O For Online Publication

O.1 Uniqueness of Optimal Information Acquisition

By Lemma 5, whenever a uniformly optimal strategy exists, it is the optimal attention strategy regardless of the form of $u(\tau, a, \omega)$. Without further assumptions on u, there could exist other optimal attention strategies. For example, consider the payoff function used in the proof of Lemma 6. Under this payoff function, the agent always stops at some fixed time t. Hence any strategy that achieves the t-optimal vector $n(t)$ would be optimal in this problem.

Nonetheless, such examples can be ruled out by an assumption on the agent's stopping rule:

Assumption 7. *Given any attention allocation strategy S, any history of signal realizations up to time t such that the agent has not stopped, and any $t' > t$, there exists a positive probability of continuation histories such that the agent optimally stops in the interval $(t, t']$.*

Proposition 5. *Suppose Assumption 2 holds strictly, and Assumption 7 is satisfied. Then, any optimal attention strategy S induces the same posterior variance as the uniformly optimal strategy S^* at every history where the agent has not stopped. Consequently, the two strategies induce the same cumulative attention vectors, and coincide at almost every time before stopping.*

Proof. Suppose S induces larger posterior variance than S^* at some time t, then by continuity the same holds from time t to some later time $t' > t$ along this history. By assumption, the agent stops between these times with positive probability. Thus there is positive probability that the agent stops with posterior variance strictly larger than the minimal variance. From the proof of Lemma 5, we see that payoff under S is strictly below S^*, contradicting the optimality of S. The second part of the result follows from the uniqueness of $n(t)$, and the fact that $\beta(t)$ integrates to $n(t)$. □

Although Assumption 7 is stated in terms of the endogenous stopping rule, it is satisfied in any problem where the agent always stops to take some action when he has an extremely high (or low) expectation about ω. This is guaranteed if extreme values of ω agree on the optimal action and the marginal cost of delay is bounded away from zero. These conditions on the primitives are rather weak, and are satisfied in many applications such as binary choice with linear waiting cost.

O.2 Non-existence of Uniformly Optimal Strategy

O.2.1 Counterexample for $K = 2$

The example below illustrates how and why Theorem 1 might fail without Assumption 3:

Example 5. There are two unknown attributes with prior distribution

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 10 & -3 \\ -3 & 1 \end{pmatrix} \right).$$

The agent wants to learn $\theta_1 + 4\theta_2$.

Given q_1 units of attention devoted to learning θ_1, and q_2 devoted to θ_2, the agent's posterior variance about ω is given by (2). Simplifying, we have $V(q_1, q_2) = \frac{2+16q_1+q_2}{(1+q_1)(10+q_2)-9}$. The t-optimal cumulative attention vectors $n(t)$ (see Section 4) are defined to minimize $V(q_1, q_2)$ subject to $q_1, q_2 \geq 0$ and the budget constraint $q_1 + q_2 \leq t$.

These vectors do not evolve monotonically: Initially, the marginal value of learning θ_1 exceeds that of learning θ_2, since the agent has greater prior uncertainty about θ_1 (even accounting for the difference in payoff weights). Thus at all times $t < 1/4$, the t-optimal vector is $(t, 0)$, and the agent learns only about attribute 1.

After a quarter-unit of time devoted to learning θ_1, the agent's posterior covariance matrix becomes $\begin{pmatrix} 20/7 & -6/7 \\ -6/7 & 5/14 \end{pmatrix}$. Note that the two sources have equal marginal values at $t = 1/4$, since $\gamma_1 = \frac{-4}{7}$ and $\gamma_2 = \frac{4}{7}$ have the same absolute value. However, to maintain equal marginal values at future instants, it would be optimal to take attention away from attribute 1 and redistribute it to attribute 2. Specifically, at all times $t \in [1/4, 1]$ the t-optimal vector is given by $n(t) = \left(\frac{-t+1}{3}, \frac{4t-1}{3}\right)$, and the optimal cumulative attention toward θ_1 is decreasing in this interval.

Consequently, there does not exist a uniformly optimal strategy in this example (Lemma 1). Hence the optimal information acquisition strategy varies according to when the agent expects to stop, and Theorem 1 cannot hold independently of the payoff criterion.

O.2.2 Necessity of Assumption 3 for Theorem 1

We show here that when $K = 2$, the assumption $cov_1 + cov_2 \geq 0$ is also *necessary* for the existence of a uniformly optimal strategy. The result generalizes Example 5 above.

Proposition 6. *Suppose $K = 2$ and Assumption 3 is violated. Then a uniformly optimal strategy does not exist.*

Proof. Suppose that $cov_1 + cov_2 < 0$. First note that one of cov_1, cov_2 is positive, because $\alpha_1 cov_1 + \alpha_2 cov_2 = \alpha' \Sigma \alpha > 0$. So without loss we can assume $cov_2 > 0 > -cov_2 > cov_1$. Moreover, from $\alpha_1 cov_1 + \alpha_2 cov_2 > 0$ we obtain $\alpha_2 > \alpha_1$ and hence $x_2 > x_1$. Below we show the t-optimal attention vector $n(t)$ is non-monotonic.

Suppose $\frac{-(cov_1+cov_2)}{x_2} < t < \frac{-(cov_1+cov_2)}{x_1}$, then by (10) we have $\partial_1 V(0,t) < \partial_2 V(0,t)$ and $\partial_1 V(t,0) > \partial_2 V(t,0)$. These imply that $n(t)$ is interior, and the first-order condition yields

$$x_1 n_2(t) + cov_1 = -(x_2 n_1(t) + cov_2),$$

where we use the fact that for t in this range $x_1 q_2 + cov_1$ is always negative. Together with $n_1(t) + n_2(t) = t$, we can solve that $n(t) = (\frac{-x_1 t - cov_1 - cov_2}{x_2 - x_1}, \frac{x_2 t + cov_1 + cov_2}{x_2 - x_1})$. Thus as t increases in this range, $n_1(t)$ actually decreases. So a uniformly optimal strategy does not exist. □

O.2.3 Counterexample for $K = 3$

We present another example where a uniformly optimal strategy does not exist. In the following example, the three attributes have positive correlation with each other, but not positive *partial correlation*. Thus the example illustrates the subtlety of Assumption 4.

Let the primitives be $K = 3$, $\alpha_1 = \alpha_2 = 1$, $\alpha_3 = 20$ and

$$\Sigma = \begin{pmatrix} 19 & 3 & 0 \\ 3 & 5 & 3 \\ 0 & 3 & 2 \end{pmatrix} \qquad \Sigma^{-1} = \begin{pmatrix} 1 & -6 & 9 \\ -6 & 38 & -57 \\ 9 & -57 & 86 \end{pmatrix}$$

The choice of $\Sigma_{11} = 19$ is such that Σ has determinant exactly 1. This makes it easier to calculate its inverse matrix, while still ensuring that Σ is positive-definite. We highlight the fact that $[\Sigma^{-1}]_{13} = 9$ is positive, so this example does not satisfy Assumption 4.

Consider the cumulative attention vector $q = (1, 14, 0)'$. Simple calculation gives $(\Sigma^{-1} + Q) \cdot (-1, 1, 1)' = (1, 1, 20)' = \alpha$. Thus $\gamma(q) = (\Sigma^{-1} + Q)^{-1}\alpha = (-1, 1, 1)'$. Since the three coordinates of γ have equal absolute value, the sources have equal marginal reduction of V at the attention vector q. This means q is the t-optimal vector for $t = 15$.

Next consider a different attention vector $\hat{q} = (0, 15, 20)$. We can similarly calculate that $(\Sigma^{-1} + \hat{Q}) \cdot (-1, 1, 1)' = (2, 2, 40)' = 2\alpha$. So $\gamma(\hat{q}) = (-1/2, 1/2, 1/2)$. By the same reasoning, $\hat{q}$ is the t-optimal vector for $t = 35$. Hence we see that when t increases from 15 to 35, the optimal amount of attention devoted to θ_1 decreases from 1 to 0. This implies that a uniformly optimal strategy does not exist in this example.

O.3 When are Sources Substitutes/Complements?

Since more information *reduces* the posterior variance V, we define two sources i and j to be substitutes if the cross-partial $\partial_{ij} V(q)$ is *non-negative* at any cumulative attention vector $q \geq 0$. The following result shows that Assumption 4 precisely characterizes when two sources are substitutes.

Proposition 7. *Given any positive payoff weight vector α. The following conditions on the prior covariance matrix Σ are equivalent:*

1. *Σ^{-1} has non-positive off-diagonal entries;*

2. *Every pair of sources $i \neq j \in \{1, \ldots, K\}$ are substitutes in the sense that $\partial_{ij} V(q) \geq 0$ at every cumulative attention vector $q \in \mathbb{R}^K_+$.*

Proof. In one direction, suppose Σ^{-1} has non-positive off-diagonal entries. Then for any vector $q \geq 0$ and corresponding diagonal matrix $Q = \operatorname{diag}(q)$, $\Sigma^{-1} + Q$ satisfies the same property. Thus the positive-definite matrix $\Sigma^{-1} + Q$ is an *M-matrix* whose inverse is known to have non-negative entries off the diagonal and strictly positive entries on the diagonal; see e.g. Plemmons (1977). It follows that $\gamma = (\Sigma^{-1} + Q)^{-1} \cdot \alpha$ is a vector with positive entries. Thus by Lemma 3,

$$\partial_{ij} V(q) = 2\gamma_i \gamma_j \cdot [(\Sigma^{-1} + Q)^{-1}]_{ij} \geq 0.$$

Conversely, suppose $\partial_{ij} V(q) \geq 0$ for every $q \geq 0$. Let us choose any vector q whose coordinates are all sufficiently large. Then

$$(\Sigma^{-1} + Q)^{-1} = [(\Sigma^{-1} \cdot Q^{-1} + I_K) \cdot Q]^{-1} = Q^{-1}(\Sigma^{-1} \cdot Q^{-1} + I_K)^{-1} = Q^{-1}(I_K + o(1)),$$

where $o(1)$ is a matrix that goes to zero as $q_1, \ldots, q_K$ all go to infinity (at any rate). Note that when the $o(1)$ term is sufficiently small, $(I_K + o(1)) \cdot \alpha$ is a vector with positive coordinates. Thus, $\gamma = (\Sigma^{-1} + Q)^{-1}\alpha = Q^{-1}(I_K + o(1))\alpha$ has positive coordinates. Together with $\partial_{12} V(q) \geq 0$ and Lemma 3, this implies $[(\Sigma^{-1} + Q)^{-1}]_{12} \geq 0$ for any such q. Using the matrix inverse formula, we further obtain that the determinant of the cofactor matrix $[\Sigma^{-1} + Q]_{-21}$ (i.e., the sub-matrix of $\Sigma^{-1} + Q$ with the second row and first column removed) must be non-positive. Expanding this determinant using permutations, it is easy to see that it contains the term $[\Sigma^{-1}]_{12} \cdot \prod_{k=3}^{K} q_k$, which is in fact the dominant term when each q_k is sufficiently large. Hence we deduce $[\Sigma^{-1}]_{12} \leq 0$, and similarly $[\Sigma^{-1}]_{ij} \leq 0$ for every pair $i \neq j$. □

We next show that Assumption 5 characterizes when two sources are complements.

Proposition 8. *Given any positive payoff weight vector α. The following conditions on the prior covariance matrix Σ are equivalent:*

1. *Σ has non-positive off-diagonal entries and $\Sigma \cdot \alpha$ has non-negative coordinates;*

2. *Every pair of sources $i \neq j \in \{1, \ldots, K\}$ are complements in the sense that $\partial_{ij} V(q) \leq 0$ at every cumulative attention vector $q \in \mathbb{R}^K_+$.*

Proof. To show the first condition implies the second, we note that since Σ is assumed to be an M-matrix, Σ^{-1} is an inverse M-matrix. By Theorem 3 in Johnson (1982), $\Sigma^{-1} + Q$ is also an inverse M-matrix. This implies that for any attention vector q, the posterior covariance matrix $(\Sigma^{-1}+Q)^{-1}$ must be an M-matrix with non-positive off-diagonal entries. We claim that the vector $\gamma = (\Sigma^{-1} + Q)^{-1} \cdot \alpha$ has non-negative coordinates. Once this is shown, Lemma 3 will imply that $\partial_{ij} V = 2\gamma_i \gamma_j [(\Sigma^{-1} + Q)^{-1}]_{ij} \leq 0$ for every pair $i \neq j$.

By assumption, if q is the zero vector, then $\gamma = \Sigma \cdot \alpha$ indeed has non-negative coordinates. Our goal below is to show this also holds at any $q \geq 0$. To do this, we first work under the stronger assumption that $\Sigma \cdot \alpha$ has strictly positive coordinates, so that $\gamma(0)$ is a positive vector. Recall from the proof of Lemma 3 that when γ_i is viewed as a function of q, its partial derivatives are given by

$$\frac{\partial \gamma_i}{\partial q_j} = -\gamma_j \cdot [(\Sigma^{-1} + Q)^{-1}]_{ij} \quad \text{for each } j \in \{1, \ldots, K\}.$$

Now since Σ^{-1} is positive-definite, we can choose $\epsilon > 0$ such that $\Sigma^{-1} \geq \epsilon I_K$ in the matrix order. Fixing this ϵ, we consider the K functions

$$f_i(q_1, \ldots, q_K) = (q_i + \epsilon) \cdot \gamma_i(q_1, \ldots, q_K).$$

For every $j \neq i$, $\partial f_i / \partial q_j = (q_i + \epsilon) \cdot -\gamma_j \cdot [(\Sigma^{-1} + Q)^{-1}]_{ij}$. This product is non-negative whenever $\gamma_j \geq 0$, since $[(\Sigma^{-1} + Q)^{-1}]_{ij} \leq 0$ as shown above. On the other hand, by the product rule,

$$\frac{\partial f_i}{\partial q_i} = \gamma_i \cdot [1 - (q_i + \epsilon)[(\Sigma^{-1} + Q)^{-1}]_{ii}].$$

This is non-negative whenever $\gamma_i \geq 0$, since $[(\Sigma^{-1} + Q)^{-1}]_{ii} \leq [(\epsilon I_K + Q)^{-1}]_{ii} = (q_i + \epsilon)^{-1}$ where the inequality uses standard properties of the matrix order.

Hence, we have shown that whenever $f_1(q), \ldots, f_K(q)$ are all non-negative, their derivatives with respect to each q_i are also non-negative. Moreover, we know from our stronger assumption that $f_1, \ldots, f_K$ are strictly positive at $q = 0$. These together imply $f_1(q), \ldots, f_K(q)$ are always strictly positive, by the following argument. Suppose for contradiction that there exists some i and some $q \geq 0$ such that $f_i(q) \leq 0$. By continuity of f and a limit argument, we may assume q is minimal in the sense that at every $q' < q$, $f_j(q')$ is positive for every j. Thus, if we let $g(t) = f_i(t \cdot q)$ be the one-variable function defined for $t \in [0, 1]$, we have $g(0) > 0$, $g'(t) \geq 0$ for $t \in [0, 1)$ and $g(1) \leq 0$. This contradicts the Mean Value Theorem.

It follows that $\gamma(0)$ being positive implies $\gamma(q)$ is positive. We now extend this result to the case of weak inequalities. Suppose $\gamma(0) = \Sigma \cdot \alpha$ has non-negative coordinates. Then by considering $\Sigma + \delta I_K$ instead of Σ, we see that the corresponding γ vector is strictly positive at $q = 0$. The above analysis applied to the M-matrix $\Sigma + \delta I_K$ thus implies $[(\Sigma + \delta I_K)^{-1} + Q]^{-1} \cdot \alpha$ is positive for any attention vector $q \geq 0$. Letting $\delta \to 0$ yields $[\Sigma^{-1} + Q]^{-1} \cdot \alpha \geq 0$, as we desire to show. This completes the proof that the first condition in the proposition guarantees complementarity.

Turning to the converse, we assume at every $q \geq 0$, $\partial_{ij} V = 2\gamma_i(q) \cdot \gamma_j(q) \cdot [(\Sigma^{-1} + Q)^{-1}]_{ij} \leq 0$. Let us choose q such that $\gamma_i(q)$ is nonzero for each i; its existence will be verified later. For this q, we claim that $\gamma_i(q)$ must be positive for each i. Indeed, since $\gamma = (\Sigma^{-1} + Q)^{-1}\alpha$, we have $\alpha'\gamma = \alpha'(\Sigma^{-1} + Q)^{-1}\alpha > 0$. Thus γ must have at least one positive coordinate. Suppose for contradiction that γ has a negative coordinate, then we can without loss assume γ_i is positive for $i \leq k$ and negative for $i > k$, where k satisfies $1 \leq k < K$. Using $\partial_{ij} V \leq 0$, we deduce $[(\Sigma^{-1} + Q)^{-1}]_{ij} \geq 0$ for every pair $i \leq k$ and $j > k$. Now let us decompose $(\Sigma^{-1} + Q)^{-1}$ into four block sub-matrices: $[(\Sigma^{-1} + Q)^{-1}]_{TL}$, $[(\Sigma^{-1} + Q)^{-1}]_{TR}$, $[(\Sigma^{-1} + Q)^{-1}]_{BL}$ and $[(\Sigma^{-1} + Q)^{-1}]_{BR}$ are the top-left $k \times k$, top-right $k \times (K-k)$, bottom-left $(K-k) \times k$ and bottom-right $(K-k) \times (K-k)$ sub-matrices of $(\Sigma^{-1} + Q)^{-1}$ respectively. Recall that $\gamma = (\Sigma^{-1} + Q)^{-1} \cdot \alpha$. By looking at the last $K - k$ coordinates, we obtain

$$\begin{aligned}(\gamma_{k+1}, \ldots, \gamma_K)' &= \left([(\Sigma^{-1} + Q)^{-1}]_{BL}, [(\Sigma^{-1} + Q)^{-1}]_{BR}\right) \cdot \alpha \\ &= [(\Sigma^{-1} + Q)^{-1}]_{BL} \cdot (\alpha_1, \ldots, \alpha_k)' + [(\Sigma^{-1} + Q)^{-1}]_{BR} \cdot (\alpha_{k+1}, \ldots, \alpha_K)'.\end{aligned}$$

The preceding analysis tells us that $[(\Sigma^{-1} + Q)^{-1}]_{BL}$ has non-negative entries, so the vector $[(\Sigma^{-1} + Q)^{-1}]_{BL} \cdot (\alpha_1, \ldots, \alpha_k)'$ is non-negative. In addition, $[(\Sigma^{-1} + Q)^{-1}]_{BR}$ is positive-definite, so the vector $[(\Sigma^{-1} + Q)^{-1}]_{BR} \cdot (\alpha_{k+1}, \ldots, \alpha_K)'$ has at least one positive coordinate. Thus, the above displayed equation contradicts the assumption that $\gamma_{k+1}, \ldots, \gamma_K$ are all negative.

We thus know that if $\gamma(0)$ has nonzero coordinates, then it is in fact a positive vector. Complementarity further requires $\Sigma_{ij} \leq 0$ for all $i \neq j$, which would complete the proof. In the general case, $\gamma(0)$ may have some coordinates equal to zero, so we instead look for q close to the zero vector such that $\gamma(q)$ has nonzero coordinates. To see why such q exists, note that we can calculate the matrix inverse $(\Sigma^{-1} + Q)^{-1}$ using the determinants of cofactor matrices. From this we see that modulo a multiplicative factor of $[\det(\Sigma^{-1} + Q)]^{-1}$, each $\gamma_i(q) = (\Sigma^{-1} + Q)^{-1}\alpha$ is a nonzero multi-linear polynomial in the $K - 1$ variables $\{q_j\}_{j \neq i}$ (with leading term $\alpha_i \cdot \prod_{j \neq i} q_j$). Thus, the set of vectors q that make $\gamma_i(q)$ equal to zero has measure zero. It follows that we can choose q with arbitrarily small coordinates, such that $\gamma_i(q)$ is nonzero for all i. By the earlier analysis, $\gamma(q)$ is a positive vector, and $[(\Sigma^{-1} + Q)^{-1}]_{ij} \leq 0$ for all $i \neq j$. Letting $q \to 0$, we conclude by continuity that $\gamma(0) \geq 0$ and $\Sigma_{ij} \leq 0$. This is what we desire to show. $\square$

O.4 Proof of Proposition 1

O.4.1 Preliminary Analysis

Suppose the agent's prior about the two unknown payoffs is normal with covariance matrix $\begin{pmatrix} \Sigma_{11} & -\Sigma_{12} \\ -\Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Throughout we assume $\Sigma_{11} \geq \Sigma_{22}$. The objective is to maximize

$$\mathbb{E}[\mathbb{E}[\max\{v_1, v_2\} \mid \mathcal{F}_\tau] - c\tau],$$

To reduce this problem to our main model, we use $\max\{v_1, v_2\} = \max\{v_1 - v_2, 0\} + v_2$ to rewrite the objective as

$$\mathbb{E}[\mathbb{E}[\max\{v_1 - v_2, 0\} \mid \mathcal{F}_\tau] - c\tau] + \mathbb{E}[\mathbb{E}[v_2 \mid F_\tau]]$$

The posterior expectations of v_2, $M_t = \mathbb{E}[v_2 \mid \mathcal{F}_t]$, form a continuous martingale with continuous paths. Moreover, the family $\{M_t\}$ are uniformly integrable because they are conditional expectations of an integrable random variable v_2. Thus we can apply Doob's Optional Sampling Theorem to deduce $\mathbb{E}[\mathbb{E}[v_2 \mid \mathcal{F}_\tau]] = \mathbb{E}[v_2]$, which is just the prior expectation of v_2 (and does not depend on the agent's strategy). It follows that the agent simply maximizes

$$\mathbb{E}[\mathbb{E}[\max\{v_1 - v_2, 0\} \mid \mathcal{F}_\tau] - c\tau].$$

As a corollary, the payoff difference $v_1 - v_2$ is a sufficient statistic for the agent's decision. Now if we let $\theta_1 = v_1$, $\theta_2 = -v_2$, then the prior covariance matrix about θ is simply $\Sigma := \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. This returns our main model with prior covariance matrix Σ and payoff-relevant state $\omega = v_1 - v_2 = \theta_1 + \theta_2$. Since $\alpha_1 = \alpha_2 = 1$, our Theorem 1 applies and yields Corollary 1.

For the subsequent analysis, we need to keep track of how fast the posterior variance of ω evolves over time. These (minimal) posterior variances are given below:

Lemma 11. *Suppose $\Sigma_{11} \geq \Sigma_{22}$. When adopting the optimal information acquisition strategy, the agent's posterior variance of $\omega = \theta_1 + \theta_2$ at time t is given by*

$$\sigma_t^2 = \begin{cases} \frac{\Sigma_{11}+\Sigma_{22}+2\Sigma_{12}+\det(\Sigma)t}{1+\Sigma_{11}t} & \text{if } t \leq t_1^*; \\ \frac{4\det(\Sigma)}{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}+\det(\Sigma)t} & \text{if } t \geq t_1^*. \end{cases}$$

Proof. At time $t \leq t_1^*$, the posterior covariance matrix is

$$\left[\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} + \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1} = \frac{1}{1+\Sigma_{11}t} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} + \det(\Sigma)t \end{pmatrix}.$$

This gives the first part of the lemma.

At time $t = t_1^*$, the levels of uncertainty about θ_1 and θ_2 have equalized. From this time on, each unit of time produces two normal signals $\theta_1 + \epsilon_1$ and $\theta_2 + \epsilon_2$ with ϵ_1 and ϵ_2 independently and identically distributed according to $\mathcal{N}(0, 2)$. These two signals are informationally equivalent to their sum and difference $\theta_1 + \theta_2 + \epsilon_1 + \epsilon_2$ and $\theta_1 - \theta_2 + \epsilon_1 - \epsilon_2$, with i.i.d. noise terms $\epsilon_1 + \epsilon_2, \epsilon_1 - \epsilon_2 \sim \mathcal{N}(0, 4)$. Note that as long as the agent's uncertainty about θ_1 and θ_2 are the same, $\theta_1 + \theta_2$ and $\theta_1 - \theta_2$ are independent. Thus, in terms of learning about $\theta_1 + \theta_2$, it is as if the agent receives only the signal $\theta_1 + \theta_2 + \mathcal{N}(0, 4)$ over each unit of time. This observation enables us to calculate the posterior variance of $\theta_1 + \theta_2$ as follows: For any $t \geq t_1^*$,

$$\sigma_t^2 = \left(\frac{1}{\sigma_{t_1^*}^2} + \frac{t - t_1^*}{4} \right)^{-1},$$

where $\frac{1}{\sigma_{t_1^*}^2}$ is the belief precision of $\theta_1 + \theta_2$ at time t_1^* and $\frac{t - t_1^*}{4}$ is the signal precision from those $\theta_1 + \theta_2 + \mathcal{N}(0, 4)$ signals between time t_1^* and time t. Plugging in $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$ and $\sigma_{t_1^*}^2 = \frac{2 \det(\Sigma)}{\Sigma_{11} - \Sigma_{12}}$ then yields the second part of the lemma. □

O.4.2 Stopping Boundaries and Choice Accuracy

Using the posterior variances characterized above, we can write down the process for the posterior expectation of ω, which we denote by $Y_t = \mathbb{E}[\theta_1 + \theta_2 \mid \mathcal{F}_t]$:

$$Y_t = Y_0 + \int_0^t \sqrt{\frac{\partial \sigma_s^2}{\partial s}} \cdot dB_s, \tag{13}$$

where B_s is a standard Brownian motion with respect to the filtration of the agent's information. The volatility term $\sqrt{\frac{\partial \sigma_s^2}{\partial s}}$ is such that the variance of posterior expectation Y_t matches the reduction in posterior variance $\sigma_0^2 - \sigma_t^2$. This representation is a direct generalization of Lemma 1 in Fudenberg et al. (2018) and follows from standard results.

Therefore, given any prior covariance matrix Σ and any prior expectation $Y_0 = y$, the agent's problem can be rewritten as

$$\max_{\tau} \mathbb{E}[\max\{Y_\tau, 0\} - c\tau], \tag{14}$$

where τ can be any stopping time adapted to the Y process given above. We now define the value function $U(y, c, \Sigma)$ to be the agent's maximal payoff in this problem. It is easy to see that U is non-negative, increasing in y and decreasing in c (we refer to weak monotonicity, unless otherwise specified). In addition, just as in Fudenberg et al. (2018), the *stopping boundary* at time $t = 0$ is symmetric and given by

$$k^*(c, \Sigma) = \min\{x > 0 : \ U(-x, c, \Sigma) = 0\}.$$

What this means is that if the prior expectation satisfies $|Y_0| \geq k^*(c, \Sigma)$, then the agent optimally stops immediately and chooses good 1 or good 2 depending on whether Y_0 is positive or negative. Whereas if $|Y_0| < k^*(c, \Sigma)$, then the optimal stopping time τ is strictly positive.

To study how the agent's choice accuracy changes over time, we need to also consider the stopping boundaries at later times $t > 0$. For this we let Σ_t be the posterior covariance matrix of θ at time t, given the prior covariance matrix Σ and given the optimal attention strategy. Then the stopping boundary at time t is simply $k^*(c, \Sigma_t)$ (this implicitly uses the fact that starting from the prior Σ_t, the posterior at time s would be Σ_{t+s}).

The next lemma (essentially Theorem 2 in Fudenberg et al. (2018)) characterizes a necessary and sufficient condition for choice accuracy to decrease over time, in terms of these stopping boundaries:

Lemma 12. *Given c and Σ, and suppose $|Y_0| < k^*(c, \Sigma)$ (so that the agent does not immediately stop). Let p_t be the conditional probability that good 1 is better than good 2 when the agent stops at time t and chooses good 1 (and vice verse, by symmetry). Then p_t decreases in t if and only if $\frac{k^*(c,\Sigma_t)}{\sigma_t}$ decreases in t.*

Proof. If the agent stops at time t and chooses good 1, then $Y_t = k^*(c, \Sigma_t)$ (and t is the earliest time this happens). So by definition, the posterior belief of ω is normal with mean $k^*(c, \Sigma_t)$ and standard deviation σ_t. Thus, the conditional probability that $\omega > 0$ (i.e., good 1 is better) is the normal c.d.f. evaluated at $\frac{k^*(c,\Sigma_t)}{\sigma_t}$. This yields the result. □

Note that σ_t^2, being the posterior variance of $\theta_1 + \theta_2$, is just the sum of all the entries in the matrix Σ_t. Thus the condition in Lemma 12 is ultimately about how $k^*(c, \Sigma)$ varies with Σ. The next section studies this change in detail.

O.4.3 Effect of Σ on Stopping Boundary

We will say two covariance matrices $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ *induce the same prior uncertainty*, if the prior variances $\widetilde{\sigma}_0^2 = \widetilde{\Sigma}_{11} + 2\widetilde{\Sigma}_{12} + \widetilde{\Sigma}_{22}$ and $\widehat{\sigma}_0^2 = \widehat{\Sigma}_{11} + 2\widehat{\Sigma}_{12} + \widehat{\Sigma}_{22}$ of $\theta_1 + \theta_2$ are equal.

The result below provides a sufficient condition for stopping boundaries under two different prior covariance matrices to be comparable:

Lemma 13. *Let $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ be two covariance matrices that induce the same prior uncertainty. Suppose further that the following two conditions holds:*

(a) $\widetilde{\Sigma}_{11} - \widetilde{\Sigma}_{22} \geq \widehat{\Sigma}_{11} - \widehat{\Sigma}_{22} \geq 0$;

(b) $(\widetilde{\sigma}_{\widetilde{t}_1^*})^2 = \frac{2\det(\widetilde{\Sigma})}{\widetilde{\Sigma}_{11} - \widetilde{\Sigma}_{12}} \leq \frac{2\det(\widehat{\Sigma})}{\widehat{\Sigma}_{11} - \widehat{\Sigma}_{12}} = (\widehat{\sigma}_{\widehat{t}_1^*})^2$.

Then the posterior variances satisfy $\widetilde{\sigma}_t^2 \le \widehat{\sigma}_t^2$ *for all* $t \ge 0$*. Consequently* $k^*(c, \widetilde{\Sigma}) \ge k^*(c, \widehat{\Sigma})$*.*

Part (a) says that there is greater asymmetry in the agent's uncertainty about the two attributes in the prior covariance matrix $\widetilde{\Sigma}$ compared to $\widehat{\Sigma}$. Part (b) says that the agent's uncertainty about $\theta_1 + \theta_2$ at the optimal switchpoint $\tilde{t}_1^*$ given prior $\widetilde{\Sigma}$ is lower than the agent's uncertainty about $\theta_1 + \theta_2$ at the optimal switchpoint $\hat{t}_1^*$ given prior $\widehat{\Sigma}$, i.e., the agent has learned more about $\theta_1 + \theta_2$ in (the endogenous) Stage 1 starting from $\widetilde{\Sigma}$. The lemma says that these conditions imply that the agent's uncertainty about $\theta_1 + \theta_2$ is lower at every moment of time starting from prior $\widetilde{\Sigma}$, i.e., the agent learns faster.

Proof. Given any prior covariance matrix Σ and resulting path of posterior variances σ_t, we can define for each $v \in [0, \sigma_0^2)$ the hitting time $T(v)$ such that $\sigma_{T(v)}^2 = \sigma_0^2 - v$. $T(v)$ is well-defined because σ_t^2 decreases strictly and continuously in t. This monotonicity also implies that the comparison $\widetilde{\sigma}_t^2 \le \widehat{\sigma}_t^2$ for each t is equivalent to $\widetilde{T}(v) \le \widehat{T}(v)$ for each $v < \widetilde{\sigma}_0^2 = \widehat{\sigma}_0^2$. Thus in what follows we study the properties of $T(v)$.

From Lemma 11, it is not difficult to derive the following formula for $T(v)$:

$$T(v) = \begin{cases} \frac{v}{(\Sigma_{11}+\Sigma_{12})^2 - \Sigma_{11} v} & \text{if } v \in [0, v^*]; \\ \frac{4}{\sigma_0^2 - v} - \frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{\det(\Sigma)} & \text{if } v \in [v^*, \sigma_0^2). \end{cases}$$

Above, the switchpoint v^* is given by

$$v^* = \sigma_0^2 - \sigma_{t_1^*}^2 = (\Sigma_{11} + \Sigma_{22} + 2\Sigma_{12}) - \frac{2\det(\Sigma)}{\Sigma_{11} - \Sigma_{12}} = \frac{(\Sigma_{11} - \Sigma_{22})(\Sigma_{11} + \Sigma_{12})}{\Sigma_{11} - \Sigma_{12}}.$$

Not surprisingly, at $v = v^*$ either formula for $T(v)$ yields the time $t_1^* = \frac{\Sigma_{11} - \Sigma_{22}}{\det(\Sigma)}$.

We now compute the (right) derivative of $T(v)$:

$$T'(v) = \begin{cases} \frac{1}{\left(\Sigma_{11}+\Sigma_{12} - \frac{\Sigma_{11}}{\Sigma_{11}+\Sigma_{12}} v\right)^2} & \text{if } v \in [0, v^*); \\ \frac{4}{(\sigma_0^2 - v)^2} & \text{if } v \in [v^*, \sigma_0^2). \end{cases} \tag{15}$$

This time perhaps more surprisingly, both formulae for $T'(v)$ yield the same value at $v = v^*$. Moreover, it can be checked that $T'(v) \le \frac{4}{(\sigma_0^2 - v)^2}$ for all v.

We claim that under the two stated conditions on $\widetilde{\Sigma}$ and $\widehat{\Sigma}$, it holds that $\widetilde{T}'(v) \le \widehat{T}'(v)$ for every v. Since $\widetilde{T}(0) = \widehat{T}(0) = 0$, this would imply the desired comparison $\widetilde{T}(v) \le \widehat{T}(v)$. To compare those derivatives, note that given $\widetilde{\Sigma}_{11} + \widetilde{\Sigma}_{22} + 2\widetilde{\Sigma}_{12} = \widehat{\Sigma}_{11} + \widehat{\Sigma}_{22} + 2\widehat{\Sigma}_{12}$, the assumption $\widetilde{\Sigma}_{11} - \widetilde{\Sigma}_{22} \ge \widehat{\Sigma}_{11} - \widehat{\Sigma}_{22}$ is equivalent to $\widetilde{\Sigma}_{11} + \widetilde{\Sigma}_{12} \ge \widehat{\Sigma}_{11} + \widehat{\Sigma}_{12}$. Thus $\widetilde{T}'(0) \le \widehat{T}'(0)$ holds. Moreover, the second assumption in the lemma translates into $\widetilde{v}^* \ge \widehat{v}^*$. Below we show these are sufficient to imply $\widetilde{T}'(v) \le \widehat{T}'(v)$.

Indeed, for $v \geq \widehat{v}^*$, we deduce from (15) that

$$\widetilde{T}'(v) \leq \frac{4}{(\sigma_0^2 - v)^2} = \widehat{T}'(v).$$

On the other hand, for $v \leq \widehat{v}^*$, $T'(v)$ is given by the first term in (15) for both $\widetilde{\Sigma}$ and $\widehat{\Sigma}$. Thus the comparison between $\widetilde{T}'(v)$ and $\widehat{T}'(v)$ reduces to a comparison between two linear functions of v: We want to show $\Sigma_{11} + \Sigma_{12} - \frac{\Sigma_{11}}{\Sigma_{11}+\Sigma_{12}} v$ is larger when $\Sigma = \widetilde{\Sigma}$ than when $\Sigma = \widehat{\Sigma}$. We already know this holds at $v = 0$ and $v = \widetilde{v}^*$, so by linearity, it also holds at any v in between, completing the proof.

Hence we have shown that $\widetilde{\sigma}_t^2 \leq \widehat{\sigma}_t^2$ for every t. It remains to show that the comparison of posterior variances implies the comparison of stopping boundaries. For this we observe that lower posterior variances under $\widetilde{\Sigma}$ imply that the value function $U(y, c, \widetilde{\Sigma})$ is weakly larger than $U(y, c, \widehat{\Sigma})$ for any cost c and any prior expectation y. This follows from the same time-change argument as in the proof of Lemma 5, and the idea is simply that any stopping time under prior $\widehat{\Sigma}$ can be replicated under prior $\widetilde{\Sigma}$ at an earlier stopping time. Thus

$$0 = U(-k^*(c, \widetilde{\Sigma}), c, \widetilde{\Sigma}) \geq U(-k^*(c, \widetilde{\Sigma}), c, \widehat{\Sigma}).$$

In fact we have equality since U is non-negative. Hence $k^*(c, \widehat{\Sigma})$, being the smallest x such that $U(-x, c, \widehat{\Sigma}) = 0$, must be smaller than $k^*(c, \widetilde{\Sigma})$. □

Two useful corollaries follow from Lemma 13 (the proofs are immediate and thus omitted):

Lemma 14 (Effect of correlation)**.** *Let $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ be two covariance matrices that induce the same prior uncertainty. Suppose further that $\widetilde{\Sigma}_{11} = \widetilde{\Sigma}_{22}$ and $\widehat{\Sigma}_{11} = \widehat{\Sigma}_{22}$ (symmetric priors). Then $k^*(c, \widetilde{\Sigma}) = k^*(c, \widehat{\Sigma})$.*

Lemma 15 (Effect of asymmetry)**.** *Let $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ be two covariance matrices that induce the same prior uncertainty. Suppose further that $\widetilde{\Sigma}_{11} > \widetilde{\Sigma}_{22}$ while $\widehat{\Sigma}_{11} = \widehat{\Sigma}_{22}$ (asymmetric versus symmetric). Then $k^*(c, \widetilde{\Sigma}) \geq k^*(c, \widehat{\Sigma})$.*

While the above results hold fixed the prior variance σ_0^2, we also need a result that considers a change in overall prior uncertainty, and characterizes its effect on the stopping boundary.

Lemma 16 (Effect of scaling Σ)**.** *For any c, Σ and any $\lambda \in (0, 1)$, it holds that $\lambda k^*(c, \Sigma) \geq k^*(c, \lambda^2 \Sigma)$.*

Proof. We will show that for any $\lambda > 0$,

$$k^*(c, \lambda^2 \Sigma) = \lambda k^*(c\lambda^{-3}, \Sigma). \tag{16}$$

This implies the lemma because the cost $c\lambda^{-3}$ is higher than c whenever $\lambda < 1$, which decreases the value function and thus also decreases the stopping boundary, resulting in $k^*(c\lambda^{-3}, \Sigma) \leq k^*(c, \Sigma)$.

We note that the identity (16) is a direct generalization of Equation (A6) in Fudenberg et al. (2018). Nonetheless, we provide a proof below for completeness. The key insight is that we can identify the belief processes under prior Σ and under $\lambda^2\Sigma$ via a scaling of time and space. Specifically, let Y_t denote the belief process under Σ as given by (13), with $\sigma_t^2(\Sigma)$ denoting the posterior variance at time t (under the optimal path). Similarly define Z_t for the belief process starting from the prior $\lambda\Sigma$, with $\sigma_t^2(\lambda^2\Sigma)$ denoting the posterior variance at time t. From Lemma 11, we have the relation

$$\sigma_t^2(\lambda^2\Sigma) = \lambda^2 \cdot \sigma_{\lambda^2 t}^2(\Sigma).$$

Assuming $Z_0 = \lambda Y_0$, then the process Z_t has the same distribution as the process $\lambda \cdot Y_{\lambda^2 t}$. Intuitively, receiving standard normal signals about $\lambda\omega$ for one unit of time is equivalent to receiving standard normal signals about ω for λ^2 units of time.

Now that we identify Z_t with $\lambda \cdot Y_{\lambda^2 t}$, we can rewrite the stopping problem with respect to Z_t in terms of the Y process instead. Specifically, the agent's problem is

$$\max_{\tau} \mathbb{E}[\max\{\lambda Y_{\lambda^2\tau}, 0\} - c\tau] = \lambda \max_{\tau'} \mathbb{E}[\max\{Y_{\tau'}, 0\} - \frac{c}{\lambda^3}\tau'],$$

with $\tau' = \lambda^2\tau$. Thus it is as if the agent chooses an optimal stopping time with respect to Y_t, but with transformed marginal cost $\frac{c}{\lambda^3}$. The agent should stop at time 0 in this problem if and only if $|Y_0| \geq k^*(\frac{c}{\lambda^3}, \Sigma)$, which is equivalent to $|Z_0| \geq \lambda k^*(\frac{c}{\lambda^3}, \Sigma)$. Thus (16) holds. □

O.4.4 Main Proof of Proposition 1

Given Lemma 12, we just need to show that for any times $t < t'$, $\frac{k^*(c,\Sigma_t)}{\sigma_t} \geq \frac{k^*(c,\Sigma_{t'})}{\sigma_{t'}}$. Define $\lambda \in (0, 1)$ such that $\sigma_{t'} = \lambda\sigma_t$. Then the inequality becomes $\lambda k^*(c, \Sigma_t) \geq k^*(c, \Sigma_{t'})$. From Lemma 16 we have $\lambda k^*(c, \Sigma_t) \geq k^*(c, \lambda^2\Sigma_t)$, so it is sufficient to show

$$k^*(c, \lambda^2\Sigma_t) \geq k^*(c, \Sigma_{t'}). \qquad \text{(to be shown)}$$

Note that, by the definition of λ, the matrices $\lambda^2\Sigma_t$ and $\Sigma_{t'}$ induce the same uncertainty about $\theta_1 + \theta_2$. There are three cases to consider:

Case 1: $t_1^* \leq t < t'$. In this case Σ_t and $\Sigma_{t'}$ are posterior covariance matrices in Stage 2, so they induce symmetric uncertainty about θ_1 and θ_2. Lemma 14 thus applies to the symmetric priors $\lambda^2\Sigma_t$ and $\Sigma_{t'}$, and yields $k^*(c, \lambda^2\Sigma_t) = k^*(c, \Sigma_{t'})$. Intuitively, the belief process in Stage 2 is the same as in Fudenberg et al. (2018) since correlation does not matter for symmetric priors. Thus the result in this case follows from the result in that paper.

Case 2: $t < t_1^* \leq t'$**.** Here Lemma 14 no longer applies. We instead apply Lemma 15 to deduce $k^*(c, \lambda^2 \Sigma_t) \geq k^*(c, \Sigma_{t'})$, since $\lambda^2 \Sigma_t$ is an asymmetric prior while $\Sigma_{t'}$ is a symmetric prior. This proves the result, and as discussed, the intuition is that asymmetry increases the stopping boundary relative to symmetric priors.

Case 3: $t < t' < t_1^*$**.** To prove the key comparison $k^*(c, \lambda^2 \Sigma_t) \geq k^*(c, \Sigma_{t'})$, we now need to invoke the more general Lemma 13 since $\lambda^2 \Sigma_t$ and $\Sigma_{t'}$ are both asymmetric. Thus, we have to check that $\widetilde{\Sigma} = \lambda^2 \Sigma_t$ and $\widehat{\Sigma} = \Sigma_{t'}$ satisfy the two conditions stated in Lemma 13.

To do this, we let $\Sigma_t = \begin{pmatrix} p & r \\ r & q \end{pmatrix}$ with $p > q > 0$ and $r^2 < pq$. The posterior covariance matrix $\Sigma_{t'}$ at the later time t' can be calculated from the "prior" covariancee matrix Σ_t, after focusing on θ_1 for $t' - t$ units of time. Thus

$$\Sigma_{t'} = \left(\Sigma_t^{-1} + \begin{pmatrix} t' - t & 0 \\ 0 & 0 \end{pmatrix} \right)^{-1} = \left(\frac{1}{pq - r^2} \begin{pmatrix} q + (t' - t)(pq - r^2) & -r \\ -r & p \end{pmatrix} \right)^{-1}.$$

Let $q' = q + (t' - t)(pq - r^2)$ with $q < q' < p$ (the latter inequality holds because $t' < t_1^*$). Then the above simplifies to

$$\Sigma_{t'} = \frac{pq - r^2}{pq' - r^2} \begin{pmatrix} p & r \\ r & q' \end{pmatrix}$$

The scaling factor λ is thus given by

$$\lambda^2 (p + q + 2r) = \frac{pq - r^2}{pq' - r^2} (p + q' + 2r). \tag{17}$$

We now check the first condition in Lemma 13, which for $\widetilde{\Sigma} = \lambda^2 \Sigma_t$ and $\widehat{\Sigma} = \Sigma_{t'}$ becomes $\lambda^2 (p - q) \geq \frac{pq - r^2}{pq' - r^2} (p - q')$. Using (17) to eliminate λ, it suffices to show that

$$\frac{p - q}{p + q + 2r} \geq \frac{p - q'}{p + q' + 2r}.$$

This inequality holds simply because $q' > q$.

We then turn to the second condition in Lemma 12, which in the current setting becomes $2\lambda^2 (\frac{pq - r^2}{p - r}) \leq 2 (\frac{pq - r^2}{pq' - r^2}) \cdot (\frac{pq' - r^2}{p - r})$. This simplifies to $\lambda^2 \leq 1$, which clearly holds. Intuitively, since Σ_t and $\Sigma_{t'}$ are both posterior beliefs following the prior Σ, they "become symmetric" at the same posterior belief $\Sigma_{t_1^*}$. Thus the second condition in Lemma 13 holds with equality when we consider Σ_t versus $\Sigma_{t'}$. It follows that when comparing $\lambda^2 \Sigma_t$ and $\Sigma_{t'}$, the former prior belief leads to lower uncertainty when entering Stage 2.

Hence Lemma 13 applies, and we again have

$$\lambda k^*(c, \Sigma_t) \geq k^*(c, \lambda^2 \Sigma_t) \geq k^*(c, \Sigma_{t'}).$$

This proves $\frac{k^*(c, \Sigma_t)}{\sigma_t} \geq \frac{k^*(c, \Sigma_{t'})}{\sigma_{t'}}$ and the proposition.

O.4.5 Generalization to Unequal Learning Speeds

In this section we show that the conclusion of Proposition 1 further generalizes to situations where the information about the two unknown payoffs arrives with different levels of precision. Formally, fix $\alpha_1, \alpha_2 > 0$ and suppose a unit of time devoted to learning about the payoff v_i produces the signal $v_i + \mathcal{N}(0, \alpha_i^2)$. Any prior covariance matrix over (v_1, v_2) can be written as $\begin{pmatrix} \alpha_1^2\Sigma_{11} & -\alpha_1\alpha_2\Sigma_{12} \\ -\alpha_1\alpha_2\Sigma_{21} & \alpha_2^2\Sigma_{22} \end{pmatrix}$ where $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ is also a positive definite matrix. This formulation of the problem will make the results below easier to state.

To transform this problem into our main model, let $\theta_1 = v_1/\alpha_1$ and $\theta_2 = -v_2/\alpha_2$, so that each unit of time devoted to θ_i produces a standard normal signal about it. Moreover, the prior covariance matrix over (θ_1, θ_2) is simply Σ, and the payoff-relevant state is $\omega = v_1 - v_2 = \alpha_1\theta_1 + \alpha_2\theta_2$. Throughout this section, we assume Σ and α satisfy Assumption 3, so that we can apply Theorem 1 to characterize optimal attention allocation.

Corollary 4. *Suppose Assumption 3 holds and* $\alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} \geq \alpha_1\Sigma_{12} + \alpha_2\Sigma_{22}$. *The agent's optimal information acquisition strategy* $(\beta_1(t), \beta_2(t))$ *in this generalized binary choice problem consists of two stages:*

- ***Stage 1:*** *At all times*

$$t < t_1^* = \frac{\alpha_1\Sigma_{11} + \alpha_2\Sigma_{12} - \alpha_1\Sigma_{21} - \alpha_2\Sigma_{22}}{\alpha_2 \det(\Sigma)},$$

  the agent optimally allocates all attention to θ_1.
- ***Stage 2:*** *At times* $t \geq t_1^*$, *the agent optimally uses the constant mixture* $\left(\frac{\alpha_1}{\alpha_1+\alpha_2}, \frac{\alpha_2}{\alpha_1+\alpha_2}\right)$.

From this we can compute the agent's posterior variance of $\alpha_1\theta_1 + \alpha_2\theta_2$ at each time t. Generalizing Lemma 11, we have

$$\sigma_t^2 = \begin{cases} \frac{\alpha_1^2\Sigma_{11}+\alpha_2^2\Sigma_{22}+2\alpha_1\alpha_2\Sigma_{12}+\alpha_2^2\det(\Sigma)t}{1+\Sigma_{11}t} & \text{if } t \leq t_1^*; \\ \frac{(\alpha_1+\alpha_2)^2\det(\Sigma)}{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}+\det(\Sigma)t} & \text{if } t \geq t_1^*. \end{cases}$$

In particular, $\sigma_{t_1^*}^2 = \frac{(\alpha_1+\alpha_2)\alpha_2\det(\Sigma)}{\Sigma_{11}-\Sigma_{12}}$. We omit the detailed calculations.

Next, note that Lemma 12 holds without change, so we just need to study how the stopping boundary $k^*(c, \Sigma)$ varies with Σ in this more general setting. Naturally, we say that two prior covariance matrices $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ *induce the same prior uncertainty* if the prior variances $\alpha_1^2\widetilde{\Sigma}_{11} + \alpha_2^2\widetilde{\Sigma}_{22} + 2\alpha_1\alpha_2\widetilde{\Sigma}_{12}$ and $\alpha_1^2\widehat{\Sigma}_{11} + \alpha_2^2\widehat{\Sigma}_{22} + 2\alpha_1\alpha_2\widehat{\Sigma}_{12}$ are equal. The key Lemma 13 above is then generalized as follows:

Lemma 17. *Let $\widetilde{\Sigma}$ and $\widehat{\Sigma}$ be two covariance matrices that induce the same prior uncertainty. Suppose further that the following two conditions holds:*

(a) $\alpha_1\widetilde{\Sigma}_{11} + \alpha_2\widetilde{\Sigma}_{12} - \alpha_1\widetilde{\Sigma}_{21} - \alpha_2\widetilde{\Sigma}_{22} \geq \alpha_1\widehat{\Sigma}_{11} + \alpha_2\widehat{\Sigma}_{12} - \alpha_1\widehat{\Sigma}_{21} - \alpha_2\widehat{\Sigma}_{22} \geq 0$*;*

(b) $(\widetilde{\sigma}_{\widetilde{t}_1^*})^2 = \frac{(\alpha_1+\alpha_2)\alpha_2\det(\widetilde{\Sigma})}{\widetilde{\Sigma}_{11}-\widetilde{\Sigma}_{12}} \leq \frac{(\alpha_1+\alpha_2)\alpha_2\det(\widehat{\Sigma})}{\widehat{\Sigma}_{11}-\widehat{\Sigma}_{12}} = (\widehat{\sigma}_{\widehat{t}_1^*})^2$.

Then the posterior variances satisfy $\widetilde{\sigma}_t^2 \leq \widehat{\sigma}_t^2$ for all $t \geq 0$. Consequently $k^*(c, \widetilde{\Sigma}) \geq k^*(c, \widehat{\Sigma})$.

That is, with potentially unequal payoff weights α_1 and α_2, prior "asymmetry" is not simply measured by the difference between Σ_{11} and Σ_{22}. Rather, it is measured by the difference in initial marginal values $cov_1 = \alpha_1\Sigma_{11} + \alpha_2\Sigma_{12}$ and $cov_2 = \alpha_1\Sigma_{21} + \alpha_2\Sigma_{22}$. Condition (a) thus requires this asymmetry to be larger under $\widetilde{\Sigma}$ than under $\widehat{\Sigma}$. This turns out to be equivalent to $\widetilde{T}'(0) \leq \widehat{T}'(0)$, where the hitting time T is the same as defined in the proof of Lemma 13.[26] Together with condition (b), this implies $\widetilde{T}'(v) \leq \widehat{T}'(v)$ for all v and thus $\widetilde{T}(v) \leq \widehat{T}(v)$, just as in that proof.

Using the same notion of asymmetry, we obtain direct generalizations of Lemma 14 and Lemma 15 as well. Finally, Lemma 16 continues to hold since its proof does not depend on payoff weights. These lemmata allow us to replicate the proof of Proposition 1 with only minor modifications. Thus, to summarize, we have the following result:

Proposition 9. *Consider the binary choice problem with general signal variances α_1^2 and α_2^2. Denote the agent's prior covariance matrix over the payoffs (v_1, v_2) as $\begin{pmatrix} \alpha_1^2\Sigma_{11} & -\alpha_1\alpha_2\Sigma_{12} \\ -\alpha_1\alpha_2\Sigma_{21} & \alpha_2^2\Sigma_{22} \end{pmatrix}$. Then whenever Σ and α satisfy Assumption 3, the agent's choice accuracy in this problem is (weakly) higher at earlier stopping times.*

We reiterate that Assumption 3 is guaranteed if $\alpha_1 = \alpha_2$ (as we assumed previously), or if $\Sigma_{12} \geq 0$ (i.e., the unknown payoffs v_1 and v_2 are negatively correlated).

O.5 Proof of Proposition 3

For $t \leq T$, $\hat{n}_i(t) = 0$ for every $i > 1$, so the result $\hat{n}_i(t) \leq n_i(t)$ trivially holds. Below we consider a fixed time $t > T$. We can further assume $n_1(t) < T$, because otherwise the proof of Proposition 2

[26] From the formulae for σ_t^2 we can compute $T(v)$ and $T'(v)$. Writing $v^* = \sigma_0^2 - \sigma_{t_1^*}^2 = \frac{(cov_1 - cov_2)cov_1}{\Sigma_{11}-\Sigma_{12}}$, we have

$$T'(v) = \begin{cases} \frac{1}{\left(cov_1 - \frac{\Sigma_{11}}{cov_1}v\right)^2} & \text{if } v \in [0, v^*]; \\ \frac{(\alpha_1+\alpha_2)^2}{(\sigma_0^2 - v)^2} & \text{if } v \in [v^*, \sigma_0^2). \end{cases}$$

(in the main text) shows that $\hat{n}(t)$ coincides with $n(t)$. Given this assumption, we claim that $\hat{n}_1(t)$ is exactly equal to T. Indeed, let $\bar{t} \geq t$ be the first time at which $n_1(\bar{t}) = T$; such a time exists by monotonicity and continuity of $n_1(\cdot)$. Then $\hat{n}(\bar{t})$ coincides with $n(\bar{t})$, which in particular implies $\hat{n}_1(\bar{t}) = n_1(\bar{t}) = T$. Monotonicity of $\hat{n}_1(\cdot)$ then implies $\hat{n}_1(t) \leq T$. But by assumption $\hat{n}_1(t) \geq T$, so equality must hold.

Next, we connect the two vectors $n(t)$ and $\hat{n}(t)$ by a continuous path. For each $x \in [n_1(t), T]$, we define q^x as the cumulative attention vector (at time t) resulting from a hypothetical attention manipulation that forces the agent to observe source 1 for x units of time. That is,

$$q^x = (q_1^x, \ldots, q_K^x) = \operatorname*{argmin}_{q_1,\ldots,q_K \geq 0:\ \sum_i q_i = t \text{ and } q_1 \geq x} V(q).$$

Clearly, $n(t) = q^{n_1(t)}$ and $\hat{n}(t) = q^T$. By the same argument as in the previous paragraph, $q_1^x = x$ holds for x in this range. So in defining q^x we can replace the constraint $q_1^x \geq x$ with equality.

To prove the proposition, it suffices to show that as x decreases from T to $n_1(t)$, q_i^x weakly increases for each $i > 1$. Similar to our proof of Theorem 2, the proof strategy here will be to use the Hessian matrix of V to compute the derivative of the vector q^x with respect to x. For this we fix $x > n_1(t)$, and assume for now that q_i^x is strictly positive for each $i > 1$. Then the first-order condition for the constrained optimality of q^x yields $\partial_2 V(q^x) = \cdots = \partial_K V(q^x)$. If the vector q^x is left-differentiable at x, then for any y slightly smaller than x, q^y must also satisfy the above equal marginal value condition (for every source $i > 1$).

Thus the left derivative of q^x is a vector $u \in \mathbb{R}^K$ that satisfies $u_1 = 1$, $u_1 + \cdots + u_K = 0$ and

$$Hess_V(q^x) \cdot u = \lambda(c, 1, \ldots, 1)', \tag{18}$$

for some $\lambda, c \in \mathbb{R}$ that will be determined later.

Under the differentiability assumption, we can solve for u as follows. Note from Lemma 3 that $\partial_{ij} V = 2\gamma_i\gamma_j[(\Sigma^{-1} + Q)^{-1}]_{ij}$, where we save notation by writing $Q = \operatorname{diag}(q^x)$ from now on. Then, we have the matrix identity

$$Hess_V(q^x) = 2\operatorname{diag}(\gamma) \cdot (\Sigma^{-1} + Q)^{-1} \cdot \operatorname{diag}(\gamma), \tag{19}$$

where γ is as usual the vector $(\Sigma^{-1} + Q)^{-1} \cdot \alpha$. As shown in the proof of Proposition 7, γ has strictly positive coordinates. Recalling $\partial_i V = -\gamma_i^2$, we thus have $\gamma_2 = \cdots = \gamma_K > 0$. γ_1 cannot be larger, since then q_1^x should be larger than x to minimize V. γ_1 cannot be equal to the other sources either, since in that case the vector q^x would satisfy the first-order condition for the unconstrained variance minimization problem, leading to $q^x = n(t)$ and $x = n_1(t)$. So $\gamma_2 = \cdots = \gamma_K > \gamma_1 > 0$, which will be useful below.

Now, using (18) and (19), we have

$$
\begin{aligned}
u = \lambda \cdot Hess^{-1} \cdot (c,1,\dots,1)' &= 2\lambda \cdot \operatorname{diag}(1/\gamma) \cdot (\Sigma^{-1}+Q) \cdot \operatorname{diag}(1/\gamma) \cdot (c,1,\dots,1)' \\
&= 2\lambda \cdot \operatorname{diag}(1/\gamma) \cdot (\Sigma^{-1}+Q) \cdot \left(\frac{c}{\gamma_1}, \frac{1}{\gamma_2}, \dots, \frac{1}{\gamma_2}\right)' \\
&= \frac{2\lambda}{\gamma_2^2} \cdot \operatorname{diag}(1/\gamma) \cdot (\Sigma^{-1}+Q) \cdot \left(\frac{c\gamma_2^2}{\gamma_1}, \gamma_2, \dots, \gamma_2\right)'.
\end{aligned}
$$

If we rewrite $\frac{c\gamma_2^2}{\gamma_1}$ as $\gamma_1 - b$ for some $b \in \mathbb{R}$, then the vector $\left(\frac{c\gamma_2^2}{\gamma_1}, \gamma_2, \dots, \gamma_2\right)'$ differs from γ only in that the first coordinate is smaller by b. Using $(\Sigma^{-1}+Q)\gamma = \alpha$, we thus obtain

$$
(\Sigma^{-1}+Q) \cdot \left(\frac{c\gamma_2^2}{\gamma_1}, \gamma_2, \dots, \gamma_2\right)' = (\alpha_1 - b[\Sigma^{-1}+Q]_{11}, \alpha_2 - b[\Sigma^{-1}+Q]_{21}, \dots, \alpha_K - b[\Sigma^{-1}+Q]_{K1})'
$$

By the symmetry of $\Sigma^{-1}+Q$, it follows that

$$
u_i = \frac{2\lambda}{\gamma_2^2} \cdot \frac{\alpha_i - b[\Sigma^{-1}+Q]_{1i}}{\gamma_i}. \tag{20}
$$

Recall $u_1 = 1$, which reflects $q_1^y = y$ for every y. Thus $\lambda \neq 0$. Moreover, recall $\sum_{i=1}^K u_i = 0$, a consequence of the fact that $\sum_{i=1}^K q_i^y = t$ for every y. Thus, we obtain

$$
b \sum_{i=1}^K \frac{[\Sigma^{-1}+Q]_{1i}}{\gamma_i} = \sum_{i=1}^K \frac{\alpha_i}{\gamma_i}. \tag{21}
$$

The RHS is clearly positive. On the LHS, using $\gamma_2 = \cdots = \gamma_K > \gamma_1 > 0$ and $[\Sigma^{-1}+Q]_{11} > 0$,

$$
\begin{aligned}
\sum_{i=1}^K \frac{[\Sigma^{-1}+Q]_{1i}}{\gamma_i} &= \frac{[\Sigma^{-1}+Q]_{11}}{\gamma_1} + \sum_{i>1} \frac{[\Sigma^{-1}+Q]_{1i}}{\gamma_i} \\
&> \frac{[\Sigma^{-1}+Q]_{11} \cdot \gamma_1}{\gamma_2^2} + \sum_{i>1} \frac{[\Sigma^{-1}+Q]_{1i} \cdot \gamma_i}{\gamma_2^2} \\
&= \frac{1}{\gamma_2^2} \sum_{i=1}^K [\Sigma^{-1}+Q]_{1i} \cdot \gamma_i \\
&= \frac{\alpha_1}{\gamma_2^2} > 0,
\end{aligned}
$$

where the last equality uses $\sum_{i=1}^K (\Sigma^{-1}+Q)_{1i} \cdot \gamma_i = \alpha_1 > 0$ as a consequence of $(\Sigma^{-1}+Q)\gamma = \alpha$.

Thus, (21) implies the crucial inequality $b > 0$. It follows that for this b and any $i > 1$, $\frac{\alpha_i - b[\Sigma^{-1}+Q]_{1i}}{\gamma_i}$ is strictly positive (by assumption $[\Sigma^{-1}+Q]_{1i} \leq 0$). From $\sum_{i=1}^K u_i = 0$, we further know that $\frac{\alpha_1 - b[\Sigma^{-1}+Q]_{11}}{\gamma_1}$ is strictly negative. We can then use (20) to determine the unique value of $\lambda < 0$ that makes $u_1 = 1$. Since $\lambda < 0$ and $b > 0$, (20) yields $u_i < 0$ for every $i > 0$. Hence, starting from q^x, any local decrease in the amount of attention towards source 1 increases the optimal amount of attention towards every other source. This is what we desire to show.

To complete the proof, we now dispense with two previous assumptions in the analysis. First, we have assumed $q_i^x > 0$ for each $i > 1$. In general, relabeling the attributes if necessary, we can assume that among the sources $2, \ldots, K$, the ones with maximal marginal value (i.e., those with maximal γ_i) are sources $2 \sim k$ for some k. Then the first-order condition for the constrained optimality of q_i^x requires $q_i^x = 0$ for $i > k$. Moreover, by the same argument as above, γ_1 is strictly smaller. By continuity, sources $2 \sim k$ also maximize the marginal value at any q_i^y where y is slightly smaller than x. So $q_i^y = 0$ for $i > k$ also holds. Thus, locally we can reduce the problem with K sources to a smaller problem with only the first k sources, similar to the proof of Theorem 2.

In this smaller problem, the payoff weight vector $\tilde{\alpha}$ is given by (11), and it is strictly positive as shown in the proof of Theorem 2. The prior covariance matrix becomes Σ_{TL}, which is the $k \times k$ top-left principal sub-matrix of Σ. Since Σ^{-1} is an M-matrix, so is $(\Sigma_{TL})^{-1}$.[27] The vector $\tilde{\gamma} = [(\Sigma_{TL})^{-1} + \tilde{Q}]^{-1} \cdot \tilde{\alpha} \in \mathbb{R}^k$ is the first k coordinates of γ, so we have $\tilde{\gamma}_2 = \cdots = \tilde{\gamma}_k > \tilde{\gamma}_1 > 0$. Hence, the above procedure for finding u directly applies to this smaller problem, and yields a vector $\tilde{u} \in \mathbb{R}^k$ with $\tilde{u}_1 = 1$, $\tilde{u}_i < 0$ for $1 < i \leq k$ and $\tilde{u}_1 + \cdots + \tilde{u}_k = 0$. In the original problem, the left-derivative of q^x is thus the vector $(\tilde{u}_1, \ldots, \tilde{u}_k, 0, \ldots, 0)'$. Once again, locally decreasing attention towards source 1 weakly increases attention towards every other source. To be fully rigorous, we also note that by the Maximum Theorem, q^x varies continuously with x. So the preceding form of local monotonicity implies global monotonicity (without continuity, q_i^x may only be piece-wise monotone).

Our second simplifying assumption is left-differentiablity, which we now argue is without loss. Let x be the infimum of those numbers $\hat{x} \in [n_1(t), T]$ such that q^y is left-differentiable in y for $y > \hat{x}$. By properties of the infimum, x is in fact the *smallest* number $\geq n_1(t)$ such that q^y is left-differentiable for $y > x$. Now suppose $x > n_1(t)$, and we will deduce a contradiction. Specifically, let sources $2 \sim k$ have the maximal marginal value at q^x. We can use the above procedure to find the vectors $\tilde{u} \in \mathbb{R}^k$ and $(\tilde{u}_1, \ldots, \tilde{u}_k, 0, \ldots, 0)' \in \mathbb{R}^K$, such that if q^x is perturbed slightly in the direction $-u$, the resulting attention vector maintains the equal marginal value property across sources $2 \sim k$. Formally, for any attention vector q with total attention t such that sources $2 \sim k$ have the maximal marginal value, we can solve for $b, \lambda \in \mathbb{R}$ and $\tilde{u} \in \mathbb{R}^k$ from (20) and (21) (with $\tilde{\alpha}$, Σ_{TL} and $\text{diag}(q)$ replacing α, Σ and Q in those equations). It is easy to see that these solutions vary continuously with q, so we can write $u = f(q)$ for a continuous function f. This allows us to define the following system of ordinary differential equations with a right boundary condition

[27] Rewriting $(\Sigma_{TL})^{-1}$ as the Schur complement of Σ^{-1} with respect to its bottom-right block, the result follows from the fact that M-matrices are closed under Schur complements. This fact can be proved by the same induction argument as in Carlson and Markham (1979).

(where the derivative at x is interpreted as the left-derivative).

$$q'(y) = f(q(y)) \text{ for every } y \in (x-\epsilon, x], \text{ and } q(x) = q^x.$$

By Peano's Existence Theorem, this system of ODE admits a solution when ϵ is sufficiently small. Note that by construction, for any y in the interval $(x-\epsilon, x]$, we have $q_1(y) = y$, $\sum_{i=1}^K q_i(y) = t$ and $q_i(y) = 0$ for $i > k$. Moreover, at the vector $q(y)$ sources $2 \sim k$ have equal marginal values, which are maximal if ϵ is sufficiently small. Hence $q(y)$ satisfies the Kuhn-Tucker conditions for minimizing $V(q)$ subject to $q_1 \geq y$, $q_i \geq 0$ and $\sum_{i=1}^K q_i = t$. Since V is convex, these conditions are sufficient, so that $q(y)$ coincides with q^y. But then we see that q^y is left-differentiable for any $y > x - \epsilon$, contradicting the definition of x.

This completes the entire proof of Proposition 3.

O.6 Proof of Proposition 4

O.6.1 Reduction to the Main Model

We first show it is without loss to consider $\sigma_b = 1$. Suppose the result holds for $\sigma_b = 1$, then for a general σ_b, we can write $b = \sigma_b \cdot \hat{b}$ and $\omega = \sigma_b \cdot \hat{\omega}$ for some random variables $\hat{b}$ and $\hat{\omega}$, where $\hat{b}$ has unit variance. Note that the two signals $\omega + \phi_1 b + \mathcal{N}(0, \zeta_1^2)$ and $\omega - \phi_2 b + \mathcal{N}(0, \zeta_2^2)$ are informationally equivalent to $\hat{\omega} + \phi_1 \hat{b} + \mathcal{N}\left(0, (\frac{\zeta_1}{\sigma_b})^2\right)$ and $\hat{\omega} - \phi_2 \hat{b} + \mathcal{N}\left(0, (\frac{\zeta_2}{\sigma_b})^2\right)$. We have thus transformed the general problem into one with payoff-relevant state $\hat{\omega}$ and unknown benefit $\hat{b}$, where $\hat{b}$ has unit variance. The result for this case then pins down equilibrium choices of ϕ_i and $\frac{\zeta_i}{\sigma_b}$, which then yield the equilibrium in the general case.

Hence, for the rest of the proof, we assume $\sigma_b = 1$. Define $\theta_1 = \frac{1}{\zeta_1}(\omega + \phi_1 b)$ and $\theta_2 = \frac{1}{\zeta_2}(\omega - \phi_2 b)$. Observe that $\omega + \phi_1 b + \mathcal{N}(0, \zeta_1^2)$ is informationally equivalent to $\frac{1}{\zeta_1}(\omega + \phi_1 b) + \mathcal{N}(0, 1)$. Thus a unit of time spent on source i produces a standard normal signal about the corresponding θ_i, which returns our main model. The prior covariance matrix for (θ_1, θ_2) is

$$\Sigma = \begin{pmatrix} \frac{\sigma_\omega^2 + \phi_1^2}{\zeta_1^2} & \frac{\sigma_\omega^2 - \phi_1 \phi_2}{\zeta_1 \zeta_2} \\ \frac{\sigma_\omega^2 - \phi_1 \phi_2}{\zeta_1 \zeta_2} & \frac{\sigma_\omega^2 + \phi_2^2}{\zeta_2^2} \end{pmatrix}$$

and the payoff-relevant state can be written as $\omega = \alpha_1 \theta_1 + \alpha_2 \theta_2$ with payoff weights $\alpha_1 = \zeta_1 \cdot \frac{\phi_2}{\phi_1 + \phi_2}$ and $\alpha_2 = \zeta_2 \cdot \frac{\phi_1}{\phi_1 + \phi_2}$.

Below we derive the reader's optimal attention allocation using our main results. Throughout we assume $\zeta_1 \leq \zeta_2$, so that source 1 is more precise. Under this assumption, we have $cov_1 = \sigma_\omega^2 / \zeta_1 \geq \sigma_\omega^2 / \zeta_2 = cov_2$, which implies that source 1 is attended to in the first stage. Moreover, by

Theorem 1 (which applies since cov_1 and cov_2 are both positive), the length of the first stage is

$$t_1^* = \frac{cov_1 - cov_2}{\alpha_2 \det(\Sigma)} = \frac{\zeta_1(\zeta_2 - \zeta_1)}{\phi_1(\phi_1 + \phi_2)}, \tag{22}$$

where we used $\det(\Sigma) = \sigma_\omega^2 \left(\frac{\phi_1+\phi_2}{\zeta_1\zeta_2}\right)^2$.

Thus, the reader optimally attends only to source 1 until time t_1^*, and afterwards gives $\frac{\alpha_1}{\alpha_1+\alpha_2}$ fraction of his attention to source 1. We can then write the two sources' payoffs as follows:

$$\begin{aligned} U_1 &= \int_0^{t_1^*} re^{-rt}dt + \int_{t_1^*}^{\infty} re^{-rt}\left(\frac{\zeta_1\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1}\right)dt - \lambda(1-\phi_1)^2 \\ &= 1 - e^{-rt_1^*}\left(\frac{\zeta_2\phi_1}{\zeta_1\phi_2 + \zeta_2\phi_1}\right) - \lambda(1-\phi_1)^2; \\ U_2 &= e^{-rt_1^*}\left(\frac{\zeta_2\phi_1}{\zeta_1\phi_2 + \zeta_2\phi_1}\right) - \lambda(1-\phi_2)^2. \end{aligned} \tag{23}$$

These payoffs define a stage game between the two sources, and our goal is to characterize its equilibrium. Our strategy below is to use first-order conditions to pin down what an equilibrium must be. We will then verify the equilibrium by checking all possible deviations.

O.6.2 Solving for Equilibrium Precisions ζ_1^*, ζ_2^*

We show here that the precision choices ζ_1, ζ_2 must be equal in any equilibrium. Suppose not, then small changes in ζ_1 and ζ_2 do not affect our standing assumption that $\zeta_1 \le \zeta_2$. Thus we can take the first-order conditions for ζ_1 and ζ_2. Observe that $\frac{\partial t_1^*}{\partial \zeta_1} = \frac{\zeta_2 - 2\zeta_1}{\phi_1(\phi_1+\phi_2)}$. So we can compute that

$$\frac{\partial U_1}{\partial \zeta_1} = e^{-rt_1^*}\left(\frac{\zeta_2\phi_1}{\zeta_1\phi_2 + \zeta_2\phi_1}\right)\left(r\frac{\zeta_2 - 2\zeta_1}{\phi_1(\phi_1+\phi_2)} + \frac{\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1}\right).$$

The FOC then requires that

$$r\frac{2\zeta_1 - \zeta_2}{\phi_1(\phi_1+\phi_2)} = \frac{\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1}. \tag{24}$$

Now consider the FOC for source 2. Observe that $\frac{\partial t_1^*}{\partial \zeta_2} = \frac{\zeta_1}{\phi_1(\phi_1+\phi_2)}$. So

$$\frac{\partial U_2}{\partial \zeta_2} = e^{-rt_1^*}\left(\frac{\zeta_1\phi_1}{\zeta_1\phi_2 + \zeta_2\phi_1}\right)\left(-r\frac{\zeta_2}{\phi_1(\phi_1+\phi_2)} + \frac{\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1}\right).$$

The FOC then requires that

$$r\frac{\zeta_2}{\phi_1(\phi_1+\phi_2)} = \frac{\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1}. \tag{25}$$

(24) and (25) together imply

$$r\frac{2\zeta_1 - \zeta_2}{\phi_1(\phi_1+\phi_2)} = r\frac{\zeta_2}{\phi_1(\phi_1+\phi_2)},$$

which simplifies to $\zeta_1 = \zeta_2$ and leads to a contradiction.

Hence $\zeta_1 = \zeta_2$ must hold in equilibrium. In this case the first-order conditions derived above need not hold with equality, because the payoffs in (23) are derived under the assumption that $\zeta_1 \le \zeta_2$, so that the same payoff expressions apply only to downward deviations of ζ_1 and upward deviations of ζ_2. Given this, the first-order conditions become inequalities $\frac{\partial U_1}{\partial \zeta_1} \ge 0$ and $\frac{\partial U_2}{\partial \zeta_2} \le 0$ (evaluated at the equilibrium choices). These translate into the following inequality versions of (24) and (25):

$$r\frac{2\zeta_1 - \zeta_2}{\phi_1(\phi_1+\phi_2)} \le \frac{\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1};$$
$$r\frac{\zeta_2}{\phi_1(\phi_1+\phi_2)} \ge \frac{\phi_2}{\zeta_1\phi_2 + \zeta_2\phi_1}.$$

Since we already know $\zeta_1 = \zeta_2$, the two inequalities above must both hold equal, and we further deduce that

$$\zeta_1 = \zeta_2 = \sqrt{\frac{\phi_1\phi_2}{r}}. \tag{26}$$

Note also that given $\zeta_2 = \sqrt{\frac{\phi_1\phi_2}{r}}$, choosing ζ_1 to be any smaller number cannot be profitable for source 1. This is because as ζ_1 decreases, the term $r\frac{\zeta_2 - 2\zeta_1}{\phi_1(\phi_1+\phi_2)} + \frac{\phi_2}{\zeta_1\phi_2+\zeta_2\phi_1}$ appearing in $\frac{\partial U_1}{\partial \zeta_1}$ increases and remains positive. So the choice $\zeta_1 = \sqrt{\frac{\phi_1\phi_2}{r}}$ is robust to any downward deviation (in this variable). Similarly, given this ζ_1, choosing any larger ζ_2 is not profitable for source 2. By symmetry, source 1 (respectively source 2) also cannot profit from upward (respectively downward) deviations in precision.

O.6.3 Solving for Equilibrium Biases ϕ_1^*, ϕ_2^*

We now fix precision choices and characterize equilibrium levels of bias. Since $\zeta_1 = \zeta_2$ in equilibrium, we have $t_1^* = 0$, meaning that there is no stage 1. Hence the two sources' payoffs simplify to

$$\begin{aligned} U_1 &= \frac{\phi_2}{\phi_1+\phi_2} - \lambda(1-\phi_1)^2; \\ U_2 &= \frac{\phi_1}{\phi_1+\phi_2} - \lambda(1-\phi_2)^2. \end{aligned} \tag{27}$$

In this smaller game, we will show that there is a (unique) pure strategy equilibrium if and only if $\lambda \ge \frac{9}{16}$, in which case the equilibrium involves $\phi_1 = \phi_2 = \frac{1}{2}\left(1 + \sqrt{1 - \frac{1}{2\lambda}}\right)$.

The first-order conditions $\frac{\partial U_i}{\partial \phi_i} = 0$ give

$$2\lambda(1-\phi_1) = \frac{\phi_2}{(\phi_1+\phi_2)^2}; \qquad 2\lambda(1-\phi_2) = \frac{\phi_1}{(\phi_1+\phi_2)^2}. \tag{28}$$

In addition, the second-order conditions $\frac{\partial^2 U_i}{\partial \phi_i^2} \le 0$ give

$$2\lambda \ge \frac{2\phi_2}{(\phi_1+\phi_2)^3}; \qquad 2\lambda \ge \frac{2\phi_1}{(\phi_1+\phi_2)^3}. \tag{29}$$

Comparing each equality in (28) with the corresponding one in (29) yields

$$2(1-\phi_1) \le \phi_1 + \phi_2; \qquad\qquad 2(1-\phi_2) \le \phi_1 + \phi_2. \tag{30}$$

Moreover, multiplying the two equalities in (28) yields $(1-\phi_1)\phi_1 = (1-\phi_2)\phi_2$. So either $\phi_1 = \phi_2$ or $\phi_1 + \phi_2 = 1$. In the former case (30) implies $\phi_1 = \phi_2 \ge \frac{1}{2}$. In the latter case (30) implies $\phi_1 = \phi_2 = \frac{1}{2}$. Thus we always have $\phi_1 = \phi_2 \ge \frac{1}{2}$.

Now since $\phi_1 = \phi_2$, we can solve from (28) that ϕ_1 satisfies $4\phi_1(1-\phi_1) = \frac{1}{2\lambda}$, which is equivalent to $(2\phi_1 - 1)^2 = 1 - \frac{1}{2\lambda}$. Using $\phi_1 \ge \frac{1}{2}$, we deduce that the only possible equilibrium is $\phi_1 = \phi_2 = \frac{1}{2}\left(1 + \sqrt{1 - \frac{1}{2\lambda}}\right)$, which we denote by ϕ^*. Clearly, for this solution to be a real number, a necessary condition is $\lambda \ge \frac{1}{2}$.

Below we show these choices are an equilibrium in this smaller game (involving only ϕ_1, ϕ_2) if and only if $\lambda \ge \frac{9}{16}$. Indeed, with these choices source 1's payoff is $\frac{1}{2} - \lambda(1-\phi^*)^2$. This must be higher than choosing ϕ_1 close to 0, which would yield a payoff close to $1-\lambda$. Thus we can deduce the inequality $\phi^*(2-\phi^*) \ge \frac{1}{2\lambda}$. But recall that ϕ^* satisfies the equation $4\phi^*(1-\phi^*) = \frac{1}{2\lambda}$. So we deduce that $2 - \phi^* \ge 4(1-\phi^*)$, or $\phi^* \ge \frac{2}{3}$. It follows that $\frac{1}{2\lambda} = 4\phi^*(1-\phi^*) \le \frac{8}{9}$, which implies $\lambda \ge \frac{9}{16}$ as a necessary condition.

Conversely, suppose $\lambda \ge \frac{9}{16}$ holds. We need to show that

$$f(\phi_1) = \frac{\phi^*}{\phi_1 + \phi^*} - \lambda(1-\phi_1)^2$$

is maximized at $\phi_1 = \phi^*$. Since we have derived the solution via the first- and second-order conditions, it holds that $f'(\phi^*) = 0$ and $f''(\phi^*) \le 0$. Moreover, note that $f'''(\phi_1) = \frac{-6\phi^*}{(\phi_1+\phi^*)^4} < 0$, so $f'(\phi_1)$ is a strictly concave function. The fact that $f'(\phi^*) = 0$ and $f''(\phi^*) \le 0$ thus imply that $f'(\phi_1) \le 0$ for all $\phi_1 \ge \phi^*$. Hence $f(\phi_1)$ is decreasing for $\phi_1 \ge \phi^*$.

On the other hand, since $f'(\phi_1)$ is concave and $f'(\phi^*) = 0$, there are two possibilities for the behavior of f' on the interval $[0, \phi^*]$: Either $f'(0) \ge 0$ and thus f' is non-negative on this whole interval, or $f'(0) < 0$ and f' crosses zero exactly once from below. This means f is either increasing on $[0, \phi^*]$, or first decreasing and then increasing. Hence the maximum of f on this interval must occur at the extreme points. When $\lambda \ge \frac{9}{16}$, we have $\phi^* = \frac{1}{2}\left(1 + \sqrt{1 - \frac{1}{2\lambda}}\right) \ge \frac{2}{3}$. Thus

$$f(\phi^*) = \frac{1}{2} - \lambda(1-\phi^*)^2 \ge \frac{1}{2} - \frac{1}{9}\lambda \ge 1 - \lambda = f(0).$$

This shows that the function f is maximized at ϕ^* whenever $\lambda \ge \frac{9}{16}$.

O.6.4 Verifying the Equilibrium

Summarizing the above analysis, we have shown that the only possible pure strategy equilibrium is $\phi_1 = \phi_2 = \phi^*$ and $\zeta_1 = \zeta_2 = \frac{\phi^*}{\sqrt{r}}$, where the latter follows from (26). We have also shown that if

$\lambda \geq \frac{9}{16}$, then given source 2's equilibrium choices, source 1 does not have a profitable deviation in ζ_1 or in ϕ_1 alone. However, without further assumptions, it is possible for source 1 to profit from choosing ζ_1 and ϕ_1 both away from the target equilibrium. To illustrate, let $\lambda = \frac{9}{16}$ and $r = 1$. Then source 2's equilibrium choices are $\phi_2 = \zeta_2 = \phi^* = \frac{2}{3}$. By also choosing $\phi_1 = \zeta_1 = \frac{2}{3}$, source 1 obtains payoff $\frac{1}{2} - \lambda(1-\phi^*)^2 = \frac{7}{16} = 0.4375$. Suppose instead source 1 chooses $\phi_1 = \frac{1}{6}$ and $\zeta_1 = \frac{1}{3}$. Then $t_1^* = \frac{\zeta_1(\zeta_2-\zeta_1)}{\phi_1(\phi_1+\phi_2)} = \frac{4}{5}$, and source 1 receives long-run attention $\frac{\zeta_1\phi_2}{\zeta_1\phi_2+\zeta_2\phi_1} = \frac{2}{3}$. In this case source 1's payoff is higher:

$$1 - \frac{1}{3}e^{-\frac{4}{5}} - \frac{9}{16}(1-\frac{1}{6})^2 \approx 0.4596.$$

So the target equilibrium would not be an equilibrium when double deviations are considered.

However, we will show that when $\lambda \geq 1.6$, then given $\phi_2 = \phi^*$ and $\zeta_2 = \frac{\phi^*}{\sqrt{r}}$, source 1 cannot profitably deviate to any ϕ_1 and $\zeta_1 \leq \frac{\phi^*}{\sqrt{r}}$. And given $\phi_1 = \phi^*$ and $\zeta_1 = \frac{\phi^*}{\sqrt{r}}$, source 2 cannot profitably deviate to any ϕ_2 and $\zeta_2 \geq \frac{\phi^*}{\sqrt{r}}$. Thanks to the symmetry of the target equilibrium, verifying these will be sufficient to show it is indeed an equilibrium.

First consider source 1. We have the elementary inequality

$$e^{-rt_1^*} \geq 1 - rt_1^* = \frac{\phi_1(\phi_1+\phi_2) - r\zeta_1(\zeta_2-\zeta_1)}{\phi_1(\phi_1+\phi_2)},$$

which is tight if and only if $\zeta_1 = \zeta_2$ and $t_1^* = 0$. Thus from (23) we have

$$U_1 \leq 1 - \frac{\phi_1(\phi_1+\phi_2) - r\zeta_1(\zeta_2-\zeta_1)}{\phi_1(\phi_1+\phi_2)} \cdot \frac{\zeta_2\phi_1}{\zeta_1\phi_2+\zeta_2\phi_1} - \lambda(1-\phi_1)^2.$$

Plugging in $\phi_2 = \phi^*$ and $\zeta_2 = \frac{\phi^*}{\sqrt{r}}$, the above simplifies to

$$\begin{aligned} U_1 &\leq 1 - \frac{\phi_1(\phi_1+\phi^*) - \sqrt{r}\zeta_1\phi^* + r\zeta_1^2}{(\phi_1+\phi^*)(\phi_1+\sqrt{r}\zeta_1)} - \lambda(1-\phi_1)^2 \\ &= \frac{(\phi_1+2\phi^*-\sqrt{r}\zeta_1)\sqrt{r}\zeta_1}{(\phi_1+\phi^*)(\phi_1+\sqrt{r}\zeta_1)} - \lambda(1-\phi_1)^2. \end{aligned}$$

We now optimize this upper-bound over ζ_1, and then over ϕ_1. Let $y = \phi_1 + \sqrt{r}\zeta_1$. Note that the range of y is $y \in [\phi_1, \phi_1+\phi^*]$, since $\sqrt{r}\zeta_1 \leq \sqrt{r}\zeta_2 = \phi^*$. Using y, we can rewrite the above inequality as

$$U_1 \leq \frac{(2\phi_1+2\phi^*-y)(y-\phi_1)}{(\phi_1+\phi^*)y} - \lambda(1-\phi_1)^2. \tag{31}$$

Thus, in terms of y, we would like to maximize

$$\frac{(2\phi_1+2\phi^*-y)(y-\phi_1)}{y} = -y - \frac{2(\phi_1+\phi^*)\phi_1}{y} + 2\phi^* + 3\phi_1$$

This is a single-peaked function in y, with global maximum occurring at $y = \sqrt{2(\phi_1 + \phi^*)\phi_1} \geq \phi_1$. We distinguish two cases. If $\sqrt{2(\phi_1 + \phi^*)\phi_1} \geq \phi_1 + \phi^*$, then the maximum of this function of y over the interval $[\phi_1, \phi_1 + \phi^*]$ occurs at $y = \phi_1 + \phi^*$. Plugging back into (31), we have

$$U_1 \leq \frac{\phi^*}{\phi_1 + \phi^*} - \lambda(1 - \phi_1)^2.$$

But in previous analysis we have already shown that the above function of ϕ_1 is maximized at ϕ^* whenever $\lambda \geq \frac{9}{16}$, so we are done in this case.

The more difficult case is $y = \sqrt{2(\phi_1 + \phi^*)\phi_1} \leq \phi_1 + \phi^*$, which corresponds to $\phi_1 \leq \phi^*$. Here the global maximum is achievable on the interval $y \in [\phi_1, \phi_1 + \phi^*]$, so we instead deduce from (31) the following:

$$U_1 \leq \frac{3\phi_1 + 2\phi^* - 2\sqrt{2(\phi_1 + \phi^*)\phi_1}}{\phi_1 + \phi^*} - \lambda(1 - \phi_1)^2.$$

We denote the function on the RHS as $g(\phi_1)$, and aim to show $g(\phi_1) \leq g(\phi^*)$ for all $\phi_1 \leq \phi^*$. Note that

$$g'(\phi_1) = \frac{\phi^*}{(\phi_1 + \phi^*)^2} \cdot \left(1 - \sqrt{\frac{2(\phi_1 + \phi^*)}{\phi_1}}\right) + 2\lambda(1 - \phi_1).$$

Thus $g'(\phi^*) = \frac{-1}{4\phi^*} + 2\lambda(1 - \phi^*) = 0$. Moreover, we claim that g' is concave on the interval $[0, \phi^*]$. To see this, let us write

$$\phi^* g'(\phi_1) = \frac{1}{(\frac{\phi_1}{\phi^*} + 1)^2} \cdot \left(1 - \sqrt{2 + \frac{2\phi^*}{\phi_1}}\right) + 2\phi^*\lambda(1 - \phi_1).$$

The second term $2\phi^*\lambda(1 - \phi_1)$ is linear in ϕ_1 and thus does not affect convexity/concavity. As for the first term, we can rewrite it as

$$h(z) = \frac{1}{(1 + z)^2} \cdot \left(1 - \sqrt{2 + \frac{2}{z}}\right)$$

with $z = \frac{\phi_1}{\phi^*}$. It thus suffices to show h is concave on the interval $[0, 1]$. For this we compute that

$$h''(z) = -3\frac{-8\sqrt{z^5(z+1)} + 8\sqrt{2}z^3 + 12\sqrt{2}z^2 + 5\sqrt{2}z + \sqrt{2}}{4z^{2.5}(z+1)^{4.5}},$$

which is negative for $z \leq 1$ because $8\sqrt{z^5(z+1)} \leq 8\sqrt{2z^5} \leq 8\sqrt{2}z^2$.

Hence $g'(\phi^*) = 0$ and g' is concave on $[0, \phi^*]$. It follows that either g' is non-negative on this whole interval, or it is first negative then positive. So g is either increasing, or first decreasing and then increasing. It thus remains to show $g(0) \leq g(\phi^*)$, which reduces to $2 - \lambda \leq \frac{1}{2} - \lambda(1 - \phi^*)^2$. Since $\lambda \geq 1.6$, we have $\phi^* = \frac{1}{2}\left(1 + \sqrt{1 - \frac{1}{2\lambda}}\right) > \frac{3}{4}$. Thus we indeed have

$$2 - \lambda \leq \frac{1}{2} - \frac{1}{16}\lambda < \frac{1}{2} - \lambda(1 - \phi^*)^2,$$

where the first inequality again uses the assumption $\lambda \geq 1.6$. This completes the proof that source 1 does not have any profitable deviation.

We now turn to source 2's incentives. We use another elementary inequality

$$e^{rt_1^*} \geq 1 + rt_1^* = 1 + r\frac{\zeta_1(\zeta_2 - \zeta_1)}{\phi_1(\phi_1 + \phi_2)}.$$

Given $\phi_1 = \phi^*$ and $\zeta_1 = \frac{\phi^*}{\sqrt{r}}$, we can simplify the above as

$$e^{rt_1^*} \geq 1 + \frac{\sqrt{r}(\zeta_2 - \zeta_1)}{\phi_1 + \phi_2} = \frac{\phi_2 + \sqrt{r}\zeta_2}{\phi_2 + \phi^*}.$$

So from (23) we obtain an upper bound for U_2:

$$\begin{aligned} U_2 &= e^{-rt_1^*}\left(\frac{\zeta_2\phi_1}{\zeta_1\phi_2 + \zeta_2\phi_1}\right) - \lambda(1-\phi_2)^2 \\ &\leq \frac{\phi_2 + \phi^*}{\phi_2 + \sqrt{r}\zeta_2} \cdot \frac{\sqrt{r}\zeta_2}{\phi_2 + \sqrt{r}\zeta_2} - \lambda(1-\phi_2)^2 \\ &= (\phi_2 + \phi^*)\frac{\sqrt{r}\zeta_2}{(\phi_2 + \sqrt{r}\zeta_2)^2} - \lambda(1-\phi_2)^2 \end{aligned} \tag{32}$$

Again we will optimize this upper bound first over ζ_2, then over ϕ_2. In terms of ζ_2, we want to minimize

$$\frac{(\phi_2 + \sqrt{r}\zeta_2)^2}{\sqrt{r}\zeta_2} = \sqrt{r}\zeta_2 + \frac{\phi_2^2}{\sqrt{r}\zeta_2} + 2\phi_2.$$

The global minimum of this single-dipped function occurs at $\sqrt{r}\zeta_2 = \phi_2$, but since $\sqrt{r}\zeta_2 \geq \sqrt{r}\zeta_1 = \phi^*$, there are two cases to consider. In the first case, $\phi_2 \leq \phi^*$. Then the minimum subject to $\sqrt{r}\zeta_2 \geq \phi^*$ occurs precisely at $\sqrt{r}\zeta_2 = \phi^*$. In this case the upper bound (32) becomes $U_2 \leq \frac{\phi^*}{\phi_2 + \phi^*} - \lambda(1-\phi_2)^2$, and as we have shown previously this is maximized at $\phi_2 = \phi^*$.

In the other case, we have $\phi_2 \geq \phi^*$. Then the optimal ζ_2 that maximizes the upper bound (32) is $\sqrt{r}\zeta_2 = \phi_2$. (32) then becomes

$$U_2 \leq \frac{\phi_2 + \phi^*}{4\phi_2} - \lambda(1-\phi_2)^2.$$

We will show that the derivative of this function of ϕ_2 is negative for any $\phi_2 \geq \phi^*$, so that $\phi_2 = \phi^*$ is again the optimal choice. This comparison reduces to $2\lambda(1-\phi_2) \leq \frac{\phi^*}{4\phi_2^2}$, which is equivalent to $8\lambda(1-\phi_2)\phi_2^2 \leq \phi^*$. Since equality obtains when $\phi_2 = \phi^*$, we just need to show $8\lambda(1-\phi_2)\phi_2^2 \leq 8\lambda(1-\phi^*)(\phi^*)^2$. After factoring out $\phi_2 - \phi^*$, this inequality reduces to

$$\phi_2 + \phi^* \leq \phi_2^2 + \phi_2\phi^* + (\phi^*)^2.$$

This holds whenever $\phi^* \geq \frac{2}{3}$, since $\phi_2 \geq \phi^*$ implies $\phi_2^2 + \phi_2\phi^* + (\phi^*)^2 \geq \frac{3}{2}\phi^*(\phi_2 + \phi^*)$. Hence whenever $\lambda \geq \frac{9}{16}$, source 2 does not have profitable deviations either.