Optimal Decision Strategies for Agents with Asymmetric Decision Thresholds

Kyle Fitzgerald

November 16, 2021

1 Single Agent

Our agent obeys the following Langevin equation:

$$dX = \mu\, dt + \sqrt{2D}\, dW, \qquad X(0) = 0,$$

where X is the current belief state, μ is the "natural" drift towards a particular decision, D is the diffusion coefficient, and W is a Wiener process. The decision thresholds are set at $x = -L < 0$ and $x = H > 0$.

1.1 Unconditional Escape Probability

(Refer to section 5.2.7(a) of Gardiner for some of the derivation.)

Suppose the agent begins at $X(0) = x$. Then the probability that the agent has not decided/escaped at time t is the *survival probability*, given by

$$S(x,t) = \int_{-L}^{H} c(y,t|x,0)\, dy,$$

where $c(y,t|x,0)$ is the probability concentration.

Now, $c(y,t|x,0)$ obeys the forward Fokker-Planck equation:

$$\begin{gathered} c_t = -\mu c_y + D c_{yy}, \\ c(-L,t|x,0) = c(H,t|x,0) = 0. \end{gathered} \tag{1}$$

A change of variables $y \mapsto y - \mu t$ transforms the PDE into a drift-free diffusion equation, from which it is apparent that the fundamental solution is

$$\Phi(y,t|x,0) = \frac{1}{\sqrt{4\pi D t}} \exp\big(-(y - x - \mu t)^2/4Dt\big).$$

Now we wish to impose the boundary conditions, which we can do via an infinite series generated by method of images.

By inspection, we arrive at

$$c(y,t|x,0) = \Phi(y,t|x,0) + \sum_{n=1}^{\infty}(-1)^n \Big[e^{-\mu\xi_n^{(L,H)}D^{-1}} \Phi\left(y + 2\xi_n^{(L,H)},\, t|x,0\right) \dots$$
$$+ e^{\mu\xi_n^{(H,L)}D^{-1}} \Phi\left(y - 2\xi_n^{(H,L)},\, t, |x,0\right) \Big],$$

where

$$\xi_n^{(a,b)} = \left\lceil \frac{n}{2} \right\rceil a + \left\lfloor \frac{n}{2} \right\rfloor b.$$

Then the survival probability is given by

$$\begin{aligned}
S(x,t) &= \int_{-L}^{H} c(y,t|x,0)\, dy \\
&= \frac{1}{2}\left[\operatorname{erf}\left(\frac{-\mu t + H - x}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{\mu t - L - x}{\sqrt{4Dt}}\right)\right] \\
&\quad + \frac{1}{2}\sum_{n=1}^{\infty}(-1)^n \Bigg\{ e^{-\mu\xi_n^{(L,H)}D^{-1}} \left[\operatorname{erf}\left(\frac{-\mu t + H + 2\xi_n^{(L,H)} - x}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{-\mu t - L + 2\xi_n^{(L,H)} - x}{\sqrt{4Dt}}\right)\right] \\
&\qquad + e^{\mu\xi_n^{(H,L)}D^{-1}} \left[\operatorname{erf}\left(\frac{-\mu t + H - 2\xi_n^{(H,L)} - x}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{-\mu t - L - 2\xi_n^{(H,L)} - x}{\sqrt{4Dt}}\right)\right] \Bigg\}
\end{aligned}$$

Of course, the escape probability is $E(x,t) = 1 - S(x,t)$, and we attain the solution to the original initial condition problem with $X(0) = 0$ by setting $x = 0$.

Alternatively, we can compute a Fourier series solution to the Focker-Planck equation in (1) with the initial condition $c(x,0) = \delta(x)$.

Separating variables as $c(x,t) = X(x)T(t)$, we have

$$\begin{aligned}
XT' &= -\mu X'T + DX''T \\
\implies \frac{T'}{T} &= \frac{DX'' - \mu X'}{X} = -\lambda.
\end{aligned}$$

If we let $X(x) = A\exp(\mu x/2D)\sin(\sqrt{4D\lambda - \mu^2}(x+L)/2D)$, then enforcing $X(H) = 0$ requires

$$\sqrt{4D\lambda - \mu^2}(L+H) = 2D\pi n \qquad \implies \qquad \lambda = \frac{D\pi^2 n^2}{(L+H)^2} + \frac{\mu^2}{4D},$$

and hence

$$c(x,t) = \sum_{n=1}^{\infty} A_n \exp(-\lambda_n t)\exp(\mu x/2D)\sin(\pi n(x+L)/(L+H)),$$

and enforcing the initial condition gives

$$\int_{-L}^{H} \exp(-\mu x/2D)\sin(\pi n(x+L)/(L+H))\delta(x)\, dx = \frac{A_n(L+H)}{2} \qquad \implies \qquad A_n = \frac{2\sin(\pi n L/(L+H))}{L+H}.$$

1.2 Mean Unconditional FPT

(Refer to section 5.2.7(a) of Gardiner for most of the derivation.)

The mean unconditional FPT is given by

$$T(x) = -\int_0^\infty t\frac{\partial}{\partial t}S(x,t)\,dt = \int_0^\infty S(x,t)\,dt$$

upon integrating by parts.

In order to avoid integrating the infinite series for $S(x,t)$, we integrate (1) from 0 to ∞ to obtain

$$\mu T_x + DT_{xx} = -1, \qquad T(-L) = T(H) = 0. \tag{2}$$

Defining $\phi(x) = \exp\big(\mu x/D\big)$, one can integrate (1.2) directly to obtain

$$\begin{aligned}
T(x) &= \frac{1}{D}\frac{\left(\int_{-L}^{x}\frac{dy}{\phi(y)}\right)\int_x^H \frac{dy'}{\phi(y')}\int_{-L}^{y'}\phi(z)dz - \left(\int_x^H \frac{dy}{\phi(y)}\right)\int_{-L}^{x}\frac{dy'}{\phi(y')}\int_{-L}^{y'}\phi(z)dz}{\int_{-L}^{H}\frac{dy}{\phi(y)}} \\
&= \frac{1}{\mu}\frac{(H-x)\phi\big(H+L\big) - (H+L)\phi\big(H-x\big) + x + L}{\phi\big(H+L\big) - 1}.
\end{aligned}$$

We then use the initial condition $X(0) = 0$ to obtain the expected FPT

$$\boxed{\langle T\rangle = \frac{1}{\mu}\frac{H\exp\big(\mu(H+L)/D\big) - (H+L)\exp\big(\mu H/D\big) + L}{\exp\big(\mu(H+L)/D\big) - 1}.}$$

1.3 Conditional Escape Probability

(Something is off with Gardiner or, more likely, my interpretation of it.)

Drawing from equations (2.3.7) and (2.3.8) in Redner, the probability of eventual exit through $-L$ (conditioned on exit through $-L$), given the initial condition $X(0) = x$, is

$$\epsilon^{(-L)}(x) = \frac{\exp(-\mu H/D) - \exp(-\mu x/D)}{\exp(-\mu H/D) - \exp(\mu L/D)},$$

and the conditional escape probability at H is

$$\epsilon^{(H)}(x) = 1 - \epsilon^{(-L)}(x) = \frac{\exp(-\mu x/D) - \exp(\mu L/D)}{\exp(-\mu H/D) - \exp(\mu L/D)}.$$

Using our initial condition $X(0) = 0$ gives

$$\boxed{\epsilon^{(-L)} = \frac{\exp(-\mu H/D) - 1}{\exp(-\mu H/D) - \exp(\mu L/D)}}$$

and

$$\boxed{\epsilon^{(H)} = \frac{1 - \exp(\mu L/D)}{\exp(-\mu H/D) - \exp(\mu L/D)}.}$$

1.4 Conditional FPT

(Refer to section 5.2.8 of Gardiner for most of the derivation.)

The mean exit time (given exit through H) is

$$T_H(x) = \int_0^\infty \frac{S_H(x,t)}{S_H(x,0)}\, dt.$$

Integrating the backwards Fokker-Planck equation for $S_H(x,t)$ gives

$$\begin{gathered} \mu[\epsilon^{(H)}(x)T_H(x)]_x + D[\epsilon^{(H)}(x)T_H(x)]_{xx} = -\epsilon^{(H)}(x), \\ \epsilon^{(H)}(-L)T_H(-L) = \epsilon^{(H)}(H)T_H(H) = 0. \end{gathered} \tag{3}$$

The method of undetermined coefficients gives us

$$\begin{aligned} \epsilon^{(H)}(x)T_H(x) = & \frac{D\phi(2H+2L) + (\mu x - H\mu)\phi(H+L) - L\mu - D - D\phi(2H+L-x)}{\mu^2(\phi(H+L)-1)^2} \\ & + \frac{-((H+L)\mu - D)\phi(H-x) - D\phi(H+2L+x) - D\phi(x+L) - \mu x}{\mu^2(\phi(H+L)-1)^2}, \end{aligned}$$

and dividing by $\epsilon^{(H)}(x)$ and using the initial condition $X(0) = 0$ gives

$$\begin{aligned} T_H = & \frac{D\phi(2H+2L) - H\mu\phi(H+L) + L\mu - D - D\phi(2H+L)}{\mu^2(\phi(H+L)-1)(\phi(L)-1)} \\ & + \frac{-((H+L)\mu - D)\phi(H) - D\phi(H+2L) - D\phi(L)}{\mu^2(\phi(H+L)-1)(\phi(L)-1)}. \end{aligned}$$

Similarly,

$$\begin{aligned} T_{-L} = & \frac{(H\mu - D)\phi(2H+2L) + L\mu\phi(H+L) + D - D\phi(H)}{\mu^2(\phi(H+L)-1)(\phi(H+L)-\phi(L))} \\ & + \frac{(D-(H+L))\phi(2H+L) + D\phi(H+2L) - D\phi(L)}{\mu^2(\phi(H+L)-1)(\phi(H+L)-\phi(L))}. \end{aligned}$$

(Wrong based on Desmos computations – need to use your corrected conditional escape probabilities!)

*****

Alternatively, we can recognize that the FPT density conditional on exit through $(\cdot)$, $f^{(\cdot)}(t)$, is given by the flux of probability density, $c(x,t|0,0)$, at the desired boundary:

$$f^{(H)}(t) = -D\frac{\partial c}{\partial x}\bigg|_{x=H}, \qquad f^{(-L)}(t) = D\frac{\partial c}{\partial x}\bigg|_{x=-L}.$$

1.5 Reward Rate

Now consider an agent who is rewarded for correct decisions and penalized for incorrect decisions in an uncertain environment. The agent must complete a sequence of TAFC tasks separated by time T_I; in each of these tasks, the agent wants to determine the "correct" choice by continuously sampling the (noisy) environmental drift, μ, which represents a rate of information flow. For simplicity, in each environment we let $D = 1$ and choose $\mu = 1$ or $\mu = -1$ with equal probability, i.e. the "+" choice (corresponding to decision threshold $x = H$) is correct as often as is the "-" choice (corresponding to $x = -L$) in the agent's environment.

We can model the reward rate, RR, by

$$RR = \frac{\langle r \rangle - \langle p \rangle}{\langle T \rangle + T_I}.$$

Here, $\langle r \rangle$ is the expected reward for a correct decision,

$$\begin{aligned}
\langle r \rangle &= R^+ \cdot P(\text{decides at } H \,|\, \mu = 1) P(\mu = 1) + R^- \cdot P(\text{decides at } -L \,|\, \mu = -1) P(\mu = -1) \\
&= \frac{R^+}{2} \cdot P(\text{decides at } H \,|\, \mu = 1) + \frac{R^-}{2} \cdot P(\text{decides at } -L \,|\, \mu = -1) \\
&= \frac{1}{2} \left(R^+ \frac{1 - e^L}{e^{-H} - e^L} + R^- \frac{e^H - 1}{e^H - e^{-L}} \right) \\
&= \frac{1}{2} \frac{R^+(e^L - 1)e^H + R^-(e^H - 1)e^L}{e^{L+H} - 1},
\end{aligned}$$

where $R^\pm$ is the reward given for crossing the $\pm$ threshold correctly. Refer to section 1.3 for the third line.

Similarly, $\langle p \rangle$ is the expected penalty for an incorrect decision, although for simplicity we will begin by having no penalty for an incorrect decision.

Finally, $\langle T \rangle$ is the unconditional mean time to make a decision,

$$\begin{aligned}
\langle T \rangle &= \langle T \,|\, \mu = 1 \rangle P(\mu = 1) + \langle T \,|\, \mu = -1 \rangle P(\mu = -1) \\
&= \frac{1}{2} \langle T \,|\, \mu = 1 \rangle + \frac{1}{2} \langle T \,|\, \mu = -1 \rangle \\
&= \frac{1}{2} \left(\frac{He^{H+L} - (H + L)e^H + L}{e^{H+L} - 1} - \frac{He^{-L-H} - (H + L)e^{-H} + L}{e^{-L-H} - 1} \right) \\
&= \frac{1}{2} \frac{(1 - e^L)(1 - e^H)(L + H)}{e^{L+H} - 1},
\end{aligned}$$

where the third line comes from section 1.2.

Putting this together, we have

$$RR = \frac{R^+(1 - e^{-L}) + R^-(1 - e^{-H})}{(1 - e^{-L})(1 - e^{-H})(L + H) + 2(1 - e^{-(L+H)})T_I}.$$

Notice that for symmetric thresholds, i.e. $L = H$, we have

$$RR_{\text{sym}} = \frac{(R^+ + R^-)/2}{H(1 - e^{-H}) + T_I(1 + e^{-H})}.$$

2 Two Agents with Asymmetric Thresholds

(Use capital vs lowercase letters to distinguish between lower and upper asymmetric thresholds.) Now our agents $1, 2$ obey the respective coupled Langevin equations,

$$
\begin{aligned}
dX_1 &= \mu_1\, dt + \sqrt{2D}\, dW_1 + G_2 q_\pm \delta(t - t_2), \qquad X_1(0) = 0,\\
dX_2 &= \mu_2\, dt + \sqrt{2D}\, dW_2 + G_1 q_\pm \delta(t - t_1), \qquad X_2(0) = 0,
\end{aligned}
$$

where $G_i = \pm 1$ if agent i decides at the $\pm$ threshold at time t_i; otherwise $G_i = 0$. $q_\pm$ is a scalar magnitude of the "kick" that an agent gets from the other making a decision at the $\pm$ threshold. Each agent's decision threshold is set at $x_i = -L_i < 0$ and $x_i = H_i > 0$.

2.1 P_{H_1,H_2}

(Derivation can be found in Caginalp.)

We are interested in the probability that both agents end up making the same decision (WLOG the $x = H_i$ threshold, whose probability we denote P_{H_1,H_2}).

First, define $c_i(x_i, t|0, 0) \equiv c_i(x, t)$ and

$$
f_i^{(\cdot)}(t) = -\text{sgn}(\cdot) D \frac{\partial c_i}{\partial x}\Big|_{x=(\cdot)}
$$

to be the probability concentration of agent i at (x, t)) and the FPT density of agent i at $(\cdot, t)$, respectively. (Before $t = t_i$, the expression for c_i is just that of a single independent agent.) Also define $f_i(t) = f_i^{(H_i)}(t) + f_i^{(-L_i)}(t)$ to be the total FPT density of agent i.

Then if we condition on agent i deciding before agent j $(t_i < t_j)$, we have

$$
\begin{aligned}
f_i^{(\cdot)}(t_i \,|\, t_i < t_j) &= \frac{f_i^{(\cdot)}(t_i) \int_{t_i}^{\infty} f_j(t_j)\, dt_j}{P(t_i < t_j)}\\
&= \frac{f_i^{(\cdot)}(t_i) \int_{t_i}^{\infty} f_j(t_j)\, dt_j}{\text{AUC}\left[\int_0^t f_j(t)\, dt \text{ vs } \int_0^t f_i(t)\, dt\right]}.
\end{aligned}
$$

The conditional probability that j eventually crosses H_j after i crosses H_i is broken into instantaneous and eventual-diffusion probabilities:

$$
\begin{aligned}
P(G_j = 1 \,|\, i \text{ crosses } H_i \text{ at } t_i) =& P(j \text{ inst. crosses } H_j \,|\, i \text{ crosses } H_i \text{ at } t_i)\\
&+ P(j \text{ diff. across } H_j \,|\, i \text{ crosses } H_i \text{ at } t_i)
\end{aligned}
$$

We have

$$
P(j \text{ inst. crosses } H_j \,|\, i \text{ crosses } H_i \text{ at } t_i) = \int_{H_j - q_+}^{H_j} \rho_j(x, t_i^-)\, dx,
$$

where

$$
\rho_j(x, t) = \frac{c_j(x, t)}{\int_{-L_j}^{H_j} c_j(x, t)\, dx}.
$$

In general, if j now has evidence x then the probability of diffusing across H_j is given by

$$\epsilon_j^{(H_j)}(x) = \frac{e^{-\mu_j L_j/(2D)} - e^{-\mu_j x/(2D)}}{e^{-\mu_j L_j/(2D)} - e^{-\mu_j H_j/(2D)}},$$

(refer to section 1.3) and so if we note that

$$c_j(x, t_i^+) = c_j(x+q, t_i^-),$$

then we have

$$P(j \text{ diff. across } H_j \,|\, i \text{ crosses } H_i \text{ at } t_i) = \left(\int_{-L_j}^{H_j-q_+} \rho_j(x, t_i^-)\, dx\right)\left(\int_{-L_j}^{H_j} \rho_j(x, t_i^+)\epsilon_j^{(H_j)}(x)\, dx\right).$$

Now, the probability of both agents deciding at $H_{i,j}$ conditioned on $t_i < t_j$ is

$$P(G_i = G_j = 1 \,|\, t_i < t_j) =$$
$$\int_0^\infty f_i^{(H_i)}(t_i|t_i < t_j)\left[\int_{H_j-q_+}^{H_j} \rho_j(x, t_i^-)\, dx + \left(\int_{-L_j}^{H_j-q_+} \rho_j(x, t_i^-)\, dx\right)\left(\int_{-L_j}^{H_j} \rho_j(x, t_i^+)\epsilon_j^{(H_j)}(x)\, dx\right)\right] dt_i,$$

and hence P_{H_1,H_2} is given by considering both $t_1 < t_2$ and $t_2 < t_1$:

$$P_{H_1,H_2} =$$
$$\sum_{\substack{i=1\\ i\neq j}}^{2} \int_0^\infty f_i^{(H_i)}(t_i|t_i < t_j)\left[\int_{H_j-q_+}^{H_j} \rho_j(x, t_i^-)\, dx + \left(\int_{-L_j}^{H_j-q_+} \rho_j(x, t_i^-)\, dx\right)\left(\int_{-L_j}^{H_j} \rho_j(x, t_i^+)\epsilon_j^{(H_j)}(x)\, dx\right)\right] dt_i.$$

2.2 Reward Rate: Agent-Symmetric, Threshold-Asymmetric Rewards with Agent-Symmetric Thresholds

Now consider the two-agent analog of the reward procedure outlined in section 1.5. Each agent completes a sequence of TAFC tasks separated by time T_I.

Assumptions: For simplicity, we let $D = 1$ and assume that agents receive the same stimulus in a given environment, i.e. $\mu_1 = \mu_2 \equiv \mu$. In each environment, we choose $\mu = 1$ or $\mu = -1$ with equal probability, i.e. the "+" choice (corresponding to decision threshold $x = H_i$) is correct as often as is the "-" choice (corresponding to $x = -L_i$). We also assume that agents are given the same reward at each "correct" decision boundary, though they may be asymmetrically rewarded, i.e. $R_1^{\pm} = R_2^{\pm} \equiv R^{\pm}$. Clearly, equivalent drifts and rewards for each agent should lead to equivalent, but possibly asymmetric, optimal decision thresholds, i.e. $H_1 = H_2 \equiv H$ and $-L_1 = -L_2 \equiv -L$. Finally, we do not have an incorrect-choice penalty for now.

Notice that maximizing reward rate will now be an optimization problem in four parameters: L, H, and $q_{\pm}$.

With the given assumptions, the agents will each have the same expected reward rate $\langle r \rangle$. We can thus model the reward rate, RR, by

$$RR = \frac{2\langle r \rangle}{\langle T_L \rangle + T_I}.$$

Here, $\langle r\rangle$ is given by

$$
\begin{aligned}
\langle r\rangle &= R^+\cdot P(i \text{ decides at } H \,|\, \mu=1)P(\mu=1) + R^-\cdot P(i \text{ decides at } -L \,|\, \mu=-1)P(\mu=-1)\\
&= \frac{R^+}{2}\cdot P(i \text{ decides at } H \,|\, \mu=1) + \frac{R^-}{2}\cdot P(i \text{ decides at } -L \,|\, \mu=-1),
\end{aligned}
$$

where, for example,

$$
\begin{aligned}
P(i \text{ decides at } H \,|\, \mu=1) =& P(i \text{ decides at } H \,|\, \mu=1 \text{ and } t_i<t_j)P(t_i<t_j \,|\, \mu=1)\\
&+P(i \text{ inst. crosses } H \,|\, \mu=1 \text{ and } t_i>t_j)\big(1-P(t_i<t_j \,|\, \mu=1)\big)\\
&+P(i \text{ diff. across } H \,|\, \mu=1 \text{ and } t_i>t_j)\big(1-P(t_i<t_j \,|\, \mu=1)\big)
\end{aligned}
$$

and

$$
\begin{gathered}
P(t_i<t_j \,|\, \mu) = \text{AUC}\left[\int_0^t f_j(t;\mu)\,dt \text{ vs } \int_0^t f_i(t;\mu)\,dt\right] = \frac{1}{2},\\
P(i \text{ decides at } H \,|\, \mu=1 \text{ and } t_i<t_j) = \int_0^\infty f_i^{(H)}(t_i;\mu \,|\, t_i<t_j)\,dt_i\,,\\
P(i \text{ inst. crosses } H \,|\, \mu=1 \text{ and } t_i>t_j) = P(j \text{ inst. crosses } H \,|\, \mu=1 \text{ and } t_i<t_j) =\\
\int_0^\infty f_i^{(H)}(t_i;\mu \,|\, t_i<t_j)\int_{H-q_+}^H \rho_j(x,t_i^-)\,dx\,dt_i\,,\\
P(i \text{ diff. across } H \,|\, \mu=1 \text{ and } t_i>t_j) = P(j \text{ diff. across } H \,|\, \mu=1 \text{ and } t_i<t_j) =\\
\int_0^\infty f_i^{(H)}(t_i;\mu \,|\, t_i<t_j)\left(\int_{-L}^{H-q_+}\rho_j(x,t_i^-)\,dx\right)\left(\int_{-L}^H \rho_j(x,t_i^+)\epsilon_j^{(H)}(x)\,dx\right)dt_i\\
+\int_0^\infty f_i^{(-L)}(t_i;\mu \,|\, t_i<t_j)\left(\int_{-L+q_-}^{H}\rho_j(x,t_i^-)\,dx\right)\left(\int_{-L}^H \rho_j(x,t_i^+)\epsilon_j^{(H)}(x)\,dx\right)dt_i,
\end{gathered}
$$

where we have used the fact that agents i and j are interchangeable (given our assumptions) in the first line's simplification as well as to exchange indices in the last two equations.

$\langle T_L\rangle$ is the unconditional mean time for *the last agent to make a decision*. With the given assumptions, this is equivalent to the unconditional mean time for agent j conditioned on agent i deciding before agent j, given by

$$
\begin{aligned}
\langle T_L\rangle &= \langle T_L \,|\, \mu=1\rangle P(\mu=1) + \langle T_L \,|\, \mu=-1\rangle P(\mu=-1)\\
&= \frac{1}{2}\langle T_j \,|\, \mu=1 \text{ and } t_i<t_j\rangle + \frac{1}{2}\langle T_j \,|\, \mu=-1 \text{ and } t_i<t_j\rangle,
\end{aligned}
$$

where

$$
\langle T_j \,|\, \mu=\pm1 \text{ and } t_i<t_j\rangle = \int_0^\infty t_j f_j(t_j;\mu|t_i<t_j)\,dt_j
$$

and

$$
f_j(t_j;\mu \,|\, t_i<t_j) = f_j(t_j;\mu)\int_0^{t_j} f_i(t_i;\mu)\,dt_i.
$$

Notice that if $L \to 0$, i.e. agents decide immediately at the lower threshold, then we simply have

$$RR \to \frac{R^-}{\langle T_L \rangle + T_I} = \frac{R^-}{T_I}.$$

2.3 Reward Rate: Agent-Symmetric, Threshold-Asymmetric Rewards with Agent-Symmetric Thresholds and Immediately-Reactive Decisions

Consider an analytical simplification to the model in the above section: once agent i decides, agent j receives a corresponding kick of strength $q_\pm$ *and then immediately decides at the nearest threshold*, which represents an interrogation-like situation imposed by the first agent's decision.

We keep the same reward rate function. What changes is the probability that i decides at H or $-L$ given $\mu = \pm 1$, and the expression for the expected time for the last agent to decide.

For example, we have

$$\begin{aligned} P(i \text{ decides at } H \,|\, \mu = 1) =& P(i \text{ decides at } H \,|\, \mu = 1 \text{ and } t_i < t_j) P(t_i < t_j \,|\, \mu = 1) \\ &+ P(i \text{ inst. crosses } H \,|\, \mu = 1 \text{ and } t_i > t_j)\big(1 - P(t_i < t_j \,|\, \mu = 1)\big) \end{aligned}$$

and

$$P(t_i < t_j \,|\, \mu) = \text{AUC}\left[\int_0^t f_j(t;\mu)\,dt \text{ vs } \int_0^t f_i(t;\mu)\,dt\right] = \frac{1}{2},$$

$$P(i \text{ decides at } H \,|\, \mu = 1 \text{ and } t_i < t_j) = \int_0^\infty f_i^{(H)}(t_i;\mu \,|\, t_i < t_j)\,dt_i\,,$$

$$P(i \text{ inst. crosses } H \,|\, \mu = 1 \text{ and } t_i > t_j) = P(j \text{ inst. crosses } H \,|\, \mu = 1 \text{ and } t_i < t_j) =$$
$$\int_0^\infty f_i^{(H)}(t_i;\mu \,|\, t_i < t_j) \int_{\max\{\frac{H+L}{2} - q_+, -L\}}^H \rho_j(x, t_i^-)\,dx\,dt_i\,.$$

$\langle T_L \rangle$ is now $\langle T_F \rangle$, the expected time for the *first* agent to make a decision. With the given assumptions, this is equivalent to the unconditional mean time for agent i conditioned on agent i deciding before agent j, given by

$$\begin{aligned} \langle T_F \rangle &= \langle T_F \,|\, \mu = 1 \rangle P(\mu = 1) + \langle T_F \,|\, \mu = -1 \rangle P(\mu = -1) \\ &= \frac{1}{2}\langle T_i \,|\, \mu = 1 \text{ and } t_i < t_j \rangle + \frac{1}{2}\langle T_i \,|\, \mu = -1 \text{ and } t_i < t_j \rangle, \end{aligned}$$

where

$$\langle T_i \,|\, \mu = \pm 1 \text{ and } t_i < t_j \rangle = \int_0^\infty t_i f_i(t_i;\mu | t_i < t_j)\,dt_i.$$

2.4 Reward Rate: Agent-Asymmetric, Threshold-Asymmetric Rewards with Agent-Asymmetric Thresholds