Online Appendix for "Uncertain Search with Knowledge Transfer"

Woonghee Tim Huh, Michael Jong Kim, Meichun Lin

Appendix A Proofs for the Finite-Alternative Problem (Section 3.1)

Proof of Proposition 1. We use the decomposition of the value function in (6) to show the properties. Specifically, by adapting the argument in Lippman and McCardle (1991), we obtain that $J(\omega; s)$ is increasing and convex in ω with $\frac{\partial}{\partial\omega}J(\omega; s) \in [0, 1]$. Since $\omega = \beta - \eta - v_i(\tau)$, we have by (6) that $V_i(\beta, \eta; s, \tau)$ is convex in β and η with derivatives satisfying

$$\frac{\partial}{\partial\beta}V_i(\beta, \eta; s, \tau) = \frac{\partial}{\partial\omega}J(\omega; s) \in [0, 1], \quad \text{and} \quad \frac{\partial}{\partial\eta}V_i(\beta, \eta; s, \tau) = -\frac{\partial}{\partial\omega}J(\omega; s) + 1 \in [0, 1],$$

which proves the desired results. □

Proof of Proposition 2. We first consider the posterior precision s. By the decomposition of the value function in (6), s affects the value function only through $J(\omega; s)$. We use an inductive argument to show that $J(\omega; s)$ is decreasing in s. (This is claimed in Theorem 2 of Lippman and McCardle (1987) without proof. Here we provide the proof for completeness.) We note that there exists $\bar{s} < \infty$ such that $J(\omega; s) = \max\{\omega, 0\}$ for any $s \geq \bar{s}$ and any $\omega \in \mathbb{R}$, implying that the monotonicity holds for $s \geq \bar{s}$. To see that, consider the value of one more sample in the single-alternative problem with a switching payoff of 0. The marginal gain from one more sample is $\mathbb{E}[\max\{\omega + Z\sigma_s, 0\}] - \max\{\omega, 0\}$, where $\sigma_s = \sqrt{\frac{s_x}{s(s+s_x)}}$. By sending $s \to \infty$, the marginal gain for any value of ω converges to 0, which is strictly less than the sampling cost c. This implies that there is no value of observing one more sample when s is large. Thus, there exists $\bar{s} < \infty$ such that for any $s \geq \bar{s}$, there is no value of continuing sampling, and $J(\omega; s) = \max\{\omega, 0\}$.

Assume by the way of induction that $J(\omega; s)$ is decreasing in s for any $s \geq \bar{s}_1$ and some $\bar{s}_1 \leq \bar{s}$. Then we consider the problem with any precision $s \in [\min\{0, \bar{s}_1 - s_x\}, \bar{s}_1)$. Specifically, let $\min\{0, \bar{s}_1 - s_x\} \leq s_1 < s_2 \leq \bar{s}_1$, and we want to show $J(\omega; s_1) \geq J(\omega; s_2)$ based on (3). By the inductive assumption, we derive that

$$\mathbb{E}_Z\left[J(\omega + Z\sigma_{s_1};\ s_1 + s_x)\right] \ \geq\ \mathbb{E}_Z\left[J(\omega + Z\sigma_{s_1};\ s_2 + s_x)\right] \ \geq\ \mathbb{E}_Z\left[J(\omega + Z\sigma_{s_2};\ s_2 + s_x)\right],$$

where the last inequality holds because $J(\cdot; s)$ is an increasing and convex function, and $\omega + Z\sigma_{s_1}$ is a mean-preserving spread of $\omega + Z\sigma_{s_2}$. (More specifically, $\omega + Z\sigma_{s_2}$ second-order stochastically dominates $\omega+Z\sigma_{s_1}$, which satisfies $\mathbb{E}_Z[u(\omega+Z\sigma_{s_2})] \leq \mathbb{E}_Z[u(\omega+Z\sigma_{s_1})]$ for any increasing and convex function $u(\cdot)$ (see, e.g., Hanoch and Levy 1975)). Since monotonicity is preserved under maximum, we have by (3) that $J(\omega; s_1) \geq J(\omega; s_2)$, which proves that $J(\omega; s)$ is decreasing in s. This implies that the value function is decreasing in s. For the threshold pair, we have by (3) that $\overline{\omega}(s) = \inf\{\omega : \omega = -c + \mathbb{E}_Z\left[J(\omega + Z\sigma_s;\ s+s_x)\right]\}$ and $\underline{\omega}(s) = \sup\{\omega : -c + \mathbb{E}_Z\left[J(\omega + Z\sigma_s;\ s+s_x)\right] = 0\}$. The inductive argument above implies that $\mathbb{E}_Z\left[J(\omega + Z\sigma_s;\ s+s_x)\right]$ is decreasing in s. It follows that $\overline{\omega}(s)$ is decreasing in s, and $\underline{\omega}(s)$ is increasing in s. Thus, Part (a) follows from (8).

We proceed to consider the hyperposterior precision τ. Similar to η, we can show that $V_i(\beta, \eta; s, \tau)$ is increasing in $v_i(\tau)$ (using an argument similar to the proof of Proposition 1). Moreover, by (6) and (8), τ affects $V_i(\beta, \eta; s, \tau)$, $\overline{\delta}_i(s, \tau)$, and $\underline{\delta}_i(s, \tau)$ only through $v_i(\tau)$. To prove the effect of τ, it suffices to show that $v_i(\tau)$ is increasing in τ. Recall that $v_2(\tau) = -k$ is increasing in τ. Assume by the way of induction that $v_{i-1}(\tau)$ is increasing in τ for some $i > 2$. Then let $0 < \tau_1 < \tau_2$, and we want to show $v_i(\tau_1) \leq v_i(\tau_2)$. By (7), we derive that

$$\begin{aligned} v_i(\tau_1) &= -k + \mathbb{E}_Z\left[J\left(Z\sigma_{\tau_1} - v_{i-1}(\tau_1+1);\ s_0\right)\right] + v_{i-1}(\tau_1+1) \\ &\leq -k + \mathbb{E}_Z\left[J\left(Z\sigma_{\tau_1} - v_{i-1}(\tau_2+1);\ s_0\right)\right] + v_{i-1}(\tau_2+1) \\ &\leq -k + \mathbb{E}_Z\left[J\left(Z\sigma_{\tau_2} - v_{i-1}(\tau_2+1);\ s_0\right)\right] + v_{i-1}(\tau_2+1) \ = \ v_i(\tau_2), \end{aligned}$$

where the first inequality follows from the inductive assumption and the fact that $v_i(\tau)$ is increasing in $v_{i-1}(\tau+1)$, and the last inequality holds because $Z\sigma_{\tau_2} - v_{i-1}(\tau_2+1)$ is a mean-preserving spread of $Z\sigma_{\tau_1} - v_{i-1}(\tau_2+1)$, where the former is preferred by the increasing and convex function $J(\cdot; s)$. This shows that $v_i(\tau)$ is increasing in τ, implying the monotonicities with respect to τ. $\square$

Proof of Proposition 3. We start with properties of the number of alternatives left i and the switching cost k in (a). Similar to the effect of τ, we only need to show that $v_i(\tau)$ is increasing in i while decreasing in k, as implied by the decomposition of the value function in (6) and the definition of the threshold pair in (8). By the definition in (7), $v_i(\tau)$ is decreasing in k. As for i, we prove the monotonicity by induction. Recall that $v_3(\tau) = -k + \mathbb{E}_Z\left[V_2\left(Z\sigma_\tau,\ 0;\ s_0,\ \tau+1\right)\right] \geq -k + \mathbb{E}_Z[Z\sigma_\tau] = -k = v_2(\tau)$. Next, we assume by the way of induction that $v_i(\tau) \geq v_{i-1}(\tau)$ for some $i-1 \geq 2$. Then we consider the case of i. Using the inductive assumption, we derive that

$$\begin{aligned} v_{i+1}(\tau) &= -k + \mathbb{E}_Z\left[V_i\left(Z\sigma_\tau,\ 0;\ s_0,\ \tau+1\right)\right] \\ &\geq -k + \mathbb{E}_Z\left[V_{i-1}\left(Z\sigma_\tau,\ 0;\ s_0,\ \tau+1\right)\right] \ = \ v_i(\tau), \quad \forall\, \tau > 0, \end{aligned}$$

which proves the monotonicity of $v_i(\tau)$ with respect to i.

Next, we consider the sample precision s_x and sampling cost c in (b). They affect the estimated payoffs associated with not only the current alternative but also alternatives that come in the future. More specifically, they affect $J(\cdot;s)$ and $v_i(\tau)$ in the decomposition of the value function in (6). Based on the definitions, we can show inductively that $J(\cdot;s)$ and $v_i(\tau)$ are increasing in s_x while decreasing in c. By Proposition 1, the value function is increasing in $v_i(\tau)$. It follows that $V_i(\beta,\eta;s,\tau)$ is increasing in s_x while decreasing in c. For the upper threshold $\overline{\delta}_i(s,\tau)=\overline{\omega}(s)+v_i(\tau)$, we note that $\overline{\omega}(s)=\inf\{\omega:\omega=-c+\mathbb{E}_Z\left[J(\omega+Z\sigma_s;\ s+s_x)\right]\}$ is increasing in s_x while decreasing in c. Thus, $\overline{\delta}_i(s,\tau)$ is also increasing in s_x while decreasing in c. For the length of the continuation interval, it satisfies $\overline{\delta}_i(s,\tau)-\underline{\delta}_i(s,\tau)=\overline{\omega}(s)-\underline{\omega}(s)$. Since $\underline{\omega}(s)=\sup\{\omega:-c+\mathbb{E}_Z\left[J(\omega+Z\sigma_s;\ s+s_x)\right]=0\}$ is decreasing in s_x while increasing in c, we obtain that the length is increasing in s_x while decreasing in c, which completes the proof. ☐

Appendix B Proofs for the Infinite-Alternative Problem (Section 3.2)

Proof of Theorem 2. We first show $V_\infty(y)=\lim_{i\to\infty}V_i(y)$ for any state $y=(\beta,\eta;s,\tau)$. Then the Bellman equations follow by sending $i\to\infty$ on both sides of (2). Recall that $\{V_i(y)\}_{i=1}^{\infty}$ is increasing in i, as shown in Proposition 3. It follows that $\{V_i(y)\}_{i=1}^{\infty}$ converges (we will show later that the limit is finite). Let $\bar{V}(y):=\lim_{i\to\infty}V_i(y)$ for any state y, and let $R_i^\pi(y)$ denote the total expected payoff under a policy π for a group of i alternatives. Then for any policy $\pi=(\pi_{1,0},\ldots)$, we have $R_i^\pi(y)\le V_i(y)$ for any i, and thus, $\liminf_{i\to\infty}R_i^\pi(y)\le\liminf_{i\to\infty}V_i(y)=\bar{V}(y)$. Thus, $V_\infty(y)=\sup_\pi\liminf_{i\to\infty}R_i^\pi(y)\le\bar{V}(y)$. On the other hand, for the optimal policy π_0 that achieves $V_i(y)$, it must accept some alternative in $\{i,\ldots,1\}$, since it is always optimal to accept the last alternative $i=1$. As such, π_0 defines a policy for the infinite-alternative problem, which implies $V_\infty(y)\ge R_i^{\pi_0}(y)=V_i(y)$. Since this inequality holds for any i, we obtain $V_\infty(y)\ge\lim_{i\to\infty}V_i(y)=\bar{V}(y)$. Combining the results above, we obtain $V_\infty(y)=\bar{V}(y)$ as desired.

Next, we show that $V_\infty(\beta,\eta;s,\tau)=\lim_{i\to\infty}V_i(\beta,\eta;s,\tau)$ is finite for any state $(\beta,\eta;s,\tau)$. Recall that for a finite number of alternatives, we denote by i the number of alternatives left (including the currently explored one), and from (6), the value function $V_i(\beta,\eta;s,\tau)$ for $i>1$ can be divided into three terms as follows:

$$V_i(\beta,\eta;s,\tau)\ =\ J\left(\omega;\ s\right)+\eta+v_i(\tau),$$

where $\omega=\beta-\eta-v_i(\tau)$, and $J(\omega;\ s)$ is the value function of a single-alternative problem with posterior $\mathcal{N}(\omega,s)$ and a switching payoff of 0, as given in (3). Note that the number of alternatives only impacts the term $v_i(\tau)$, which is given by the recursive equations in (7). The sequence $\{v_i(\tau)\}_{i=2}^{\infty}$ also converges as it is increasing in i (as shown in the proof of Proposition 3(a)). Thus, it suffices to show that $\bar{v}(\tau):=\lim_{i\to\infty}v_i(\tau)$ is finite. We claim that $v_i(\tau)$ for any $i>1$ and any $\tau>0$ is bounded by a finite constant $\bar{v}_\infty$, which is the unique solution to (10). To prove the claim, we adopt an inductive argument and use the recursive equations in (7). Specifically, recall that

$v_2(\tau) = -k$ for any $\tau > 0$. By definition of $\bar{v}_\infty$, we have $k = \mathbb{E}_Z\left[J\left(Z - \bar{v}_\infty;\ s_0\right)\right] \geq \mathbb{E}_Z\left[Z - \bar{v}_\infty\right] = -\bar{v}_\infty$. Thus, $v_2(\tau) = -k \leq \bar{v}_\infty$, which proves the case of $i = 2$. Suppose by the way of induction that $v_i(\tau) \leq \bar{v}_\infty$ for any $\tau > 0$ and some $i \geq 2$. Then consider $i + 1$, and we have by (7) that

$$v_{i+1}(\tau) \ = -k + \mathbb{E}_Z\left[J\left(Z - v_i(\tau+1);\ s_0\right)\right] + v_i(\tau+1) \ \leq -k + \mathbb{E}_Z\left[J\left(Z - \bar{v}_\infty;\ s_0\right)\right] + \bar{v}_\infty \ = \bar{v}_\infty,$$

where the inequality follows from the inductive assumption and $\frac{\partial}{\partial v}\left\{\mathbb{E}_Z\left[J\left(Z - v;\ s_0\right)\right] + v\right\} = \mathbb{E}_Z\left[\frac{\partial}{\partial v} J\left(Z - v;\ s_0\right)\right] + 1 \geq 0$. This proves our claim that $v_i(\tau)$ is bounded, implying $\bar{v}(\tau) = \lim_{i\to\infty} v_i(\tau) \leq \bar{v}_\infty$. Thus, $V_\infty(\beta, \eta; s, \tau) = \lim_{i\to\infty} V_i(\beta, \eta; s, \tau)$ is finite, which completes our proof. ☐

Proof of Proposition 4. We use properties of $v_i(\tau)$ established in the proof of Proposition 2 to prove the results. Since $v_i(\tau)$ is increasing in τ for any i, $\bar{v}(\tau) = \lim_{i\to\infty} v_i(\tau)$ is also increasing in τ. Since $v_i(\tau)$ is increasing in τ and bounded by $\bar{v}_\infty$, $\bar{v}(\tau)$ is also increasing in τ and bounded by $\bar{v}_\infty$. This implies that $\bar{v}(\tau)$ converges as $\tau \to \infty$. The recursive equations in (12) follow by sending $i \to \infty$ in (7). Moreover, by sending $\tau \to \infty$ in (12), we obtain that $\lim_{\tau\to\infty} \bar{v}(\tau)$ satisfies (10), and thus, $\lim_{\tau\to\infty} \bar{v}(\tau) = \bar{v}_\infty$.

Now we show that $\bar{v}(\tau)$ is continuous in $\tau > 0$. Since $v_i(\tau)$ is continuous in τ, it suffices to show that $v_i(\tau)$ uniformly converges to $\bar{v}(\tau)$ (by the uniform limit theorem). By definition, we need to prove that for any $\varepsilon > 0$, there exists a finite constant I_0 such that for any $i \geq I_0$, $|\bar{v}(\tau) - v_i(\tau)| < \varepsilon$ for any $\tau > 0$. Since $v_i(\tau)$ is increasing in i and $\bar{v}(\tau) = \lim_{i\to\infty} v_i(\tau)$, $|\bar{v}(\tau) - v_i(\tau)| = \bar{v}(\tau) - v_i(\tau)$. In addition, we have $\lim_{\tau\to\infty} \bar{v}(\tau) = \bar{v}_\infty$, which implies that there exists a finite constant τ_0 such that $\bar{v}_\infty - \bar{v}(\tau) < \frac{1}{2}\varepsilon$ for any $\tau \geq \tau_0$. We will then bound $\bar{v}(\tau) - v_i(\tau)$ for $\tau \geq \tau_0$ and $\tau < \tau_0$, separately. For $\tau \geq \tau_0$, we use $\lim_{i\to\infty} v_i(\tau_0) = \bar{v}(\tau_0)$ and $v_i(\tau_0)$ increasing in i. It follows that there exists a finite constant i_0 such that $\bar{v}(\tau_0) - v_i(\tau_0) < \frac{1}{2}\varepsilon$ for any $i \geq i_0$. Based on the analysis above, we can derive that for any $\tau \geq \tau_0$ and any $i \geq i_0$,

$$\bar{v}(\tau) - v_i(\tau) \ \leq \ \bar{v}_\infty - v_i(\tau_0) \ = \ \left[\bar{v}_\infty - \bar{v}(\tau_0)\right] + \left[\bar{v}(\tau_0) - v_i(\tau_0)\right] \ < \ \varepsilon.$$

Next, we consider the case of $0 < \tau < \tau_0$. For any $0 < \tau < \tau_0$, let $i_\tau := \lceil \tau_0 - \tau \rceil$. Using the recursive equations in (7) and (12), we obtain

$$\begin{aligned}\bar{v}(\tau) - v_i(\tau) &= \mathbb{E}_Z\left[J\left(Z\sigma_\tau - \bar{v}(\tau+1);\ s_0\right)\right] + \bar{v}(\tau+1) - \mathbb{E}_Z\left[J\left(Z\sigma_\tau - v_{i-1}(\tau+1);\ s_0\right)\right] - v_{i-1}(\tau+1)\\ &\leq \ \bar{v}(\tau+1) - v_{i-1}(\tau+1) \ \leq \ \cdots \ \leq \ \bar{v}(\tau+i_\tau) - v_{i-i_\tau}(\tau+i_\tau),\end{aligned}$$

where the first inequality holds because $\frac{\partial}{\partial \omega} J(\omega; s_0) \in [0,1]$ implies $\frac{\partial}{\partial v}\left\{\mathbb{E}_Z\left[J\left(Z\sigma_\tau - v;\ s_0\right)\right] + v\right\} \in [0,1]$. It follows that $\bar{v}(\tau) - v_i(\tau) \leq \bar{v}(\tau+i_\tau) - v_{i-i_\tau}(\tau+i_\tau) < \varepsilon$ for any $i - i_\tau \geq i_0$. Let $I_0 := \lceil \tau_0 \rceil + i_0$. Since $i_\tau \leq \lceil \tau_0 \rceil$, any $i \geq I_0$ satisfies $i - i_\tau \geq i_0$. Thus, $\bar{v}(\tau) - v_i(\tau) < \varepsilon$ for any $0 < \tau < \tau_0$ and any $i \geq I_0$. Combining the two cases above, we can conclude that $\bar{v}(\tau) - v_i(\tau) < \varepsilon$ for any $i \geq I_0$, which proves that $v_i(\tau)$ uniformly converges to $\bar{v}(\tau)$.

Finally, we show that $\bar{v}(\tau)$ is the unique solution to (12). Suppose by the way of contradiction that another function $g(\tau)$ satisfies (12). Then $g(\tau_0) \neq \bar{v}(\tau_0)$ for some $\tau_0 > 0$. It follows that

$$\begin{aligned}|\bar{v}(\tau_0) - g(\tau_0)| &= |[\mathbb{E}_Z [J (Z\sigma_\tau - \bar{v}(\tau_0+1);\ s_0)] + \bar{v}(\tau_0+1)] - [\mathbb{E}_Z [J (Z\sigma_\tau - g(\tau_0+1);\ s_0)] - g(\tau_0+1)]| \\ &\leq\ |\bar{v}(\tau_0+1) - g(\tau_0+1)|\ \leq\ \cdots\ \leq\ |\bar{v}(\tau_0+K) - g(\tau_0+K)|,\end{aligned}$$

where the first inequality holds because $\frac{\partial}{\partial\omega}J(\omega; s_0) \in [0,1]$, and thus the terms in the first and second square bracket have partial derivatives in $\bar{v}(\tau)$ and $g(\tau)$ less than or equal to 1, respectively. However, (12) implies that $\bar{v}(\tau)$ and $g(\tau)$ converge to $\bar{v}_\infty$ as $\tau \to \infty$. It follows that $|\bar{v}(\tau_0) - g(\tau_0)| \leq \lim_{K\to\infty} |\bar{v}(\tau_0+K) - g(\tau_0+K)| = 0$. Thus, $\bar{v}(\tau_0) = g(\tau_0)$, and we reach a contradiction. Hence, $\bar{v}(\tau)$ is the unique solution to (12). This completes our proof. □

Proof of Proposition 5. Let T be the stopping time at which the DM stops and accepts the alternative in search. We note that T is a random variable due to the uncertainty in the observed utilities and noisy samples of individual values. Each type of observation follows its underlying distribution, which is in contrast to the posterior distribution used before in the Bayesian dynamic program. Decisions are made according to the optimal policy characterized by the threshold pair in (13). We then consider the expected stopping time, where the expectation is taken over all potential observations of utilities and samples. In particular, we will show that $\mathbb{E}[T]$ is bounded, which implies that T is finite almost surely.

We first bound the number of sampling for each alternative. Recall that there is no value of sampling the current alternative when the posterior precision exceeds some constant $\bar{s} < \infty$ independent of the posterior mean. It follows that the DM will stop sampling and make a final decision about j before the alternative has been sampled $N = \lceil \frac{1}{s_x}(\bar{s} - s_0) \rceil$ times. Recall that J_0 is the number of alternatives that have been searched until acceptance. Then the expected stopping time can be bounded by

$$\mathbb{E}(T)\ \leq\ \sum_{j=1}^{\infty}(1+N)j \cdot \mathbb{P}(J_0 = j)\ =\ (1+N)\cdot\mathbb{E}[J_0].$$

Thus, we only need to bound $\mathbb{E}[J_0]$, the expected number of alternatives that have been searched until acceptance.

Next, we bound $\mathbb{E}[J_0]$. By the tail sum formula, we have $\mathbb{E}[J_0] = \sum_{j=0}^{\infty} \mathbb{P}(J_0 > j)$. Intuitively, $\mathbb{P}(J_0 > j)$, the probability of rejecting the past j alternatives, decreases as the number j increases. For $\mathbb{E}[J_0]$ to be bounded, we need this probability to decrease at a fast rate. This is not obvious since the switching region expands as j increases and the DM knows more about the population (i.e., τ increases). To show that, we consider how the decision of rejecting an alternative depends on past observations. This enables us to establish a recursive bound on $\mathbb{P}(J_0 > j)$. Recall that the DM starts with a hyperprior $\mathcal{N}(\eta_0, \tau_0)$ and updates the hyperprior/hyperposterior as more alternatives are explored and their utilities are observed (which are i.i.d. from $\mathcal{N}(\mu, 1)$). After observing utilities $\{U_\ell\}_{\ell=1}^{j}$, the hyperposterior is updated as $\mathcal{N}\left(\frac{\eta_0\tau_0+\sum_{\ell=1}^{j} U_\ell}{\tau_0+j}, \tau_0 + j\right)$. Following the optimal policy, the DM would accept the j-th alternative if the mean difference is above the upper threshold given in (13). This implies that the DM would not accept the j-th alternative immediately as long as $U_j - \frac{\eta_0\tau_0+\sum_{\ell=1}^{j} U_\ell}{\tau_0+j} < \overline{\omega}(s_0) + \bar{v}(\tau_0 + j)$. Since rejecting the j-th alternative implies rejecting the past $j - 1$ alternative and not accepting j immediately, we obtain

$$\begin{aligned}
\mathbb{P}(J_0 > j) \le\ & \mathbb{P}\left(\left\{U_j - \frac{\eta_0\tau_0 + \sum_{\ell=1}^{j} U_\ell}{\tau_0 + j} < \overline{\omega}(s_0) + \bar{v}(\tau_0 + j)\right\} \cap \{J_0 > j - 1\}\right) \\
= \mathbb{P}&\left(\left\{U_j - \mu < \frac{(\eta_0 - \mu)\tau_0 + \sum_{\ell=1}^{j-1}(U_\ell - \mu)}{\tau_0 + j - 1} + \frac{\tau_0 + j}{\tau_0 + j - 1}\left[\overline{\omega}(s_0) + \bar{v}(\tau_0 + j)\right]\right\} \cap \{J_0 > j - 1\}\right).
\end{aligned}$$

For any small $\varepsilon > 0$, let $E_0 := \left\{\frac{1}{j-1}\sum_{\ell=1}^{j-1}(U_\ell - \mu) \le \varepsilon\right\}$. If E_0 occurs, the event in the first curly bracket above implies $U_j - \mu < A_j$, where $A_j := \frac{(\eta_0-\mu)\tau_0+(j-1)\varepsilon}{\tau_0+j-1} + \frac{\tau_0+j}{\tau_0+j-1}\left[\overline{\omega}(s_0) + \bar{v}(\tau_0 + j)\right]$. Denote by E_0^C the complementary event of E_0. Considering two cases depending on whether E_0 occurs, we can derive that

$$\begin{aligned}
\mathbb{P}(J_0 > j) \le\ & \mathbb{P}\left(\{U_j - \mu < A_j\} \cap \{J_0 > j - 1\} \cap E_0\right) + \mathbb{P}\left(E_0^C\right) \\
\le\ & \mathbb{P}\left(\{U_j - \mu < A_j\} \cap \{J_0 > j - 1\}\right) + \mathbb{P}\left(E_0^C\right) \\
=\ & \Phi(A_j) \cdot \mathbb{P}\left(J_0 > j - 1\right) + \left[1 - \Phi(\varepsilon\sqrt{j-1})\right],
\end{aligned}$$

where $\Phi(\cdot)$ is the standard Normal CDF, and the last equality holds because $\{U_\ell\}_{\ell=1}^{j}$ are i.i.d. from $\mathcal{N}(\mu, 1)$ and $\frac{1}{\sqrt{j-1}}\sum_{\ell=1}^{j-1}(U_\ell - \mu) \sim \mathcal{N}(0, 1)$. For the second term above, we can show that $1 - \Phi(\varepsilon\sqrt{j-1}) \le \phi(\varepsilon\sqrt{j-1})/(\varepsilon\sqrt{j-1}) \le \phi(\varepsilon\sqrt{j-1})/\varepsilon$ for any $j > 1$, where $\phi(\cdot)$ is the standard Normal PDF, and the first inequality follows from $1 - \Phi(z) = \int_z^\infty \phi(x)\,dx \le \int_z^\infty (x/z)\phi(x)\,dx = \phi(z)/z$. Since $A_j \to \varepsilon + \overline{\omega}(s_0) + \bar{v}_\infty$ as $j \to \infty$, there exists a finite constant $\tilde{J}$ such that $A_j \le 2\varepsilon + \overline{\omega}(s_0) + \bar{v}_\infty$ holds for any $j > \tilde{J}$. Let $p_\varepsilon := \Phi(2\varepsilon + \overline{\omega}(s_0) + \bar{v}_\infty) < 1$. It follows that for any $j \ge \tilde{J}$,

$$\begin{aligned}
\mathbb{P}(J_0 > j) \le\ & p_\varepsilon \cdot \mathbb{P}\left(J_0 > j - 1\right) + \frac{1}{\varepsilon}\phi(\varepsilon\sqrt{j-1}) \\
\le\ & p_\varepsilon^{j-\tilde{J}} \cdot \mathbb{P}\left(J_0 > \tilde{J}\right) + p_\varepsilon^{j-\tilde{J}-1} \cdot \frac{1}{\varepsilon}\phi(\varepsilon\sqrt{\tilde{J}}) + \cdots + p_\varepsilon \cdot \frac{1}{\varepsilon}\phi(\varepsilon\sqrt{j-2}) + \frac{1}{\varepsilon}\phi(\varepsilon\sqrt{j-1}) \\
=\ & p_\varepsilon^{j-\tilde{J}} \cdot \mathbb{P}\left(J_0 > \tilde{J}\right) + \sum_{\ell=0}^{j-\tilde{J}} p_\varepsilon^{\ell} \cdot \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left(-\frac{1}{2}(j - 1 - \ell)\varepsilon^2\right),
\end{aligned}$$

where the last inequality follows from recursively applying the first inequality. Note that the above inequality holds for $j = \tilde{J}$. Also, if we choose ε sufficiently small, we have $p_\varepsilon \exp\left(\frac{1}{2}\varepsilon^2\right) < 1$. Thus, the summation above can be bounded, which yields

$$\begin{aligned}\mathbb{P}(J_0 > j) \leq\ & p_\varepsilon^{j-\tilde{J}} \cdot \mathbb{P}\left(J_0 > \tilde{J}\right) + \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left(-\frac{1}{2}(j-1)\varepsilon^2\right) \sum_{\ell=0}^{j-\tilde{J}} p_\varepsilon^\ell \cdot \exp\left(\frac{1}{2}\ell\varepsilon^2\right) \\ \leq\ & p_\varepsilon^{j-\tilde{J}} + \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left(-\frac{1}{2}(j-1)\varepsilon^2\right) \cdot \frac{1}{1 - p_\varepsilon \exp\left(\frac{1}{2}\varepsilon^2\right)}, \quad \forall\, j \geq \tilde{J}.\end{aligned}$$

Let $c_\varepsilon := \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left(\frac{1}{2}\varepsilon^2\right) \cdot \frac{1}{1-p_\varepsilon \exp\left(\frac{1}{2}\varepsilon^2\right)}$. It follows that $\mathbb{P}(J_0 > \tilde{J}) \leq p_\varepsilon^{j-\tilde{J}} + c_\varepsilon \exp\left(-\frac{1}{2}j\varepsilon^2\right)$ for any $j \geq \tilde{J}$.

We have shown that $\mathbb{P}(J_0 > j)$ decreases exponentially when j is large. Recall that $\mathbb{E}[T] \leq (1+N)\mathbb{E}[J_0] = (1+N)\sum_{j=0}^\infty \mathbb{P}(J_0 > j)$. Combining the above bound with $\mathbb{P}(J_0 > j) \leq 1$ for $j < \tilde{J}$, we obtain

$$\begin{aligned}\frac{\mathbb{E}[T]}{1+N} \leq\ & \tilde{J} + \sum_{j=\tilde{J}}^{\infty} \left[p_\varepsilon^{j-\tilde{J}} + c_\varepsilon \exp\left(-\frac{1}{2}j\varepsilon^2\right)\right] \\ =\ & \tilde{J} + \sum_{\ell=0}^{\infty} p_\varepsilon^\ell + c_\varepsilon \sum_{j=\tilde{J}}^{\infty} \exp\left(-\frac{1}{2}j\varepsilon^2\right) \ = \ \tilde{J} + \frac{1}{1-p_\varepsilon} + c_\varepsilon \frac{\exp\left(-\frac{1}{2}\tilde{J}\varepsilon^2\right)}{1 - \exp\left(-\frac{1}{2}\varepsilon^2\right)}.\end{aligned}$$

Thus, the expected stopping time is bounded, and the stopping time T is finite almost surely. This completes the proof. □

Appendix C Proofs for Problem with More Knowledge Transfer (Section 4)

Proof of Theorem 3. When $i = 1$, we always have $W_i(\omega; s, \tau) = \omega$. As in the base model, the thresholds are $\underline{\delta}_i^W(s,\tau) = \overline{\delta}_i^W(s,\tau) = -\infty$.

Now we consider $i > 1$. Denote by $w_i^S(\tau)$ and $w_i^C(\omega; s, \tau)$ the payoffs of continuation and switching, respectively. Then by (16), the optimal decisions can be found by solving $W_i(\omega; s, \tau) = \max\{\omega, w_i^C(\omega; s, \tau), w_i^S(\tau)\}$. We note that there exists $\bar{s} < \infty$ such that $W_i(\omega; s, \tau) = \max\{\omega, w_i^S(\tau)\}$ for any $s \geq \bar{s}$, $\omega \in \mathbb{R}$, and $\tau > 0$. This is because the value of collecting one more sample for the current alternative is $-c + \mathbb{E}[\max\{\omega + Z\sigma_s \cdot \frac{\tau_{i,0}}{\tau_{i,0}+1}, w_i^S(\tau + \tau_x)\}] - \max\{\omega, w_i^S(\tau)\}$, which converges to $-c < 0$ as $s \to \infty$ (since $\sigma_s, \tau_x \to 0$ as $s \to \infty$). Then we can show by induction on s that $W_i(\omega; s, \tau)$ is increasing and convex in ω with $\frac{\partial}{\partial\omega} W_i(\omega; s, \tau) \in [0, 1]$. It follows that the payoff of continuation satisfies $\frac{\partial}{\partial\omega} w_i^C(\omega; s, \tau) \in [0, 1]$. Thus, if $w_i^C(\omega; s, \tau) \leq \max\{\omega, w_i^S(\tau)\}$ at $\omega = w_i^S(\tau)$, the payoff of continuation is always smaller than the other two, and the two thresholds are $\underline{\delta}_i^W(s,\tau) = \overline{\delta}_i^W(s,\tau) = w_i^S(\tau)$. Otherwise, $w_i^C(\omega; s, \tau) > \max\{\omega, w_i^S(\tau)\}$ at $\omega = w_i^S(\tau)$, and the two thresholds are different, as given by

$$\underline{\delta}_i^W(s,\tau) = \inf\left\{\omega : w_i^S(\tau) = w_i^C(\omega; s,\tau)\right\} \quad \text{and} \quad \overline{\delta}_i^W(s,\tau) = \sup\left\{\omega : \omega = w_i^C(\omega; s,\tau)\right\}. \quad \text{(C.1)}$$

We note that the first set above can be empty if $w_i^S(\tau) \leq \lim_{\omega\to-\infty} w_i^C(\omega; s,\tau) = -c + w_i^S(\tau+\tau_x)$, and we set $\underline{\delta}_i^W(s,\tau) = -\infty$. By contrast, $\overline{\delta}_i^W(s,\tau) < \infty$, since $\omega - w_i^C(\omega; s,\tau) \to c > 0$ as $\omega \to \infty$. This completes the proof. □

Proof of Proposition 6. We start with Part (a). Similar to the function $J(\cdot)$ used in the base model, we can show by induction on s that $W_i(\omega; s,\tau)$ is increasing and convex in ω and satisfies $\frac{\partial}{\partial\omega}W_i(\omega; s,\tau) \in [0,1]$. Recall that $\tilde{V}_i(\beta,\eta; s,\tau) = W_i(\beta-\eta; s,\tau)+\eta$. It follows that $\frac{\partial}{\partial\beta}\tilde{V}_i(\beta,\eta; s,\tau) = \frac{\partial}{\partial\omega}W_i(\beta-\eta; s,\tau) \geq 0$ and $\frac{\partial}{\partial\eta}\tilde{V}_i(\beta,\eta; s,\tau) = -\frac{\partial}{\partial\omega}W_i(\beta-\eta; s,\tau)+1 \geq 0$. Thus, $\tilde{V}_i(\beta,\eta; s,\tau)$ is increasing in β and η.

For Part (b) and (c), we also adopt an inductive argument to show that $W_i(\omega; s,\tau)$ is decreasing in s while increasing in τ. When $i = 1$, we have $W_i(\omega; s,\tau) = \omega$, and (b) and (c) hold. Assume by the way of induction that (b) and (c) hold for some $i-1$. We proceed to consider the case of i. As discussed in the proof of Theorem 3, $W_i(\omega; s,\tau) = \max\{\omega, w_i^S(\tau)\}$ for any $s \geq \bar{s}$. We claim that $w_i^S(\tau)$ is increasing in τ. Then (b) and (c) hold for $s \geq \bar{s}$. Assume by induction that the properties hold for any $s \geq \tilde{s}$ and some $\tilde{s}$. Then we consider any $s \in [\min\{0, \tilde{s}-s_x\}, \tilde{s})$. Recall that the payoff of continuation is

$$w_i^C(\omega; s,\tau) \;=\; -c + \mathbb{E}_Z\left[W_i\left(\omega + Z\sigma_s \cdot \frac{\tau_{i,0}}{\tau_{i,0}+1};\; s+s_x, \tau+\tau_x\right)\right],$$

where the right hand side is increasing in $\sigma_s \cdot \frac{\tau_{i,0}}{\tau_{i,0}+1}$ (as this yields a mean-preserved spread for an increasing and convex function). By the inductive assumption, the function $W_i(\cdot)$ is decreasing in the second argument while increasing in the third argument. We note that σ_s is decreasing in s and independent of τ, while $\tau_{i,0}$ and τ_x are decreasing in s and increasing in τ (as defined in (15)). It follows that $w_i^C(\omega; s,\tau)$ is decreasing in s while increasing in τ. Note that the payoff of switching $w_i^S(\tau)$ is independent of s. If the claim holds, we have $W_i(\omega; s,\tau)$ decreasing in s while increasing in τ for any $s \in [\min\{0, \tilde{s}-s_x\}, \tilde{s})$, which completes the inductive proof for (b) and (c).

Now we prove the claim that $w_i^S(\tau)$ is increasing in τ. By definition, the payoff of switching is

$$w_i^S(\tau) \;=\; -k + \mathbb{E}_Z\left[W_{i-1}(Z\tilde{\sigma}_\tau;\; \frac{\tau}{\tau+1}+s_x,\; \tau+\frac{s_x}{s_x+1})\right],$$

where the function $W_i(\cdot)$ is decreasing in the second argument while increasing in the third argument. As τ increases, the second and third arguments increase, resulting in opposite effects. To analyze the combined effect, we consider an alternate formulation. Recall that there is a one-to-one correspondence between the pairs (s, τ) and $(\tau_{i,0}, n)$, where the variable $\tau_{i,0}$ is the hyperposterior precision when the DM switches to the current alternative i, and n is the number of samples that have been observed for i. Using the state variables $(i, \omega; \tau_{i,0}, n)$, we have the following alternate formulation of the value function:

$$\begin{aligned}\tilde{W}_i(\omega; \tau_{i,0}, n) \ &:= \ \max\Big\{\omega,\ -c + \mathbb{E}_Z\left[\tilde{W}_i\left(\omega + Z\sigma_s \cdot \frac{\tau_{i,0}}{\tau_{i,0}+1};\ \tau_{i,0},\ n+1\right)\right], \\ &\qquad -k + \mathbb{E}_Z\left[\tilde{W}_{i-1}\left(Z\tilde{\sigma}_\tau;\ \tau + \frac{s_x}{s_x+1},\ 0\right)\right]\Big\}, \end{aligned} \tag{C.2}$$

where the precisions at the current stage are given by $s = \frac{\tau_{i,0}}{\tau_{i,0}+1} + ns_x$ and $\tau = \tau_{i,0} + \frac{ns_x}{ns_x+1}$. Thus, we obtain $w_i^S(\tau) = -k + \mathbb{E}_Z\left[\tilde{W}_{i-1}(Z\tilde{\sigma}_\tau;\ \tau + \frac{s_x}{s_x+1},\ 0)\right]$. Note that the right hand side is increasing in $\tilde{\sigma}_\tau$, and $\tilde{\sigma}_\tau = \sqrt{\frac{s_x}{1+s_x\left(1+\frac{1}{\tau}\right)}}$ is increasing in τ. If we can show $\tilde{W}_{i-1}(\cdot)$ is increasing in the second argument, then we have $w_i^S(\tau)$ increasing in τ. To prove that, we use the formulation in (C.2) and show inductively that $\tilde{W}_\ell(\omega; \tau_{\ell,0}, n)$ is increasing in $\tau_{\ell,0}$ for any ℓ (including $\ell = i-1$).

For $\ell = 1$, we have $\tilde{W}_\ell(\omega; \tau_{j,0}, n) = \omega$ increasing in $\tau_{\ell,0}$. Assume by the way of induction that the property holds for some $\ell - 1$. We proceed to the case of ℓ. Since $s = \frac{\tau_{\ell,0}}{\tau_{\ell,0}+1} + ns_x$, we have for any $n \geq \bar{s}/s_x$ and any $\tau_{\ell,0} > 0$ that

$$\tilde{W}_\ell(\omega; \tau_{\ell,0}, n) \ = \ \max\left\{\omega, -k + \mathbb{E}_Z\left[\tilde{W}_{\ell-1}\left(Z\tilde{\sigma}_\tau;\ \tau + \frac{s_x}{s_x+1},\ 0\right)\right]\right\}.$$

The right hand side above is increasing in $\tilde{\sigma}_\tau$ (since it yields a mean-preserved spread for an increasing and convex function) and $\tau + \frac{s_x}{s_x+1}$ (by the inductive assumption). Since $\tau = \tau_{\ell,0} + \frac{ns_x}{ns_x+1}$ and $\tilde{\sigma}_\tau = \sqrt{\frac{s_x}{1+s_x\left(1+\frac{1}{\tau}\right)}}$ are increasing in $\tau_{\ell,0}$, we have $\tilde{W}_\ell(\omega; \tau_{\ell,0}, n)$ increasing in $\tau_{\ell,0}$ for any $n \geq \bar{s}/s_x$. Assume by induction that this property holds for some $n+1$, and we want to show for the case of n. Compared to the case of $n \geq \bar{s}/s_x$, we need to show the monotonicity for the payoff of continuation, which is given by $-c+\mathbb{E}_Z\left[\tilde{W}_\ell\left(\omega + Z\sigma_s \cdot \frac{\tau_{\ell,0}}{\tau_{\ell,0}+1};\ \tau_{\ell,0},\ n+1\right)\right]$. This payoff is increasing in $\sigma_s \cdot \frac{\tau_{\ell,0}}{\tau_{\ell,0}+1}$ and the second argument. Recall that

$$\sigma_s \cdot \frac{\tau_{\ell,0}}{\tau_{\ell,0}+1} \ = \ \sqrt{\frac{s_x}{s(s+s_x)}} \cdot \frac{\tau_{\ell,0}}{\tau_{\ell,0}+1} \ = \ \sqrt{\frac{s_x}{\left(1 + ns_x\left(1 + \frac{1}{\tau_{\ell,0}}\right)\right)\left(1 + (n+1)s_x\left(1 + \frac{1}{\tau_{\ell,0}}\right)\right)}},$$

which is also increasing in $\tau_{\ell,0}$. Thus, the payoff of continuation is increasing in $\tau_{\ell,0}$. This completes the induction that shows $\tilde{W}_\ell(\omega; \tau_{\ell,0}, n)$ is increasing in $\tau_{\ell,0}$. By setting $\ell = i-1$, we obtain that $w_i^S(\tau) = -k + \mathbb{E}_Z\left[\tilde{W}_{i-1}(Z\tilde{\sigma}_\tau;\ \tau + \frac{s_x}{s_x+1},\ 0)\right]$ is increasing in τ, which proves our claim. This completes the proof. □

Appendix D Numerical Experiments

We conduct numerical experiments based on a real-world dataset to illustrate the value of the two levels of learning. In particular, we use a speed dating dataset introduced in Fisman et al. (2006) and apply our model to solve a dating problem. We consider a finite group of alternatives with the size $I \in \{3\ell\}_{\ell=1}^{10}$. Data are used to estimate the utility distribution $N(\mu, \tau_u)$, the population-level prior $N(\eta_0, \tau_0)$, and the prior on each idiosyncratic value $N(0, r_0)$. (Note that in our base model, we assume without loss of generality that $\tau_u = 1$, yet our model can be extended to incorporate any given utility precision $\tau_u > 0$.) The sample precision s_x and the sampling cost c are chosen such that the DM only samples a few times. The switching cost k is set as twice the sampling cost. Details of the dataset and estimation can be found in Appendix D.1.

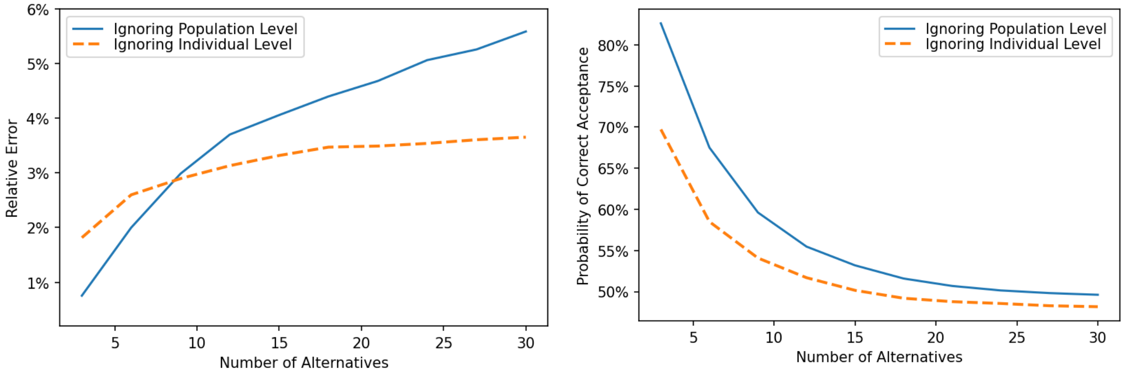
To shed light on the two levels of learning, we consider policies that ignore one level and compare their performance to the optimal policy given by our Bayesian dynamic program. The policy that ignores the population level of learning keeps using the hyperprior to estimate the utility of each alternative. The policy that ignores the individual level of learning ignores learning of the idiosyncratic values and uses the prior mean to estimate them. As such, there is no sampling for each alternative, and the DM accepts or rejects an alternative based on its utility observation. Each one of the two policies can be found by solving a corresponding dynamic program.

In Figure D.1, we present the performance of the policies that ignore one level of learning. We consider two performance measures. The first one is the relative error, which is defined as $1 - \frac{R^\pi}{R^*}$, where R^π and R^* are the total expected payoffs under a policy π and our optimal policy, respectively. We note that the total expected payoff of accepting the i-th alternative is given by $\mathbb{E}[\theta_i - \sum_{j=1}^{i}(k + cN_j)]$, where N_j is the number of sampling for the j-th alternative given by the policy. The second measure is the probability of correct acceptance compared to our optimal policy. This is a performance measure typically studied in the literature of the secretary problem. It can be viewed as the expectation of a binary payoff, i.e., $\mathbb{E}[\mathbb{1}\{i^\pi = i^*\}]$, where i^π and i^* are the alternatives accepted by a policy π and our optimal policy. We generate 100 decision makers, each with 200 groups of alternatives and 1000 trajectories of sample observations. Then we take the average to approximate the expectations.

In Figure D.1(a), we plot the relative errors versus the number of alternatives. We observe that as the number of alternatives increases, the relative errors for both policies grow, and their performance becomes worse. This is intuitive because it becomes harder to identify the "best" alternative, and ignoring one level of learning incurs a higher cost. When the number of alternative is small, ignoring the individual level performs worse. When the number of alternative is large, however, ignoring the population level performs worse. This is because there is more value of knowledge transfer within a larger group. In Figure D.1(b), we present the probability of correct acceptance versus the number of alternatives. We also observe that as the number of alternatives increases, the probabilities decrease, and both policies perform worse, as identifying the "best" alternative becomes harder in a larger group. In addition, we find that ignoring the individual level always has a smaller probability of correct acceptance than ignoring the population level. Nevertheless, it may have a smaller relative error since it ignores the option of sampling and saves some costs. We note that ignoring the individual level can have a better probability of correct acceptance when the sample precision is small and learning the individual has less value. We refer to Appendix D.2 for an example.

Figure D.1: Performance of Policies that Ignore One Level of Learning.

(a) Relative Error (b) Probability of Correct Acceptance



We also vary the utility and sample precisions to illustrate how the performance changes with the information provided by each observation at the two levels. In Figure D.2, we plot the relative errors for both policies that ignore one level of learning under different values of the utility precision τ_u. As τ_u increases, the utility observation provides more information, and it becomes worse to ignore the population level of learning, as shown in Figure D.2(a). Interestingly, ignoring the individual level also performs worse as τ_u increases, which is shown in Figure D.2(b). This may be because more information improves both payoffs of the optimal policy and the policy that ignores the individual level of learning, while the former improves more significantly. We note that when $\tau_u = 100$, each utility observation is highly informative, and the corresponding relative error can be viewed approximately as an upper bound.

In Figure D.3, we plot the relative error under different values of the sample precision s_x. As shown in Figure D.3(b), ignoring the individual level of learning performs worse as s_x increases and sampling provides more information, which is intuitive. However, ignoring the population level may perform better as s_x increases, as shown in Figure D.3(a). More information from sampling improves both payoffs of the optimal policy and the policy that ignores the population level of learning, but it is unclear which improvement is more significant.

Figure D.2: Relative Error Under Different Utility Precisions.

(a) Ignoring Population Level (b) Ignoring Individual Level

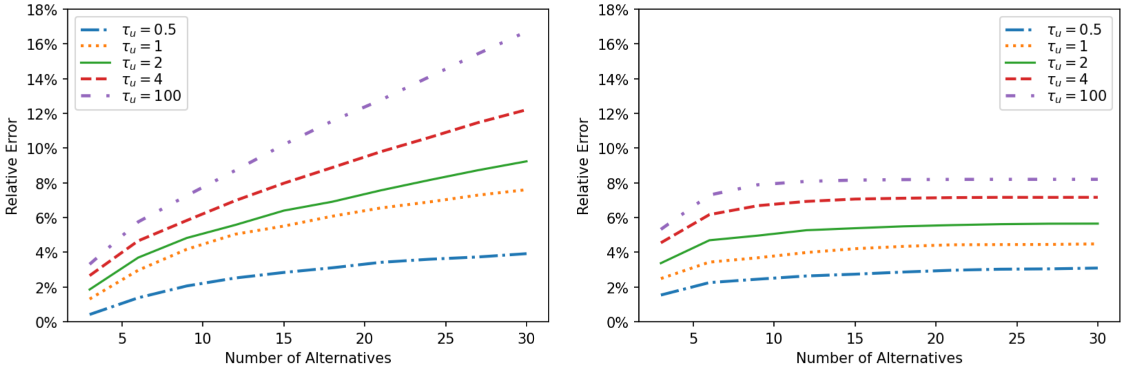
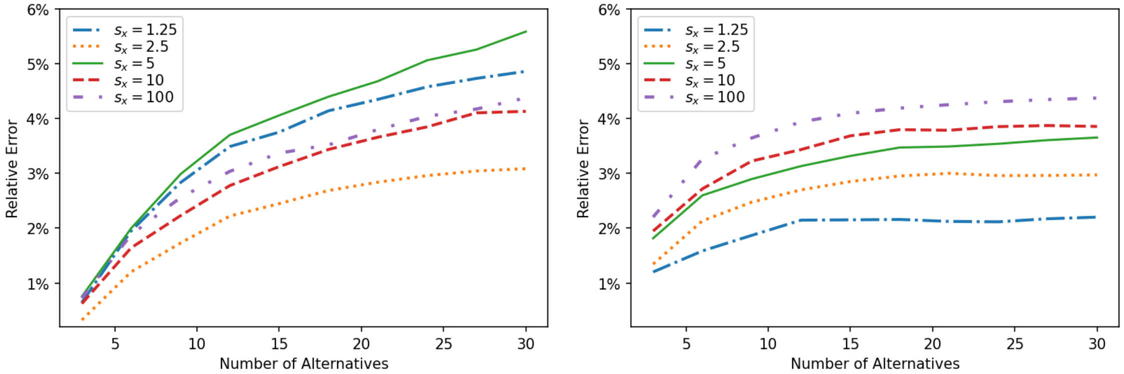


Figure D.3: Relative Error Under Different Sample Precisions.

(a) Ignoring Population Level (b) Ignoring Individual Level



D.1 Dataset and Estimation

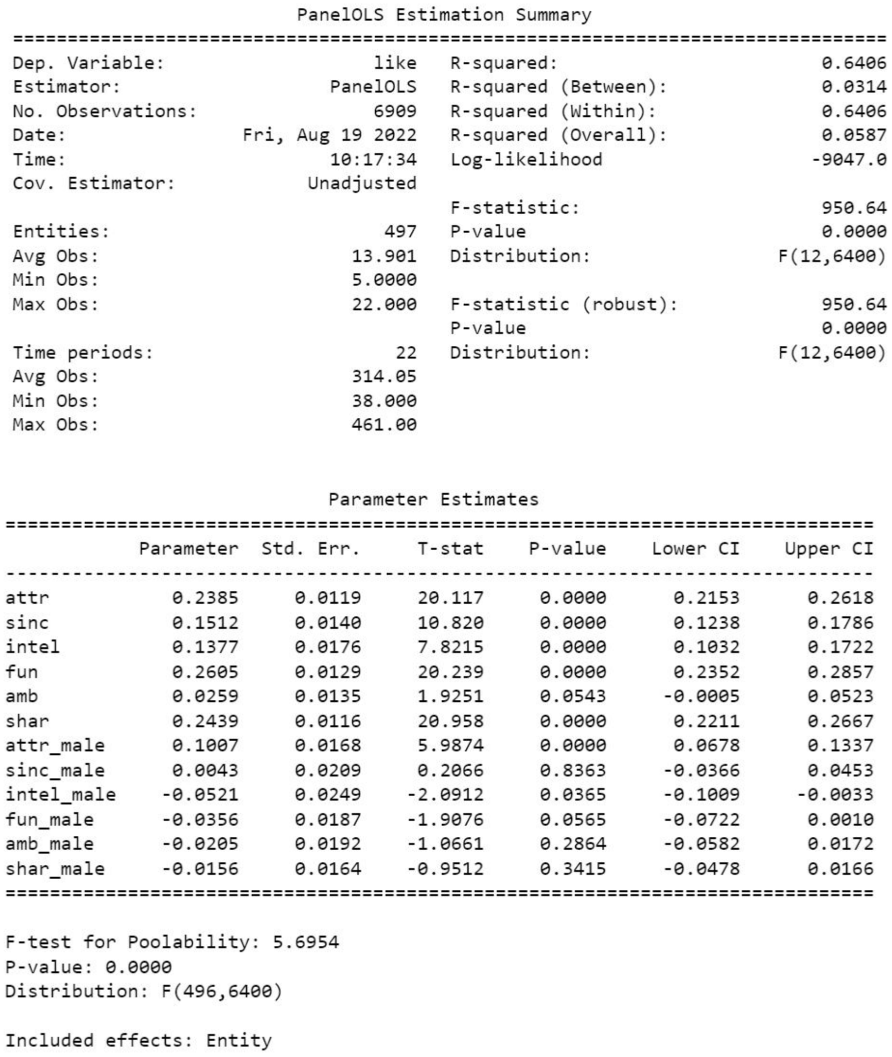
In this section, we introduce the speed dating dataset in Fisman et al. (2006), illustrate how we utilize it in our numerical experiments, and present more numerical results. Data were gathered from 552 participants in experimental speed dating events from 2002 to 2004. In each event, subjects met a number of potential mates (between 9 and 21) for four minutes each. At the end of the event, they rated each partner on a 1-10 scale. They were also asked to rate their date on six attributes on a 1-10 scale: attractiveness, sincerity, intelligence, fun, ambition, and shared interests.

To apply our model, we consider the rating on each partner as a proxy of the individual value. Similar to Fisman et al. (2006), we adopt a fixed-effect regression model to estimate the rating by the attribute ratings. Specifically, for each participant m (i.e., DM) and his/her partner i (i.e., alternative), the rating, denoted by Like_{mi}, can be estimated by

$$\text{Like}_{mi} \;=\; \alpha_m + \sum_c \beta_c \text{Rating}_{mic} + \sum_c \gamma_c \text{Rating}_{mic} \times \text{Male}_m + \varepsilon_{mi},$$

where c denotes one of the six attributes, Rating_{mic} is the attribute rating participant m assigns to partner i, and $\text{Male}_m \in \{0, 1\}$ is an indicator on whether the participant m is a male. The distributions of the attribute ratings are adjusted to be centered at zero. We run a regression to estimate the unknown parameters $\{\alpha_m\}_m$, $\{\beta_c\}_c$, and $\{\gamma_c\}_c$. The estimation results are summarized in Figure D.4. We observe that all attributes are statistically significant (although ambition is significant at 10 percent). We also find a clear gender difference as in Fisman et al. (2006): males weight more on attractiveness but less on intelligence than females do.

Figure D.4: Estimation Summary.



```
                          PanelOLS Estimation Summary
================================================================================
Dep. Variable:                   like   R-squared:                        0.6406
Estimator:                   PanelOLS   R-squared (Between):              0.0314
No. Observations:                6909   R-squared (Within):               0.6406
Date:                Fri, Aug 19 2022   R-squared (Overall):              0.0587
Time:                        10:17:34   Log-likelihood                   -9047.0
Cov. Estimator:            Unadjusted
                                        F-statistic:                      950.64
Entities:                         497   P-value                           0.0000
Avg Obs:                       13.901   Distribution:                 F(12,6400)
Min Obs:                       5.0000
Max Obs:                       22.000   F-statistic (robust):             950.64
                                        P-value                           0.0000
Time periods:                      22   Distribution:                 F(12,6400)
Avg Obs:                       314.05
Min Obs:                       38.000
Max Obs:                       461.00


                             Parameter Estimates
==============================================================================
            Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
------------------------------------------------------------------------------
attr           0.2385     0.0119     20.117     0.0000      0.2153      0.2618
sinc           0.1512     0.0140     10.820     0.0000      0.1238      0.1786
intel          0.1377     0.0176     7.8215     0.0000      0.1032      0.1722
fun            0.2605     0.0129     20.239     0.0000      0.2352      0.2857
amb            0.0259     0.0135     1.9251     0.0543     -0.0005      0.0523
shar           0.2439     0.0116     20.958     0.0000      0.2211      0.2667
attr_male      0.1007     0.0168     5.9874     0.0000      0.0678      0.1337
sinc_male      0.0043     0.0209     0.2066     0.8363     -0.0366      0.0453
intel_male    -0.0521     0.0249    -2.0912     0.0365     -0.1009     -0.0033
fun_male      -0.0356     0.0187    -1.9076     0.0565     -0.0722      0.0010
amb_male      -0.0205     0.0192    -1.0661     0.2864     -0.0582      0.0172
shar_male     -0.0156     0.0164    -0.9512     0.3415     -0.0478      0.0166
==============================================================================

F-test for Poolability: 5.6954
P-value: 0.0000
Distribution: F(496,6400)

Included effects: Entity
```

As discussed above, Like_{mi} is a proxy of the true individual value. The speed dating event can be considered as the first date where the participant finds the values of the six attributes and assesses the utility of accepting the partner. Thus, we consider a linear function of the attribute ratings as a proxy of the observable utility. The remaining term, the noise, is considered as the idiosyncratic value that needs to be learned gradually in later dates. If we add an index m to denote the participant, our model parameters and their proxies are summarized below:

$$\theta_{mi} = \text{Like}_{mi}, \quad u_{mi} = \alpha_m + \sum_c (\beta_c + \gamma_c \text{Male}_m) \cdot \text{Rating}_{mic}, \quad \text{and} \quad \Delta_{mi} = \varepsilon_{mi}.$$

For each m, the distribution of the utilities $\{u_{mi}\}_i$ can be approximated by a Normal distribution with mean μ_m, which corresponds to the hyperparameter in our model. The mean is given by

$$\mu_m \;=\; \alpha_m + \sum_c (\beta_c + \gamma_c \text{Male}_m) \cdot \overline{\text{Rating}}_{mc},$$

where $\overline{\text{Rating}}_{mc}$ is the mean of the attribute distribution for the group. The parameter α_m captures the personal preference of the participant, which may be known or unknown. $\overline{\text{Rating}}_{mc}$ captures the quality of the group and is unknown *a priori*. As participants date with more partners, they can learn more about the quality of the group (in terms of the six attributes) and/or their own preferences.

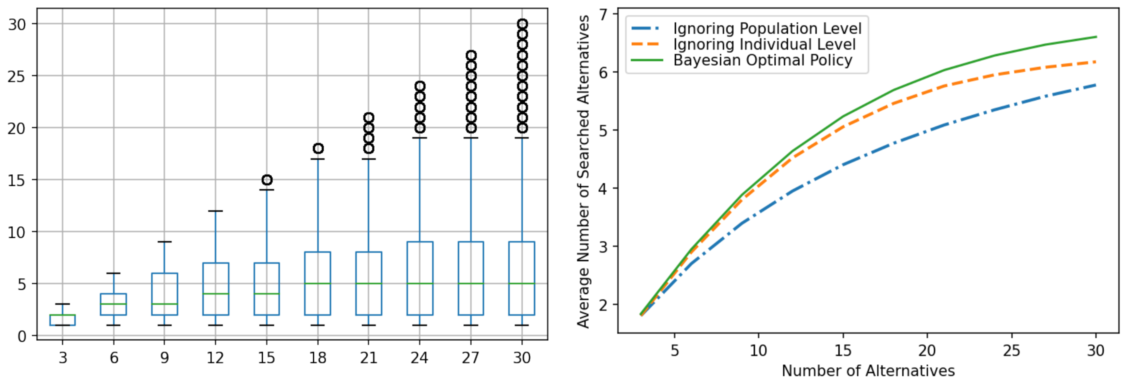
Problem parameters are estimated using method of moments. Specifically, we estimate the hyperparameter μ_m by the sample mean, i.e., $\hat{\mu}_m = \frac{1}{I}\sum_{i=1}^{I} u_{mi}$. For each gender, we fit a Normal distribution for the set $\{\hat{\mu}_m\}_m$, which is used to generate the hyperparameter μ in our numerical experiments. This is also the population-level prior when the participant starts searching, as denoted by $N(\eta_0, \tau_0)$. We also fit a Normal distribution for $\{u_{mi} - \mu_m\}_{m,i}$, which yields a zero mean and a utility precision $\hat{\tau}_u$. Thus, uitilities can be generated by $N(\mu, \hat{\tau}_u)$. Note that while our model assumes $\tau_u = 1$, the formulation and structural results extend to any value of $\tau_u > 0$. For the idiosyncratic values, we fit a Normal distribution to estimate the precision s_0 of the individual-level prior, which is then used to generate Δ_i in the experiments. It is also set as the prior distribution of Δ_i when the participant starts sampling. Starting from the second date, each date with the partner i provides a noisy sample of the true individual value θ_i, which provides information for the unknown Δ_i. The sample precision is set to be $s_x = 5$, and we also test different values of s_x in our experiments. The cost of sampling is set as $c = 0.05$ so that each DM only samples each alternative for a few times under the optimal policy. The cost of switching is set as $k = 2c = 0.1$. After cleaning the data, there are 250 female participants and 247 male participants. Each one has ratings for a group of partners ranging from 5 to 22. In Table D.1, we summarize values of the estimated parameters used in the experiments.

Table D.1: Estimated Parameters.

	Estimated Value	
Parameter	Male	Female
η_0	6.28	6.03
τ_0	0.95	0.80
τ_u	0.75	0.65
r_0	1.24	

In addition to the numerical results presented in Section D, it is also interesting to consider the number of alternatives that have been searched before acceptance. The number depends on the policy and the utility and sample observations. In Figure D.5(a), we present the number of alternatives that have been searched by the optimal policy. For a group of size $I \in \{3\ell\}_{\ell=1}^{10}$, the optimal number is no more than 10 in most cases and no more than 5 in average. In Figure D.5(b), we plot the average number for the optimal policy and the policies that ignore one level of learning. For each policy, the average number of searched alternatives increases as the group size grows, yet the increasing rate becomes slower when the group size is large. We also observe that the average numbers are close for the three policies.

Figure D.5: The Number of Alternatives Searched.



(a) The Optimal Number

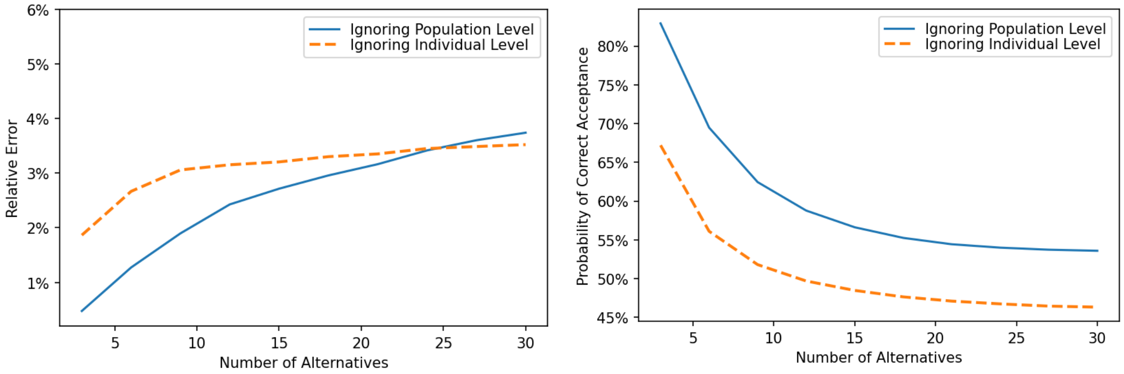
(b) The Average Number

D.2 More about Numerical Experiments

Note that we use estimated values for female participants in our experiments, and similar results can be obtained for male participants. For example, we present in Figure D.6 the performance of policies that ignore one level of learning for male participants, which is similar to Figure D.1.

In Figure D.1(b) and Figure D.6(b), ignoring the individual level of learning always has a worse probability of correct acceptance than ignoring the population level. However, this is not necessarily

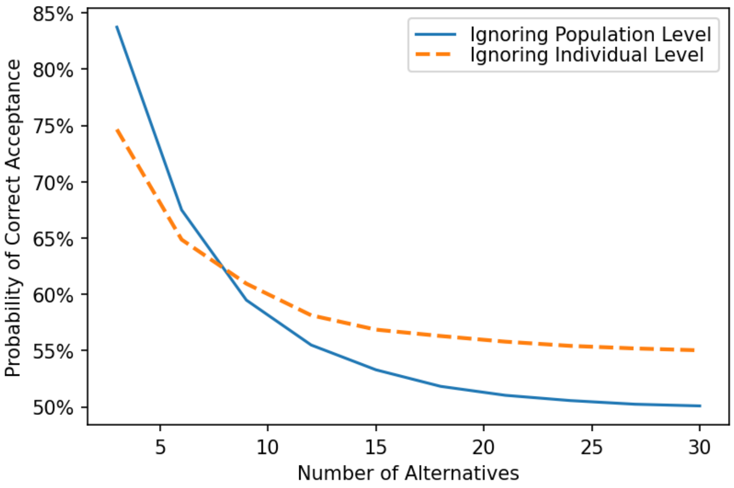
Figure D.6: Performance of Policies that Ignore One Level of Learning for Male Participants.



(a) Relative Error (b) Probability of Correct Acceptance

the case. In Figure D.7, we present the probability with a smaller sample precision. In that case, each sample provides a small amount of information, rendering it less costly to ignore learning at the individual level. We can observe that ignoring the individual level may perform better when the number of alternatives is large and there is a high value of knowledge transfer.

Figure D.7: Ignoring Individual Level Not Always Worse



Appendix E Proofs of Results in Extensions (Section 5)

Proof of Proposition 7. We prove for $K = 2$, and the case of $K > 2$ can be shown in a similar manner. In what follows, we ignore the superscript 2 for simplicity. Similar to the base model, we can construct a finite-alternative version of (20) with $K = 2$. Specifically, similar to (7), we define for any $\tau > 0$ and $i > 3$ that

$$u_i(\tau) \;:=\; -k + \mathbb{E}_Z \left[J\left(Z\sigma_\tau + v_{i-1}(\tau+1) - u_{i-1}(\tau+1);\; s_0 \right) \right] + u_{i-1}(\tau+1), \tag{E.1}$$

where $v_i(\tau) = u_i^1(\tau)$ is as defined in (7) for the base model. We note that $u_3(\tau) := -2k$ because te DM has to accept the last two alternatives if nothing has been accepted before. Since $\frac{\partial}{\partial\omega} J(\omega; s_0) \in [0,1]$, the right hand side above is increasing in $v_{i-1}(\tau+1)$ and $u_{i-1}(\tau+1)$. Since $v_{i-1}(\tau+1)$ is increasing in τ and i (as shown in the proofs of Proposition 2 and Proposition 3), we can show inductively that $u_i(\tau)$ is increasing in i and τ.

Next, we show that $u_i(\tau)$ is bounded. Recall that $\bar{v}_\infty$ satisfies $0 = -k + \mathbb{E}_Z \left[J\left(Z - \bar{v}_\infty;\; s_0 \right) \right]$ by the definition in (10). As discussed in the proof of Proposition 1, $J(\omega; s_0)$ is convex in ω, which implies that $k = \mathbb{E}_Z \left[J\left(Z - \bar{v}_\infty;\; s_0 \right) \right] \geq J\left(-\bar{v}_\infty;\; s_0 \right) \geq -\bar{v}_\infty$. Thus, $2\bar{v}_\infty \geq -2k = u_3(\tau)$. We can show inductively that $u_i(\tau) \leq 2\bar{v}_\infty$. Specifically, if the inequality holds for some $i-1$ with $i > 3$, we derive that

$$\begin{aligned} u_i(\tau) &= \; -k + \mathbb{E}_Z \left[J\left(Z\sigma_\tau + v_{i-1}(\tau+1) - u_{i-1}(\tau+1);\; s_0 \right) \right] + u_{i-1}(\tau+1) \\ &\leq \; -k + \mathbb{E}_Z \left[J\left(Z\sigma_\tau + \bar{v}_\infty - 2\bar{v}_\infty;\; s_0 \right) \right] + 2\bar{v}_\infty \\ &= \; -k + \mathbb{E}_Z \left[J\left(Z\sigma_\tau - \bar{v}_\infty;\; s_0 \right) \right] + 2\bar{v}_\infty \;=\; 2\bar{v}_\infty, \end{aligned}$$

where the first inequality follows from $v_{i-1}(\tau+1) \leq \bar{v}_\infty$ and the inductive assumption. Thus, $u_i(\tau)$ is bounded.

Since $u_i(\tau)$ is increasing in i and $u_i(\tau) \leq 2\bar{v}_\infty$, $\lim_{i\to\infty} u_i(\tau)$ exists by the monotone convergence theorem. Sending $i \to \infty$ in (E.1), we find that $\lim_{i\to\infty} u_i(\tau)$ satisfies the recursive equations in (10). Indeed, the solution to (10) is unique, which can be shown using a similar argument to the proof of Proposition 4. This implies that $\lim_{i\to\infty} u_i(\tau) = \bar{u}(\tau)$. Similar to the proof of Proposition 4, we can use properties of $u_i(\tau)$ to prove the desired results. $\square$

Proof of Proposition 8. Denote by $\gamma^K(\tau) := \bar{u}^K(\tau) - \bar{u}^{K-1}(\tau)$ for any $K \geq 1$ and $\tau > 0$. In what follows, we show $\gamma^K(\tau) \leq \gamma^{K+1}(\tau)$ and $\gamma^K(\tau) \leq \gamma^K(\tau+1)$ by induction on K.

For $K = 1$, we have $\gamma^1(\tau) = \bar{u}^1(\tau) - \bar{u}^0(\tau) = \bar{v}(\tau)$, which is increasing in τ, as shown in Proposition 4. Thus, $\gamma^1(\tau) \leq \gamma^1(\tau+1)$. Then we show $\gamma^1(\tau) \leq \gamma^2(\tau)$ by the way of contradiction. Suppose that $\gamma^1(\tau_0) > \gamma^2(\tau_0)$ for some $\tau_0 > 0$. Then $\gamma^1(\tau_0) - \gamma^2(\tau_0) > \varepsilon_1$ for some $\varepsilon_1 > 0$. Since $\bar{v}(\tau)$ is increasing in τ, we have

$$\varepsilon_1 < \; \gamma^1(\tau_0) - \gamma^2(\tau_0) \;=\; \bar{v}(\tau_0) - [\bar{u}^2(\tau_0) - \bar{v}(\tau_0)] \;\leq\; \bar{v}(\tau_0) - [\bar{u}^2(\tau_0) - \bar{v}(\tau_0+1)]. \tag{E.2}$$

Define $F(x) := \mathbb{E}_Z \left[J\left(Z\sigma_{\tau_0} - x;\; s_0 \right) \right] + x$ for any $x \in \mathbb{R}$. Using (10) and (20), the right hand side in the equation above satisfies

$$
\begin{aligned}
& \bar{v}(\tau_0) - [\bar{u}^2(\tau_0) - \bar{v}(\tau_0+1)] \\
= \; & -k + \mathbb{E}_Z \left[J\left(Z\sigma_{\tau_0} - \bar{v}(\tau_0+1);\; s_0 \right) \right] + \bar{v}(\tau_0+1) \\
& \quad - \left[-k + \mathbb{E}_Z \left[J\left(Z\sigma_{\tau_0} + \bar{v}(\tau_0+1) - \bar{u}^2(\tau_0+1);\; s_0 \right) \right] + \bar{u}^2(\tau_0+1) - \bar{v}(\tau_0+1) \right] \\
= \; & F(\gamma^1(\tau_0+1)) - F(\gamma^2(\tau_0+1)),
\end{aligned}
$$

where the last equality follows from $\gamma^1(\tau_0+1) = \bar{v}(\tau_0+1)$ and $\gamma^2(\tau_0+1) = \bar{u}^2(\tau_0+1) - \bar{v}(\tau_0+1)$. Plugging it into (E.2), we obtain

$$
\varepsilon_1 < \; F(\gamma^1(\tau_0+1)) - F(\gamma^2(\tau_0+1)) \; \leq \; \gamma^1(\tau_0+1) - \gamma^2(\tau_0+1),
$$

where the last inequality holds because $\frac{\partial}{\partial \omega} J(\omega;\; s_0) \in [0,1]$ implies $\frac{d}{dx} F(x) \in [0,1]$. Repeating the argument above, we obtain $\varepsilon_1 < \gamma^1(\tau_0 + M) - \gamma^2(\tau_0 + M)$ for any integer $M \geq 0$. Recall that as $\tau \to \infty$, $\gamma^1(\tau) = \bar{v}(\tau) \to \bar{v}_\infty$ and $\gamma^2(\tau_0) = \bar{u}^2(\tau) - \bar{v}(\tau) \to \bar{v}_\infty$. Then sending $M \to \infty$ yields $\varepsilon_1 \leq \bar{v}_\infty - \bar{v}_\infty = 0$, which contradicts with $\varepsilon_1 > 0$. Thus, $\gamma^1(\tau) \leq \gamma^2(\tau)$ holds for any $\tau > 0$, and the properties hold for $K = 1$.

Assume, by the way of induction, that $\gamma^K(\tau) \leq \gamma^{K+1}(\tau)$ and $\gamma^K(\tau) \leq \gamma^K(\tau+1)$ hold for some $K \geq 1$ and any $\tau > 0$. We now consider the case of $K+1$. Our goal is to show (a) $\gamma^{K+1}(\tau) \leq \gamma^{K+2}(\tau)$ and (b) $\gamma^{K+1}(\tau) \leq \gamma^{K+1}(\tau+1)$ for any $\tau > 0$. To prove (b), we derive from the recursive equations in (20) that

$$
\begin{aligned}
\gamma^{K+1}(\tau) = \; & \bar{u}^{K+1}(\tau) - \bar{u}^K(\tau) \\
= \; & -k + \mathbb{E}_Z \left[J\left(Z\sigma_\tau + \bar{u}^K(\tau+1) - \bar{u}^{K+1}(\tau+1);\; s_0 \right) \right] + \bar{u}^{K+1}(\tau+1) \\
& \quad - \left[-k + \mathbb{E}_Z \left[J\left(Z\sigma_\tau + \bar{u}^{K-1}(\tau+1) - \bar{u}^K(\tau+1);\; s_0 \right) \right] + \bar{u}^K(\tau+1) \right] \\
\leq \; & \bar{u}^{K+1}(\tau+1) - \bar{u}^K(\tau+1) \; = \; \gamma^{K+1}(\tau+1),
\end{aligned}
$$

where the inequality holds because $J(\omega; s_0)$ is increasing in ω and the inductive assumption implies $\bar{u}^K(\tau+1) - \bar{u}^{K+1}(\tau+1) = -\gamma^{K+1}(\tau+1) \leq -\gamma^K(\tau+1) = \bar{u}^{K-1}(\tau+1) - \bar{u}^K(\tau+1)$. This proves (b). Similar to the case of $K = 1$, we can show (a) by contradiction. More specifically, suppose that $\gamma^{K+1}(\tau_0) > \gamma^{K+2}(\tau_0)$ for some $\tau_0 > 0$. Then there exists $\varepsilon_0 > 0$ such that $\gamma^{K+1}(\tau_0) - \gamma^{K+2}(\tau_0) > \varepsilon_0$.

Since (b) implies $\gamma^{K+1}(\tau_0) \leq \gamma^{K+1}(\tau_0+1)$, we further have

$$\begin{aligned}
\varepsilon_0 <\ & \gamma^{K+1}(\tau_0+1) - \gamma^{K+2}(\tau_0) \\
=\ & [\bar{u}^{K+1}(\tau_0+1) - \bar{u}^{K}(\tau_0+1)] - [\bar{u}^{K+2}(\tau_0) - \bar{u}^{K+1}(\tau_0)] \\
=\ & [\bar{u}^{K+1}(\tau_0) - \bar{u}^{K}(\tau_0+1)] - [\bar{u}^{K+2}(\tau_0) - \bar{u}^{K+1}(\tau_0+1)] \\
=\ & \left\{-k + \mathbb{E}_Z\left[J\left(Z\sigma_{\tau_0} + \bar{u}^{K}(\tau_0+1) - \bar{u}^{K+1}(\tau_0+1);\ s_0\right)\right] + \bar{u}^{K+1}(\tau_0+1) - \bar{u}^{K}(\tau_0+1)\right\} \\
& - \left\{-k + \mathbb{E}_Z\left[J\left(Z\sigma_{\tau_0} + \bar{u}^{K+1}(\tau_0+1) - \bar{u}^{K+2}(\tau_0+1);\ s_0\right)\right] + \bar{u}^{K+2}(\tau_0+1) - \bar{u}^{K+1}(\tau_0+1)\right\} \\
=\ & F(\gamma^{K+1}(\tau_0+1)) - F(\gamma^{K+2}(\tau_0+1)) \\
\leq\ & \gamma^{K+1}(\tau_0+1) - \gamma^{K+2}(\tau_0+1),
\end{aligned}$$

where $F(x) = \mathbb{E}_Z\left[J\left(Z\sigma_{\tau_0} - x;\ s_0\right)\right] + x$ and the inequality follows from $\frac{d}{dx}F(x) \in [0,1]$, as discussed above. We can repeat the above argument and obtain $\varepsilon_0 < \gamma^{K+1}(\tau_0+M) - \gamma^{K+2}(\tau_0+M)$ for any integer $M \geq 1$. By Proposition 7, we have $\gamma^{K+1}(\tau) \to \bar{v}_\infty$ and $\gamma^{K+2}(\tau_0) \to \bar{v}_\infty$ as $\tau \to \infty$. It follows that $0 < \varepsilon_0 \leq \lim_{M\to\infty}[\gamma^{K+1}(\tau_0+M) - \gamma^{K+2}(\tau_0+M)] = \bar{v}_\infty - \bar{v}_\infty = 0$. We reach a contradiction, which proves (a). This completes our inductive proof. □

Appendix F Learning Individual Values with Unknown Sample Precisions

In the paper, we have considered learning each individual value θ_i by sampling and observing a noisy signal $X_i|\theta_i \sim \mathcal{N}(\theta_i, s_{x,i})$, where the precision satisfies $s_{x,i} \equiv s_x$ for any i and is known *a priori*. In practice, the sample precisions may be different and even unknown to the DM, which introduces the challenge of learning $s_{x,i}$ in addition to θ_i for each alternative. In this section, we extend the base model to incorporate unknown sample precisions by taking the Bayesian approach. We let the DM assign a prior belief to $s_{x,i}$ and update it as more samples are collected.

In the base model with a known sample precision, learning of the individual value can be separated from learning of the common prior, as shown in (11). The optimal decisions can be found by solving a single-alternative problem with a switching payoff of 0, and its value function is given by $J(\omega; s)$ in (3). When the precision $s_{x,i}$ is unknown, learning of the individual value can be captured by a similar single-alternative problem where the precision of each sample is unknown. In that case, the DM needs to learn a Normal distribution with unknown mean and precision, which we denote by $\mathcal{N}(\theta, s_x)$, and the conjugate prior should follow a Normal-gamma distribution (see, e.g., DeGroot 2005). Assume that $\theta|s_x \sim \mathcal{N}(\omega, \lambda s_x)$, where λ is a positive number and s_x follows a Gamma distribution with shape parameter α and mean ν. Then (θ, s_x) follows a Normal-gamma distribution, which we denote by $\mathcal{NG}(\omega, \lambda, \alpha, \nu)$. The density function of (θ, s_x) is given as

$$f(\theta, s_x) \ = \ \frac{(\alpha/\nu)^\alpha \sqrt{\lambda}}{\Gamma(\alpha)\sqrt{2\pi}} s_x^{\alpha-\frac{1}{2}} \exp\left(-\frac{\alpha}{\nu}s_x\right) \exp\left(-\frac{1}{2}\lambda s_x(\theta-\omega)^2\right),$$

where $\Gamma(\cdot)$ is the Gamma function. (For parameters of the Gamma distribution, a rate parameter or a scale parameter is typically used instead of the mean. However, we use the mean ν in our problem because we are interested in how the decisions depend on the mean estimate ν of the unknown precision.)

Now we consider the problem with a single alternative and a switching payoff of 0. Suppose that at the current stage, the posterior/prior of the unknown pair (θ, s_x) is $\mathcal{NG}(\omega, \lambda, \alpha, \nu)$. The DM may accept the current alternative and receive a payoff estimated by ω, or reject it and take the switching payoff of 0. Another option is to pay a sampling cost of c and observe a sample $X|(\theta, s_x) \sim \mathcal{N}(\theta, s_x)$. Based on the posterior/prior $(\theta, s_x) \sim \mathcal{NG}(\omega, \lambda, \alpha, \nu)$, it is estimated that X follows a nonstandard Student's t-distribution. More specifically, we derive that $X = \omega + T_{2\alpha}\sqrt{\frac{\lambda+1}{\lambda\nu}}$, where $T_{2\alpha}$ follows the standard t-distribution with degree of freedom 2α, and its density function is given by

$$f_\alpha(t) \;=\; \frac{\Gamma(\alpha+\frac{1}{2})}{\sqrt{2\pi\alpha}\,\Gamma(\alpha)}\left(1+\frac{t^2}{2\alpha}\right)^{-(\alpha+\frac{1}{2})}.$$

Then the sample X is used to update the posterior as $\mathcal{NG}\left(\frac{\lambda\omega+X}{\lambda+1}, \lambda+1, \alpha+\frac{1}{2}, \frac{\alpha+\frac{1}{2}}{\frac{\alpha}{\nu}+\frac{\lambda}{2(\lambda+1)}(X-\omega)^2}\right)$ by Bayes' rule (see, e.g., DeGroot 2005). Since $X = \omega + T_{2\alpha}\sqrt{\frac{\lambda+1}{\lambda\nu}}$, the updated posterior can be expressed as $\mathcal{NG}\left(\omega + T_{2\alpha}\sqrt{\frac{1}{\lambda(\lambda+1)\nu}},\ \lambda+1,\ \alpha+\frac{1}{2},\ \nu\cdot\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2}\right)$. Let $J^u(\omega;\lambda,\alpha,\nu)$ denote the maximal expected payoff for any $\omega \in \mathbb{R}$ and $\lambda, \alpha, \nu > 0$. Then the value function satisfies the Bellman equations

$$\begin{aligned} J^u(\omega;\lambda,\alpha,\nu) \;=&\; \\ &\max\left\{\omega,\quad 0,\quad -c+\mathbb{E}_{T_{2\alpha}}\left[J^u\left(\omega+T_{2\alpha}\sqrt{\frac{1}{\lambda(\lambda+1)\nu}};\ \lambda+1,\ \alpha+\frac{1}{2},\ \nu\cdot\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2}\right)\right]\right\}. \quad \text{(F.1)} \end{aligned}$$

We note that when λ is large, there is no value of continuation, that is, there exists $\bar{\lambda} < \infty$ such that $J^u(\omega;\lambda,\alpha,\nu) = \max\{\omega, 0\}$ for any $\lambda \geq \bar{\lambda}$, any ω, $\alpha > 0$, and $\nu > 0$. (This can be shown if we consider the marginal gain from observing exactly one more sample, which turns out to be zero and strictly less than the sampling cost c as $\lambda \to \infty$.) By induction on λ, we can show the following structural properties of the value function. The proof is deferred to Appendix E.

Proposition F.1. For the single-alternative problem with unknown sample precision and a switching payoff of 0, the value function $J^u(\omega;\lambda,\alpha,\nu)$ satisfies the following properties.

(a) $J^u(\omega;\lambda,\alpha,\nu)$ is increasing and convex in ω with derivatives less than or equal to 1, i.e., $\frac{\partial}{\partial\omega}J^u(\omega;\lambda,\alpha,\nu) \leq 1$.

(b) $J^u(\omega;\lambda,\alpha,\nu)$ is decreasing in λ and ν.

(c) As $\alpha \to \infty$ and $\nu \to s_x$, $J^u(\omega;\lambda,\alpha,\nu) \to J(\omega;\lambda s_x)$.

Proof. We show (a) and (b) by backward induction on λ. Recall that there exists $\bar{\lambda}$ such that for any $\lambda \geq \bar{\lambda}$, $J^u(\omega;\lambda,\alpha,\nu) = \max\{\omega, 0\}$, which satisfies (a) and (b). Assume by the way of induction that (a) and (b) hold for some $\lambda+1 \leq \bar{\lambda}$. Now we proceed to consider the stage with λ. Since these properties can be preserved under maximum, we only need to show for the expectation in the payoff of continuation, i.e., $\mathbb{E}_{T_{2\alpha}}\left[J^u\left(\omega + T_{2\alpha}\sqrt{\frac{1}{\lambda(\lambda+1)\nu}};\ \lambda+1,\ \alpha+\frac{1}{2},\ \nu\cdot\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2}\right)\right]$. By the inductive assumption, $J^u\left(\cdot\,;\lambda+1,\ \alpha+\frac{1}{2},\ \nu\cdot\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2}\right)$ is increasing and convex with derivatives less than or equal to 1, which is preserved under expectation. This proves (a). As for the monotonicities in (b), we note that the function $\mathbb{E}_{T_{2\alpha}}\left[J^u\left(\omega + T_{2\alpha}\cdot\sigma^u;\ \lambda+1,\ \alpha+\frac{1}{2},\ \nu\cdot\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2}\right)\right]$ is increasing in σ^u since this yields a mean-preserving spread. Let $\sigma^u = \sqrt{\frac{1}{\lambda(\lambda+1)\nu}}$. Then as λ or ν increases, both σ^u and $J^u\left(\omega';\ \lambda+1,\ \alpha+\frac{1}{2},\ \nu\cdot\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2}\right)$ for any ω' decrease. Thus, the expectation decreases, which proves (b).

To prove (c), we note that as $\alpha \to \infty$, $T_{2\alpha} \to Z \sim \mathcal{N}(0,1)$ and $\frac{2\alpha+1}{2\alpha+T_{2\alpha}^2} \to 1$. Then (c) follows, which completes the proof. ☐

Part (a) in Proposition F.1 shows the monotonicity and convexity of the value function in the posterior mean ω, which is the same as in the base model where the sample precision is known. This implies that the optimal policy for a single-alternative problem has a threshold structure with respect to ω. Specifically, it can be characterized by a threshold pair $(\underline{\omega}^u(\lambda,\alpha,\nu), \overline{\omega}^u(\lambda,\alpha,\nu))$. It is optimal to (i) accept the alternative if $\omega \geq \overline{\omega}^u(\lambda,\alpha,\nu)$, (ii) continue sampling if $\underline{\omega}^u(\lambda,\alpha,\nu) < \omega < \overline{\omega}^u(\lambda,\alpha,\nu)$, and (iii) stop sampling and take the switching payoff of 0 if $\omega \leq \underline{\omega}^u(\lambda,\alpha,\nu)$. To understand Part (b), we note that the conditional distribution of θ given s_x is $\mathcal{N}(\omega, \lambda s_x)$, and the marginal distribution of s_x has mean ν in the posterior. Thus, a larger value of λ or ν suggests less uncertainty about the estimate of the individual value, and the DM would expect a smaller gain from continuing sampling. It follows that as λ or ν increases, the value function decreases, and the continuation interval shrinks with $\overline{\omega}^u(\lambda,\alpha,\nu)$ decreasing and $\underline{\omega}^u(\lambda,\alpha,\nu)$ increasing. We note that ν affects the estimate of the sample precision, in addition to the posterior precision of the individual value. A higher sample precision implies more value of continuation in the base model (as shown in Proposition 3), yet this effect is eclipsed by that of the posterior precision. As such, the value of continuation becomes less as ν increases, and the corresponding set shrinks. Properties (a) and (b) show how the value function depends on the state variables. We note that the effect of the shape parameter α is unclear as it is involved in several terms of the density function f_α of $T_{2\alpha}$ and appears in the updating rule of ω and ν. Part (c) suggests that the marginal distribution of the unknown sample precision becomes concentrated on s_x as $\alpha \to \infty$ and $\nu \to s_x$. In that case, the

DM becomes certain about the sample precision s_x, and the posterior of the unknown individual value becomes $\mathcal{N}(\omega, \lambda s_x)$. Thus, we can apply the formulation of the base model, and the value function reduces to $J(\omega; \lambda s_x)$.

The analysis above paves a way to solve our uncertain search problem with multiple alternatives. For an infinite-alternative problem, let $V^u_\infty(\beta, \eta; \lambda, \alpha, \nu, \tau)$ denote the maximal expected payoff given the posterior $\mathcal{NG}(\beta, \lambda, \alpha, \nu)$ of the current alternative and the hyperposterior $\mathcal{N}(\eta, \tau)$. Similar to (11) for the base model, the value function can be separated as

$$V^u_\infty(\beta, \eta;\ \lambda, \alpha, \nu, \tau) \ = \ J^u\left(\beta - \eta - \bar{v}^u(\tau);\ \lambda, \alpha, \nu\right) + \eta + \bar{v}^u(\tau),$$

where $\eta + \bar{v}^u(\tau)$ is the payoff of switching, and $\bar{v}^u(\tau)$ satisfies recursive equations similar to $\bar{v}(\tau)$ in (12). Specifically, if we starts with a prior $\mathcal{NG}(0, \lambda_0, \alpha_0, \nu_0)$ of each idiosyncratic value Δ_i, we can extend the analysis in Section 3.2 and show that $\bar{v}^u(\tau)$ is the unique solution to

$$\bar{v}^u(\tau) \ = \ -k + \mathbb{E}_Z[J^u(Z\sigma_\tau - \bar{v}^u(\tau+1);\ \lambda_0, \alpha_0, \nu_0)] + \bar{v}^u(\tau+1).$$

The decomposition of the value function suggests that the optimal policy is determined by the first term $J^u\left(\beta - \eta - \bar{v}^u(\tau);\ \lambda, \alpha, \nu\right)$ for the single-alternative problem. By the analysis above, the optimal policy can be characterized by a threshold pair with respect to $\beta - \eta$, the difference between the posterior and hyperposterior means, which is defined as

$$\underline{\delta}^u(\lambda, \alpha, \nu, \tau) \ = \ \underline{\omega}^u(\lambda, \alpha, \nu) + \bar{v}^u(\tau) \quad \text{and} \quad \overline{\delta}^u(\lambda, \alpha, \nu, \tau) \ = \ \overline{\omega}^u(\lambda, \alpha, \nu) + \bar{v}^u(\tau). \tag{F.2}$$

Similar to Proposition 2 and Proposition 3, we can use inductive arguments to show the following monotonicities with respect to the state variables λ, ν, and τ, and the costs c and k.

Proposition F.2. For the infinite-alternative problem with unknown sample precisions, suppose that the DM has a posterior $\mathcal{NG}(\beta, \lambda, \alpha, \nu)$ of the current alternative and a hyperposterior $\mathcal{N}(\eta, \tau)$ of the common prior. Then the value function and the optimal thresholds satisfy the following properties:

(a) $V^u_\infty(\beta, \eta; \lambda, \alpha, \nu, \tau)$ and $\overline{\delta}^u(\lambda, \alpha, \nu, \tau)$ are decreasing in λ, ν, and c, while $\underline{\delta}^u(\lambda, \alpha, \nu, \tau)$ is increasing in λ and ν.

(b) $V^u_\infty(\beta, \eta; \lambda, \alpha, \nu, \tau)$, $\overline{\delta}^u(\lambda, \alpha, \nu, \tau)$, and $\underline{\delta}^u(\lambda, \alpha, \nu, \tau)$ are increasing in τ while decreasing in k.

Properties in Proposition F.2 differ from those in the base model only due to the unknown sample precision. As discussed above, increasing λ or ν has the same effect as increasing s in the base model. If we plot the threshold pair as functions of λ or ν, the curves are similar to those in Figure 2(a) and divide the space into three decision regions.

Appendix G Feature-Based Learning

In our base model, we consider learning the utility distribution for the group of alternatives. Once the DM switches to a new alternative, the observable utility is fully revealed, which is used to update the posterior and learn the unknown mean of the common prior. In practice, the utility may depend on a set of common features, and the DM may observe the common features and assess the utility. In that case, she knows the function mapping features to utility, which captures her preferences, but the population distribution of the features is typically unknown and needs to be learned via online observations. This motivates us to consider learning the distribution of the high-dimensional feature vector rather than its corresponding utility as a single parameter. We focus on the problem with a known sample precision.

To model the feature-based learning, we assume that the feature vector is drawn from a multivariate Normal distribution with an unknown mean vector and a given precision matrix, as denoted by $\mathcal{N}(\mu^f, \Lambda)$. At each stage, the DM has a hyperposterior/hyperprior $\mathcal{N}(\eta^f, \Gamma)$ of the unknown mean vector μ^f and a posterior $\mathcal{N}(\beta, s)$ of the individual value θ_i (we assume both precision matrices Γ and Λ to be positive definite so that they are invertible and the updated precision/covariance matrices are also invertible). For the infinite-alternative problem with a known sample precision, let $V^f_\infty(\beta, \eta^f; s, \Gamma)$ denote the maximal expected payoff at the current stage. The DM can estimate the payoff of the three potential actions based on the information about the current alternative and the feature distribution. If the DM accepts the current alternative, an estimated payoff of β is obtained. If the DM continues sampling, a cost of c incurs, and a sample of the individual value can be observed. As in the base model, the sample distribution is estimated as $X \sim \mathcal{N}(\beta, \frac{ss_x}{s+s_x})$, and the posterior is then updated as $N\left(\frac{\beta s+Xs_x}{s+s_x}, s+s_x\right)$. Thus, the payoff of continuation is

$$-c + \mathbb{E}_{X\sim N\left(\beta, \frac{ss_x}{s+s_x}\right)}\left[V^f_\infty\left(\frac{\beta s + X s_x}{s+s_x},\ \eta;\ s+s_x,\ \Gamma\right)\right] = -c + \mathbb{E}_Z\left[V^f_\infty\left(\beta + Z\sigma_s,\ \eta;\ s+s_x,\ \Gamma\right)\right],$$

where $Z \sim \mathcal{N}(0,1)$ and $\sigma_s = \sqrt{\frac{s_x}{s(s+s_x)}}$. If the DM switches to a new alternative, a cost of k incurs, and we would expect to observe a feature vector $X^f \sim \mathcal{N}(\eta^f, (\Gamma^{-1} + \Lambda^{-1})^{-1})$, which can be derived from $X^f|\mu^f \sim \mathcal{N}(\mu^f, \Lambda)$ and $\mu^f \sim \mathcal{N}(\eta^f, \Gamma)$. We consider the utility as a weighted average of the feature values, i.e., $u_i = a^\top X^f$, where the weight vector a captures preferences of the DM and is known to be the same for all alternatives. Then the hyperposterior would be updated as $\mathcal{N}\left((\Gamma+\Lambda)^{-1}(\Gamma\eta^f + \Lambda X^f), \Gamma+\Lambda\right)$. For the next alternative $i-1$, the DM assigns a prior $\mathcal{N}(0, s_0)$ to the idiosyncratic value Δ_{i-1}. Then the individual value $\theta_{i-1} = a^\top X^f + \Delta_{i-1}$ has a prior $\mathcal{N}\left(a^\top X^f, s_0\right)$ after observing the feature vector. Thus, the payoff of switching is

$$\begin{aligned}
-k \;&+\; \mathbb{E}_{X^f}\left[V^f_\infty\left(a^\top X^f,\ (\Gamma+\Lambda)^{-1}(\Gamma\eta^f+\Lambda X^f);\ s_0,\ \Gamma+\Lambda\right)\right] \\
&= \; -k+\mathbb{E}_{\bar{X}}\left[V^f_\infty\left(a^\top\eta^f+a^\top\bar{X},\ \eta^f+(\Gamma+\Lambda)^{-1}\Lambda\bar{X};\ s_0,\ \Gamma+\Lambda\right)\right],
\end{aligned}$$

where $\bar{X}:=X^f-\eta^f\sim\mathcal{N}(0,(\Gamma^{-1}+\Lambda^{-1})^{-1})$, and we replace X^f with $\eta^f+\bar{X}$ to separate η^f and Γ. Summarizing the analysis above, we obtain the following dynamic programming formulation for any $\beta\in\mathbb{R}$, $\eta^f\in\mathbb{R}^d$, $s>0$, and any positive definite matrix $\Gamma\in\mathbb{R}^{d\times d}$:

$$\begin{aligned}
V^f_\infty(\beta,\eta^f;s,\Gamma) \;=\; \max\Big\{\beta,\quad & -c+\mathbb{E}_Z\left[V^f_\infty\left(\beta+Z\sigma_s,\ \eta^f;\ s+s_x,\ \Gamma\right)\right], \\
& -k+\mathbb{E}_{\bar{X}}\left[V^f_\infty\left(a^T\eta^f+a^T\bar{X},\ \eta^f+(\Gamma+\Lambda)^{-1}\Lambda\bar{X};\ s_0,\ \Gamma+\Lambda\right)\right]\Big\}. \quad (\text{G.1})
\end{aligned}$$

Note that when $d=1$ and $a\equiv 1$, this feature-based learning problem reduces to our base model without features.

For the problem with features, we need to learn a high-dimensional vector μ^f rather than a single parameter μ. As such, the decisions depend on the hyperprior/hyperposterior of the vector μ^f, which complicates the analysis. Nevertheless, we can generalize the properties shown in the base model. Similar to the base model, eventually the DM has to accept the current alternative or switch to another alternative in the group. Then the invariance transition property given in (5) extends as follows:

$$V^f_\infty(\beta-a^\top x,\eta^f-x;s,\Gamma) \;=\; V^f_\infty(\beta,\eta^f;s,\Gamma)-a^\top x,\quad \forall\, x\in\mathbb{R}^d.$$

This implies that we can still focus on the the difference between the estimated values of the current alternative and the population, which is now given by $\beta-a^\top\eta^f$. Moreover, we can extend (11) and decompose the value function into three terms below:

$$V^f_\infty(\beta,\eta^f;s,\Gamma) \;=\; J\left(\omega;\ s\right)+a^\top\eta^f+\bar{v}^f(\Gamma),$$

where $\omega:=\beta-a^\top\eta^f-\bar{v}^f(\Gamma)$ is the difference between the acceptance payoff β and the switching payoff $a^\top\eta^f+\bar{v}^f(\Gamma)$. The first term $J(\omega;s)$ is the value function for a single-alternative problem defined in (3) and determines the threshold structure of the optimal policy. Specifically, the optimal policy can be characterized by a threshold pair of the difference $\beta-a^\top\eta^f$, which is given by

$$\underline{\delta}^f(s,\Gamma) \;=\; \underline{\omega}(s)+\bar{v}^f(\Gamma),\quad \text{and}\quad \overline{\delta}^f(s,\Gamma) \;=\; \overline{\omega}(s)+\bar{v}^f(\Gamma),$$

where $\underline{\omega}(s)$ and $\overline{\omega}(s)$ are the threshold pair for the single-alternative problem. It is optimal to (i) accept the current alternative if $\beta - a^\top \eta^f \geq \overline{\delta}^f(s,\Gamma)$, (ii) continue sampling if $\underline{\delta}^f(s,\Gamma) < \beta - a^\top \eta^f < \overline{\delta}^f(s,\Gamma)$, and (iii) switch to a new alternative if $\beta - a^\top \eta^f \leq \underline{\delta}^f(s,\Gamma)$. We note that $\bar{v}^f(\Gamma)$ satisfies

$$\begin{aligned}\bar{v}^f(\Gamma) :=&\ -k + \mathbb{E}_{\bar{X}}\left[V_\infty^f\left(a^\top\eta + a^\top\bar{X},\ \eta^f + (\Gamma+\Lambda)^{-1}\Lambda\bar{X};\ s_0,\ \Gamma+\Lambda\right)\right] - a^\top\eta^f \\ =&\ -k + \mathbb{E}_{\bar{X}}\left[V_\infty^f\left(a^\top(\Gamma+\Lambda)^{-1}\Gamma\bar{X},\ 0;\ s_0,\ \Gamma+\Lambda\right)\right] \\ =&\ -k + \mathbb{E}_Z\left[V_\infty^f\left(Z\sigma_\Gamma,\ 0;\ s_0,\ \Gamma+\Lambda\right)\right] \\ =&\ -k + \mathbb{E}_Z\left[J\left(Z\sigma_\Gamma - \bar{v}^f(\Gamma+\Lambda);\ s_0\right)\right] + \bar{v}^f(\Gamma+\Lambda),\end{aligned}$$

where $\sigma_\Gamma := \sqrt{a^\top(\Lambda^{-1} - (\Gamma+\Lambda)^{-1})a}$ since $\bar{X} \sim \mathcal{N}(0, (\Gamma^{-1}+\Lambda^{-1})^{-1})$. As such, $\bar{v}^f(\Gamma)$ can be computed independent of the individual-level estimate and the population-level mean estimate η^f.

Similar to Proposition 1, 2, and 3, we can show how the value function and the threshold pair change with state variables and problem parameters. These structural properties extend to the feature-based learning problem. In particular, recall that the value function and both thresholds are increasing in the precision τ of the hyperparameter in the base model without features. This implies that as the DM learns more about the common prior, the payoff of switching is estimated to be higher, and the switching interval expands. The intuition is also valid for the extension with features. If we consider $\Gamma = \Gamma_0 + \tau I$ for any $\tau > 0$ and some symmetric and positive definite matrix $\Gamma_0 \in \mathbb{R}^{d\times d}$, we can show that $\bar{v}^f(\Gamma)$ is increasing in τ, and thus $V_\infty^f(\beta, \eta^f; s, \Gamma)$, $\underline{\delta}^f(s,\Gamma)$, and $\overline{\delta}^f(s,\Gamma)$ are increasing in τ, which is consistent with the intuition.

We remark that our setting is not a typical setting of learning with features as in parameter estimation, where the feature distribution is given while the weight associated with each feature is unknown (e.g., the classical linear regression). The DM in our uncertain search problem starts with limited information about the population yet has her own metric in selecting the alternative. We show that structural results in the base model extend when learning the feature distribution.

We note that multiple attributes/features have been incorporated in the classical secretary problem, although in a different way. For example, Bearden et al. (2005) consider an extension where the DM observes the relative ranks of the attributes for each alternative (rather than the relative rank of each candidate). The payoff function is separable in the absolute ranks of the attributes. They provide a way to compute the optimal policy. They also conduct two experiments and find that subjects tend to stop earlier than the optimal policy. Attributes have also been used to model the information gathering process (see, e.g., Lim et al. 2006, Ke et al. 2016, Sanjurjo 2017). In that case, the individual value of each alternative is divided into values of multiple attributes that need to be learned sequentially and costly. Different from our problem with Bayesian learning, the attribute values are unknown *a priori*, but their distributions are known with mean zero. In each period of search, exactly one attribute value is observed. As time grows to infinity, all attribute values can be obtained, which constitute the total value of the alternative on search.

References

Bearden JN, Murphy RO, Rapoport A (2005) A multi-attribute extension of the secretary problem: Theory and experiments. *Journal of Mathematical Psychology* 49(5):410–422.

DeGroot MH (2005) *Optimal statistical decisions* (John Wiley & Sons Inc., Hoboken, N.J.).

Fisman R, Iyengar SS, Kamenica E, Simonson I (2006) Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics* 121(2):673–697.

Hanoch G, Levy H (1975) The efficiency analysis of choices involving risk. Ziemba W, Vickson R, eds., *Stochastic Optimization Models in Finance*, 89–100 (Academic Press).

Ke TT, Shen ZJM, Villas-Boas JM (2016) Search for information on multiple products. *Management Science* 62(12):3576–3603.

Lim C, Bearden JN, Smith JC (2006) Sequential search with multiattribute options. *Decision Analysis* 3(1):3–15.

Lippman SA, McCardle KF (1987) Does cheaper, faster, or better imply sooner in the timing of innovation decisions? *Management Science* 33(8):1058–1064.

Lippman SA, McCardle KF (1991) Uncertain search: A model of search among technologies of uncertain values. *Management Science* 37(11):1474–1490.

Sanjurjo A (2017) Search with multiple attributes: Theory and empirics. *Games and Economic Behavior* 104:535–562.