepsilon} t^q e^{-C_+/t} \, \mathrm{d}t \\
&\quad - (j-1) \int_0^\delta (A_{+\varepsilon} t^p e^{-C_0/t})^{j-1} (pt^{-1} + C_0 t^{-2}) \left(1 - A_{-\varepsilon} t^p e^{-C_0/t}\right)^{N-j} B_{-\varepsilon} t^q e^{-C_+/t} \, \mathrm{d}t,
\end{aligned}
$$

$$(37)$$

where we have used Eq. (36) in the final inequality. The first term in the righthand side of Eq. (37) vanishes exponentially fast as $N \to \infty$. Using Proposition 2 to find the large N behavior of the second two terms in the righthand side of Eq. (37) and the fact that $I_{\delta,\infty}$ vanishes exponentially fast as $N \to \infty$ yields

$$
\limsup_{N \to \infty} \frac{j \binom{N}{j} I_{0,\infty}}{\eta_j (\ln N)^{p\beta - q} N^{1-\beta}} \leq \frac{(1+\varepsilon)}{(1-\varepsilon)^\beta}.
$$

The analogous argument yields the lower bound

$$
\liminf_{N \to \infty} \frac{j \binom{N}{j} I_{0,\infty}}{\eta_j (\ln N)^{p\beta - q} N^{1-\beta}} \geq \frac{(1-\varepsilon)}{(1+\varepsilon)^\beta}.
$$

Since $\varepsilon \in (0,1)$ is arbitrary, the proof is complete. $\qquad\square$

Proof of Theorem 4. Define the integral from $t = a$ to $t = b$,

$$
I_{a,b} := \int_a^b [F(t)]^{j-1} [1 - F(t)]^{N-j} \, \mathrm{d}F_+(t).
$$

By Eq. (15), there exists a $\delta > 0$ so that

$$
e^{-(C_0+\varepsilon)/t} \leq F(t) \leq e^{-(C_0-\varepsilon)/t} \quad \text{for all } t \in (0,\delta), \tag{38}
$$

$$
F_+(t) \leq e^{-(C_+ -\varepsilon)/t} \quad \text{for all } t \in (0,\delta). \tag{39}
$$

Using Eq. (38) and integrating by parts yields

$$
\begin{aligned}
I_{0,\delta} &\leq \int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t}\left(1-e^{-(C_0+\varepsilon)/t}\right)^{N-j}\,\mathrm{d}F_+(t) \\
&= F_+(\delta)e^{-(j-1)(C_0-\varepsilon)/\delta}\left(1-e^{-(C_0+\varepsilon)/\delta}\right)^{N-j} \\
&\quad + (N-j)(C_0+\varepsilon)\int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t}t^{-2}e^{-(C_0+\varepsilon)/t}\left(1-e^{-(C_0+\varepsilon)/t}\right)^{N-j-1}F_+(t)\,\mathrm{d}t \\
&\quad - \int_0^\delta (j-1)(C_0-\varepsilon)t^{-2}e^{-(j-1)(C_0-\varepsilon)/t}\left(1-e^{-(C_0+\varepsilon)/t}\right)^{N-j}F_+(t)\,\mathrm{d}t.
\end{aligned}
\tag{40}
$$

The first term in the righthand side of Eq. (40) vanishes exponentially fast as $N \to \infty$. To handle the second term in the righthand side of Eq. (40), note that Eq. (39) implies that

$$
\begin{aligned}
&\int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t}t^{-2}e^{-(C_0+\varepsilon)/t}\left(1-e^{-(C_0+\varepsilon)/t}\right)^{N-j-1}F_+(t)\,\mathrm{d}t \\
&\leq \int_0^\delta e^{-(j-1)(C_0-\varepsilon)/t}t^{-2}e^{-(C_0+\varepsilon)/t}\left(1-e^{-(C_0+\varepsilon)/t}\right)^{N-j-1}e^{-(C_+-\varepsilon)/t}\,\mathrm{d}t.
\end{aligned}
\tag{41}
$$

Since the third term in the righthand side of Eq. (40) is nonpositive, applying Proposition 2 to Eq. (41) and using Eq. (40) and the fact that $I_{\delta,\infty}$ vanishes exponentially fast as $N \to \infty$ completes the proof of Eq. (16).

If Eq. (17) holds, then there exists a $\delta > 0$ so that

$$
\begin{aligned}
e^{-(C_0+\varepsilon)/t} &\leq F(t) \leq e^{-(C_0-\varepsilon)/t} \quad \text{for all } t \in (0,\delta), \\
e^{-(C_++\varepsilon)/t} &\leq F_+(t) \leq e^{-(C_+-\varepsilon)/t} \quad \text{for all } t \in (0,\delta).
\end{aligned}
\tag{42}
$$

Using Eq. (42) and integrating by parts yields

$$
\begin{aligned}
I_{0,\delta} &\geq \int_0^\delta e^{-(j-1)(C_0+\varepsilon)/t}\left(1-e^{-(C_0-\varepsilon)/t}\right)^{N-j}\,\mathrm{d}F_+(t) \\
&= F_+(\delta)e^{-(j-1)(C_0+\varepsilon)/\delta}\left(1-e^{-(C_0-\varepsilon)/\delta}\right)^{N-j} \\
&\quad + (N-j)(C_0-\varepsilon)\int_0^\delta e^{-(j-1)(C_0+\varepsilon)/t}t^{-2}e^{-(C_0-\varepsilon)/t}\left(1-e^{-(C_0-\varepsilon)/t}\right)^{N-j-1}F_+(t)\,\mathrm{d}t \\
&\quad - \int_0^\delta (j-1)(C_0+\varepsilon)t^{-2}e^{-(j-1)(C_0+\varepsilon)/t}\left(1-e^{-(C_0-\varepsilon)/t}\right)^{N-j}F_+(t)\,\mathrm{d}t \\
&\geq F_+(\delta)e^{-(j-1)(C_0+\varepsilon)/\delta}\left(1-e^{-(C_0-\varepsilon)/\delta}\right)^{N-j} \\
&\quad + (N-j)(C_0-\varepsilon)\int_0^\delta e^{-(j-1)(C_0+\varepsilon)/t}t^{-2}e^{-(C_0-\varepsilon)/t}\left(1-e^{-(C_0-\varepsilon)/t}\right)^{N-j-1}e^{-(C_++\varepsilon)/t}\,\mathrm{d}t \\
&\quad - \int_0^\delta (j-1)(C_0+\varepsilon)t^{-2}e^{-(j-1)(C_0+\varepsilon)/t}\left(1-e^{-(C_0-\varepsilon)/t}\right)^{N-j}e^{-(C_+-\varepsilon)/t}\,\mathrm{d}t.
\end{aligned}
\tag{43}
$$

The first term in the righthand side of Eq. (43) vanishes exponentially as $N \to \infty$. Using Proposition 2 to estimate the second two terms in the righthand side of Eq. (43) completes the proof. □

Lemma 6. *For fixed $j \in \{0, 1, \dots\}$, $c > 0$, $\lambda > 0$, and $\delta > 0$, we have that*

$$(N - j)\binom{N}{N - j} \int_{1/\delta}^{\infty} [1 - ce^{-\lambda t}]^{N-j-1} e^{-(j+1)\lambda t} \, dt \to \frac{1}{\lambda c^{j+1}} \quad as \ N \to \infty.$$

Proof of Lemma 6. Changing variables

$$u = 1 - ce^{-\lambda t}, \quad du = \lambda c e^{-\lambda t} \, dt$$

yields

$$\int_{1/\varepsilon}^{\infty} [1 - ce^{-\lambda t}]^{N-j-1} e^{-(j+1)\lambda t} \, dt = \frac{1}{\lambda c^{j+1}} \int_{1-ce^{-\lambda t}}^{1} u^{N-j-1}(1-u)^j \, du$$

$$= \frac{1}{\lambda c^{j+1}} \left[\frac{(N-j-1)!j!}{N!} - \int_0^{1-ce^{-\lambda t}} u^{N-j-1}(1-u)^j \, du \right],$$

(44)

where we have used that $\int_0^1 u^{a-1}(1-u)^{b-1} \, du = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. Since the integral in Eq. (44) vanishes exponentially fast, the proof is complete. □

Proof of Theorem 5. Let $\varepsilon \in (0, 1)$. By assumption, there exists $\delta > 0$ so that

$$1 - (1+\varepsilon)ce^{-\lambda t} \le F(t) \le 1 - (1-\varepsilon)ce^{-\lambda t} \quad \text{for all } t \ge 1/\delta,$$

$$\lambda(1-\varepsilon)c_i e^{-\lambda t} \le f_i(t) \le \lambda(1+\varepsilon)c_i e^{-\lambda t} \quad \text{for all } t \ge 1/\delta.$$

Defining the integral from $t = a$ to $t = b$,

$$I_{a,b} := \int_a^b [F(t)]^{N-j-1}[1 - F(t)]^j f_i(t) \, dt,$$

we therefore have that

$$(1-\varepsilon)^{j+1} \lambda c_i c^j \int_{1/\delta}^{\infty} \left[1 - (1+\varepsilon)ce^{-\lambda t}\right]^{N-j-1} e^{-(j+1)\lambda t} \, dt \le I_{1/\delta, \infty}$$

$$\le (1+\varepsilon)^{j+1} \lambda c_i c^j \int_{1/\delta}^{\infty} \left[1 - (1-\varepsilon)ce^{-\lambda t}\right]^{N-j-1} e^{-(j+1)\lambda t} \, dt.$$

Since $I_{0,1/\delta}$ vanishes exponentially fast as $N \to \infty$, Lemma 6 implies that

$$\left(\frac{1-\varepsilon}{1+\varepsilon}\right)^{j+1} \frac{c_i}{c} \le \liminf_{N \to \infty}(N-j)\binom{N}{N-j} I_{0,\infty}$$

$$\le \limsup_{N \to \infty}(N-j)\binom{N}{N-j} I_{0,\infty} \le \left(\frac{1+\varepsilon}{1-\varepsilon}\right)^{j+1} \frac{c_i}{c}.$$

Since $\varepsilon \in (0, 1)$ is arbitrary, the proof is complete. □

27

NUMERICAL SOLUTIONS

Numerical solutions were computed via trapezoidal quadrature on Eq. (9) in Proposition 1. In each set of dynamics, we rescaled the drift-diffusion process on $[-\theta, \theta]$ to the interval $[0, \ell]$. The probability density function for hitting the left boundary in this system is [24]

$$f_0(t) := \frac{\mathrm{d}}{\mathrm{d}t} F_0(t) = \exp\left(-\frac{\mu x_0}{2D} - \frac{\mu^2 t}{4D} \right) \frac{D}{\ell^2} \phi\left(\frac{Dt}{\ell^2}, \frac{x_0}{\ell} \right), \tag{45}$$

where

$$\phi(s, w) := \begin{cases} \sum_{k=1}^{\infty} \exp(-k^2 \pi^2 s) 2k\pi \sin(k\pi w), \\ (4\pi s^3)^{-1/2} \sum_{k=-\infty}^{\infty} (w + 2k) \exp\left(-\frac{(w+2k)^2}{4s} \right). \end{cases} \tag{46}$$

The expressions in Eq. (46) are equivalent but have distinct utility: the top expansion converges quickly for large s while the bottom expansion converges quickly for small s. Hence, we utilize both expressions to more accurately compute probabilities associated with slow and fast deciders, respectively.

Integrating Eq. (45) yields

$$F_0(t) = \int_0^t f_0(t') \, \mathrm{d}t' = \exp\left(-\frac{\mu x_0}{2D} \right) \Phi\left(\frac{Dt}{\ell^2}, \frac{x_0}{\ell} \right)$$

with long- and short-time expansions of $\Phi(s, w)$ given by

$$\Phi(s, w) = \int_0^s \phi(s', w) \, \mathrm{d}s'$$
$$= \begin{cases} \sum_{k=1}^{\infty} \left(1 - \exp[-(b + k^2 \pi^2)s] \right) \frac{2k\pi}{b + k^2 \pi^2} \sin(k\pi w), \\ \sum_{k=-\infty}^{\infty} \frac{\mathrm{sgn}(2k+w)}{2} \left(e^{-\sqrt{\frac{b}{D}}|2k+w|} \mathrm{erfc}\left(\frac{|2k+w|}{\sqrt{4s}} - \sqrt{bs} \right) + e^{\sqrt{\frac{b}{D}}|2k+w|} \mathrm{erfc}\left(\frac{|2k+w|}{\sqrt{4s}} + \sqrt{bs} \right) \right) \end{cases}$$

where $b = (\mu\ell/2D)^2$. By symmetry one can determine the corresponding probability density and cumulative distribution functions for hitting the right boundary. Altogether, we acquire long- and short-time expressions for the cumulative distribution function of an agent making a decision,

$$F(t) := F_0(t) + F_1(t).$$

Where numerical solutions are illustrated, we use the short-time expressions of ϕ and Φ for $10^{-10} \leq t \leq 1$ and the complementary long-time expressions for $1 < t \leq 100$, discretizing

each time interval into 10^3 log-spaced points. We consider 10^3 terms in each series expansion. Moreover, we take $\ell = 1$ and unless otherwise stated $D = 1$. Finally, where more than one but finitely many initial beliefs are considered, we scale the probability functions according to the corresponding initial distribution as outlined in Section .

Specific details of figures with numerical solutions are as follows: In Fig 1B we illustrate in color Eq. (10) where $F_E = F$ as defined above with $X_{n(1)}(0) = y$. The black curve, which contains the remaining mass of the total probability, is computed as the sum of the colored curves subtracted from one. In Fig 2A-C we illustrate the probability that the first decider chooses the decision at $X(T_1) = \theta$ conditioned on having a particular initial bias. Hence, by definition of conditional probability, the numerical solutions are produced from quadrature on ratios of Eq. (10) with $F_E = F_1$ in the numerator and $F_E = F$ in the denominator with $X_{n(1)} = y$. The inset of Fig 2C is one minus the outset. In Fig 2D we illustrate Eq. (9) where $F_E = F$ and $S_{n(j)} = s$. In Fig 3B we illustrate the probability that the last decider chooses the decision at $X(T_N) = \theta$ conditioned on having a particular initial bias. Similar to Fig 2B, the numerical solutions are produced from quadrature on ratios of Eq. (11) with $F_E = F_1$ in the numerator and $F_E = F$ in the denominator with $X_{n(N)}(0) = y$.

AGENT-BASED STOCHASTIC SIMULATIONS

a. One-dimensional drift diffusion equation. To test the analytical solutions, we solved Eq. (1) in the main text using the Euler-Maruyama method, which describes the evidence accumulation process preceding binary decisions. In this approximation scheme, the true solution to the stochastic differential equation is approximated by a Markov chain Y constructed by setting $Y_0 = X(0)$ and updating Y according to the iterative scheme

$$Y_{n+1} = Y_n + \mu \Delta t + \sqrt{2D} \Delta W$$

where $Y_n \equiv Y(n\Delta t)$ is the value of the Markov chain after the nth update, and the random variables ΔW are independent and identically distributed Gaussian random variables with mean 0 and variance Δt. The equations were integrated until the value of Y_n exceeded $\pm\theta$.

The temporal discretization, Δt, is user-defined. As N grows, the time to first decision decays slowly. Thus, for large N, Δt must be taken to be sufficiently small for accurate representation of decision dynamics. For simulations here, we chose $\Delta t = 10^{-3}$ for $1 \leq N \leq$

1000. For $N > 1000$, we chose $\Delta t = N^{-1}$.

b. *Two-dimensional drift diffusion equation.* Decisions between three choices require a drift-diffusion model evolving on a planar domain [21]. Updating the discrete-time approximation of Eq. (21) for each observer (dropping the i subscript) using Euler-Maruyama provides the following iterative scheme

$$Y_{n+1}^1 = Y_n^1 + \mu^1 \Delta t + \sqrt{2D} \Delta W^1,$$
$$Y_{n+1}^2 = Y_n^2 + \mu^2 \Delta t + \sqrt{2D} \Delta W^2,$$

where $Y_n^j = Y^j(n\Delta t)$ is the value of the belief after the nth update, the random variables ΔW^j are Gaussian random variables with mean 0 and variance σ^2. Equations are integrated until the vector $\begin{pmatrix} Y_n^1 \\ Y_n^2 \end{pmatrix}$ departs the triangular domain

$$\{(Y^1, Y^2) | Y^2 < h \ \& \ Y^2 > -2h(3Y^1 + 1) \ \& \ Y^2 > 2h(3Y^1 - 1)\}$$

where $h = 1/2/\sqrt{3}$. Choices of each agent are determined by whether the agent crosses the $Y^2 = h$ or $Y^2 = -2h(3Y^1 + 1)$ or $Y^2 = 2h(3Y^1 - 1)$ boundary. For simulations again we use $\Delta t = 10^{-3}$ for $1 \le N \le 1000$ and $\Delta t = N^{-1}$ for $N > 1000$.

For the 2D case in an equilateral triangle, the threshold θ is taken to be equal to the length of the apothem—defined as a line from the center of a regular polygon at right angles to any of its sides. Hence, an unbiased agent begins at the centre of the equilateral triangle. We prescribe initial data for biased agents to be anywhere along an apothem except the centre of the triangle.

———————————

[*] samantha.linn@utah.edu

[†] lawley@math.utah.edu

[‡] bkaramched@fsu.edu

[§] zpkilpat@colorado.edu

[¶] kresimir.josic@gmail.com

[1] Daron Acemoglu, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

[2] CM Bender and SA Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory.* Springer Science & Business Media, 2013.

[3] Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D. Cohen. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700–765, 2006.

[4] Jerome R Busemeyer and James T Townsend. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432, 1993.

[5] Lars Chittka, Adrian G Dyer, Fiola Bock, and Anna Dornhaus. Bees trade off foraging speed for accuracy. *Nature*, 424(6947):388–388, 2003.

[6] Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature neuroscience*, 14(7):811–819, 2011.

[7] Joshua I Gold and Michael N Shadlen. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, 36(2):299–308, 2002.

[8] Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annu. Rev. Neurosci.*, 30:535–574, 2007.

[9] Denis S Grebenkov, Ralf Metzler, and Gleb Oshanin. From single-particle stochastic kinetics to macroscopic reaction rates: fastest first-passage time of n random walkers. *New Journal of Physics*, 22(10):103004, 2020.

[10] Bhargav Karamched, Megan Stickler, William Ott, Benjamin Lindner, Zachary P Kilpatrick, and Krešimir Josić. Heterogeneity improves speed and accuracy in social networks. *Physical review letters*, 125(21):218302, 2020.

[11] Bhargav Karamched, Simon Stolarczyk, Zachary P Kilpatrick, and Krešimir Josić. Bayesian evidence accumulation on social networks. *SIAM Journal on Applied Dynamical Systems*, 19(3):1884–1919, 2020.

[12] Roozbeh Kiani, Timothy D Hanks, and Michael N Shadlen. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12):3017–3029, 2008.

[13] Asher Koriat. When are two heads better than one and why? *Science*, 336(6079):360–362, 2012.

[14] Kenneth W Latimer, Jacob L Yates, Miriam LR Meister, Alexander C Huk, and Jonathan W

Pillow. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244):184–187, 2015.

[15] S D Lawley. Distribution of extreme first passage times of diffusion. *Journal of Mathematical Biology*, 2020.

[16] S. D. Lawley. Universal formula for extreme first passage statistics of diffusion. *Physical Review E*, (1):012413, 2020.

[17] Samantha Linn and Sean D Lawley. Extreme hitting probabilities for diffusion. *Journal of Physics A: Mathematical and Theoretical*, 55(34):345002, 2022.

[18] Jacob B Madrid and Sean D Lawley. Competition between slow and fast regimes for extreme first passage times of diffusion. *Journal of Physics A: Mathematical and Theoretical*, 53(33):335002, 2020.

[19] Richard P Mann. Collective decision making by rational individuals. *Proceedings of the National Academy of Sciences*, 115(44):E10387–E10396, 2018.

[20] James AR Marshall, Gavin Brown, and Andrew N Radford. Individual confidence-weighting and group decision-making. *Trends in ecology & evolution*, 32(9):636–645, 2017.

[21] Tyler McMillen and Philip Holmes. The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1):30–57, 2006.

[22] Elchanan Mossel and Omer Tamuz. Opinion exchange dynamics. *arXiv preprint arXiv:1401.4770*, 2014.

[23] Martijn J Mulder, Eric-Jan Wagenmakers, Roger Ratcliff, Wouter Boekel, and Birte U Forstmann. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, 32(7):2335–2343, 2012.

[24] Daniel J Navarro and Ian G Fuss. Fast and accurate calculations for first-passage times in wiener diffusion models. *Journal of Mathematical Psychology*, 53(4):222–230, 2009.

[25] William T Newsome, Kenneth H Britten, and J Anthony Movshon. Neuronal correlates of a perceptual decision. *Nature*, 341(6237):52–54, 1989.

[26] For an ideal observer if only measurement noise is present then drift and diffusion are equal in the continuum limit, $\mu_i = D_i = m_i$ [3, 38]. We consider the more general case where internal noise can increase $D_i > \mu_i$.

[27] This is the distribution of beliefs conditioned on the absence of a decision after a long time.

[28] We interpret beliefs as log-likelihood ratios. Therefore, with k alternatives beliefs evolve in

$k - 1$ dimensions.

[29] G A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*, volume 60. Springer, 2014.

[30] Roger Ratcliff. A theory of memory retrieval. *Psychological review*, 85(2):59, 1978.

[31] Andreagiovanni Reina, Thomas Bose, Vaibhav Srivastava, and James AR Marshall. Asynchrony rescues statistically optimal group decisions from information cascades through emergent leaders. *Royal Society Open Science*, 10(3):230175, 2023.

[32] Michael N Shadlen and William T Newsome. Motion perception: seeing and deciding. *Proceedings of the national academy of sciences*, 93(2):628–633, 1996.

[33] Megan Stickler, William Ott, Zachary P Kilpatrick, Krešimir Josić, and Bhargav R Karamched. Impact of correlated information on pioneering decisions. *Physical Review Research*, 5(3):033020, 2023.

[34] John A Swets, Wilson P Tanner Jr, and Theodore G Birdsall. Decision processes in perception. *Psychological review*, 68(5):301, 1961.

[35] Satohiro Tajima, Jan Drugowitsch, Nisheet Patel, and Alexandre Pouget. Optimal policy for multi-alternative decisions. *Nature neuroscience*, 22(9):1503–1511, 2019.

[36] Alan N Tump, Max Wolf, Pawel Romanczuk, and Ralf HJM Kurvers. Avoiding costly mistakes in groups: the evolution of error management in collective decision making. *PLoS Computational Biology*, 18(8):e1010442, 2022.

[37] Naoshige Uchida and Zachary F Mainen. Speed and accuracy of olfactory discrimination in the rat. *Nature neuroscience*, 6(11):1224–1229, 2003.

[38] A. Veliz-Cuba, Z. P. Kilpatrick, and K. Josić. Stochastic models of evidence accumulation in changing environments. *SIAM Review*, 58(2):264–289, 2016.

[39] Xiao-Jing Wang. Neural dynamics and circuit mechanisms of decision-making. *Current opinion in neurobiology*, 22(6):1039–1046, 2012.

[40] Bram Zandbelt, Braden A Purcell, Thomas J Palmeri, Gordon D Logan, and Jeffrey D Schall. Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences*, 111(7):2848–2853, 2014.