

Conflict and Strategy in Collective Computation

Eleanor Brush, David Krakauer, Jessica Flack

June 30, 2015

Contents

1	Sequential Probability Ratio Test	1
2	Derivation of PDEs for decision time, error rate, and probability of reaching decision preference	2
2.1	Decision time	2
2.2	Error rate	3
2.3	Our model	6
3	Application of model to neural systems	8
4	Dimensionality and initial conditions of the SDEs	8
5	Details on Nash equilibria	9
6	Calculation of mutual information	10
7	Explanation for most informative measures of consensus	10
8	Explanation of skewness being maximized at intermediate waiting costs	11
9	Skewness of DSP across species	11
10	Supplemental figures	12

1 Sequential Probability Ratio Test

In this section, we describe the the sequential probability ratio test (SPRT) and provide more details on why we chose to use a leaky integrator model (LIM) rather than an SPRT. The SPRT is a sequential hypothesis test. We are interested in testing the hypotheses $H_1 : Y$ is drawn from $p_1(y)$ and $H_2 : Y$ is drawn from $p_2(y)$. If we sample Y repeatedly and observe the sequence $\{y_1, \dots, y_k\}$, then we form the decision variable

$$Z_k = \sum_{i=1}^k \log \left(\frac{p_1(y_i)}{p_2(y_i)} \right).$$

The procedure dictated by the SPRT is to continue sampling until it happens that $Z_k > T_1$ or $Z_k < -T_2$, where T_1 and T_2 are thresholds determined by the desired error rate of the hypothesis test. If $Z_k > T_1$ first, then we accept H_1 and, if $Z_k < -T_2$, then we accept H_2 .

The SPRT is the optimal way to decide between two alternatives in a noisy environment, in the sense that, for a given decision time, the SPRT is the most accurate algorithm and, for a given error rate, the SPRT reaches a decision most quickly [? ?]. The SPRT is mathematically equivalent to the LIM when the leak rate $\ell = 0$ [?]. The LIM has the advantage over the SPRT of describing a more realistic leaky decision process and of being more analytically tractable. While the SPRT has been applied to animal conflicts [?], to our knowledge the leaky integrator model has not been.

2 Derivation of PDEs for decision time, error rate, and probability of reaching decision preference

In this section, we derive partial differential equations that describe the decision time, error rate, and probability of reaching decision preference for the LIM described in the main text. The derivation below follows closely that given in [?], which provides more details. For the sake of generality, we will write our system of stochastic differential equations as

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{W}_t,$$

where $\mathbf{x} \in \mathbb{R}^N$, $A : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N$, $\mathbf{B} : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}^N \times \mathbb{R}^N$, and \mathbf{W} is an N -dimensional Wiener process. The forward equation describing the probability density $p(\mathbf{x}, t|\mathbf{y}, s)$ is

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^N \frac{\partial (A_i(\mathbf{x}, t)p)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N \frac{\partial^2 ([\mathbf{B}(\mathbf{x}, t)\mathbf{B}^T(\mathbf{x}, t)]_{ij}p)}{\partial x_i \partial x_j} \quad (1)$$

and the backward equation is

$$- \frac{\partial p}{\partial s} = \sum_{i=1}^N A_i(\mathbf{y}, s) \frac{\partial p}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, s)\mathbf{B}^T(\mathbf{y}, s)]_{ij} \frac{\partial^2 p}{\partial y_i \partial y_j}. \quad (2)$$

2.1 Decision time Let $G(\mathbf{y}, t)$ be the probability that the particle is still in a region R at time t , given that it started at \mathbf{y} so that

$$G(\mathbf{y}, t) = \int_R p(\mathbf{x}, t|\mathbf{y}, 0) d\mathbf{x}.$$

Claim 2.1 If the system is time homogeneous,

$$\frac{\partial G}{\partial t} = \sum_{i=1}^N A_i(\mathbf{y}) \frac{\partial G}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y})\mathbf{B}^T(\mathbf{y})]_{ij} \frac{\partial^2 G}{\partial y_i \partial y_j}. \quad (3)$$

Proof. If the system is time homogeneous, then $p(\mathbf{x}, t|\mathbf{y}, 0) = p(\mathbf{x}, 0|\mathbf{y}, -t)$ so that we can rewrite G as

$$\begin{aligned} G(\mathbf{y}, t) &= \int_R p(\mathbf{x}, 0|\mathbf{y}, -t) d\mathbf{x}, \\ \Rightarrow \frac{\partial G}{\partial t} \Big|_{t=t} &= \int_R \frac{\partial p(\mathbf{x}, 0|\mathbf{y}, -t)}{\partial t} \Big|_{t=t} d\mathbf{x} \\ &= \int_R - \frac{\partial p(\mathbf{x}, 0|\mathbf{y}, s)}{\partial s} \Big|_{s=t} d\mathbf{x} \\ &= \int_R \left(\sum_{i=1}^N A_i(\mathbf{y}) \frac{\partial p}{\partial y_i} \Big|_{s=t} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y})\mathbf{B}^T(\mathbf{y})]_{ij} \frac{\partial^2 p}{\partial y_i \partial y_j} \Big|_{s=t} \right) d\mathbf{x} \text{ using (2)} \\ &= \sum_{i=1}^N A_i(\mathbf{y}) \left(\int_R \frac{\partial p}{\partial y_i} \Big|_{s=t} d\mathbf{x} \right) + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y})\mathbf{B}^T(\mathbf{y})]_{ij} \left(\int_R \frac{\partial^2 p}{\partial y_i \partial y_j} \Big|_{s=t} d\mathbf{x} \right) \\ &= \sum_{i=1}^N A_i(\mathbf{y}) \frac{\partial G}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y})\mathbf{B}^T(\mathbf{y})]_{ij} \frac{\partial^2 G}{\partial y_i \partial y_j} \end{aligned}$$

■

We can now use Equation (3) to find an equation that the expected hitting time satisfies. If T is the time at which the particle hits the boundary $S = \partial R$, then $P(T \geq t|\mathbf{y}) = G(\mathbf{y}, t)$. Therefore

$$\begin{aligned}
E_{\mathbf{y}}(T) &= \int_0^\infty t P(T \in [t, t+dt]|\mathbf{y}) \\
&= \int_0^\infty t \left(P(T \geq t|\mathbf{y}) - P(T \geq t+dt|\mathbf{y}) \right) \\
&= \int_0^\infty t \left(G(\mathbf{y}, t) - G(\mathbf{y}, t+dt) \right) \\
&= \int_0^\infty t \frac{G(\mathbf{y}, t) - G(\mathbf{y}, t+dt)}{dt} dt \\
&= - \int_0^\infty t \frac{\partial G}{\partial t} dt \\
&= - \left([tG(\mathbf{y}, t)]_0^\infty - \int_0^\infty G(\mathbf{y}, t) dt \right) \text{ by integration by parts} \\
&= \int_0^\infty G(\mathbf{y}, t) dt \text{ if } \lim_{t \rightarrow \infty} G(\mathbf{y}, t) = 0 \text{ quickly enough.}
\end{aligned}$$

If we let $f(\mathbf{y}) = E_{\mathbf{y}}(T)$, then

$$\begin{aligned}
&\sum_{i=1}^N A_i(\mathbf{y}, t) \frac{\partial f}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, t) \mathbf{B}^T(\mathbf{y}, t)]_{ij} \frac{\partial^2 f}{\partial y_i \partial y_j} \\
&= \int_0^\infty \left(\sum_{i=1}^N A_i(\mathbf{y}, t) \frac{\partial G}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, t) \mathbf{B}^T(\mathbf{y}, t)]_{ij} \frac{\partial^2 G}{\partial y_i \partial y_j} \right) dt \\
&= \int_0^\infty \frac{\partial G}{\partial t} dt \text{ using (3)} \\
&= G(\mathbf{y}, \infty) - G(\mathbf{y}, 0) \\
&= -1
\end{aligned}$$

Finally, this gives us a PDE for the expected hitting time of the boundary of the region R :

$$\sum_{i=1}^N A_i(\mathbf{y}) \frac{\partial f}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}) \mathbf{B}^T(\mathbf{y})]_{ij} \frac{\partial^2 f}{\partial y_i \partial y_j} = -1$$

with boundary conditions $f(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial R$.

2.2 Error rate Let $\mathbf{J}(\mathbf{x}, t|\mathbf{y}, s) : (\mathbb{R}^N \times \mathbb{R}) \times (\mathbb{R}^N \times \mathbb{R}) \rightarrow \mathbb{R}^N$ be such that

$$J_k(\mathbf{x}, t|\mathbf{y}, s) = A_k(\mathbf{x}, t)p(\mathbf{x}, t|\mathbf{y}, s) - \frac{1}{2} \sum_{l=1}^N \frac{\partial ([\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{kl} p(\mathbf{x}, t|\mathbf{y}, s))}{\partial x_j}.$$

If p satisfies the forward equation (1), then

$$\frac{\partial p}{\partial t} = - \sum_{i=1}^N \frac{\partial (A_i(\mathbf{x}, t)p)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N \frac{\partial^2 ([\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{ij} p)}{\partial x_i \partial x_j} = - \sum_{i=1}^N \frac{\partial J_i(\mathbf{x}, t|\mathbf{y}, s)}{\partial x_i}. \quad (4)$$

For this reason, \mathbf{J} is called the probability current.

Claim 2.2 Each J_k satisfies the backward equation.

Proof.

$$\begin{aligned}
& \sum_{i=1}^N A_i(\mathbf{y}, s) \frac{\partial J_k(\mathbf{x}, t|\mathbf{y}, s)}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, s) \mathbf{B}^T(\mathbf{y}, s)]_{ij} \frac{\partial^2 J_k(\mathbf{x}, t|\mathbf{y}, s)}{\partial y_i \partial y_j} \\
&= \sum_{i=1}^N A_i(\mathbf{y}, s) \left[A_k(\mathbf{x}, t) \frac{\partial p}{\partial y_i} - \frac{1}{2} \sum_{l=1}^N \frac{\partial \left([\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{kl} \frac{\partial p(\mathbf{x}, t|\mathbf{y}, s)}{\partial y_i} \right)}{\partial x_l} \right] \\
&+ \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, s) \mathbf{B}^T(\mathbf{y}, s)]_{ij} \left[A_k(\mathbf{x}, t) \frac{\partial^2 p}{\partial y_i \partial y_j} - \frac{1}{2} \sum_{l=1}^N \frac{\partial \left([\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{kl} \frac{\partial^2 p(\mathbf{x}, t|\mathbf{y}, s)}{\partial y_i \partial y_j} \right)}{\partial x_l} \right] \\
&= A_k(\mathbf{x}, t) \left[\sum_{i=1}^N A_i(\mathbf{y}, s) \frac{\partial p}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, s) \mathbf{B}^T(\mathbf{y}, s)]_{ij} \frac{\partial^2 p}{\partial y_i \partial y_j} \right] \\
&- \frac{1}{2} \sum_{l=1}^N \frac{\partial}{\partial x_l} \left([\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{kl} \left[\sum_{i=1}^N A_i(\mathbf{y}, s) \frac{\partial p}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, s) \mathbf{B}^T(\mathbf{y}, s)]_{ij} \frac{\partial^2 p}{\partial y_i \partial y_j} \right] \right) \\
&= A_k(\mathbf{x}, t) \left(-\frac{\partial p}{\partial s} \right) - \frac{1}{2} \sum_{l=1}^N \frac{\partial}{\partial x_l} \left(-[\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{kl} \frac{\partial p}{\partial s} \right) \text{ using (2)} \\
&= -\frac{\partial}{\partial s} \left(A_k(\mathbf{x}, t) p(\mathbf{x}, t|\mathbf{y}, s) - \frac{1}{2} \sum_{l=1}^N \frac{\partial ([\mathbf{B}(\mathbf{x}, t) \mathbf{B}^T(\mathbf{x}, t)]_{kl} p(\mathbf{x}, t|\mathbf{y}, s))}{\partial x_l} \right) \\
&= -\frac{\partial J_k}{\partial s}.
\end{aligned}$$

■

It follows that any linear combination $q(\mathbf{x}, t|\mathbf{y}, s) = \sum_{k=1}^N c_k J_k(\mathbf{x}, t|\mathbf{y}, s)$ also satisfies the backward equation:

$$-\frac{\partial q}{\partial s} = \sum_{i=1}^N A_i(\mathbf{y}, s) \frac{\partial q}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{y}, s) \mathbf{B}^T(\mathbf{y}, s)]_{ij} \frac{\partial^2 q}{\partial y_i \partial y_j}. \quad (5)$$

Claim 2.3 Consider two adjacent regions of space R_1 and R_2 , where their shared boundary is denoted S_{12} . The net flow of probability from R_2 to R_1 can be written as $\int_{S_{12}} dS \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t)$ where \mathbf{n} is a normal vector from R_2 to R_1 .

Proof. Let S_1 be the part of the boundary of R_1 not shared with R_2 and S_2 be the part of the boundary of R_2 not shared with R_1 so that $\partial R_1 = S_1 \cup S_{12}$, $\partial R_2 = S_2 \cup S_{12}$, and $S_1 \cap S_{12} = S_2 \cap S_{12} = \emptyset$. The probability of crossing from R_2 to R_1 across S_{12} in the interval of time $[t, t + dt)$ is

$$\int_{R_1} d\mathbf{x} \int_{R_2} d\mathbf{y} p(\mathbf{x}, t + dt|\mathbf{y}, t).$$

The probability of crossing from R_1 to R_2 across S_{12} in the interval of time $[t, t + dt)$ can be found analogously so the net flow F from R_2 to R_1 at time t is

$$F = \lim_{dt \rightarrow 0} \frac{1}{dt} \int_{R_1} d\mathbf{x} \int_{R_2} d\mathbf{y} (p(\mathbf{x}, t + dt|\mathbf{y}, t) - p(\mathbf{y}, t + dt|\mathbf{x}, t)).$$

Since $p(\mathbf{x}, t|\mathbf{y}, t) = p(\mathbf{y}, t|\mathbf{x}, t) = 0$ for $\mathbf{y} \in R_2$ and $\mathbf{x} \in R_1$,

$$\begin{aligned}
F &= \lim_{dt \rightarrow 0} \frac{1}{dt} \int_{R_1} d\mathbf{x} \int_{R_2} d\mathbf{y} (p(\mathbf{x}, t+dt|\mathbf{y}, t) - p(\mathbf{x}, t|\mathbf{y}, t) - p(\mathbf{y}, t+dt|\mathbf{x}, t) + p(\mathbf{y}, t|\mathbf{x}, t)) \\
&= \int_{R_1} d\mathbf{x} \int_{R_2} d\mathbf{y} \left(\left. \frac{\partial p(\mathbf{x}, t'|\mathbf{y}, t)}{\partial t'} \right|_{t'=t} - \left. \frac{\partial p(\mathbf{y}, t'|\mathbf{x}, t)}{\partial t'} \right|_{t'=t} \right) \\
&= \int_{R_1} d\mathbf{x} \int_{R_2} d\mathbf{y} \left(- \sum_{i=1}^N \frac{\partial J_i(\mathbf{x}, t|\mathbf{y}, t)}{\partial x_i} + \sum_{i=1}^N \frac{\partial J_i(\mathbf{y}, t|\mathbf{x}, t)}{\partial y_i} \right) \text{ using (4)} \\
&= \int_{R_1} d\mathbf{x} \int_{R_2} d\mathbf{y} \sum_{i=1}^N \frac{\partial J_i(\mathbf{y}, t|\mathbf{x}, t)}{\partial y_i} - \int_{R_2} d\mathbf{y} \int_{R_1} d\mathbf{x} \sum_{i=1}^N \frac{\partial J_i(\mathbf{x}, t|\mathbf{y}, t)}{\partial x_i} \\
&= \int_{R_1} d\mathbf{x} \left[\int_{S_2} dS \mathbf{n}_2 \cdot \mathbf{J}(\mathbf{y}, t|\mathbf{x}, t) + \int_{S_{12}} dS \mathbf{n}_2 \cdot \mathbf{J}(\mathbf{y}, t|\mathbf{x}, t) \right] \\
&\quad - \int_{R_2} d\mathbf{y} \left[\int_{S_1} dS \mathbf{n}_1 \cdot \mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) + \int_{S_{12}} dS \mathbf{n}_1 \cdot \mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) \right]
\end{aligned}$$

by the divergence theorem,

where \mathbf{n}_2 is a normal pointing out of R_2 and similarly for \mathbf{n}_1 .

For $\mathbf{y} \in S_2$ and $\mathbf{x} \in R_1$, $J(\mathbf{y}, t|\mathbf{x}, t) = 0$. Similarly, for $\mathbf{x} \in S_1$ and $\mathbf{y} \in R_2$, $\mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) = 0$. Therefore, the integrals over S_2 and S_1 disappear and F simplifies to

$$\begin{aligned}
F &= \int_{R_1} d\mathbf{x} \left[\int_{S_{12}} dS \mathbf{n}_2 \cdot \mathbf{J}(\mathbf{y}, t|\mathbf{x}, t) \right] - \int_{R_2} d\mathbf{y} \left[\int_{S_{12}} dS \mathbf{n}_1 \cdot \mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) \right] \\
&= \int_{S_{12}} dS \left[\int_{R_1} d\mathbf{y} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) + \int_{R_2} d\mathbf{y} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) \right] \\
&\quad \text{where } \mathbf{n} \text{ is a normal pointing from } R_2 \text{ to } R_1 \\
&= \int_{S_{12}} dS \left[\int_{R_1 \cup R_2} d\mathbf{y} \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t|\mathbf{y}, t) \right] \\
&= \int_{S_{12}} dS \mathbf{n} \cdot \mathbf{J}(\mathbf{x}, t).
\end{aligned}$$

■

Now consider a region R with boundary S . Let $g(\mathbf{a}, \mathbf{x}, t)|d\mathbf{S}(\mathbf{a})|$ be the probability of exiting through an element $d\mathbf{S}(\mathbf{a})$ of the boundary S after time t .

Claim 2.4 If the system is time homogeneous,

$$\frac{\partial g}{\partial t} = \sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial g}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}. \quad (6)$$

Proof. By the definition of g ,

$$\begin{aligned}
g(\mathbf{a}, \mathbf{x}, t) |d\mathbf{S}(\mathbf{a})| &= \int_t^\infty dt' \mathbf{J}(\mathbf{a}, t' | \mathbf{x}, 0) \cdot d\mathbf{S}(\mathbf{a}) \text{ using Claim 2.3} \\
&= \int_t^\infty dt' \mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, -t') \cdot d\mathbf{S}(\mathbf{a}) \\
&= \int_{-\infty}^{-t} dt' \mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, t') \cdot d\mathbf{S}(\mathbf{a}) \\
&\quad \text{using integration by substitution} \\
\text{so that } \sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial g}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j} &= \\
&\quad \int_{-\infty}^{-t} dt' \left(\sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial (\mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, t') \cdot d\mathbf{S}(\mathbf{a}))}{\partial x_i} \right. \\
&\quad \left. + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 (\mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, t') \cdot d\mathbf{S}(\mathbf{a}))}{\partial x_i \partial x_j} \right) \\
&= \int_{-\infty}^{-t} dt' \left(- \frac{\partial (\mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, t') \cdot d\mathbf{S}(\mathbf{a}))}{\partial t'} \right) \text{ using Eq. (5)} \\
&= \left(- \mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, t') \cdot d\mathbf{S}(\mathbf{a}) \right) \Big|_{-\infty}^{-t} \\
&= \left(- \mathbf{J}(\mathbf{a}, -t' | \mathbf{x}, 0) \cdot d\mathbf{S}(\mathbf{a}) \right) \Big|_{-\infty}^{-t} \\
&= -\mathbf{J}(\mathbf{a}, t | \mathbf{x}, 0) \cdot d\mathbf{S}(\mathbf{a}) \\
&= \frac{\partial g}{\partial t}.
\end{aligned}$$

■

We are interested in finding the total probability of exiting through $d\mathbf{S}(\mathbf{a})$, $h(\mathbf{a}, \mathbf{x}) = g(\mathbf{a}, \mathbf{x}, 0)$. In the course of the proof of the last claim, we found that

$$\sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial g(\mathbf{a}, \mathbf{x}, t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 g(\mathbf{a}, \mathbf{x}, t)}{\partial x_i \partial x_j} = -\mathbf{J}(\mathbf{a}, t | \mathbf{x}, 0) \cdot d\mathbf{S}(\mathbf{a}).$$

If we let $t \rightarrow 0$ and notice that $\mathbf{J}(\mathbf{a}, 0 | \mathbf{x}, 0) = \delta(|\mathbf{a} - \mathbf{x}|)$, we find that

$$\sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial h}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 h}{\partial x_i \partial x_j} = 0$$

with boundary conditions $h(\mathbf{a}, \mathbf{x}) = \delta(|\mathbf{a} - \mathbf{x}|)$ for \mathbf{x} in S . To find the total probability of exiting through a section $S' \subset S$, $H(S', \mathbf{x})$, we can simply integrate the solutions $h(\mathbf{a}, \mathbf{x})$ for $\mathbf{a} \in S'$ to find

$$\sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial H}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 H}{\partial x_i \partial x_j} = 0$$

with boundary conditions $H(S', \mathbf{x}) = \delta(\mathbf{x} \in S')$ for $\mathbf{x} \in S$.

2.3 Our model We now introduce the operator \mathbf{L} such that

$$\mathbf{L}(f) = \sum_{i=1}^N A_i(\mathbf{x}) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^N [\mathbf{B}(\mathbf{x}) \mathbf{B}^T(\mathbf{x})]_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Then the expected time to hit the boundary S satisfies $\mathbf{L}f = -1$ with boundary conditions $f(\mathbf{x}) = 0$ for $\mathbf{x} \in S$ and the probability of hitting a subset $S' \subset S$ satisfies $\mathbf{L}H = 0$ with boundary conditions $H(\mathbf{x}) = \delta(\mathbf{x} \in S')$ for $\mathbf{x} \in S$. The stochastic differential equations in our model are

$$\begin{aligned} dX_1 &= \left(-\ell X_1(t) + br(2c-1) \right) dt + \left(b\sqrt{rc} \right) dW_{At} - \left(b\sqrt{r(1-c)} \right) dW_{Bt} \\ dX_2 &= \left(-\ell X_2(t) - br(2c-1) \right) dt - \left(b\sqrt{rc} \right) dW_{At} + \left(b\sqrt{r(1-c)} \right) dW_{Bt}, \end{aligned}$$

This gives

$$\begin{aligned} \mathbf{A}(\mathbf{x}, t) &= \begin{pmatrix} -\ell x_1 + br(2c-1) \\ -\ell x_2 - br(2c-1) \end{pmatrix}, \text{ and} \\ \mathbf{B}(\mathbf{x}, t) &= \begin{pmatrix} b\sqrt{rc} & -b\sqrt{r(1-c)} \\ -b\sqrt{rc} & b\sqrt{r(1-c)} \end{pmatrix}, \text{ so that} \\ \mathbf{B}(\mathbf{x}, t)\mathbf{B}^T(\mathbf{x}, t) &= \begin{pmatrix} b^2r & b^2r \\ b^2r & b^2r \end{pmatrix} \end{aligned}$$

Thus, the operator \mathbf{L} is given by

$$\begin{aligned} \mathbf{L}(f) &= (-\ell x_1 + br(2c-1)) \frac{\partial f}{\partial x_1} + (-\ell x_2 - br(2c-1)) \frac{\partial f}{\partial x_2} \\ &\quad + \frac{1}{2} b^2 r \frac{\partial^2 f}{\partial x_1^2} + \frac{1}{2} b^2 r \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{2} b^2 r \frac{\partial^2 f}{\partial x_2 \partial x_1} + \frac{1}{2} b^2 r \frac{\partial^2 f}{\partial x_2^2} \\ &= (-\ell x_1 + br(2c-1)) \frac{\partial f}{\partial x_1} + (-\ell x_2 - br(2c-1)) \frac{\partial f}{\partial x_2} \\ &\quad + b^2 r \left(\frac{1}{2} \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{2} \frac{\partial^2 f}{\partial x_2^2} \right) \end{aligned}$$

The region we are interested in is the one in which both individuals' opinions (x_1, x_2) are greater than or equal to their respective thresholds, i.e.

$$R = \{(x_1, x_2) : x_1 \geq -T_1 \text{ and } x_2 \geq -T_2\}$$

so that $S = \partial R = \{(x_1, -T_2) : x_1 > -T_1\} \cup \{(-T_1, x_2) : x_2 > -T_2\}$. The subset of the boundary we are interested in is the one that indicates a correct decision has been made, i.e. $x_2 = -T_2$, so that $S' = \{(x_1, -T_2) : x_1 > -T_1\}$. The expected decision time is the expected hitting time of S . The probability of a correct decision is the probability of exiting R through S' . The first component prefers that the second component reaches its threshold first, so the probability of reaching its decision preference is the same as the probability of a correct decision. The probability of reaching the decision preference of the second component is 1 minus that probability.

There are analytical solutions to these equations when the system of SDEs is reduced to one dimension. However, we were not able to find an analytical solution for the full two-dimensional system and we therefore used numerical methods to solve the PDEs in this case. In order to numerically solve the PDEs for decision time and error rate, we artificially close the region R so that we can impose boundary conditions on a finite subset of the plane. In order to this, we pick two upper bounds B_1 and B_2 and restrict R to

$$R_2 = \{(x_1, x_2) : x_1 \in [-T_1, B_1] \text{ and } x_2 \in [-T_2, B_2]\},$$

with boundary S_2 . The subset of the boundary we are interested in becomes

$$S'_2 = \{(x_1, -T_2) : x_1 \in [-T_1, B_1]\} \cup \{(B_1, x_2) : x_2 \in [-T_2, B_2]\}$$

The boundary condition $f(\mathbf{x}) = 0$ for $\mathbf{x} \in S$ becomes $f(\mathbf{x}) = 0$ for $\mathbf{x} \in S_2$ and the condition $H(\mathbf{x}) = \delta(\mathbf{x} \in S')$ for $\mathbf{x} \in S$ becomes $H(\mathbf{x}) = \delta(\mathbf{x} \in S'_2)$ for $\mathbf{x} \in S_2$.

3 Application of model to neural systems

In this section, we provide details on how the LIM has been applied to neural decision-making. The experimental paradigm the LIM is often applied to is one in which an experimental subject, usually a monkey, is shown a visual stimulus in which some dots are moving either left or right and the rest are flashing randomly. The LIM describes the firing rates of two neuronal populations, one of which fires in response to left motion and the other of which fires in response to right motion. Our derivation of the LIM for our model system could also be used to derive the LIM describing the firing rates of the two populations. Instead of wins and losses, the two types of inputs would be dots moving left and dots moving right. Instead of an asymmetry in fighting ability, c would describe the coherence of the moving dots. In models of the firing rates of neuronal populations, there are usually two independent sources of noise affecting the two neuronal populations, whereas in our model there is a single source of noise affecting both decision variables. However, we could derive independent sources of noise using the same framework. Table 2 summarizes the interpretation of the model in the two systems. The similarity in the two models highlights the similarities between the two systems.

4 Dimensionality and initial conditions of the SDEs

In this section, we discuss the sensitivity of our model to the initial conditions and we show that the two-dimensional and one-dimensional decision processes are nearly equivalent. In our analyses, we assume that both decision variables start at 0, i.e. $X_1(0) = X_2(0) = 0$. If a component were to have previous information, it might start with a different initial condition. In our model system, a monkey could gain information about another monkey by observing its fights against other opponents, without ever encountering it in a fight directly. If the monkey has prior information that the other monkey has a higher fighting ability, it might start with its decision variable below 0 and conversely if it has information that the other monkey has a lower fighting ability. However, incorporating this into our model would require additional assumptions about how observations affect the decision variables, so for simplicity we assume that both decision variables start at 0.

Regardless of initial conditions, we can compare the error rate and decision time from a two-dimensional process and a one-dimensional process. The two-dimensional process ends when $X_1 < -T_1$ or $X_2 < -T_2$, for given thresholds T_1 and T_2 . If $Y = X_1 - X_2$, the one-dimensional process ends when $Y > T_1$ or $Y < -T_2$, for (possibly different) thresholds T_1 and T_2 . If we let $Z = X_1 + X_2$, then we can use our model to write down the dynamics of Y and Z . Specifically, from the dynamics of X_1 and X_2 ,

$$\begin{aligned} dX_1 &= \left(-\ell X_1(t) + br(2c - 1) \right) dt + b\sqrt{r}dW_t \\ dX_2 &= \left(-\ell X_2(t) - br(2c - 1) \right) dt - b\sqrt{r}dW_t, \end{aligned}$$

we find that

$$\begin{aligned} dY &= \left(-\ell Y + 2br(2c - 1) \right) dt + 2b\sqrt{r}dW_t \\ dZ &= -\ell Z dt. \end{aligned}$$

Since X_1 and X_2 experience the same noise with opposite signs, their sum Z is described by a deterministic differential equation with solution $Z(t) = Z(0)\exp(-\ell t)$. With the initial conditions $X_1(0) = X_2(0) = 0$, $Z(t) = 0$ for all t and the dynamics of X_1 and X_2 are entirely described by the dynamics of Y . Otherwise, Y only gives one dimension of the dynamics of (X_1, X_2) .

To make the processes directly comparable, we consider symmetric thresholds, i.e. $T_1 = T_2$. With this constraint, for a given level of difficulty c , error rate becomes a well-defined function of decision time. With the initial conditions $X_1(0) = X_2(0) = 0$, Y captures the full dynamics of X_1 and X_2 . Accordingly, the error rate given by analyzing either the two-dimensional or the one-dimensional process are identical (Figure S1a). With initial conditions biased toward the incorrect decision (i.e. $c > 0.5$ and $X_1(0) < X_2(0)$), the two-dimensional process actually makes a decision with slightly lower error rate for a given decision time (Figure S1b). Conversely, with initial conditions biased toward the correct

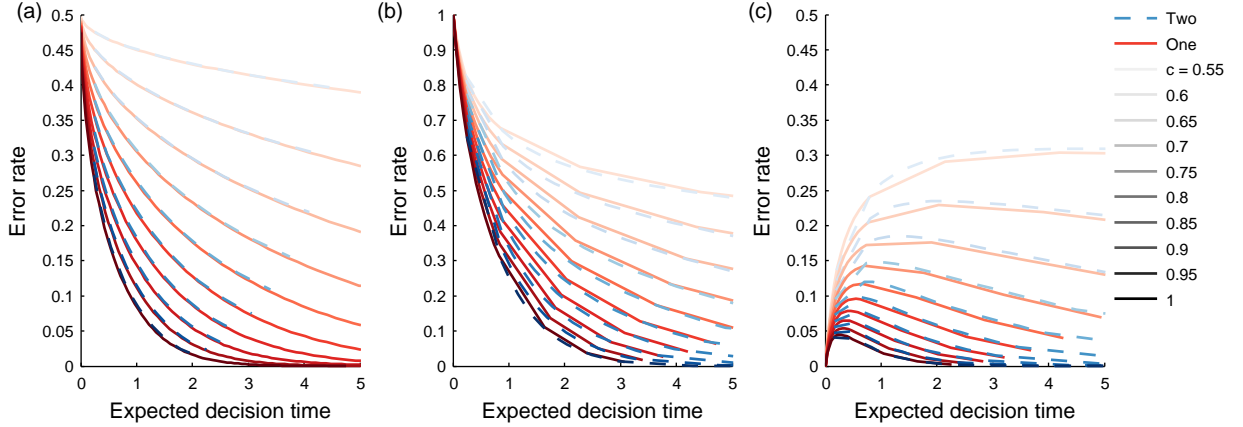


Figure S1: The error rate decreases as decision time increases, as long as the initial conditions are not biased toward the correct decision. Here we show the probability of a correct output as a function of the expected time to a decision, as the decision thresholds are varied, with the constraint that $T_1 = T_2$. The red lines correspond to the reduced one-dimensional model and the dashed blue lines correspond to the full two-dimensional process. Decisions of different difficulties (c) are represented with lines of different intensity. In (a), both decision variables start at 0 and the two processes are equally accurate. In (b), $X_1(0) = 1$, $X_2(0) = 0.1$ (incorrect bias) and the two-dimensional process has a slightly lower error rate. In (c), $X_1(0) = 0.1$, $X_2(0) = 1$ (correct bias) and the two-dimensional process has a slightly higher error rate. (Parameters: $N = 20$, $b = 1$, $r = 1$, $\ell = 0.1$).

decision (i.e. $c > 0.5$ and $X_1(0) > X_2(0)$), the two-dimensional process makes a decision with slightly higher error rate for a given decision time (Figure S1c). However, the differences between the two processes are negligible and we use results from an analysis of the full two-dimensional process because it better describes our model system.

The initial conditions affect the tradeoff between error rate and decision time, and therefore the absolute value of the Nash thresholds. To avoid including additional mechanisms in our model, we focus on the Nash thresholds when the initial conditions are not biased. However, we expect the Nash thresholds with bias toward the incorrect decision to respond similarly to how the Nash thresholds without bias respond to the optimization weights: they should increase when decision preference is more important and decrease when decision time is more important. The situation is slightly different if the initial conditions are biased toward the correct decision. In this case, the pair can optimize both error rate and decision time by setting their thresholds low (Figure S1c).

5 Details on Nash equilibria

In this section, we discuss our procedure for finding the Nash equilibrium thresholds and how they might be reached in the model system. The Nash equilibrium thresholds are those such that no component can improve its utility by changing its threshold. The thresholds we allow in our model are $0.5, 0.6, \dots, 1.9, 2$. To find the Nash thresholds, we assign all individuals in a group the highest possible strategy (although the procedure is robust to this choice of initial conditions). Each component, in turn, chooses the threshold that gives it the best utility, given the thresholds that the other components are using, until the components equilibrate on a set of Nash thresholds. If the Nash thresholds are $\{T_i\}$, there may be a set of thresholds $\{T'_i\}$ such that some components' utilities are better when the group uses the second set. However, no single individual has an incentive to change its strategy from T_i to T'_i , and when the group uses the thresholds $\{T'_i\}$, at least one component will have an incentive to change its strategy. This sub-optimality is highlighted by our result that low-valued components would do better if all components raised their thresholds, even though no individual component has an incentive to do so. Our result agrees with previous findings that evolutionary dynamics of strategies

for gathering social information can lead a population to use suboptimal strategies [?]. By assuming that the components use Nash strategies, we are able to analyze the effects of the optimization weights, but we do not address the interesting question of how those strategies might be reached. Finding the appropriate timescale separation between the dynamics of how strategies are learned and the decision dynamics, and the consequences of this additional level of learning, are left for future work.

6 Calculation of mutual information

In this section, we explain how we calculate the mutual information between each measure of consensus and the underlying fighting abilities. As explained in the main text and illustrated in Figure S2, to generate a signaling network, we first draw a set of N values $\{a_1, \dots, a_N\}$ from a uniform distribution. Then we find the Nash thresholds $\{T_1, \dots, T_N\}$, and then we find the expected decision times and probabilities of the decision outcomes for each pair of individuals. We then form an edge between each pair according to these probabilities, with a weight that depends on the decision time. Once we have this signaling network, we apply each of our four consensus measures to it, generating consensus scores $\{s_1, \dots, s_N\}$. To measure how well each measure performs, we find the mutual information between the consensus scores and the values. The mutual information between two random variables X and Y is

$$I(X, Y) = \int_y \int_x p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (7)$$

where $p(x, y)$ is the joint probability of X and Y and $p(x)$ and $p(y)$ are the marginal probabilities of X and Y respectively. For us, this becomes,

$$I(\{a_i\}, \{s_i\}) = \int_{\{a_i\} \in [0,1]^N} \int_{\{s_i\} \in \mathbb{R}^N} p(\{a_i\}, \{s_i\}) \log \left(\frac{p(\{a_i\}, \{s_i\})}{p(\{a_i\})p(\{s_i\})} \right). \quad (8)$$

If a measure produces consensus scores with high mutual information between the scores and the values, then knowing the consensus scores of the components in the system tells us (and possibly the components themselves) about the underlying values. However, calculating the mutual information requires many more samples than are computationally feasible. To reduce the number of states whose probabilities we want to evaluate, we discretize both the values and the scores. Rather than trying to use the full set of consensus scores to infer the values, this is more like using the order of the consensus scores to infer the order of the values. Specifically, given a set of values $\{a_1, \dots, a_N\}$, we bin them into 10 bins and assign each component a new value $a'_i \in \{1, \dots, 10\}$ corresponding to the bin in which its value falls. Similarly, given a set of consensus score $\{s_1, \dots, s_N\}$, we bin them into 10 bins and assign each component a new score $s'_i \in \{1, \dots, 10\}$ corresponding to the bin in which its score falls. The mutual information between the discretized values and scores is

$$I(\{a'_i\}, \{s'_i\}) = \sum_{\{a'_i\} \in \{1, \dots, 10\}^N} \sum_{\{s'_i\} \in \{1, \dots, 10\}^N} p(\{a'_i\}, \{s'_i\}) \log \left(\frac{p(\{a'_i\}, \{s'_i\})}{p(\{a'_i\})p(\{s'_i\})} \right). \quad (9)$$

7 Explanation for most informative measures of consensus

In this section, we provide further explanation for why different measures of consensus are most informative with different optimization weights. Which measure is the most informative about the individuals' true values depends on how accurately pairwise decisions tend to be and what kind of errors are most frequent. When the error rate is quite low ($w_2 < 0.4$), the finer measures of consensus—weighted in-degree and eigenvector centrality—can use all of the information in the decision network to make fine distinctions between individuals, leading to consensus scores that are more informative than those from the coarser measures—in-degree and entropy. Whether weighted in-degree or eigenvector centrality is more informative depends on what type of error is common. When there are no waiting costs ($w_2 = 0$) and the individuals value their personal preferences over error rate ($w_1 < 0.8$), low-valued individuals can sometimes wait out higher valued individuals, leading to an incorrect decision being made in favor

of a low individual at the expense of a high individual. Since eigenvector centrality takes the identity of the source of decisions into account, it will incorrectly assign the low individual a high score, whereas weighted in-degree will ignore the source and assign the low individual an appropriately low score. Therefore, in these circumstances, weighted in-degree is more informative. When error rate matters ($w_1 \geq 0.8$, $w_2 = 0$) or when there are low non-zero waiting costs ($0 < w_2 \leq 0.3$) low-valued individuals are less likely to wait out high-valued individuals, but there are a moderate amount of mistakes between pairs of individuals in the middle and bottom of the group. Consider two intermediate-valued individuals, with A having slightly higher value than B . Another individual might make a mistake with respect to B so that weighted in-degree gives B a slightly higher score than A . But A and B will mostly decide correctly so that, despite errors, B will tend to send more decisions than A . Since eigenvector centrality is affected by the decisions sent as well as received, it will correctly give A a higher score. Therefore, in these circumstances, eigenvector centrality is the most informative. As waiting costs increase ($0.4 \leq w_2 \leq 0.5$) and decisions become even less accurate, even the high-valued individuals start to make mistakes. Now consider two high-valued individuals, with A having slightly higher value than B . If A errs and decides strongly in favor of a lower value individual, eigenvector centrality will give A a lower score because of those sent decisions and weighted in-degree will give A a lower score because of the decisions it has prevented itself from receiving. However, A is likely to make at most one more mistake than B so that the number of individuals deciding in favor of either will be the same. Thus the coarseness of unweighted in-degree becomes an advantage by allowing it to group high-valued individuals together when finer measures would make incorrect fine distinctions. When decisions are even less accurate ($0.5 < w_2 < 0.7$), weighted in-degree, unweighted in-degree, and eigenvector centrality can all get confused by incorrect decisions between any pair of individuals. What becomes especially difficult is distinguishing between low-value individuals, one of whom might incorrectly receive decisions from higher individuals. If a low-value individual receives correct decisions, they will be of similar strengths from even lower individuals. However, if it receives decisions from a higher individual, even though its score from any of the other measures will be increased, entropy decreases its score because of the difference in the strengths of signals coming from different types of individuals. In these circumstances, entropy does well at measuring the lack of consensus of decisions with respect of low individuals. When decisions are very inaccurate ($w_2 > 0.6$), weighted in-degree becomes the most reliable.

8 Explanation of skewness being maximized at intermediate waiting costs

In this section, we provide further explanation for why the skewness of the consensus scores is maximized at intermediate waiting costs. When waiting costs are low and accuracy is high, the decisions accurately represent the individuals' true abilities, so the distribution of number of signalers accurately represents the distribution of abilities, which is not very right-skewed. When waiting costs are high, decisions are so noisy that all individuals receive a similar number of decisions, giving a distribution that is quite uniform and not right-skewed. At intermediate waiting costs, the decisions between individuals with low to intermediate values are very noisy, so that the individuals in the bottom and middle of the group receive a similar number of signals, but all individuals make accurate decisions with the top few individuals, which gives them high scores and results in a right-skewed distribution.

9 Skewness of DSP across species

In this section, we explain how our model could explain differences in the DSP across species of macaques. Different species of macaque have stereotypical power structures with different amounts of skewness [? ? ? ?]. One explanation might be that the distribution of fighting abilities is different from species to species, for instance if there is more or less variation in body size. However, our model suggests that even if the underlying distribution of fighting abilities are identical, the priorities given to preference, error rate, and waiting costs can affect the distribution of power. The benefit of having

a partner agree to be subordinate, the benefit of reaching the “correct” agreement, and the costs of fighting may vary from species to species, causing variation in the shape of the power structure.

10 Supplemental figures

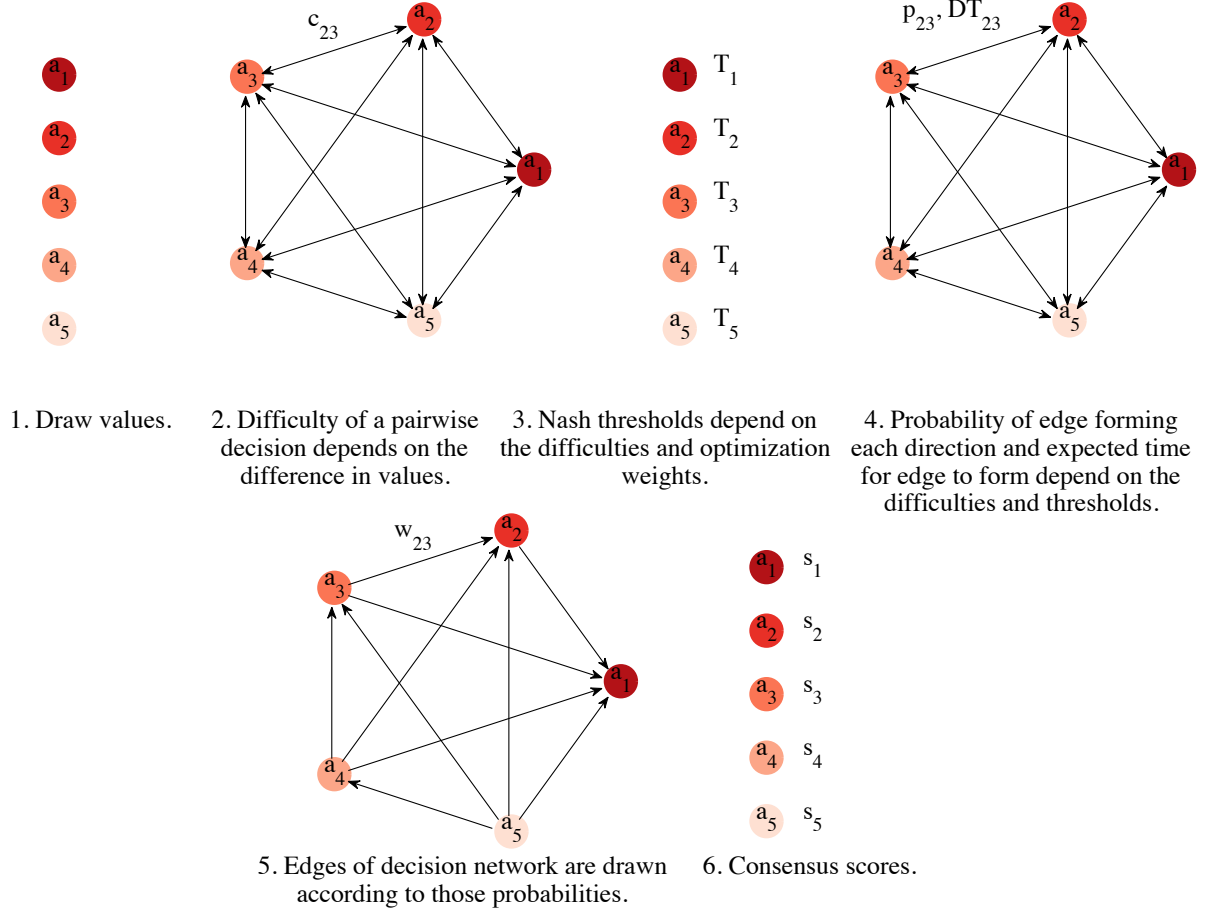


Figure S2: Schematic of the model. 1. First, each individual's value is drawn from a uniform distribution. 2. The difficulty of each pairwise decision increases with the similarity of values, so that c_{ij} increases as $|a_i - a_j|$ increases. 3. Given these difficulties, each individual has a Nash threshold strategy T_i . 4. Given the set of Nash thresholds, a pair will reach each decision with some probability and in an expected amount of time. 5. We draw the direction of each edge according to the probabilities of the two possible outcomes; the strength of the decision decreases with increasing decision time. 6. We apply a set of measures of consensus to each decision network, giving each individual a consensus score. We repeat this process for many random draws of values to find average Nash thresholds and the mutual information between the consensus scores and the values.

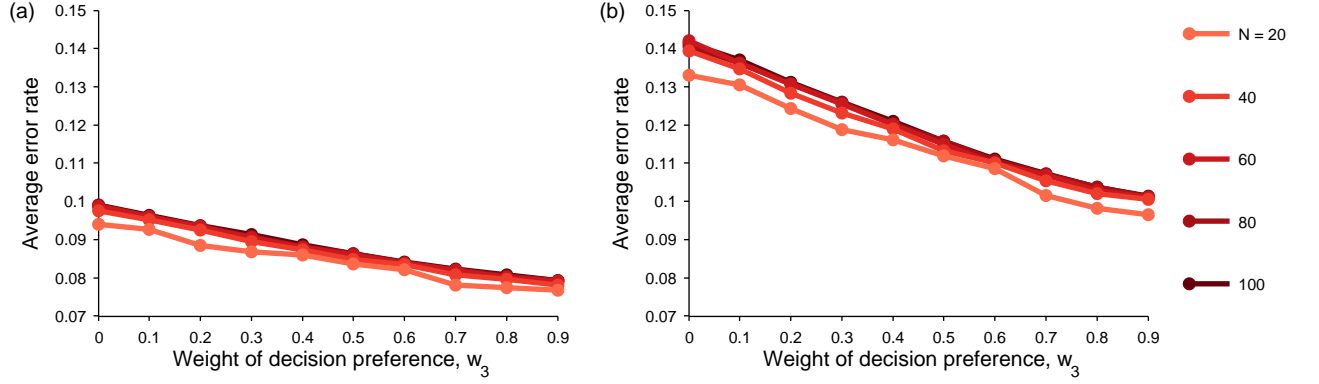


Figure S3: Increasing the weight given to decision preference improves the average error rate, regardless of the size of the group. In each panel, the weight given to decision preference, w_3 , is on the horizontal axis and average error rate is on the vertical axis. In (a) we show the average error rate of the whole group, and in (b) we show the average error rate of decisions made by the bottom quartile. Each line corresponds to a different group size N . Parameters: $b = 1$, $r = 1$, $\ell = 0.1$, $w_2 = 0.1$, $w_1 = 1 - w_3$.

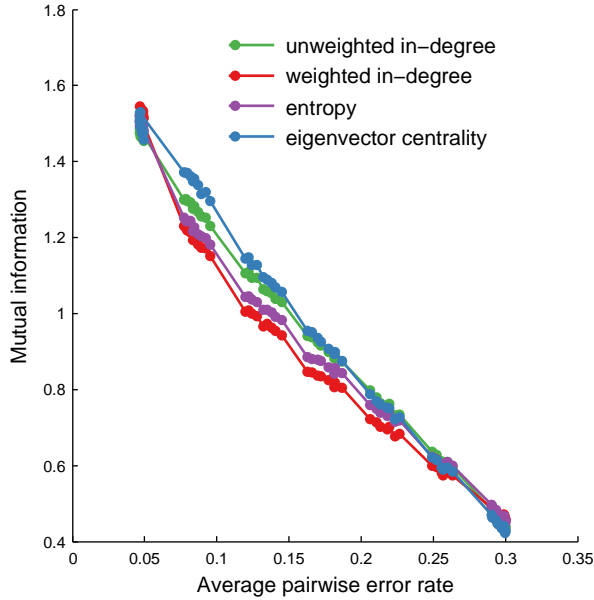


Figure S4: The mutual information of each consensus algorithm is a decreasing function of the average pairwise error rate. This allows us to use average pairwise error rate as a proxy for the overall information content of the decision network, regardless of the algorithm being used. Parameters: $N = 20$, $b = 1$, $r = 1$, $\ell = 0.1$.

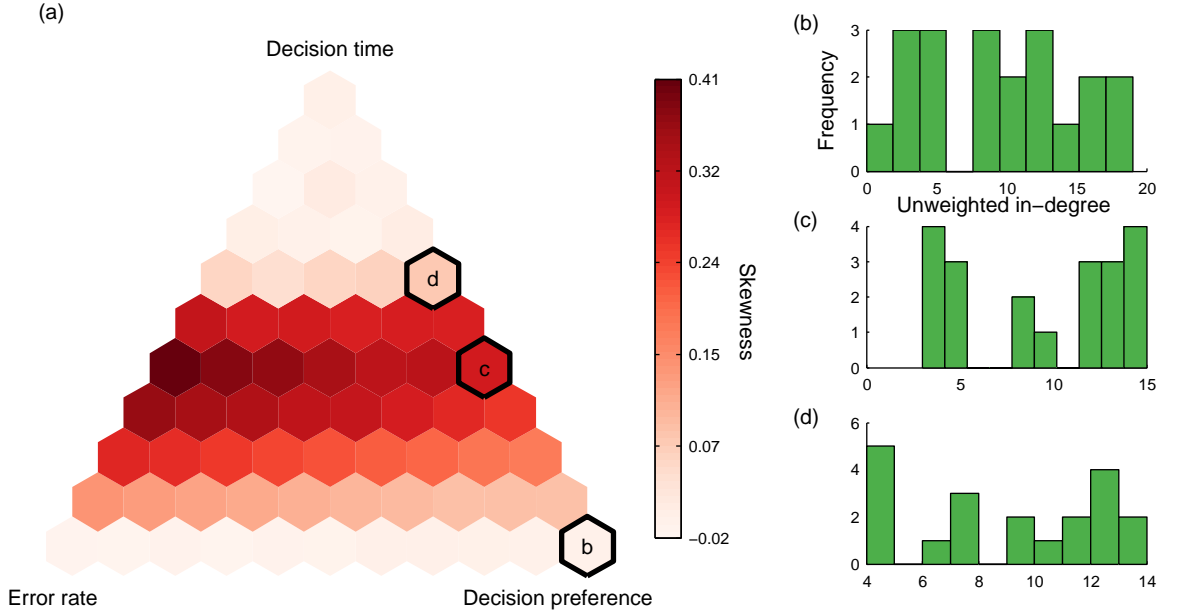


Figure S5: The average skewness of the distribution of unweighted in-degree is maximized at intermediate waiting costs. In (a), the color indicates the average skewness of the distribution of weighted in-degree for a group using Nash thresholds, as a function of the optimization weights, w_1, w_2, w_3 . In the lower left corner of the simplex, only error rate matters ($w_1 = 1$). In the upper corner, only decision time matters ($w_2 = 1$). In the lower right corner, only preference matters ($w_3 = 1$). In (b)-(d) we show a histogram for a representative distribution of unweighted in-degree scores for a group using Nash thresholds. The optimization weights for each panel are indicated in the simplex with the corresponding letter. Parameters: $w_1 = 0, w_2 = 0, w_3 = 1$ (c), $w_1 = 0, w_2 = 0.4, w_3 = 0.6$ (d), $w_1 = 0, w_2 = 0.6, w_3 = 0.4$ (e), $N = 20, b = 1, r = 1, \ell = 0.1$.

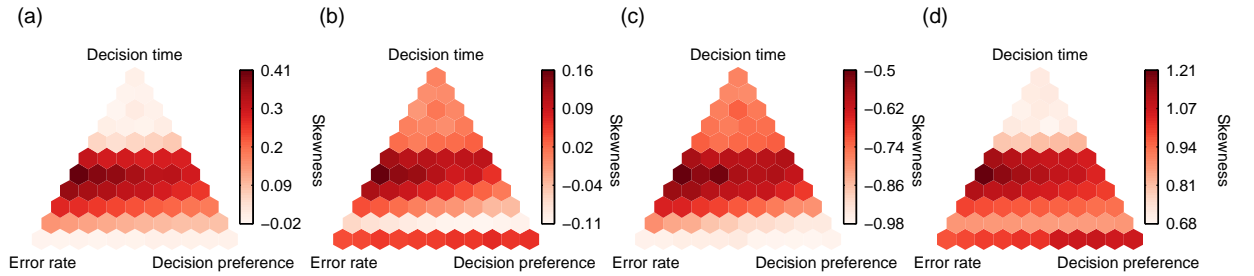


Figure S6: The average skewness of the distribution of consensus scores from each measure is maximized at intermediate waiting costs. In each panel, the color indicates the average skewness of the consensus scores of a group using Nash thresholds, as a function of the optimization weights, w_1, w_2, w_3 . In the lower left corner of the simplex, only error rate matters ($w_1 = 1$). In the upper corner, only decision time matters ($w_2 = 1$). In the lower right corner, only preference matters ($w_3 = 1$). In (a) consensus is given by unweighted in-degree, in (b) by weighted in-degree, in (c) by entropy, and in (d) by eigenvector centrality. Parameters: $N = 20, b = 1, r = 1, \ell = 0.1$.