Optimal Learning before Choice [*]

T. Tony Ke
MIT
kete@mit.edu

J. Miguel Villas-Boas
University of California, Berkeley
villas@haas.berkeley.edu

This version: June 2017
First version: May 2016

[*]We thank Jerker Denrell, Drew Fudenberg, Yuichiro Kamada, Giuseppe Moscarini, Aniko Öry, Jiwoong Shin, Philipp Strack, and Jidong Zhou for helpful comments and suggestions. We also thank Steven Evans, Nicolae Gârleanu, Ioannis Karatzas, Wenpin Tang, and Johan Walden for helpful comments and suggestions regarding the behavior of the stochastic process under the optimal learning strategy. All remaining errors are our own.

Optimal Learning before Choice

Abstract

A Bayesian decision maker is choosing among multiple alternatives with uncertain payoffs and an outside option with known payoff. Before deciding which alternative to adopt, the decision maker can purchase sequentially multiple informative signals on each of the available alternatives. To maximize the expected payoff, the decision maker solves the problem of optimal dynamic allocation of learning efforts as well as optimal stopping of the learning process. We show that the optimal learning strategy is of the type of consider-then-decide. The decision maker considers an alternative for learning or adoption if and only if the expected payoff of the alternative is above a threshold. Given several alternatives in the decision maker's consideration set, we find that sometimes it is optimal for the decision maker to learn information from an alternative that has a lower expected payoff and less uncertainty, given all other characteristics of all the alternatives being the same. If the decision maker subsequently receives enough positive informative signals, the decision maker will switch to learning the better alternative; otherwise the decision maker will rule out this alternative from consideration and adopt the currently most preferred alternative. We find that this strategy works because it minimizes the decision maker's learning efforts. It becomes the optimal strategy when the outside option is weak, and the decision maker's beliefs about the different alternatives are in an intermediate range.

Keywords: *Information; Bayesian Learning; Search Theory; Dynamic Allocation; Optimal Stopping; Consideration Set*

JEL Classification: D83, L15, M31

1. Introduction

In many circumstances, agents have the opportunity to learn sequentially about multiple different alternatives before making a choice. Specifically, consider a decision maker who is deciding among several alternatives with uncertain payoffs and an outside option with a known payoff. Before deciding which one to adopt, he can gather some information on each alternative. After learning more about an alternative, the decision maker gains more precise knowledge about its payoff, with initial learning being more informative than later learning. It is costly to gather and process information. Therefore, at some point, the decision maker will decide to stop learning and make a choice on which alternative to adopt or to take the outside option.

For example, consider a consumer in the market for a car. The consumer could first learn information about a certain model A, then choose to get information about model B, then go back to get more information on model A again, and so on, until the consumer decides to either purchase model A, model B, or some other model, or not to purchase a car for now. With the recent development of information technologies, it becomes more and more important to understand this information gathering behavior of consumers, as there is further information collected about how consumers gather information by web browsing the Internet.

Information gathering is not unique to the consumers' purchase process. In fact, many other important economic activities involve similar costly gradual information acquisition: companies allocating resources to R&D, individuals looking for jobs, firms recruiting job candidates, politicians evaluating better public policies, manufacturers considering alternative suppliers, etc.

We study a decision maker's optimal information gathering problem under a Bayesian framework. We consider a setting in which the payoff of each alternative follows a two-point distribution, being either high or low. The decision maker has a prior belief on the probability that an alternative is of high payoff. Each time he learns some information about an alternative, he incurs a cost and gets a noisy signal on its payoff. This signal does not reveal the payoff of this alternative completely; rather, it updates the decision maker's belief of the distribution of the payoff. The decision maker could buy another signal on the same alternative if he would like to learn more about it or, alternatively, he could buy signals on other alternatives. In this setting the precision of the decision maker's belief is a function of his belief. Therefore, we only need to account for one state variable per alternative during the learning process—the probability that each alternative is of high payoff. Despite its simplicity, this setup captures the ideas that the decision maker gains more precise information during the learning process, and that the marginal information per learning effort decreases with the cumulative information gathered so far on an alternative. This two-point distribution assumption also implies that the agent becomes more certain about the payoff of an alternative when the payoff is either relatively high or low. The decision maker solves the problem of optimal dynamic allocation of learning efforts as well as optimal stopping of the learning process, so as to maximize the expected payoff.

We consider a two-alternative continuous-time model with infinite time horizon, which enables a time-stationary characterization of the optimal learning strategy when the outside option takes relatively large or small values.[1] The result is a partition of the belief space into regions, within which it is optimal to learn each alternative, adopt each alternative, or take the outside option.

We find that if the belief on an alternative is below a threshold, the decision maker will never consider that alternative—he will neither learn nor adopt that alternative. This then provides an endogenous formation of a consideration set, given the decision maker's belief on the payoff distribution, his learning costs, the value of the outside option, and the noisiness of the signals received. If more than one alternative has the beliefs above that threshold, then the decision maker chooses to stop learning and adopt one alternative if the expected payoff of that alternative is sufficiently larger than that of the next best alternative. That is, when the expected payoffs from the alternatives are not too different, the decision maker will continue to learn.

We also investigate the effect of the value of the outside option on optimal learning behavior. We find that when the value of the outside option is relatively high, the decision maker always chooses to learn about the alternative with the higher expected payoff first. More interestingly, we find that this is not necessarily the case when the value of the outside option is sufficiently low. In this case it may be optimal to first learn about the alternative that has a lower expected payoff, all else being equal among alternatives, with the purpose of possibly ruling it out early. If the decision maker receives sufficiently poor signals on this alternative, he will stop learning and immediately adopt the other alternative. We find this strategy works because it saves learning costs.

It is interesting to understand the intuition about why a decision maker's optimal learning strategy depends on the value of the outside option. When the outside option has a relatively high value, it will be relevant for the decision maker's ultimate choice at the end of the learning process. In this case, by learning the alternative with higher expected payoff, the decision maker keeps both the outside option and the other alternative as reservation options. Therefore, the decision maker prefers to learn the alternative with higher expected payoff. On the other hand, when the outside option has a relatively low value, it is no longer as relevant for the decision maker's ultimate choice. In this case, the decision maker is basically deciding between the two uncertain alternatives, and it may be better to learn about the alternative with lower expected payoff first to potentially rule it

[1] See, for example, Bolton and Harris (1999), Moscarini and Smith (2001), for the use of the same learning process in continuous time.

out early, so as to make a clear distinction between the two uncertain alternatives with minimum learning effort.

We also compute the decision maker's likelihood to adopt each alternative given his current belief, when he follows the optimal learning strategy. We find that the probability of choosing either one of the alternatives can fall as there are more alternatives available. This is because having more alternatives makes it harder for the decision maker to make a choice, and thus leads to more learning, which can ultimately result in no adoption of any of the alternatives. Specifically, information is ex ante neutral. It is possible that with more learning, the decision maker gets positive information on either alternative, in which case he will, at best, adopt one of the alternatives. On the other hand, it is equally possible that with more learning, the decision maker gets negative information, in which case he may choose to take the outside option. We also compute the decision maker's expected probability of being correct ex post given his current belief, when he follows the optimal learning strategy. Finally, we consider the optimal learning problem when alternatives have heterogeneous learning costs or payoff distributions, when the number of alternatives are greater than two, and when there is time discounting instead of learning costs.

The problem considered here is related to the multi-armed bandit problem (e.g., Rothschild 1974a, Gittins 1979, Whittle 1980, Bergemann and Välimäki 1996, Felli and Harris 1996, Bolton and Harris 1999), where a decision maker learns about different options by trying them, one in each period, while earning some stochastic rewards along the way. That problem has an elegant result that the optimal policy is to choose the arm with the highest Gittins index, which for each arm only depends on what the decision maker knows about that arm until then. However, the problem considered here is different from the bandit problem in one major aspect. In our setting, a decision maker optimally decides when to stop learning and make an adoption. Therefore, the decision horizon is endogenous, and optimally determined by the decision maker. In contrast, multi-arm bandit problems generally presume an exogenously given decision horizon, which could be either finite or infinite. In fact, our problem belongs to a general class of the *stoppable bandits problem*, first introduced by Glazebrook (1979), which generalizes simple bandit processes to allow for two arms per bandit. By adding a second "stopping arm" to each bandit, extra payoff can be generated when the stopping arm is pulled. In general, index policies are not optimal for this type of problems (Gittins et al. 2011, Chapter 4). Glazebrook (1979) shows that an index policy can be optimal under certain conditions. However, as shown below, Glazebrook's sufficient conditions are not satisfied in our setting, and the index policy is sub-optimal. We will compare our optimal policy with that of Glazebrook (1979) below. Forand (2015) considers a revised multi-armed bandit problem where the decision maker must incur maintenance costs to keep an alternative available for choice and any positive signal is fully revealing. Forand finds that it may be optimal to first try the worst alternative to rule it out, and continue then searching the other alternative. In contrast, we consider a decision maker that receives gradual signals on an alternative searched, either positive or negative. If he receives enough negative signals on the worst alternative, the decision maker chooses the other alternative immediately, without learning more about it.

The literature on search theory is also related to the results presented here. Although the problem considered here is central to choice in a market environment, it is quite under-researched when all of its dimensions are included. For the simpler case in which all learnable information about an alternative can be learned in one search occasion, there is a large literature on optimal search and some of its market implications (e.g., McCall 1970, Diamond 1971, Weitzman 1979, Doval 2014). This literature, however, does not consider the possibility of gradual revelation of information throughout the search process. There is also some literature on gradual learning when a single alternative is considered or information is gathered to uncover a single uncertain value (e.g., Roberts and Weitzman 1981, Moscarini and Smith 2001, Branco et al. 2012, Fudenberg et al. 2015), and the choice there is between adopting the alternative or not. In the face of more than one uncertain alternative (as is the case considered in this paper) the problem becomes more complicated. This is because opting for one alternative in a choice set means giving up potential high payoffs from other alternatives about which the decision maker has yet to learn more information. This paper can then be seen as combining these two literatures, with gradual search for information on multiple alternatives. Kuksov and Villas-Boas (2010) and Doval (2014) consider a search problem where everything that can be learned about one alternative is learned in one search occasion, but the decision maker can choose an alternative without search. This feature is also present here by including an outside option and allowing for adoption of an alternative without gaining full information about it. Doval, in particular, finds that in this case the decision maker may be indifferent to search an alternative that does not have the highest reservation price, which can be seen as related to the case in this paper of choosing to learn an alternative other than the best one under some conditions. Different from Doval, this paper considers gradual learning of each alternative, and we find that it could be strictly optimal to learn an alternative that has both lower expected payoff and lower uncertainty.[2]

Ke et al. (2016) considers a stationary problem of gradual search for information with multiple products, where the precision of beliefs does not increase with search, and where earlier learning is not more informative than later learning. One important result in this paper is that it can be optimal to learn about an inferior alternative, which is never optimal in Ke et al. (2016). Another

[2]In a concurrent paper, Che and Mierendorff (2016) consider which type of information to collect in a Poisson-type model, when the decision maker has to choose between two alternatives with one, and only one, alternative having a high payoff.

important difference is that the possible payoffs are bounded here, while they are unbounded in Ke et al. (2016).[3] The problem considered here also relates to the literature studying search while learning the payoff distribution (e.g., Rothschild 1974b, Adam 2001), but there what can be learned about each alternative is learned in one search occasion, so there is no gradual learning on each alternative.

The remainder of the paper is organized as follows. Section 2 presents the optimal learning problem. Section 3 solves the problem and presents the optimal learning strategy, and Section 4 looks at the implications of the results for adoption likelihood and probability of being correct. Section 5 presents several extensions of the main model, including heterogeneous alternatives, the effect of the number of alternatives, and time discounting. Section 6 presents concluding remarks, and the proofs are presented in the Appendix.

2. Optimal Learning Problem

A decision maker (DM) is uncertain about the payoffs of n alternatives. It is assumed that the payoff of alternative i, π_i, is either "high" as $\overline{\pi}_i$ or "low" as $\underline{\pi}_i$, with $\overline{\pi}_i > \underline{\pi}_i$ for $i = 1, \cdots, n$. Besides the n alternatives, there is an outside option with known deterministic payoff as π_0. The DM is assumed to be Bayesian and risk-neutral, or that the payoffs are in utils. To avoid trivialities, let us assume that $\pi_0 < \overline{\pi}_i$ for all i, so that no alternative i will be dominated by the outside option in all circumstances. Before deciding which alternative to adopt, the decision maker can gather information sequentially on the n alternatives. Specifically, we consider a continuous-time setup. At any time point t, when the DM spends extra time dt in learning some information on alternative i, he pays cost $c_i dt$ and gets an informative signal on π_i. This signal does not reveal π_i completely; instead, it updates the DM's belief of the distribution of π_i according to Bayes' rule. The DM could buy another signal on the same alternative if he would like to learn more about it or, alternatively, he could buy signals on other alternatives. Signals are assumed to be i.i.d. for each alternative and across all alternatives. It is assumed that the flow learning cost c_i, once paid, is sunk.[4]

Consider a dynamic decision process with infinite time horizon. Let us define the DM's posterior

[3] This paper also provides proof of existence and uniqueness of the optimal solution, and a more complete characterization of the stochastic process, which is not presented in Ke et al. (2016). With the possibility of a single observation process one can also consider the case of choosing one among several alternatives whose payoffs depend on the same state variable (see, for example, Moscarini and Smith 2001, Décamps et al. 2006).

[4] Another possible dimension, not explored here, is that the DM can also choose the intensity of effort in the search process at each moment in time, with greater effort leading to a more precise signal (as in Moscarini and Smith 2001).

belief of π_i being high at time $t \in [0, \infty)$ as $x_i(t)$.

$$x_i(t) = \Pr(\pi_i = \overline{\pi}_i | \mathscr{F}_t), \tag{1}$$

where $\{\mathscr{F}_t\}_{t=0}^T$ is a filtration that represents all the observed signals by time t. Then we have that $x_i(0)$ is the DM's prior belief at time zero before obtaining any signal. For the sake of simplicity, we will just call $x_i(t)$ the *belief* of alternative i at time t below, and in the cases without confusion, we will drop the argument t and write it as x_i. The DM's conditional expected payoff from adopting alternative i at time t is then,

$$\mathrm{E}[\pi_i|\mathscr{F}_t] = x_i(t)\overline{\pi}_i + (1 - x_i(t))\,\underline{\pi}_i = \Delta\pi_i \cdot x_i(t) + \underline{\pi}_i, \tag{2}$$

where $\Delta\pi_i \equiv \overline{\pi}_i - \underline{\pi}_i > 0$. Below, we will use $\Delta\pi_i$ and $\underline{\pi}_i$ to replace $\overline{\pi}_i$ and $\underline{\pi}_i$ as our primary parameters. When we consider homogeneous alternatives below, we will drop the subscript i and use $\Delta\pi$ and $\underline{\pi}$, respectively.

At any time t, the decision maker makes a choice from $2n + 1$ actions: to learn alternative $1, \cdots,$ to learn alternative n, to adopt alternative $1, \cdots,$ to adopt alternative n, or to take the outside option. The decision process terminates when the DM adopts either one alternative or the outside option. He makes the choice of which action to take based on the information acquired so far, $\mathscr{F}_t$. Since π_i follows a two-point distribution, intuitively, all information about π_i can be summarized by one variable—the posterior belief x_i. Formally, it is straightforward to prove that $\mathbf{x}(t) \equiv (x_1(t), \cdots, x_n(t))$ is a sufficient statistic for the observed history $\mathscr{F}_t$. Therefore, the DM's learning problem is essentially for each time t, to find a mapping from the belief space where $\mathbf{x}(t)$ lives in to the set of $2n + 1$ actions, so as to maximize the expected payoff.

We can actually classify the $2n + 1$ actions into two categories: to continue learning, and to stop learning and make an adoption decision. From this perspective, the decision maker's optimal learning problem can be equivalently viewed as dynamically deciding which alternative to learn at each time, and when to stop learning and make an adoption decision. In other words, the DM needs to solve the problem of *optimal allocation* (which alternative to learn) and *optimal stopping* (when to stop learning) at the same time, so as to maximize his expected payoff.

To formalize the DM's optimal learning problem, let us introduce the allocation policy as $I_t(\mathbf{x}) \in \{1, \cdots, n\}$ that specifies which alternative to learn at time t, based on the DM's current posterior belief $\mathbf{x}$. Formally speaking, $\mathscr{I} \equiv \{I_t(\mathbf{x})\}_{t=0}^T$ is a progressively measurable process adapted with respect to the natural filtration of $\{\mathbf{x}(t)\}_{t=0}^T$. Let us also introduce τ as the stopping time adapted with respect to the natural filtration of $\{\mathbf{x}(t)\}_{t=0}^T$. Given any allocation policy $\mathscr{I}$ and stopping time

τ, let us define $T_i(t)$ as the cumulated time that alternative i has been engaged in learning until time $t \leq \tau$, or formally,

$$T_i(t) \equiv \mu\left(0 \leq z \leq t; I(\mathbf{x}(z)) = i\right), \tag{3}$$

where $\mu(\cdot)$ is the Lebesgue measure on $\mathbb{R}$. The observation process for the i-th alternative is assumed to follow the stochastic differential equation (SDE),

$$ds_i(t) = \pi_i dT_i(t) + \sigma_i dW_i\left(T_i(t)\right), \tag{4}$$

where $W_i(t)$ is a standard Brownian motion, and $\sigma_i^2 > 0$ specifies the noise level for signals of alternative i (see, for example, Roberts and Weitzman 1981, Karatzas 1984, Bolton and Harris 1999, Moscarini and Smith 2001, Weeds 2002, for a similar formulation). As an analog to the discrete-time counterpart, the continuous-time signal $ds_i(t)$ consists of two parts: the "true value" $\pi_i dT_i(t)$, and a white noise term $\sigma_i dW_i\left(T_i(t)\right)$. Under this framework, if the DM keeps learning information from one alternative continuously, his observation will be a Brownian motion with the drift as the (unknown) payoff.

The decision maker's expected payoff given his allocation policy $\mathscr{I}$ and stopping time τ can be written as the following expression:

$$J(\mathbf{x}; \mathscr{I}, \tau) = \mathrm{E}\Bigg[\max\left\{\Delta\pi_1 \cdot x_1(\tau) + \underline{\pi}_1, \cdots, \Delta\pi_n \cdot x_n(\tau) + \underline{\pi}_n, \pi_0\right\} - \sum_{i=1}^{n} c_i \int_0^{\tau} \mathbb{1}_{\{I_t(\mathbf{x}(t))=i\}} dt \,\Bigg|\, \mathbf{x}(0) = \mathbf{x} \Bigg], \tag{5}$$

where $\mathbb{1}_{\{I_t(\mathbf{x}(t))=i\}}$ equals to one when $I_t(\mathbf{x}(t)) = i$ and zero otherwise. The allocation policy $\mathscr{I}$ not only enters into the cost term in equation (5) directly, but also influences the belief updating process, and thus $\mathbf{x}(\tau)$. Starting from any belief $\mathbf{x}$, it is obvious that the DM's expected payoff is no greater than $\max_i \overline{\pi}_i$, so the supremum of the objective function in equation (5) exists. Therefore, we can write the DM's optimal learning problem as the following.

$$V(\mathbf{x}) = \sup_{\mathscr{I}, \tau} J(\mathbf{x}; \mathscr{I}, \tau), \tag{6}$$

where $V(\mathbf{x})$ is the so-called *value function*. Similarly, starting from any time $t \geq 0$, we can define $V_t(\mathbf{x})$ as the supremum of DM's expected payoff, given his belief as $\mathbf{x}$ at time t.

Now let us formulate the DM's optimal learning problem in equation (6) as a dynamic decision problem. Given the signal generation process in equation (4), by applying Theorem 9.1 from Liptser and Shiryaev (2001), we can write down the following SDE for the belief updating process,

$$\begin{aligned} dx_i(t) &= \frac{\Delta\pi_i}{\sigma_i^2} x_i(t)\left[1 - x_i(t)\right]\left\{ds_i(t) - \mathrm{E}\left[\pi_i|\mathscr{F}_t\right] dT_i(t)\right\} \\ &= \frac{\Delta\pi_i}{\sigma_i^2} x_i(t)\left[1 - x_i(t)\right]\left\{\left[\pi_i - \underline{\pi}_i\left(1 - x_i(t)\right) - \overline{\pi}_i x_i(t)\right] dT_i(t) + \sigma_i dW_i(T_i(t))\right\}. \end{aligned} \tag{7}$$

The equation above can be understood intuitively. The first multiplicative term on the right hand side, $\frac{\Delta\pi_i}{\sigma_i^2}$, is the signal-to-noise ratio for alternative i. The second term, $x_i(t)\left[1 - x_i(t)\right]$, takes the maximal value when $x_i(t) = 1/2$, which implies that the DM will update his belief of an alternative most significantly when he is most uncertain about its payoff. On the other hand, $x_i(t)\left[1 - x_i(t)\right]$ takes the minimal value of zero when $x_i(t) = 0$ or 1, i.e., when he is most certain about the payoff. For the third term, we find that $dx_i(t) > 0$ if and only if $ds_i > \mathrm{E}\left[\pi_i|\mathscr{F}_t\right] dT_i(t)$. This implies that, a DM will update his belief upward if and only if the observation process rises faster than his current expectation. We can also see that $dx_i(t)$ does not depend on t explicitly due to time stationarity, and therefore we can drop the argument of t in the above equation. Note also that $\mathrm{E}[dx_i(t)|\mathscr{F}_t] = 0$, and $\mathrm{E}[dx_i(t)^2|\mathscr{F}_t] = \frac{(\Delta\pi_i)^2}{\sigma_i^2} x_i(t)^2[1 - x_i(t)]^2 dt$.

Based on the belief updating process, we can approximate continuous time by discrete time periods of length $dt > 0$ to obtain the following Bellmam equation that is associated with the optimal learning problem in equation (6).

$$\begin{aligned} \widehat{V}(\mathbf{x}) = \max\Big\{ &- c_1 dt + \mathrm{E}\left[\widehat{V}(x_1 + dx_1, x_2, \cdots, x_n)|\mathbf{x}\right], \cdots, \\ &-c_n dt + \mathrm{E}\left[\widehat{V}(x_1, \cdots, x_{n-1}, x_n + dx_n)|\mathbf{x}\right], \\ &\mathrm{E}[\pi_1|\mathbf{x}], \cdots, \mathrm{E}[\pi_n|\mathbf{x}], \pi_0\Big\}. \end{aligned} \tag{8}$$

where the $2n+1$ terms in the maximization correspond to the DM's $2n+1$ choices of actions respectively at any time t. By defining $g(\mathbf{x})$ as the maximum payoff that can be obtained if search were not allowed given the beliefs $\mathbf{x}$, $g(\mathbf{x}) \equiv \max\left\{\mathrm{E}[\pi_1|\mathbf{x}], \cdots, \mathrm{E}[\pi_n|\mathbf{x}], \pi_0\right\}$, the approximate Bellman equation (8) can be formally represented by the following *Hamilton-Jacobi-Bellman* (*HJB*) equation:

$$\max\left\{\max_{1\le i\le n}\left\{\frac{(\Delta\pi_i)^2}{2\sigma_i^2} x_i^2(1 - x_i)^2 \widehat{V}_{x_i x_i}(\mathbf{x}) - c_i\right\}, g(\mathbf{x}) - \widehat{V}(\mathbf{x})\right\} = 0, \tag{9}$$

where $\widehat{V}(\mathbf{x})$ is bounded, continuous, and with continuous first and second derivatives (see, for example, Karatzas 1984), and $\widehat{V}_{x_i x_i}(\mathbf{x})$ represents the second-order partial derivative of $\widehat{V}(\mathbf{x})$ with respect to x_i. Let us provide some intuition on (9). The first term in the max operator can be obtained by using equation (7), a Taylor expansion on $\widehat{V}(\mathbf{x})$ in equation (8), and Itô's lemma. It says that the DM's maximum expected payoff should be no less than the expected payoff from continuing learning one alternative. The second term of the max operator says that the DM's maximum expected payoff should be no less than the expected payoff from adopting the current best alternative. Finally, the max operator says that at any time, the DM will choose to either stop learning or continue learning. Applying the Itô's lemma to $\widehat{V}(\mathbf{x})$ we can get that $\widehat{V}(\mathbf{x})$ is an upper bound on the DM's expected payoff, $\widehat{V}(\mathbf{x}) \geq J(\mathbf{x}; \mathscr{I}, \tau)$ for any allocation policy $\mathscr{I}$, any stoping time τ, and any belief vector $\mathbf{x}$.

We will construct, under some parameter conditions, an allocation policy $\mathscr{I}^*$ and a stopping time policy τ^* that generates a value function $\widehat{V}$ that satisfies the partial differential equation (9), with $\widehat{V}(\mathbf{x})$ bounded, continuous, and with continuous first derivatives. Under these policies $\widehat{V}(\mathbf{x}) = J(\mathbf{x}; \mathscr{I}^*, \tau^*)$, $\forall \mathbf{x}$, which means the optimality of $\mathscr{I}^*$ and τ^* and we obtain the value function $V(\mathbf{x}) = \widehat{V}(\mathbf{x})$, $\forall \mathbf{x}$. To simplify the presentation, in the following, we will use the notation $V(\mathbf{x})$ to represent both $V(\mathbf{x})$ and $\widehat{V}(\mathbf{x})$, as we concentrate on the case of optimality of the allocation and stopping time policies.[5] As the solution for the value function is continuous, we obtain what is called a *viscosity solution* (see Crandall et al. 1992 for an extensive description of this type of solutions). In order to show that the viscosity solution to the *HJB* equation (9) exists and is unique, we show how the *HJB* equation (9) can be adjusted to satisfy *degenerate ellipticity* (one possible property of *HJB* equations related to some particular monotonicity conditions), and then apply the results in Lions (1983a) and Lions (1983b) to show existence and uniqueness of the value function. We state next the existence and uniqueness result, and the details and sketch of the proof are presented in the Appendix.[6]

LEMMA 1: *Under the belief updating process in equation (7), $V(\mathbf{x})$ exists and is the unique solution to the HJB equation (9).*

In the next section, we present our main results on two homogeneous alternatives ($n = 2$). We will study the case with heterogeneous alternatives in Section 5.1, and the case with more than two alternatives in Section 5.2.

3. Optimal Learning Strategy

We consider two homogeneous alternatives in this section with $\Delta\pi_i = \Delta\pi$, $\underline{\pi}_i = \underline{\pi}$, $\sigma_i = \sigma$, and $c_i = c$. To solve the DM's optimal learning problem, we will first propose a solution, and then verify that it satisfies equation (9). Then, by Lemma 1, it must be the only solution.

[5] See also Peskir and Shiryaev (2006), pp. 40-49, on how to obtain the value function in stopping time problems.

[6] We thank Wenpin Tang for this proof. Any remaining errors are our own.

Single Alternative

To solve the problem with two alternatives, it is illustrative to first work out the simpler case with only one alternative, i.e. $n = 1$. In this case, the decision maker is solving an optimal stopping problem—when to stop learning information and make an adoption decision. This problem, in similar settings, has been studied by Roberts and Weitzman (1981), Moscarini and Smith (2001), and Branco et al. (2012). For simplicity, we will drop the subscript i, since we only have one alternative to consider. The optimal learning strategy is that there exist $0 \leq \underline{x} < \overline{x} \leq 1$ such that the DM continues learning when $\underline{x} \leq x \leq \overline{x}$, and he stops learning to take the outside option when $x < \underline{x}$ or to adopt the alternative when $x > \overline{x}$. Therefore learning is preferred only when the DM's belief is in the middle range, i.e., when the DM is uncertain between the outside option and the alternative. For this one alternative case, we denote the value function as $U(x)$, so $V(\mathbf{x})$ is used only for $n \geq 2$, to make the presentation clearer.

To solve for the value function $U(x)$, let us consider three cases. In the first case, when $\underline{x} \leq x \leq \overline{x}$, equation (9) is reduced to the following second-order ordinary differential equation (ODE)

$$\frac{(\Delta\pi)^2}{2\sigma^2} x_i^2 (1-x)^2 U''(x) - c = 0. \tag{10}$$

The general solution to this ODE is

$$U(x) = \frac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x)\ln\left(\frac{1-x}{x}\right) + C_1 x + C_2, \tag{11}$$

where C_1 and C_2 are two undetermined constants. In the second case, when $x < \underline{x}$, we have $U(x) = \pi_0$. Finally, in the third case, when $x > \overline{x}$, we have $U(x) = \Delta\pi \cdot x + \underline{\pi}$. To summarize the three cases, we have the following expression for the value function

$$U(x) = \begin{cases} \pi_0, & x < \underline{x}, \\ \dfrac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x)\ln\left(\dfrac{1-x}{x}\right) + C_1 x + C_2, & \underline{x} \leq x \leq \overline{x}, \\ \Delta\pi \cdot x + \underline{\pi}, & x > \overline{x}. \end{cases} \tag{12}$$

To complete the computation of the optimal policy, we need a set of boundary conditions to determine $\underline{x}$, $\overline{x}$, C_1 and C_2. In fact, $U(x)$ has to be continuous at $\underline{x}$ and $\overline{x}$, which implies the following so-called *value matching* conditions,

$$U(\underline{x}^+) = U(\underline{x}^-), \tag{13}$$

$$U(\overline{x}^+) = U(\overline{x}^-), \tag{14}$$

where $U(\underline{x}^+)$ denotes the right limit of $U(x)$ at $x = \underline{x}$, and $U(\underline{x}^-)$ denotes the left limit of $U(x)$ at $x = \underline{x}$. We also have that $U(x)$ has to be smooth at $\underline{x}$ and $\overline{x}$ (i.e., $U'(x)$ has to be continuous at $\underline{x}$ and $\overline{x}$), which implies the following so-called *smooth pasting* conditions,

$$U'(\underline{x}^+) = U'(\underline{x}^-), \tag{15}$$

$$U'(\overline{x}^+) = U'(\overline{x}^-). \tag{16}$$

Value matching and smooth pasting are standard conditions for optimal stopping problems of diffusion processes. See, e.g., Dixit (1993), Dixit and Pindyck (1994), Chapter 9.1 of Peskir and Shiryaev (2006) and Strulovici and Szydlowski (2015) for more discussion. We will provide more details on these smoothness conditions when we analyze the case with two alternatives below.

By combining equations (12)-(16), we can determine $\underline{x}$, $\overline{x}$, C_1 and C_2, and thus determine $U(x)$. Unfortunately, this set of equations cannot be solved explicitly, and $\underline{x}$, $\overline{x}$ can only be determined implicitly by the following set of equations,

$$\left(\frac{1}{\overline{x}} + \frac{1}{1-\overline{x}}\right) - \left(\frac{1}{\underline{x}} + \frac{1}{1-\underline{x}}\right) = -\frac{(\Delta\pi)^2(\Delta\pi + 2\underline{\pi} - 2\pi_0)}{2\sigma^2 c}, \tag{17}$$

$$\Phi(\overline{x}) - \Phi(\underline{x}) = -\frac{(\Delta\pi)^3}{2\sigma^2 c}, \tag{18}$$

where,

$$\Phi(x) \equiv 2\ln\left(\frac{1-x}{x}\right) + \frac{1}{x} - \frac{1}{1-x}, \tag{19}$$

which is a strictly decreasing function in $(0, 1)$. By equation (18), we immediately get that $\overline{x} > \underline{x}$. The constants C_1 and C_2 can be explicitly expressed in terms of $\overline{x}$ and $\underline{x}$.

Next, we will proceed to analyze the optimal learning problem with two homogeneous alternatives. We find that the solution structure critically depends on the value of the outside option. We will start with the case with sufficiently high outside option. This is a relatively simpler case which allows us to clearly demonstrate the solving techniques. Then, we will also fully characterize the case with sufficiently low outside option, and finally, we will present some results on the complicated case with intermediate outside option.

Two Alternatives with High-Value Outside Option

In this subsection, we consider the case of two homogeneous alternatives with the value of the outside option sufficiently high. We will propose a solution, and then verify that it satisfies the *HJB* equation (9). After that, we will come back and determine the exact threshold for the outside option to be considered sufficiently high.

First, we note that the two alternatives are symmetric in the belief space of x_1-x_2, so we only need to consider the case with $x_1 \geq x_2$ below. We consider two cases: In the first case, when $x_2 \leq \underline{x}$, we propose that alternative 2 will not be considered for either learning or adoption, and one goes back to the case with a single uncertain alternative—alternative 1. Correspondingly, we have

$$V(x_1, x_2) = U(x_1), \text{ if } x_2 \leq \underline{x}. \tag{20}$$

In the second case, when $x_1 \geq x_2 > \underline{x}$, we propose that there exists a smooth function $\overline{X}(\cdot)$ such that the DM learns alternative 1 when $x_2 \leq x_1 \leq \overline{X}(x_2)$, and adopts alternative 1 when $x_1 > \overline{X}(x_2)$. Intuitively, we call $\overline{X}(x_2)$ the *adoption boundary* for alternative 1. Correspondingly, by applying similar analysis for the single-alternative case above, we have $V(x_1, x_2)$ that satisfies the following partial differential equation (PDE) when $x_2 \leq x_1 \leq \overline{X}(x_2)$,

$$\frac{(\Delta\pi)^2}{2\sigma^2} x_1^2(1-x_1)^2 V_{x_1x_1}(x_1, x_2) - c = 0. \tag{21}$$

Thanks to its simple parabolic form, equation (21) has the following general solution,

$$V(x_1, x_2) = \frac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x_1)\ln\left(\frac{1-x_1}{x_1}\right) + C_1(x_2)x_1 + C_2(x_2), \text{ if } x_2 \leq x_1 \leq \overline{X}(x_2), \tag{22}$$

where $C_1(\cdot)$ and $C_2(\cdot)$ are two undetermined functions.[7] On the other hand, when $x_1 > \overline{X}(x_2)$, we have

$$V(x_1, x_2) = \Delta\pi \cdot x_1 + \underline{\pi}, \text{ if } x_1 > \overline{X}(x_2). \tag{23}$$

Thus far we have proposed a solution of $V(x_1, x_2)$ for each region in the belief space of x_1-x_2. Next, along the same line of argument of the value matching and smooth pasting conditions in equations (13)-(16), we will consider all the boundary conditions connecting all the regions. It remains an open question to prove the necessity of the smoothness conditions for a two-dimensional mixed optimal stopping and optimal control problem like ours.[8] Fortunately, we do not need to prove the

[7] When $x_1 = 1$, $V(x_1, x_2)$ is not well defined, but the optimal learning problem is trivial. Under $x_1 = 1$, the payoff of alternative 1 is known to be high, therefore, the DM's optimal strategy is to adopt alternative 1 immediately, and $V(x_1, x_2) = \overline{\pi}$. Similarly we can solve the case where $x_2 = 1$. For the sake of convenience, we only consider the case that $x_1 < 1$ and $x_2 < 1$ below. Later, we will show that $V(x_1, x_2)$ is continuous at $x_i = 1$.

[8] See Strulovici and Szydlowski (2015) for recent progress on smoothness conditions for one-dimensional mixed optimal stopping and optimal control problems. Peskir and Shiryaev (2006), p.144, note that for optimal stopping problems if the boundary is Lipschitz-continuous, then the smooth pasting conditions will hold at the optimum. Caffarelli (1977), Theorem 2, shows that if the payoff for stopping is continuously differentiable, which occurs in this case, then the boundary is Lipschitz-continuous. Caffarelli (1977), Theorem 3, also shows that if the payoff of stopping is continuously differentiable, then the boundary is continuously differentiable. In our case, that means that

necessity of these conditions, since we only use them to guess a solution. We follow Chernoff (1968) and Chapter 9.1 of Peskir and Shiryaev (2006) to provide in the Appendix an intuitive argument on the smooth pasting conditions based on a Taylor expansion.

First, we require $V(x_1,x_2)$, $V_{x_1}(x_1,x_2)$ and $V_{x_2}(x_1,x_2)$ to be continuous at the adoption boundary $x_1=\overline{X}(x_2)$. By substituting equations (22) and (23) into these three conditions, we find that one of them is redundant (see also Mandelbaum et al. 1990, p. 1016, for a similar result), and the three conditions are equivalent to the following two equations,

$$
\begin{aligned}
C_1(x) &= \frac{4\sigma^2 c}{(\Delta\pi)^2}\ln\left(\frac{1-\overline{X}(x)}{\overline{X}(x)}\right)+\frac{2\sigma^2 c}{(\Delta\pi)^2}\frac{1-2\overline{X}(x)}{\overline{X}(x)\left(1-\overline{X}(x)\right)}+\Delta\pi && (24)\\
C_2(x) &= -\frac{2\sigma^2 c}{(\Delta\pi)^2}\ln\left(\frac{1-\overline{X}(x)}{\overline{X}(x)}\right)-\frac{2\sigma^2 c}{(\Delta\pi)^2}\frac{1-2\overline{X}(x)}{1-\overline{X}(x)}+\underline{\pi}. && (25)
\end{aligned}
$$

The redundancy of one condition is not a mere coincidence. Instead, it is due to the learning problem structure that the DM can learn only one alternative at a time. Consequently, only one Brownian motion can move at a time and we have a parabolic PDE in equation (21).[9] In the Appendix, we show that due to the parabolic PDE, we will always end up with one redundant condition.

By substituting the expressions of $C_1(x)$ and $C_2(x)$ in equations (24) and (25) into equation (22), we have

$$
\begin{aligned}
V(x_1,x_2) &= \frac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right)-\ln\left(\frac{1-\overline{X}(x_2)}{\overline{X}(x_2)}\right)\right]\\
&- \frac{2\sigma^2 c}{(\Delta\pi)^2}\frac{1-2\overline{X}(x_2)}{1-\overline{X}(x_2)}\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)}+\Delta\pi\cdot x_1+\underline{\pi},\ \text{if } x_2\le x_1\le\overline{X}(x_2). && (26)
\end{aligned}
$$

Second, we require $V(x_1,x_2)$, $V_{x_1}(x_1,x_2)$ and $V_{x_2}(x_1,x_2)$ to be continuous at the boundary $x_1=x_2$. Due to the symmetry of $V(x_1,x_2)$ in x_1-x_2 space, it is not difficult to recognize that the only necessary boundary condition is the continuity of $V_{x_1}(x_1,x_2)$. Based on equation (26), we can

$\overline{X}(x_2)$ is continuously differentiable.

[9]We could generalize our learning problem to allow the DM at any time, to allocate a fraction of his effort to learn alternative 1 and the complementary fraction of the effort to learn alternative 2. If we impose a capacity constraint on the total number of "signals" per unit of time from the two alternatives, we can show that the optimal learning strategy will be exactly the same as that in our main model, i.e., the DM will *optimally* allocate his effort to one alternative at a time. That is to say, sequential learning is optimal. On the other hand, if we do not impose this capacity constraint on the number of total signals per time, depending on the model setting, we may end up with the case where it is optimal for the DM to learn multiple alternatives simultaneously. This would correspond to an elliptic PDE instead of the parabolic PDE that occurs here.

simplify this boundary condition as the following ODE,

$$\overline{X}'(x) = -\frac{\Phi\left(\overline{X}(x)\right) - \Phi(x) + (\Delta\pi)^3/(2\sigma^2 c)}{\Phi'\left(\overline{X}(x)\right)\left(\overline{X}(x) - x\right)}. \tag{27}$$

Third, we also require $V(x_1, x_2)$, $V_{x_1}(x_1, x_2)$ and $V_{x_2}(x_1, x_2)$ to be continous at $x_2 = \underline{x}$. We can show that this set of three boundary conditions are equivalent to

$$\overline{X}(\underline{x}) = \overline{x}. \tag{28}$$

$\overline{X}(x)$ can then be determined by combining equations (27) and (28) as a boundary value problem of an ODE. We are unable to solve the ODE analytically, but we prove its existence and uniqueness, and some useful properties of the solution in the following lemma.

LEMMA 2: *There exists a unique continuous function $\overline{X}(x)$ in $[\underline{x}, 1]$ that solves the boundary value problem of an ODE in equations (27) and (28). $\overline{X}(x)$ is smooth, increasing, and $\overline{X}(x) \geq x$ with $\overline{X}(1^-) = 1$.*

Thus far we have considered all the boundary conditions which, together with the expressions of $V(x_1, x_2)$ given by equations (20), (23) and (26), entirely characterize the solution to equation (9). Lastly, to verify this is indeed the solution, we need to guarantee that the DM prefers to learn alternative 1 for $\underline{x} \leq x_2 \leq x_1 \leq \overline{X}(x_2)$. By equation (9) and (21), we require that

$$x_2^2(1 - x_2)^2 V_{x_2 x_2}(x_1, x_2) \leq \frac{2\sigma^2 c}{(\Delta\pi)^2}, \text{ for } \forall \underline{x} \leq x_2 \leq x_1 \leq \overline{X}(x_2). \tag{29}$$

Without solving $\overline{X}(x)$, it is difficult to translate the inequality above into expressions in terms of model primitives. Nevertheless, we are able to prove the following necessary and sufficient condition for inequality (29).

LEMMA 3: *The DM prefers learning alternative 1 for $\underline{x} \leq x_2 \leq x_1 \leq \overline{X}(x_2)$, if and only if*

$$\overline{X}'(x_2) \leq \frac{\overline{X}(x_2)\left(1 - \overline{X}(x_2)\right)}{x_2\left(1 - x_2\right)}, \text{ for } \forall x_2 \in [\underline{x}, 1). \tag{30}$$

There exists a $x_0^* < 1/2$, such that this condition holds if and only if $x_0 \equiv (\pi_0 - \underline{\pi})/(\Delta\pi) \geq x_0^*$.

The variable x_0 measures the relative value of the outside option compared with the uncertain alternatives. This lemma implies that the solution that we just proposed exists only when the outside option is sufficiently high.

Finally, we have the following theorem to summarize everything so far for this subsection on the value function (the optimal learning strategy is characterized in Theorem 2 below). The existence and uniqueness of the solution has been established in Lemma 1 and, based on the way we construct the solution, we have already verified that it satisfies equation (9).

THEOREM 1: *Consider the decision maker's optimal learning problem in equation (6) with two homogeneous alternatives. When the value of the outside option is relatively high, i.e., $x_0 \geq x_0^*$, the value function $V(x_1, x_2)$ is given by*

$$V(x_1,x_2)=\begin{cases} \frac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right)-\ln\left(\frac{1-\overline{X}(x_2)}{\overline{X}(x_2)}\right)\right] & \\ -\frac{2\sigma^2 c}{(\Delta\pi)^2}\frac{1-2\overline{X}(x_2)}{1-\overline{X}(x_2)}\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)}+\Delta\pi\cdot x_1+\underline{\pi}, & \text{if } x_2\leq x_1\leq \overline{X}(x_2), x_1\geq \underline{x} \\ \frac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x_2)\left[\ln\left(\frac{1-x_2}{x_2}\right)-\ln\left(\frac{1-\overline{X}(x_1)}{\overline{X}(x_1)}\right)\right] & \\ -\frac{2\sigma^2 c}{(\Delta\pi)^2}\frac{1-2\overline{X}(x_1)}{1-\overline{X}(x_1)}\frac{\overline{X}(x_1)-x_2}{\overline{X}(x_1)}+\Delta\pi\cdot x_2+\underline{\pi}, & \text{if } x_1\leq x_2\leq \overline{X}(x_1), x_2\geq \underline{x} \\ \Delta\pi\cdot x_1+\underline{\pi}, & \text{if } x_1>\overline{X}(x_2) \\ \Delta\pi\cdot x_2+\underline{\pi}, & \text{if } x_2>\overline{X}(x_1) \\ \pi_0, & \text{otherwise,} \end{cases} \tag{31}$$

where $\overline{X}(x)$ is determined by the boundary value problem in equations (27) and (28) for $x \in [\underline{x}, 1]$, and $\overline{X}(x) = \overline{x}$ for $x \in [0, \underline{x}]$.

The function $\overline{X}(\cdot)$ was defined only on the support $[\underline{x}, 1)$ when first introduced. In the theorem above, we have abused the notation by expanding its support to $[0, 1]$. By defining $\overline{X}(x) \equiv \overline{x}$ for $x \in [0, \underline{x})$ and $\overline{X}(1) \equiv 1$, we are able to define the adoption boundary and write down the value function in a uniform way. It is easy to verify that $\overline{X}(x)$ is still smooth, increasing, and $\overline{X}(x) \geq x$ for $x \in [0, 1]$.

Because $x_0^* < 1/2$ and $x_0 = (\pi_0 - \underline{\pi})/\Delta\pi$, we have the following corollary.

COROLLARY 1: *Consider the decision maker's optimal learning problem in equation (6) with two homogeneous alternatives. A sufficient condition for the function defined by (31) to be the value function is that $\pi_0 \geq (\underline{\pi} + \overline{\pi})/2$.*

Figure 1 shows the value function $V(x_1, x_2)$, as well as the payoff from learning $L(x_1, x_2)$ under some parameter setting. The payoff from learning is defined as $L(x_1, x_2) \equiv V(x_1, x_2) - \max\{\Delta\pi \cdot x_1 + \underline{\pi}, \Delta\pi \cdot x_1 + \underline{\pi}, \pi_0\}$, which is basically the difference between a DM's value function and his expected payoff in the case that learning is not allowed. As expected, $V(x_1, x_2)$ is increasing and convex and $L(x_1, x_2)$ is always non-negative. $L(x_1, x_2)$ peaks at $x_1 = x_2 = x_0$, at which point the

DM is most uncertain between adopting alternative i and taking the outside option. This is the point where the DM benefits most from learning.

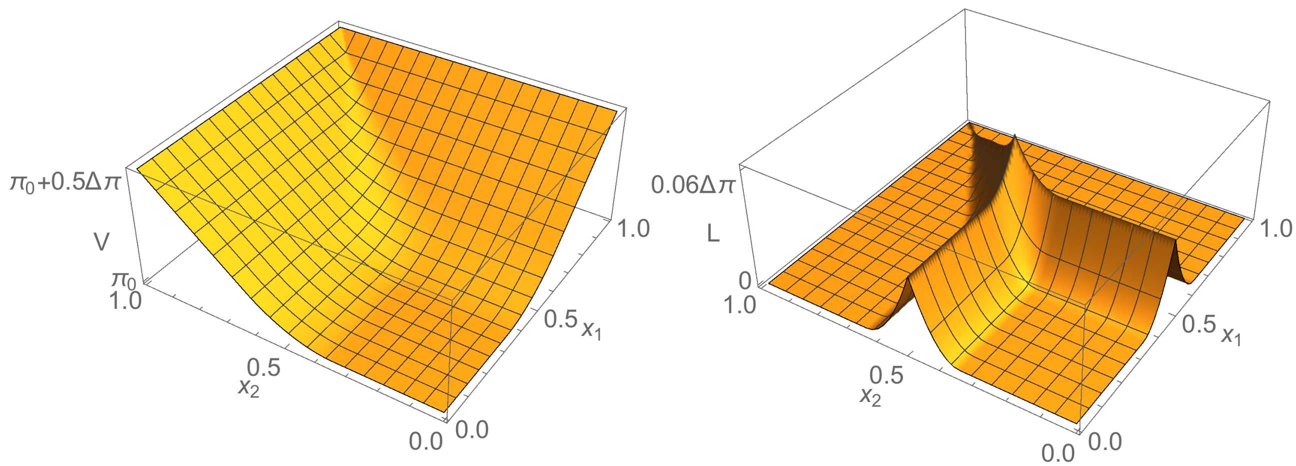

Figure 1: Maximum expected payoff (left panel), payoff from learning (right panel), given a DM's current belief, with relatively high value of outside option, $x_0 = 0.5$. $c\sigma^2/\Delta\pi = 0.1$.

Based on the value function given by Theorem 1, we can construct the optimal learning strategy that achieves the value function, and the corresponding belief updating process under the optimal strategy. Figure 2 shows the optimal learning strategy under some parameter setting. The belief space is divided into five regions by black lines, corresponding to the DM's five choices of actions. The optimal learning strategy shown in Figure 2 can be understood intuitively. When the DM's belief of alternative i is quite high, it is optimal to adopt the alternative immediately without learning. When the DM's beliefs of both alternatives are quite low, it is optimal to take the outside option. Roughly speaking, learning becomes preferred when the DM's beliefs of the two alternatives are similar or take medium values.

Note that when $x_i < \underline{x}$, the DM decides only among learning j, adopting j, and taking the outside option ($j \neq i$), and he never considers alternative i for learning or adoption. This means the DM optimally uses a simple *cutoff strategy* when forming a *consideration set* of alternatives for learning or adoption. Given both alternatives in his consideration set, i.e., $x_1, x_2 \geq \underline{x}$, the DM always prioritizes learning about the alternative with higher belief, and adopts it when the belief is sufficiently high. As we will show below, this learning strategy critically depends on the assumption that the value of the outside option is sufficiently high.

Formally, the following theorem summarizes the DM's optimal learning strategy. In the Appendix, we construct the belief updating process under the optimal learning strategy, which com-

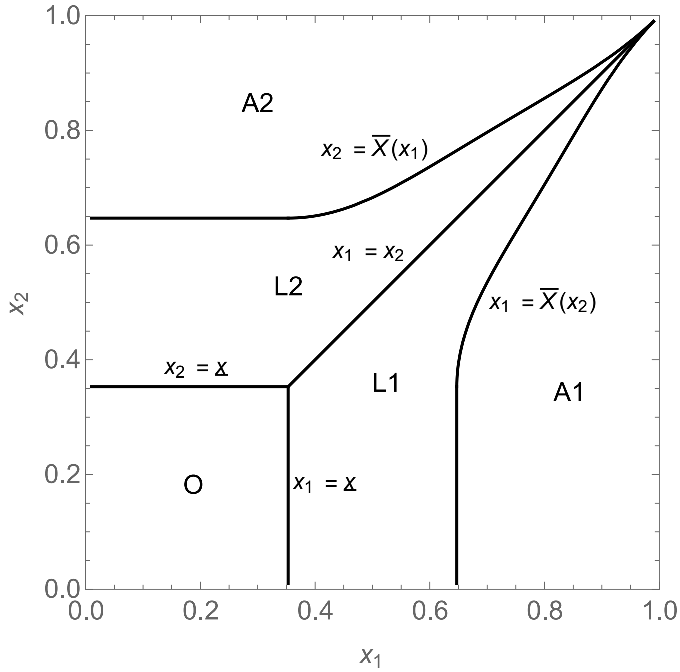

Figure 2: Optimal learning strategy with relatively high value of the outside option, $x_0 = 0.5$, and $c\sigma^2/\Delta\pi = 0.1$.

pletes the analysis to show that the value function $V(x_1, x_2)$ given by Theorem 1 is attainable. Following Karatzas (1984), we show that the belief updating process has a weak, unique in distribution, solution.[10] As the value function is obtained by taking an expected value over the changes in beliefs, uniqueness in distribution allows us to obtain an unique value function. Furthermore, we show that during the learning process, $\min\{x_1, x_2\}$ falls through time with positive probability and that it never increases. This means that starting from a prior belief in the regions that it is optimal to learn, there is a positive probability of the outside option being chosen.

To see how $\min\{x_1, x_2\}$ falls through time with positive probability and it never increases, note that $\min\{x_1, x_2\} = \frac{1}{2}\left(x_1 + x_2 - |x_1 - x_2|\right)$. By Tanaka's formula (e.g., Karatzas and Shreve 1991, p.205) one can obtain $|x_1(t) - x_2(t)| = |x_1(0) - x_2(0)| + x_1(t) + x_2(t) - x_1(0) - x_2(0) + 2L^{x_1 - x_2}(t)$ where $L^{x_1-x_2}(t)$ is the local time of $x_1(t) - x_2(t)$ at $x_1(t) = x_2(t)$ and is defined as $L^{x_1-x_2}(t) \equiv \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{\{0 \le x_1(s) - x_2(s) < \varepsilon\}} ds$. We can then obtain that $\min\{x_1(t), x_2(t)\} = \min\{x_1(0), x_2(0)\} - L^{x_1-x_2}(t)$, which shows that $\min\{x_1, x_2\}$, conditional on the DM being in the region where it is optimal to learn more information, falls through time under the optimal learning strategy. The Appendix presents further details.

[10] As discussed in the Appendix, if the belief updating processes had equal constant variance through time across alternatives (which is not the case here), and no trend, then we would know that the learning strategy presented here would have a strong, pathwise unique solution, using the results in Fernholz et al. (2013).

Theorem 2: *When the value of the outside option is relatively high, i.e., $x_0 \geq x_0^*$, a decision maker considers alternative i for learning or adoption if and only if his belief of i is sufficiently high, i.e., $x_i \geq \underline{x}$. Given one alternative i under his consideration, the DM keeps learning it if $\underline{x} \leq x_i \leq \overline{x}$, adopts it if $x_i > \overline{x}$, and takes the outside option if $x_i < \underline{x}$. Given two alternatives under his consideration, the DM always learns information from the alternative with higher belief.*[11] *He stops learning and adopts it if and only if his belief of this alternative is above a threshold. This threshold (i.e., the adoption boundary) increases with the DM's belief of the other alternative, and gets closer to the line $x_1 = x_2$ as x_2 increases, with the threshold converging to x_1 as $x_2 \to 1$.*

It is interesting to revisit the index policy proposed by Glazebrook (1979) as the optimal policy for stoppable bandits problem under some conditions. Under the index policy, a DM chooses alternative i to act on (either learn or adopt) if it has the greatest index, which, like the Gittins index, only depends on the state of alternative i. Given that alternative i is chosen, the DM chooses to learn or to adopt it based on some strategy that only depends on the state of alternative i. Here, we have shown that the DM chooses to stop learning and adopt alternative i if and only if $x_i \geq \overline{X}(x_j)$, which depends on x_j as well. Therefore, Glazebrook's index policy is suboptimal for our problem.

It is also interesting to compare our optimal learning strategy with the Pandora's rule in Weitzman (1979), which consists of a selection rule and a stopping rule. The selection rule states that it is optimal to learn the alternative with the highest reservation value; the stopping rule states that it is optimal to terminate learning whenever the maximum sampled alternative exceeds the reservation value of every unsampled alternative. In our setting, by Theorem 2, we know that it is optimal for the DM to learn an alternative if its belief is higher than the other alternative and sufficiently high relative to the outside option. Therefore, by properly defining the reservation value for each alternative including the outside option, the selection rule in Weitzman (1979) can still apply to our setting. However, the stopping rule in Weitzman (1979) will not apply to our setting in general, because in our setting, it is optimal to stop learning and adopt alternative i if and only if $x_i \geq \overline{X}(x_j)$, which not only depends on x_j but also depends on the value of the outside option, π_0. Thus, the stopping rule in our setting is not a simple pairwise comparison of two alternatives.[12] Furthermore, we will show below that, when the outside option is sufficiently low, both selection and stopping rules in Weitzman (1979) will fail for our setting.

[11] In the case of equal beliefs, the DM is indifferent about which alternative to search, and we assume that the DM learns alternative 1.

[12] This is also in contrast with Doval (2014), who considers optional learning in the setting of Weitzman (1979) and finds that the rule derived there remains optimal when the outside option is sufficiently high (see Proposition 1 in Doval 2014).

Lastly, we can compute the comparative statics of the optimal learning strategy with respect to the learning costs, noise of the signals, and value of the outside option. The results are intuitive.

PROPOSITION 1: *As learning cost c increases, $\underline{x}$ increases, $\overline{X}(x)$ decreases, and $V(x_1, x_2)$ decreases. As signal noise σ increases, $\underline{x}$ increases, $\overline{X}(x)$ decreases, and $V(x_1, x_2)$ decreases; as $\underline{\pi}$ increases, $\underline{x}$ decreases, $\overline{X}(x)$ decreases, and $V(x_1, x_2)$ increases; as π_0 increases, $\underline{x}$ increases, $\overline{X}(x)$ increases, and $V(x_1, x_2)$ increases.*

Two Alternatives with Low-Value Outside Option

In this subsection, we consider the opposite case where the value of the outside option is sufficiently low such that there is essentially no outside option, i.e., $\pi_0 \leq \underline{\pi}$, which means $x_0 \leq 0$. Similarly, we propose a solution and verify that it satisfies the *HJB* equation (9). The following theorem presents the solution to the DM's optimal learning problem, with proof and value function presented in the Appendix.

THEOREM 3: *Consider the decision maker's optimal learning problem in equation (6) with two homogeneous alternatives, and suppose that the value of the outside option is relatively low, i.e., $x_0 \leq 0$. Then there is a function $\overline{X}(x)$ with $\overline{X}'(x) > 0, \overline{X}(x) > x$ for all $x \in (0, 1)$, $\overline{X}(0) = 0$, and $\overline{X}(1) = 1$, such that: (1) the DM adopts alternative i if $x_i \geq \overline{X}(x_{3-i})$ and continues learning if that condition does not hold for any alternative; (2) When learning, the DM learns about the alternative with the higher belief if $x_1 + x_2 \geq 1$, and learns about the alternative with the lower belief if $x_1 + x_2 < 1$.*

The optimal learning strategy can be understood intuitively by Figure 3. We find that it is optimal for the DM to learn the worse alternative when $x_1 + x_2 \leq 1$, as shown in the starred regions. If this alternative turns out to be really poor, the DM will adopt the other alternative immediately; otherwise, if this alternative turns out to be not as poor as thought earlier, the DM will switch to learn the other alternative.

It is interesting to understand why sometimes it is optimal for a DM to prioritize learning on the worse alternative. Let us consider a DM with belief in L2* in Figure 3. One may suspect that while alternative 2 has a lower expected payoff than alternative 1, the DM may be more uncertain about alternative 2 and thus prefers to learn alternative 2 first. This speculation turns out to be incorrect. In fact, given (x_1, x_2) in L2*, we have $x_1 > x_2$ and $x_1 + x_2 \leq 1$, which implies that

$$\mathrm{E}[\pi_1] = \Delta\pi x_1 + \underline{\pi} \quad > \quad \Delta\pi x_2 + \underline{\pi} = \mathrm{E}[\pi_2], \tag{32}$$

$$\mathrm{Var}[\pi_1] = (\Delta\pi)^2 x_1(1 - x_1) \quad \geq \quad (\Delta\pi)^2 x_2(1 - x_2) = \mathrm{Var}[\pi_2]. \tag{33}$$

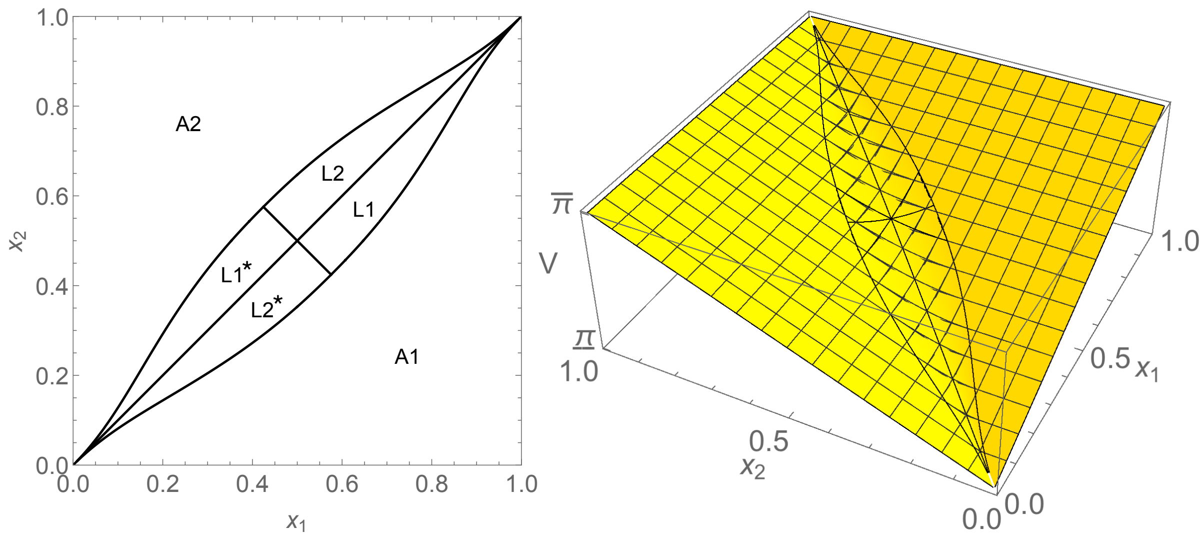

Figure 3: Optimal learning strategy with relatively low value of the outside option, $x_0 \leq 0$, and $c\sigma^2/\Delta\pi = 0.1$.

Therefore, the DM not only expects a higher payoff from alternative 1 but also is more uncertain about the payoff of alternative 1.

It turns out that the reason for the DM to learn a worse and more certain alternative first is to save expected learning costs. Here is the intuition. Given the DM's belief (x_1, x_2) in region L2*, if he learns alternative 2, he will either end the learning process by adopting alternative 1 if he accumulates enough bad news on alternative 2, or continue to learn alternative 1 if he accumulates enough good news on alternative 2. As the DM aims to save learning costs (the outside option is dominated, by assumption), he would like to end the learning process as soon as possible, so he cares about the probability of getting bad news on alternative 2, which is equal to $1 - x_2$. On the other hand, if the DM instead chooses to learn alternative 1, he will either end the learning process by adopting alternative 1 if he accumulates enough good news on alternative 1, or continue to learn alternative 2 if he accumulates enough bad news on alternative 1. With the aim of saving learning costs, the DM would like to end the learning process as soon as possible, so he cares about the probability of getting good news on alternative 1, which is equal to x_1. To minimize future learning costs, the DM chooses to learn alternative 2 if $1 - x_2 \geq x_1$, i.e., $x_1 + x_2 \leq 1$, which is exactly the upper-right boundary condition for L2*.[13] To summarize, in the stared regions, the DM finds it optimal to learn the worse alternative first with the possibility of ruling it out early so as to choose the other alternative immediately.

Lastly, it is interesting to revisit the optimal learning rule under "complete learning" in Weitzman

[13]We have also analyzed a two-period discrete time learning problem, and find that the intuition of saving learning costs here becomes the exact condition there. The analysis is available upon request.

(1979), where everything learnable about an alternative is learned with the first signal. For complete learning, one can show that the optimal rule is to learn about the alternative with the highest belief if everything else is the same. Our result highlights that with the possibility of gradual learning, a decision maker's optimal learning strategy can be quite different and complex—sometimes, it is not optimal to learn the best alternative first. It is true that under complete learning, it can be optimal for the DM to first learn an alternative that has low expected payoff but high variance, in the hopes of capturing its high payoff. Strictly speaking, in that possibility, the alternative with low expected payoff is "seemingly" worse, but is, in fact, different from the other alternatives in terms of different payoff distributions. In contrast, we find that a DM may optimally learn a worse alternative first, given everything else between the two alternatives the same in this binary environment. Note also that this worse alternative is also the alternative with lower variance, so that the DM may optimally learn an alternative with lower expected value and lower variance. To put it more simply, in our setting in region L2*, alternative 2 is indeed worse than alternative 1 and the DM optimally chooses to learn alternative 2 first. Also, our underlying mechanism works in a different way. With complete learning, after learning an alternative with low expected payoff and high variance, the DM will either adopt that alternative if he gets a very positive signal, or learn other alternative if he gets a negative signal. In contrast, in our setting with gradual learning, when learning alternative 2, the DM will either potentially switch to learning alternative 1 if he gets enough positive signals on alternative 2, or adopt alternative 1 immediately if he gets enough negative signals on alternative 2.

Two Alternatives with Medium-Value Outside Option

So far, we have completely solved the optimal learning problem for the cases where $x_0 \geq x_0^*$ and $x_0 \leq 0$. For the intermediate case with $0 < x_0 < x_0^*$, we are not able to characterize the value function analytically, but can compute it numerically. Figure 4 illustrates the optimal policy of the learning problem under some parameter setting where the outside option takes a medium value. The optimal learning strategy becomes more complex.[14] For both high and low beliefs, it is optimal for the DM to learn the better alternative first; however, for the medium beliefs, it is optimal for the DM to learn the worse alternative first. For low beliefs, it is optimal for the DM to learn about the superior alternative because the beliefs about the worse alternative could fall below the consideration set threshold, $\underline{x}$.

By comparing the optimal learning strategies under high, medium, and low value of the outside options, we can evaluate numerically that the higher the outside option, the smaller the stared

[14]The features of this more complex learning strategy can also be obtained analytically in a two-period model.

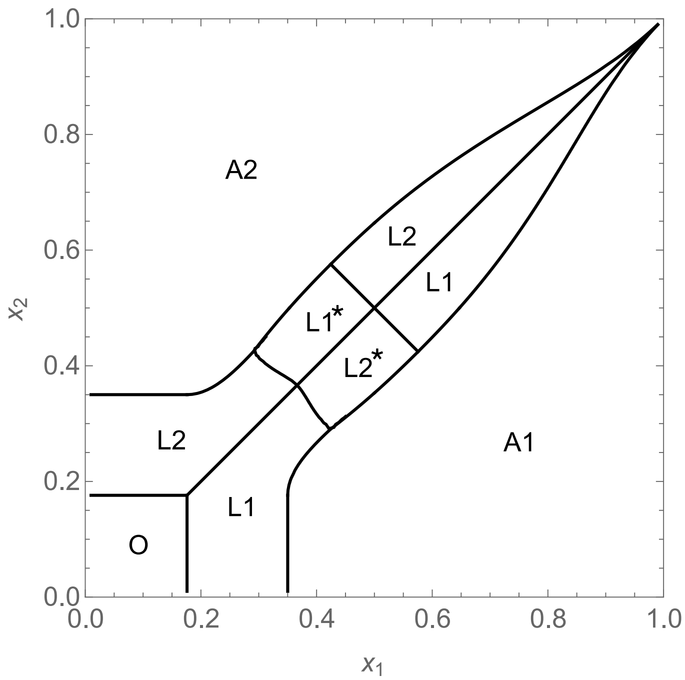

Figure 4: Optimal learning strategy with medium value of the outside option, $x_0 = 0.25$, and $c\sigma^2/\Delta\pi = 0.1$.

regions. That is to say, a good outside option makes the better alternative more attractive for learning. The reason is that with a good outside option, if the DM learns the better alternative and ends up with bad news on it, he can use the outside option instead of the worse alternative as the backup option. In contrast, if the DM chooses to learn the worse alternative, the outside option may not be useful as a backup option, since it can be dominated by the better alternative.

4. Adoption Likelihood and Probability of Being Correct

In this section, we continue the analysis on the case with two homogeneous alternatives and a high-value outside option, i.e., $x_0 \geq x_0^*$. We derive a decision maker's adoption likelihood of each alternative as well as expected probability of being ex-post correct, based on his optimal learning strategy.

It is interesting to understand the adoption likelihood in many applications. For example, in product markets, at the aggregated level, the adoption likelihood corresponds to the demand function of multiple products under consumer learning. Given a DM's current posterior beliefs as x_1 and x_2, we denote his adoption likelihood of alternative i as $P_i(x_1, x_2)$. By symmetry, we have $P_2(x_1, x_2) = P_1(x_2, x_1)$, therefore, we only need to focus on $P_1(x_1, x_2)$ below. We calculate $P_1(x_1, x_2)$ by invoking the *Optional Stopping Theorem* for martingales, and solving an ordinary differential equation. The adoption likelihood is presented in the following theorem, with proof in the Appendix.

THEOREM 4: *Consider two homogeneous alternatives. Under the optimal learning strategy in the case that $x_0 \geq x_0^*$, given his current posterior beliefs as (x_1, x_2), a decision maker's probability of adopting alternative 1 is*

$$P_1(x_1,x_2) = \begin{cases} 0, & \text{if } x_1 \leq \underline{x} \text{ or } x_2 \geq \overline{X}(x_1) \\ \frac{1}{2}\frac{\overline{X}(x_1)-x_2}{\overline{X}(x_1)-x_1}\left[1-e^{-\int_{\underline{x}}^{x_1}\frac{2d\xi}{\overline{X}(\xi)-\xi}}\right], & \text{if } \underline{x} < x_1 \leq x_2 < \overline{X}(x_1) \\ \frac{x_1-x_2}{\overline{X}(x_2)-x_2}+\frac{1}{2}\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)-x_2}\left[1-e^{-\int_{\underline{x}}^{x_2}\frac{2d\xi}{\overline{X}(\xi)-\xi}}\right], & \text{if } \underline{x} < x_2 < x_1 < \overline{X}(x_2) \\ \frac{x_1-\underline{x}}{\overline{x}-\underline{x}}, & \text{if } \underline{x} < x_1 < \overline{x} \text{ and } x_2 \leq \underline{x} \\ 1, & \text{if } x_1 \geq \overline{X}(x_2). \end{cases} \tag{34}$$

The left panel in Figure 5 presents $P_1(x_1, x_2)$ under some parameter setting. From the figure, we can see the intuitive result that a DM is more likely to adopt an alternative if his belief of the alternative is higher, or his belief of the other alternative is lower.

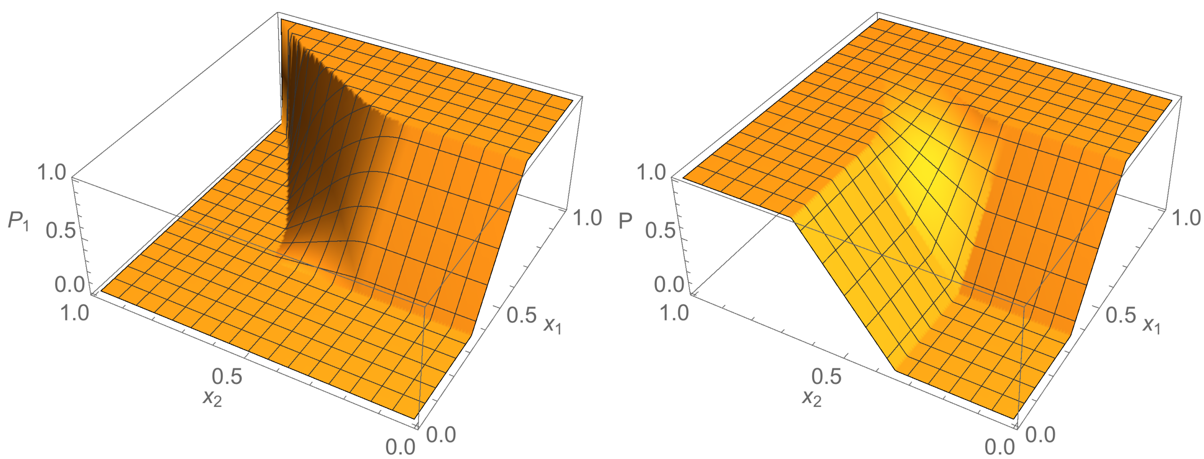

Figure 5: Adoption likelihood of alternative 1 (left panel) and either one alternative (right panel), given a DM's current belief, under the parameter setting that $x_0 = 0.5$ and $c\sigma^2/\Delta\pi = 0.1$.

We are also interested in the adoption likelihood of (either) one alternative, which is defined as $P(x_1, x_2) \equiv P_1(x_1, x_2) + P_2(x_1, x_2)$. The right panel in Figure 5 presents $P(x_1, x_2)$ under some parameter setting. It is interesting to note that $P(x_1, x_2)$ does not always increase with x_1 or x_2. This means that a higher belief of one alternative may lead to a lower adoption likelihood of the two alternatives combined. To understand the intuition, let us consider a special case. Given a DM's beliefs of the two alternatives as x_1 and x_2, if x_1 is high enough such that it exceeds the adoption boundary $\overline{X}(x_2)$, the DM will adopt alternative 1 immediately. In this case, the adoption likelihood is one. Now consider an exogenous increase of x_2 such that x_1 is now below the adoption boundary $\overline{X}(x_2)$. In this case, the DM will optimally learn more information before making an adoption decision. After obtaining more information, it is possible that the DM likes the alternatives more, in which case, he will adopt at most one of them; it is also possible that the DM gets some negative signals and decides to take the outside option. In general, the adoption likelihood will be lower than one after the increase of x_2.

The introduction of a new alternative can be equivalently viewed as increasing its belief from below $\underline{x}$ to some level above $\underline{x}$. By the same argument, we can show that more alternatives available for learning and adoption may decrease the adoption likelihood. Applying this result to product markets, this provides a rational explanation to *consumer choice overload*,[15] under the circumstance that a consumer engages in costly information gathering before making a choice. More options to choose from may lead a consumer to exert a greater effort to distinguish the best from the rest, resulting in a lower probability of choosing anything.

It is also interesting to investigate a decision maker's expected probability of being correct ex post. The DM decides whether to learn more information and which alternative to adopt based on his current imperfect information. Therefore, it is possible that he will make mistakes in an ex post point of view, when all the uncertainties about both alternatives have been resolved. The following theorem characterizes a decision maker's expected probability of being correct ex post, given his current beliefs of the two alternatives (with proof in the Appendix).

THEOREM 5: *Consider two homogeneous alternatives. Under the optimal learning strategy in the case that $x_0 \geq x_0^*$, given his current posterior beliefs as (x_1, x_2), a decision maker's expected proba-*

[15] For lab and field experiments on choice overload, refer to a review by Scheibehenne et al. (2010). Different from our setting, Fudenberg and Strzalecki (2015) consider a dynamic logit model with choice aversion where a consumer may prefer to have a smaller choice set ex ante.

bility of being correct ex post is

$$Q(x_1,x_2) =$$
$$\begin{cases} \frac{\overline{x}(x_1-\underline{x})+(1-\underline{x})(\overline{x}-x_1)(1-x_2)}{\overline{x}-\underline{x}}, & \textit{if } x_2 \le \underline{x} < x_1 < \overline{x} \\ \frac{\overline{x}(x_2-\underline{x})+(1-\underline{x})(\overline{x}-x_2)(1-x_1)}{\overline{x}-\underline{x}}, & \textit{if } x_1 \le \underline{x} < x_2 < \overline{x} \\ \frac{(x_1-x_2)\overline{X}(x_2)}{\overline{X}(x_2)-x_2} + \frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)-x_2}\left[(1-\underline{x})^2 e^{-\int_{\underline{x}}^{x_2}\frac{2d\xi}{\overline{X}(\xi)-\xi}} + \int_{\underline{x}}^{x_2} e^{-\int_{\eta}^{x_2}\frac{2d\xi}{\overline{X}(\xi)-\xi}}\frac{2\overline{X}(\eta)}{\overline{X}(\eta)-\eta}d\eta\right], & \textit{if } \underline{x} < x_2 < x_1 < \overline{X}(x_2) \\ \frac{(x_2-x_1)\overline{X}(x_1)}{\overline{X}(x_1)-x_1} + \frac{\overline{X}(x_1)-x_2}{\overline{X}(x_1)-x_1}\left[(1-\underline{x})^2 e^{-\int_{\underline{x}}^{x_1}\frac{2d\xi}{\overline{X}(\xi)-\xi}} + \int_{\underline{x}}^{x_1} e^{-\int_{\eta}^{x_1}\frac{2d\xi}{\overline{X}(\xi)-\xi}}\frac{2\overline{X}(\eta)}{\overline{X}(\eta)-\eta}d\eta\right], & \textit{if } \underline{x} < x_1 < x_2 < \overline{X}(x_1) \\ x_1, & \textit{if } x_1 \ge \overline{X}(x_2) \\ x_2, & \textit{if } x_2 \ge \overline{X}(x_1) \\ (1-x_1)(1-x_2), & \textit{otherwise.} \end{cases} \tag{35}$$

Figure 6 presents $Q(x_1,x_2)$ and $\Delta Q(x_1,x_2)$ under some parameter setting, where $\Delta Q(x_1,x_2) \equiv Q(x_1,x_2) - Q_0(x_1,x_2)$ represents the improvement in probability of being correct ex post due to learning, and where $Q_0(x_1,x_2)$ is the probability of being correct ex post when learning is not allowed before adoption. It is straightforward to show that,

$$Q_0(x_1,x_2) = \begin{cases} x_1, & x_1 \ge \max\{x_2, x_0\} \\ x_2, & x_2 \ge \max\{x_1, x_0\} \\ (1-x_1)(1-x_2), & x_0 \ge \max\{x_1, x_2\} \end{cases} . \tag{36}$$

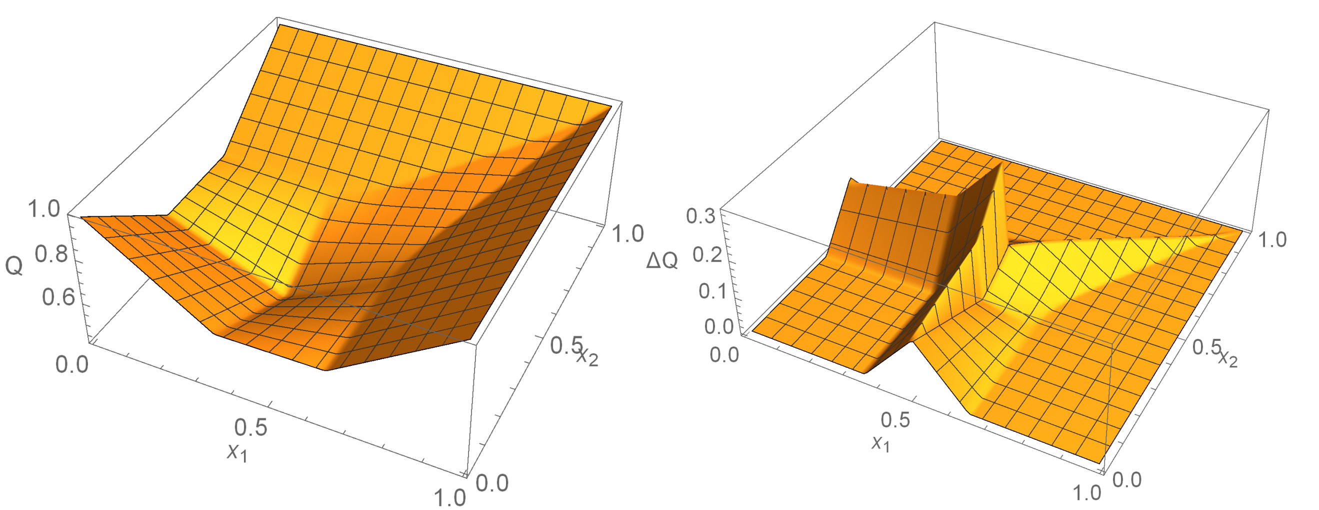

Figure 6: Probability of being correct ex post (left panel) and the improvement in probability of being correct ex post due to learning (right panel) given a DM's current belief, under the parameter setting that $x_0 = 0.5$ and $c\sigma^2/\Delta\pi = 0.1$.

Figure 6 illustrates that it is easier to make mistakes ex post if two alternatives have similar intermediate beliefs ex ante. This is also the case where optimal learning before adoption reduces the probability of ex-post mistakes the most, similar to Figure 1.

5. Extensions

In this section we study several extensions, including heterogeneous alternatives, the effect of the number of alternatives, and time discounting.

5.1. Heterogeneous Alternatives

We consider the case that the outside option has a sufficiently high value. We allow the two alternatives to differ in their payoff distributions, noise levels of signals, as well as learning costs. We aim to solve for the decision maker's optimal learning strategy or, equivalently, we want to understand when to switch learning between the two alternatives, and when to stop learning and make an adoption decision. We should expect that for the heterogeneous case the solution structure will be similar to that of the homogeneous case. We must have the adoption boundaries for the two alternatives, $X_1(x_2)$ and $X_2(x_1)$, different from each other. Moreover, the boundary separating "Learn 1" and "Learn 2" will no longer be a 45-degree line, and thus needs to be determined at the same time. This is a more difficult problem from a technical point of view.

Let us first look at the case that one alternative is "better" than the other, in terms of a higher information-to-noise ratio or a lower learning cost. Without loss of generality, let us assume that alternative 1 is better, i.e., we consider the case that $\Delta\pi_1 > \Delta\pi_2$, or $c_1 < c_2$.[16] Intuitively, in this case, we expect that alternative 1 will be preferred for learning, and thus the boundary separating "Learn 1" and "Learn 2" should be above the 45-degree line. We further assume that there is a point (x_1^*, x_2^*) such that the boundary separating "Learn 1" and "Learn 2" intersects with the adoption boundary of alternative 2 under this belief. Based on this assumption, we can characterize the solution to the *HJB* equation (9) analytically, and we present the solution in the Appendix. Figure 7 illustrates the optimal learning strategy under some parameter settings. The left panel shows the case of different learning costs between the two alternatives, $c_1 < c_2$, with everything else being the same. The right panel shows the case with different payoff ranges between the two alternatives, $\Delta\pi_1 > \Delta\pi_2$, with everything else the same.

[16] Mathematically, $\sigma_1^2 < \sigma_2^2$ will be exactly the same as $c_1 < c_2$, because in our model only $\sigma_i^2 c_i$ is identified. To see this note that (9) can be equivalently written with the elements in the internal max operator divided by c_i (see also the proof of Proposition 1).

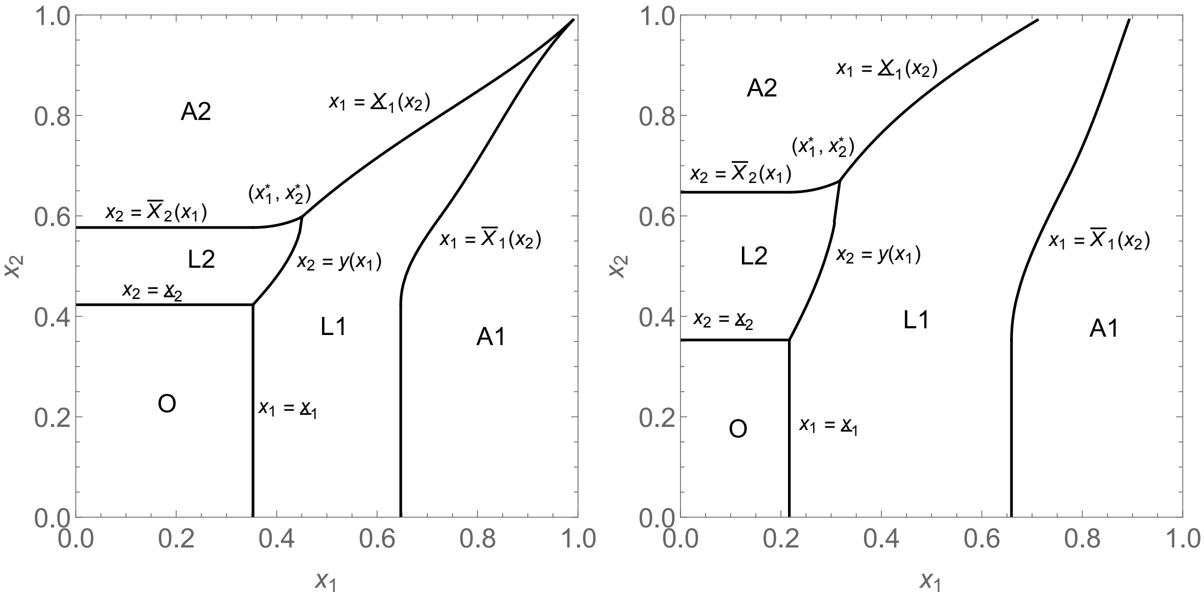

Figure 7: Optimal learning strategy with heterogeneous alternatives. For the left panel: $\Delta\pi_1 = \Delta\pi_2$, $(\pi_0 - \underline{\pi}_1)/\Delta\pi_1 = (\pi_0 - \underline{\pi}_2)/\Delta\pi_2 = 0.5$, and $c_1\sigma_1^2/\Delta\pi_1 = 0.1 < c_2\sigma_2^2/\Delta\pi_2 = 0.2$. For the right panel: $\Delta\pi_1 = 1.2\Delta\pi_2$, $(\pi_0 - \underline{\pi}_1)/\Delta\pi_2 = (\pi_0 - \underline{\pi}_2)/\Delta\pi_2 = 0.5$, and $c_1\sigma_1^2/\Delta\pi_2 = c_2\sigma_2^2/\Delta\pi_2 = 0.1$.

Consistent with the homogeneous case, when $x_i < \underline{x}_i$, alternative i is not considered for learning or adoption. When $x_1 \geq \underline{x}_1$ and $x_2 \geq \underline{x}_2$, both alternatives are in the consideration set, and the decision maker's optimal learning strategy is now slightly more complicated. When $x_2 \geq x_2^*$, he only learns alternative 1. The DM keeps on learning alternative 1 until either x_1 exceeds the adoption boundary of alternative 1, $\overline{X}_1(x_2)$, in which case he adopts alternative 1, or x_1 drops below the adoption boundary of alternative 2, $\underline{X}_1(x_2)$, in which case he adopts alternative 2. On the other hand, when $x_2^* > x_2 \geq \underline{x}_2$, the DM learns alternative 1 if and only if

$$\left(\overline{X}_1(x_2) - x_1\right)\left[\frac{2\sigma_1^2 c_1}{(\Delta\pi_1)^2}\left(\Phi\left(\overline{X}_1(x_2)\right) - \Phi(x_1)\right) + \Delta\pi_1\right]$$
$$\leq \left(\overline{X}_2(x_1) - x_2\right)\left[\frac{2\sigma_2^2 c_2}{(\Delta\pi_2)^2}\left(\Phi\left(\overline{X}_2(x_1)\right) - \Phi(x_2)\right) + \Delta\pi_2\right], \quad (37)$$

where $\overline{X}_1(x_2)$ and $\overline{X}_2(x_1)$ are the adoption boundaries for alternatives 1 and 2, respectively, for the region that $x_1 \geq \underline{x}_1$ and $x_2^* > x_2 \geq \underline{x}_2$. While it is not intuitive to understand the condition above, it does show that a simple index policy is not optimal. Lastly, we want to point out that the boundary separating "Learn 1" and "Learn 2", denoted as $x_2 = y(x_1)$ in the figure, cannot be obtained with value matching and smooth pasting conditions alone. This is intuitive, because even with the boundary exogenously given, we find that both value matching and smooth pasting conditions are necessary to pin down the solution. Now, to maximize the expected payoff by choosing the boundary optimally is equivalent to imposing an extra set of conditions, so called *super contact* conditions (Dumas 1991, Strulovici and Szydlowski 2015), which guarantee that $V_{x_1x_1}(x_1, x_2)$, $V_{x_2x_2}(x_1, x_2)$ and $V_{x_1x_2}(x_1, x_2)$ are continuous across the boundary. We find that two of the three super contact conditions are redundant, and we only need to guarantee $V_{x_1x_2}(x_1, x_2)$ to be continuous across the boundary that separates "Learn 1" and "Learn 2". Similarly, we can show that the redundancy is due to the learning problem structure that the DM can only learn one alternative at a time and the resulting equation (21) is a parabolic PDE.

On the right panel of Figure 7, note also that because $\Delta\pi_1 > \Delta\pi_2$, the DM is more likely to learn and adopt alternative 1 than alternative 2, and the adoption thresholds do not converge anymore when the beliefs converge to $(1, 1)$.

So far, we have characterized the DM's optimal learning strategy when one alternative is "better" than the other in terms of a higher information-to-noise ratio and a lower learning cost. More generally, there are also cases where one alternative has a higher information-to-noise ratio, but at a higher learning cost. This case is complicated with discrete "jumps" of the DM's optimal learning strategy with respect to the parameters (e.g., learning costs). We briefly discuss this case in the Appendix. We have also considered heterogeneous alternatives when the outside option is very low. In that case, it is difficult to characterize the optimal solution analytically, but with numerical analysis, we can find that with small heterogeneity between the two alternatives, we still have the starred regions in the optimal learning strategy where it is optimal to learn an alternative with lower belief first.

5.2. More Than Two Alternatives

In this subsection, we consider the optimal learning problem with more than two alternatives. Although it is difficult to solve the general problem analytically with more than two alternatives (except for the infinite number of alternatives case presented below), we can see numerically how the optimal policy is affected. Consider three homogeneous alternatives without an outside option. By symmetry, we only need to obtain the optimal learning strategy under the belief $x_1 \geq x_2 \geq x_3$. Figure 8 illustrates the optimal learning strategy. In this case it is optimal to learn the second best alternative (alternative 2) under some belief only when x_3 is relatively low. Comparing the three panels in Figure 8 with Figure 2, 4, and 3 respectively, we find the the impact of the third alternative on the optimal learning strategy to be very similar to that of the outside option.

Moreover, we could not find cases where it was optimal to learn the third best alternative (alternative 3). It is interesting that the DM only considers learning from the top two alternatives, and sometimes does not learn about the top alternative, in order to possibly rule out the second best alternative, to then decide to adopt the top alternative.

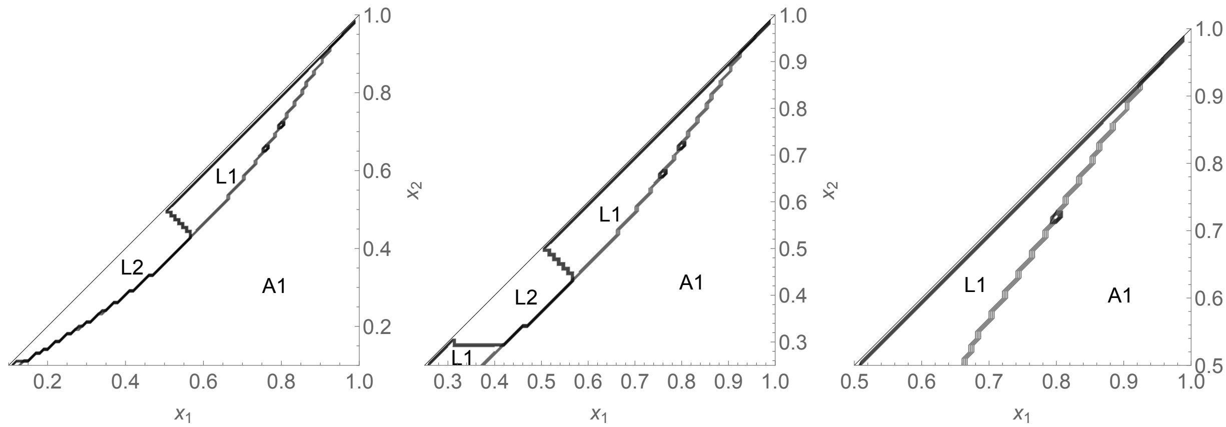

Figure 8: Optimal learning strategy with three alternatives and no outside option. $c\sigma^2/\Delta\pi = 0.1$. $x_3 = 0.1, 0.25, 0.5$ from left to the right.

We have also numerically solved the optimal learning problem with three alternatives and an outside option, and in that case the cutoff strategy is still valid—a decision maker considers an alternative for learning or adoption if and only if his belief of this alternative is sufficiently high.

Lastly, we consider the optimal learning problem of infinite homogeneous alternatives with the same prior belief $\hat{x}$. By the standard analysis, we can show that the optimal learning strategy in this setting follows a simple *reservation value* policy. A DM continues to learn information on alternative i until either x_i exceeds x_R, in which case he adopts alternative i, or x_i drops below $\hat{x}$, in which case he moves to learn the next alternative. For $\hat{x} \le x \le x_R$, the value function $U(x)$ is given by equation (11). We can determine x_R by the following set of boundary conditions:

$$U'(\hat{x}) = 0, \tag{38}$$

$$U(x_R) = \Delta\pi \cdot x_R + \underline{\pi}, \tag{39}$$

$$U'(x_R) = \Delta\pi. \tag{40}$$

By combining the three equations above, we have x_R as:

$$x_R = \Phi^{-1}\left(\Phi(\hat{x}) - \frac{(\Delta\pi)^3}{2\sigma^2 c}\right). \tag{41}$$

We also get that x_R is greater than $\overline{X}(\hat{x})$, illustrating that with more alternatives the DM is more demanding on what an alternative has to deliver before the DM settles on it.

5.3. Time Discounting

In this subsection, we consider a similar problem setting with no flow cost for learning but with time discounting. Similar to equation (8), we can write down the discrete-time-analog Bellman equation associated with the learning problem as,

$$\begin{aligned} V(x_1, x_2) \quad = \quad \max \Big\{ & e^{-rdt}\mathrm{E}\left[V(x_1 + dx_1, x_2)|x_1, x_2\right], \\ & e^{-rdt}\mathrm{E}\left[V(x_1, x_2 + dx_2)|x_1, x_2\right], \\ & \mathrm{E}[\pi_1|x_1, x_2], \mathrm{E}[\pi_1|x_1, x_2], \pi_0 \Big\}, \end{aligned} \tag{42}$$

where r is the time discounting instantaneous interest rate. Similarly, we can reduce the Bellman equation as the following *HJB* equation.

$$\max\left\{\max_{1\leq i\leq n}\{\frac{(\Delta\pi_i)^2}{2\sigma_i^2}x_i^2(1-x_i)^2V_{x_ix_i}(\mathbf{x}) - rV(\mathbf{x})\}, g(\mathbf{x}) - V(\mathbf{x})\right\} = 0. \tag{43}$$

Along the same lines as above, we can show that there exists a unique solution to this problem, and the value function is given by Theorem A2 in the Appendix. Following an almost identical analysis as in Section 3, we can construct the optimal allocation and stopping rule and the corresponding belief updating process to attain the value function. We find that it is always optimal for the DM to learn the alternative with higher belief whenever it is optimal to learn. Formally, we have the following theorem that characterizes the optimal learning strategy.

THEOREM 6: *Consider the DM's optimal learning problem defined in equation (6) but with time discounting instead of a constant flow learning cost. Given one alternative i under his consideration, the DM keeps learning it if $\underset{\sim}{x} \leq x_i \leq \tilde{x}$, adopts it if $x_i > \tilde{x}$, and takes the outside option if $x_i < \underset{\sim}{x}$. Given two alternatives under his consideration, the DM always learns information from the alternative with higher belief. He stops learning and adopts it if and only if his belief of this alternative is above a threshold, which increases with his belief of the other alternative.*

Figure 9 presents a DM's optimal learning strategy under some parameter setting. Now it is always optimal to learn the best alternative first. The reason is that it does not help to minimize the learning costs by learning an inferior alternative. Specifically, with time discounting, the learning cost depends on the current expected payoff. By learning the inferior alternative first, if the DM receives negative signals, he will adopt the other alternative immediately. On the other hand, if he receives positive signals, he will continue learning. However, his learning cost is also high because he just received positive signals. Therefore, the DM will not save any cost by not learning the best alternative first.

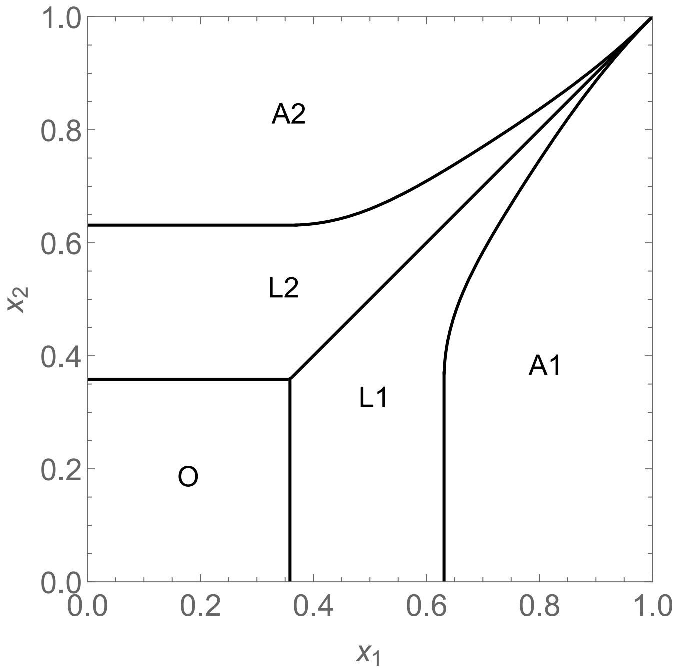

Figure 9: Optimal learning strategy with time discounting under the parameter setting that $\underline{\pi}/\Delta\pi = 0$, $\pi_0/\Delta\pi = 0.5$, $\sigma/\Delta\pi = 1$, and $r = 0.2$.

6. Conclusion

This paper examines a general problem of optimal information gathering. We allow gradual learning on multiple alternatives before one's choice decision. The decision maker solves the problem of optimal dynamic allocation of learning efforts as well as optimal stopping of the learning process, so as to maximize the expected payoff. We show that the optimal learning strategy is of the consider-then-decide type. The decision maker considers an alternative for learning or adoption if and only if the expected payoff of the alternative is above a threshold. Given several alternatives in the decision maker's consideration set, we find that sometimes it is not optimal for him to learn information from the alternative with the highest expected payoff, given that all other characteristics of the other alternatives are the same. If the decision maker subsequently receives enough positive informative signals, the decision maker will switch to learn the better alternative; otherwise the decision maker will rule out this inferior alternative from consideration and adopt the better alternative. We find that this strategy works because it minimizes the decision maker's learning efforts. It becomes the optimal strategy when the outside option is weak, and the decision maker's beliefs about different alternatives are in an intermediate range.

APPENDIX

Sketch of the Proof of Lemma 1: Existence and Uniqueness of the Solution:

Consider a nonlinear second-order PDE of the form

$$H(\mathbf{x}, u, Du, D^2u) = 0, \tag{i}$$

where $Du \in \mathbb{R}^n$ is the differential of u (at $\mathbf{x}$), $D^2u \in \mathcal{S}_n(\mathbb{R})$ is the Hessian matrix of u (at $\mathbf{x}$), and $H : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times \mathcal{S}_n(\mathbb{R}) \to \mathbb{R}$, with $\mathcal{S}_n(\mathbb{R})$ the set of symmetric $n \times n$ matrices. The function H is said to satisfy *degenerate ellipticity* if and only if $H(\mathbf{x}, u, p, X) \leq H(\mathbf{x}, v, p, Y)$ for all $u \leq v$ and (X, Y) such that $X - Y$ is positive semi-definite or $X = Y$. To see that the HJB equation (9) satisfies *degenerate ellipticity*, it is equivalent to consider

$$\max\left\{\max_{1\leq i\leq n}\left\{-\widehat{a}_i(\mathbf{x})\overline{V}_{x_ix_i}(\mathbf{x}) - c_i\right\}, g(\mathbf{x}) + \overline{V}(\mathbf{x})\right\} = 0. \tag{ii}$$

where $\widehat{a}_i(\mathbf{x}) \equiv \frac{(\Delta\pi_i)^2}{2\sigma_i^2}x_i^2(1 - x_i)^2$ and $\overline{V}(\mathbf{x}) \equiv -V(\mathbf{x})$, for all $\mathbf{x}$.

A continuous function u is said to be a *viscosity solution* to (i) if for all twice continuously differentiable functions $\phi : \mathbb{R}^n \to \mathbb{R}$: (1) if $u - \phi$ attains a local maximum at $x_0 \in \mathbb{R}^n$ and $u(x_0) = \phi(x_0)$ then $H(x_0, \phi(x_0), D\phi(x_0), D^2\phi(x_0)) \leq 0$; (2) if $u - \phi$ attains a local minimum at $x_0 \in \mathbb{R}^n$ and $u(x_0) = \phi(x_0)$, then $H(x_0, \phi(x_0), D\phi(x_0), D^2\phi(x_0)) \geq 0$. Note that this definition uses the fact that (i) satisfies degenerate ellipticity.

Lions (1983a,b) showed that if the HJB equation (i) satisfies degenerate ellipticity then it has a unique viscosity solution. Let us outline it in our situation. Observe that the HJB equation (ii) can be rewritten as:

$$\max_{1\leq i\leq n+1}\{-\widehat{a}_i(\mathbf{x})\overline{V}_{x_ix_i}(\mathbf{x}) - c_i + g_i(\mathbf{x}) + \lambda_i\overline{V}V(\mathbf{x})\} = 0, \tag{iii}$$

with $\widehat{a}_{n+1}(\mathbf{x}) \equiv 0$, $c_{n+1} \equiv 0$, and

$$g_i(\mathbf{x}) \equiv 0 \text{ for all } 1 \leq i \leq n \quad \text{and} \quad g_{n+1}(\mathbf{x}) \equiv g(\mathbf{x}),$$

$$\lambda_i \equiv 0 \text{ for all } 1 \leq i \leq n \quad \text{and} \quad \lambda_{n+1} \equiv 1.$$

We can further simplify the HJB equation (iii) as:

$$\max_{1\leq i\leq n+1}\{\mathcal{A}_i\overline{V}(\mathbf{x})-f_i(\mathbf{x})\}=0, \tag{iv}$$

where $\mathcal{A}_i \equiv -\widehat{a}_i\partial_{x_ix_i}+\lambda_i$, and $f_i(\mathbf{x})\equiv c_i-g_i(\mathbf{x})$.

Consider the stochastic differential equation

$$dX_t=\sigma(X_t,I_t)dW_t \quad \text{and} \quad X_0=\mathbf{x}, \tag{v}$$

where $I_t=\widetilde{I}(X_t)\in\{1,2,\cdots,n+1\}$ is the *control*, with $\widetilde{I}:\mathbb{R}^n\to\{1,2,\cdots,n+1\}$, $(W_t;t\geq 0)$ is a n-dimensional Brownian motion, and σ is a $n\times n$ diagonal matrix, where element i in the diagonal, $\sigma_{ii}:\mathbb{R}^n\times\{1,2,\cdots,n+1\}\to\mathbb{R}_+$. If $I_t=n+1$ it means that the DM stopped learning. Take

$$\sigma_{ii}(\mathbf{x},j)\equiv\frac{\Delta\pi_i}{\sigma_i}x_i(1-x_i) \quad \text{for all } 1\leq i\leq n, 1\leq j\leq n+1 \quad \text{if} \quad i=j$$

and $\sigma_{ii}(\mathbf{x},j)=0$ otherwise. Let $\lambda(\mathbf{x},i)\equiv\lambda_i$, $f(\mathbf{x},i)\equiv f_i(x)$, for all $1\leq i\leq n+1$ and

$$J(\mathbf{x},\widetilde{I})\equiv E\left[\int_0^\infty f(X_t,\widetilde{I}(X_t))\exp\left(\int_0^t -\lambda(X_s,\widetilde{I}(X_s))ds\right)\,dt\right].$$

Lions (1983a) shows that the value function $\overline{V}(x)\equiv\min_{\widetilde{I}}J(x,\widetilde{I})$ is continuous (Theorem II.2, p. 1136), and is a viscosity solution to the HJB equation (iv) (Theorem I.2, p. 1120), when this equation satisfies degenerate ellipticity. Furthermore, Lions (1983b) shows that this solution is unique (Theorem II.1, p. 1250).

Smooth Pasting Conditions at the Adoption Boundary and Derivation of Equations (24) and (25):

We will provide an intuitive argument below for the smooth-pasting condition at the adoption boundary of alternative 1. Consider a DM spending time dt in learning alternative 1 at the boundary $\big(\overline{X}(x_2),x_2\big)$. The corresponding belief update dx_1 can be either positive or negative, with equal odds. If $dx_1\geq 0$, the DM will adopt alternative 1 immediately; otherwise if $dx_1<0$, the DM will stay in the market learning more information on alternative 1. Therefore, the DM's expected payoff given that he will spend dt to learn alternative 1 would be:

$$\begin{aligned} V_1\left(\overline{X}(x_2), x_2\right) &\equiv -cdt + \frac{1}{2}\left[\Delta\pi\left(\overline{X}(x_2) + \mathrm{E}[dx_1|dx_1 \geq 0]\right) + \underline{x}\right] + \frac{1}{2}\mathrm{E}\left[V(\overline{X}(x_2) + dx_1, x_2)|dx_1 < 0\right] \\ &= V\left(\overline{X}(x_2), x_2\right) + \frac{\Delta\pi}{2\sigma}\overline{X}(x_2)\left[1 - \overline{X}(x_2)\right]\sqrt{\frac{dt}{2\pi}}\left[\Delta\pi - V_{x_1}\left(\overline{X}(x_2), x_2\right)\right] + o(\sqrt{dt}), \end{aligned}$$

where we have used that (1) $\mathrm{E}[dW|dW \geq 0] = -\mathrm{E}[dW|dW < 0] = \sqrt{\frac{dt}{2\pi}}$ for $\{W(t)|t \geq 0\}$ being a standard Brownian motion; (2) the value matching condition $V\left(\overline{X}(x_2), x_2\right) = \Delta\pi\overline{X}(x_2) + \underline{\pi}$.

On the other hand, let us consider a DM who spends dt in learning alternative 2 at the boundary $\left(\overline{X}(x_2), x_2\right)$. If the resulting belief update $dx_2 \geq 0$, the DM's adoption threshold for alternative 1 increases, so he will continue to learn alternative 1; otherwise, if $dx_2 < 0$, the DM will adopt alternative 1 immediately. Therefore, the DM's expected payoff given that he will spend dt to learn alternative 2 would be:

$$\begin{aligned} V_2\left(\overline{X}(x_2), x_2\right) &\equiv -cdt + \frac{1}{2}\left[\Delta\pi\overline{X}(x_2) + \underline{x}\right] + \frac{1}{2}\mathrm{E}\left[V(\overline{X}(x_2), x_2 + dx_2)|dx_2 < 0\right] \\ &= V\left(\overline{X}(x_2), x_2\right) - \frac{\Delta\pi}{2\sigma}x_2\left(1 - x_2\right)\sqrt{\frac{dt}{2\pi}}V_{x_2}\left(\overline{X}(x_2), x_2\right) + o(\sqrt{dt}), \end{aligned}$$

The DM chooses which alternative to learn on based on expected payoff maximization. By definition, his value function should satisfy:

$$V\left(\overline{X}(x_2), x_2\right) = \max\{V_1\left(\overline{X}(x_2), x_2\right), V_2\left(\overline{X}(x_2), x_2\right)\}.$$

By substituting the expression of $V_1\left(\overline{X}(x_2), x_2\right)$ and $V_2\left(\overline{X}(x_2), x_2\right)$ into the above equation, we have

$$\max\{\overline{X}(x_2)\left[1 - \overline{X}(x_2)\right]\left[\Delta\pi - V_{x_1}\left(\overline{X}(x_2), x_2\right)\right], x_2\left(1 - x_2\right)V_{x_2}\left(\overline{X}(x_2), x_2\right)\} = 0.$$

Meanwhile, by taking derivative of both sides of the value matching condition $V\left(\overline{X}(x_2), x_2\right) = \Delta\pi\overline{X}(x_2) + \underline{\pi}$ with respect to x_2, we have

$$\overline{X}'(x_2)\left[\Delta\pi - V_{x_1}\left(\overline{X}(x_2), x_2\right)\right] = V_{x_2}\left(\overline{X}(x_2), x_2\right). \tag{vi}$$

Combining the above two equations, we obtain that $V_{x_1}\left(\overline{X}(x_2), x_2\right) = \Delta\pi$ and $V_{x_2}\left(\overline{X}(x_2), x_2\right) = 0$, which are the smooth pasting conditions at the adoption boundary of alternative 1. It is straightforward to derive equations (24) and (25) based on the value matching and smooth pasting conditions.

Redundancy of One Smoothness Condition:

As shown in equation (22), the general solution to the parabolic PDE in equation (21) can be written as the following:

$$V(x_1, x_2) = C_0(x_1) + C_1(x_2)x_1 + C_2(x_2),$$

where $C_0(x_1) = \frac{2\sigma^2 c}{(\Delta\pi)^2}(1-2x_1)\ln\left(\frac{1-x_1}{x_1}\right)$. The value matching and smooth pasting conditions at the adoption boundary of alternative 1 $x_1 = \overline{X}(x_2)$ are:

$$C_0(\overline{X}(x_2)) + C_1(x_2)\overline{X}(x_2) + C_2(x_2) = \Delta\pi\overline{X}(x_2) + \overline{\pi}, \tag{vii}$$

$$C_0'(\overline{X}(x_2)) + C_1(x_2) = \Delta\pi, \tag{viii}$$

$$C_1'(x_2)\overline{X}(x_2) + C_2'(x_2) = 0. \tag{ix}$$

We will show that given equations (vii) and (viii), equation (ix) is redundant. In fact, taking derivative of both sides of equation (viii) with respect to x_2, we have,

$$C_0'(\overline{X}(x_2))\overline{X}'(x_2) + C_1'(x_2)\overline{X}(x_2) + C_1(x_2)\overline{X}'(x_2) + C_2'(x_2) = \Delta\pi\overline{X}'(x_2).$$

i.e.,

$$\left[C_0'(\overline{X}(x_2)) + C_1(x_2) - \Delta\pi\right]\overline{X}'(x_2) + C_1'(x_2)\overline{X}(x_2) + C_2'(x_2) = 0.$$

By substituting equation (viii) to the above equation, we get equation (ix).

Proof of Lemma 2: Existence and Uniqueness of a Solution to the ODE:

Proof. According to the Picard-Lindelöf theorem, to prove the existence and uniqueness of the solution in the neighborhood of $[\underline{x}, \underline{x}+\varepsilon_0]$ for a small enough $\varepsilon_0 > 0$, we only need to show that the right hand side of equation (27) is uniformly Lipschitz continuous in $\overline{X}$ and continuous in x. This is obviously true, because $\overline{X}(\underline{x}) = \overline{x} > \underline{x}$.

Our objective is to show that the solution exists uniquely for all $x \in [\underline{x}, 1)$. To prove this, we can apply the argument above iteratively. Consequently, there are two possibilities: (1) $\overline{X}(x) > x$ for $x \in [\underline{x}, 1)$, and in this case, the solution exists uniquely and the proof is complete; (2) there exists $x' \in [\underline{x}, 1)$ such that the solution does not exist or there are multiple solutions at x'.

Let us understand the implication of the case (2); we will show that this case is impossible. First, let us define $\hat{x}$ as the infimum of all x' that satisfies the condition in case (2). By definition, we must have that $\overline{X}(x) > x$ for $x \in [\underline{x}, \hat{x})$. Moreover, $\lim_{x\to\hat{x}^-} \overline{X}(x) = \hat{x}$; otherwise, the solution will exist uniquely at $\hat{x}$, and $\overline{X}(\hat{x}) > \hat{x}$. Meanwhile, by equation (27), we have that $\lim_{x\to\hat{x}^-} \overline{X}'(x) = +\infty$. This implies that there exists $\varepsilon > 0$, such that $\overline{X}'(x) > 1$ for $x \in [\hat{x}-\varepsilon, \hat{x})$. As a result, we have,

$$\lim_{x\to\hat{x}^-} \overline{X}(x) - \hat{x} = \overline{X}(\hat{x}-\varepsilon) - (\hat{x}-\varepsilon) + \int_{\hat{x}-\varepsilon}^{\hat{x}^-} (\overline{X}'(x) - 1)dx > 0, \tag{x}$$

which contradicts the fact that $\lim_{x\to\hat{x}^-} \overline{X}(x) = \hat{x}$. Therefore, we have proved that the solution exists uniquely for $x \in [\underline{x}, 1)$.

Smoothness of $\overline{X}(x)$ is easy to verify by taking derivatives. Let us prove the monotonicity of $\overline{X}(x)$. By equation (31) and the monotonicity of $V(x_1, x_2)$, we have

$$\frac{\partial V(x_1,x_2)}{\partial x_2} = \frac{2\sigma^2 c}{(\Delta\pi)^2} \frac{\left(\overline{X}(x_2) - x_1\right)\overline{X}'(x_2)}{\overline{X}(x_2)^2\left(1-\overline{X}(x_2)\right)^2} \geq 0, \text{ for } \underline{x} \leq x_2 \leq x_1 \leq \overline{X}(x_2)$$

which implies $\overline{X}'(x_2) \geq 0$. Now let us prove $\overline{X}(1) = 1$. In fact, by definition, $1 \geq \overline{X}(x) \geq x$ for any $x \in [\underline{x}, 1]$. This implies that $1 \geq \overline{X}(1) \geq 1$, so $\overline{X}(1) = 1$. It is easy to verify that $\overline{X}(1) = 1$ also satisfies the differential equation (27). ■

Proof of Lemma 3:

Proof. Let us first derive inequality (30) from inequality (29). In fact, when $\underline{x} \leq x_2 \leq x_1 \leq \overline{X}(x_2)$, $V(x_1, x_2)$ is given by equation (26). We have

$$V_{x_2}(x_1,x_2) = \frac{2c\sigma^2}{(\Delta\pi)^2}\frac{(\overline{X}(x_2)-x_1)\overline{X}'(x_2)}{\overline{X}(x_2)^2(1-\overline{X}(x_2))^2} = \frac{2c\sigma^2}{(\Delta\pi)^2}\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)-x_2}\left(\Phi\left(\overline{X}(x)\right) - \Phi(x) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right),$$

where the second equality is by equation (28). By taking the partial derivative with respect to x_2 and using equation (28) again, we have

$$\begin{aligned} V_{x_2x_2}(x_1,x_2) &= \frac{2c\sigma^2}{(\Delta\pi)^2}\left[\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)-x_2}\frac{1}{x_2^2(1-x_2)^2}\right. \\ &+ \left.\frac{(x_1-x_2)\overline{X}(x_2)^2\left(1-\overline{X}(x_2)\right)^2}{\left(\overline{X}(x_2)-x_2\right)^3}\left(\Phi\left(\overline{X}(x)\right)-\Phi(x)+\frac{(\Delta\pi)^3}{2\sigma^2 c}\right)^2\right] \end{aligned}$$

Using this expression of $V_{x_2x_2}(x_1, x_2)$, inequality (29) can be written as

$$\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)-x_2} + \frac{(x_1-x_2)x_2^2(1-x_2)^2\overline{X}(x_2)^2\left(1-\overline{X}(x_2)\right)^2}{\left(\overline{X}(x_2)-x_2\right)^3}\left(\Phi\left(\overline{X}(x)\right)-\Phi(x)+\frac{(\Delta\pi)^3}{2\sigma^2 c}\right)^2 \leq 1$$

By multiplying $\left(\overline{X}(x_2) - x_2\right)$ and rearranging, the inequality above is equivalent to

$$(x_1 - x_2)\left[\frac{x_2^2(1-x_2)^2\overline{X}(x_2)^2\left(1-\overline{X}(x_2)\right)^2}{\left(\overline{X}(x_2)-x_2\right)^2}\left(\Phi\left(\overline{X}(x)\right) - \Phi(x) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right)^2 - 1\right] \leq 0.$$

Because $x_1 \geq x_2$, this is equivalent to

$$\frac{x_2^2(1-x_2)^2\overline{X}(x_2)^2\left(1-\overline{X}(x_2)\right)^2}{\left(\overline{X}(x_2)-x_2\right)^2}\left(\Phi\left(\overline{X}(x)\right) - \Phi(x) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right)^2 \leq 1,$$

which is equivalent to

$$\frac{x_2(1-x_2)\overline{X}(x_2)\left(1-\overline{X}(x_2)\right)}{\overline{X}(x_2)-x_2}\left(\Phi\left(\overline{X}(x)\right) - \Phi(x) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right) \leq 1,$$

which in turn, is equivalent to

$$H(x_2) \equiv x_2(1-x_2)\left(\Phi\left(\overline{X}(x_2)\right) - \Phi(x_2) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right) - \frac{\overline{X}(x_2) - x_2}{\overline{X}(x_2)\left(1-\overline{X}(x_2)\right)} \leq 0. \tag{xi}$$

By using equation (28) again, the inequality above can be equivalently written as inequality (30).

Now we prove the existence of x_0^* from inequality (xi). To guarantee inequality (xi) for any $x_2 \in [\underline{x}, 1]$, we need

$$\max_{\underline{x} \leq x_2 \leq 1} H(x_2) \leq 0. \tag{xii}$$

Note that $H(\underline{x}) = -(\overline{x} - \underline{x})/\left[\overline{x}(1-\overline{x})\right] < 0$ and $H(x)$ is continuous, so the inequality (xii) will be violated if and only if there exists $x_2^* \equiv \min_{\underline{x} \leq x_2 \leq 1}\{x_2 | H(x_2) = 0\}$, such that $H'(x_2^*) > 0$. By using the equation (27) and the condition that $H(x_2^*) = 0$, we can simplify and get the following expression,

$$H'(x_2^*) = -2\left(\Phi\left(\overline{X}(x_2^*)\right) - \Phi(x_2^*) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right)\left(x_2^* + \overline{X}(x_2^*) - 1\right).$$

From the proof of Lemma 2, we know that $\Phi\left(\overline{X}(x_2^*)\right) - \Phi(x_2^*) + \frac{(\Delta\pi)^3}{2\sigma^2 c} \geq 0$. Let us define $h(x) \equiv x + \overline{X}(x) - 1$. Therefore, the equality above implies that

$$H'(x_2^*) > 0 \Leftrightarrow h(x_2^*) < 0.$$

Meanwhile, from equations (17) and (18), it is easy to show that

$$\underline{x} + \overline{x} \geq 1 \Leftrightarrow x_0 \geq 1/2.$$

Therefore, when $x_0 \geq 1/2$, $h(\underline{x}) = \underline{x} + \overline{x} - 1 \geq 0$. According to Lemma 2, we also know that $h(x)$ is strictly increasing. This implies that when $x_0 \geq 1/2$, $h(x) \geq 0$ for any $x \in [\underline{x}, 1]$. This in turn implies that $h(x_2^*) \geq 0$, and inequality (xii) must hold. Therefore, if x_0^* exists, it must be less than 1/2.

Next, we will prove the existence of x_0^*. We also note that

$$
\begin{aligned}
\frac{dh(x_2^*)}{d\pi_0} &= \left.\frac{\partial \overline{X}(x)}{\partial \pi_0}\right|_{x=x_2^*} + \left[1 + \overline{X}'(x_2^*)\right] \frac{dx_2^*}{d\pi_0} \\
&= \left.\frac{\partial \overline{X}(x)}{\partial \pi_0}\right|_{x=x_2^*} - \frac{1 + \overline{X}'(x_2^*)}{H'(x_2^*)} \times \left.\frac{dH}{d\overline{X}}\right|_{x=x_2^*} \times \left.\frac{\partial \overline{X}(x)}{\partial \pi_0}\right|_{x=x_2^*} \\
&= \left\{ 1 - \frac{\left[(\overline{X}(x_2^*) - x_2^*)^2 + 2x_2^*(1 - x_2^*)\right] \left[\overline{X}(x_2^*)\left(1 - \overline{X}(x_2^*)\right) + x_2^*(1 - x_2^*)\right]}{2h(x_2^*)\left(\overline{X}(x_2^*) - x_2^*\right)\overline{X}(x_2^*)\left(1 - \overline{X}(x_2^*)\right)} \right\} \left.\frac{\partial \overline{X}(x)}{\partial \pi_0}\right|_{x=x_2^*}
\end{aligned}
$$

As shown in Proposition 1, $\partial \overline{X}(x)/\partial \pi_0 \geq 0$, so $dh(x_2^*)/d\pi_0 \geq 0$ when $h(x_2^*) < 0$. Consider $x_0' > x_0''$. If under x_0', inequality (xii) is violated, then there exists $x_2^* \in [\underline{x}, 1]$ such that $H'(x_2^*) > 0$, which implies that $h(x_2^*) < 0$, which in turn implies that $dh(x_2^*)/d\pi_0 \geq 0$. This implies that under $x_0'' < x_0'$, $h(x_2^*)$ will get even smaller, and inequality (xii) will also be violated. In summary, to guarantee inequality (xii), we must have $x_0 \geq x_0^*$ for some x_0^*.

The threshold x_0^* can be obtained as the x_0 that makes $\max_{x_2} H(x_2) = 0$, which is a function of x_0 as $\overline{X}(.)$ is a function of both $\underline{x}$ and $\overline{x}$ and both $\underline{x}$ and $\overline{x}$ are a function of x_0. ■

Construction of The Belief Updating Process under The Optimal Learning Strategy:

Consider the case of Theorem 2 when the value of the outside option is relatively high. We have already divided the belief space of x_1-x_2 into five regions corresponding to the DM's five choices of action. Updating the beliefs of both alternatives ceases when entering the regions where it is optimal to adopt 1, or adopt 2, or take the outside option. Therefore, we only need to construct the belief updating process for the regions where it is optimal to learn 1 or learn 2. Equation (7) can be equivalently rewritten as the following:

$$
dx_i(t) = \frac{\Delta\pi}{\sigma^2} x_i(t)\left[1 - x_i(t)\right] \mathbb{1}_{L_i}\left(x_1(t), x_2(t)\right) \left\{\left[\pi_i - \underline{\pi}\left(1 - x_i(t)\right) - \overline{\pi} x_i(t)\right] dt + \sigma dW_i(t)\right\}, \quad \text{(xiii)}
$$

where $\mathbb{1}_{L_i}\left(x_1(t), x_2(t)\right)$ is an indicator function. It is equal to one when $(x_1(t), x_2(t)) \in L_i$, i.e., when the DM's beliefs are in the region where it is optimal to learn alternative i; it takes zero otherwise. We have $L_1 \equiv \left\{x_1, x_2 | x_2 \leq x_1 \leq \overline{X}(x_2), x_1 \geq \underline{x}\right\}$ and $L_2 \equiv \left\{x_1, x_2 | x_1 < x_2 \leq \overline{X}(x_1), x_2 \geq \underline{x}\right\}$.

$\{W_i(t); t \geq 0\}$ is a standard Brownian motion, and $W_1(\cdot)$ is independent of $W_2(\cdot)$. Now in equation (xiii), we have a set of SDE's in the standard form of $dx_i(t) = a_i(x_1(t), x_2(t))dt + b_i(x_1(t), x_2(t))dW_i(t)$, which is first studied formally by Itô (1951). As shown in Stroock and Varadhan (1979), solving the SDE in equation (xiii) amounts to solving a martingale problem for a diffusion operator. Karatzas (1984) considers this problem for multi-arm bandits for diffusion processes. Because here we only consider the regions where it is optimal to learn alternative 1 or 2, by defining the "allocation index" in their setting as the belief in our setting, we can directly apply Theorem 6.1 in Karatzas (1984) to show that there exists a unique-in-distribution solution to equation (xiii).[17] See also Bass and Pardoux (1987) for a general solution to this problem.

It is also interesting to see what happens to the stochastic process when we are at $x_1 = x_2$. In particular, given the optimal learning strategies, let us investigate what happens to $\min\{x_1(t), x_2(t)\}$ conditional on the DM's belief being in L_1 or L_2. In terms of the SDE in equation (xiii), let us define $\widetilde{a}_i(x_i(t)) \equiv \frac{\Delta\pi}{\sigma^2} x_i(t)[1 - x_i(t)][\pi_i - \underline{\pi}(1 - x_i(t)) - \overline{\pi} x_i(t)]$, for $i = 1, 2$, and $\widetilde{b}(x_i(t)) \equiv \frac{\Delta\pi}{\sigma} x_i(t)[1 - x_i(t)]$ such that we can write equation (xiii) as

$$dx_i(t) = \mathbb{1}_{L_i}(x_1(t), x_2(t))[\widetilde{a}_i(x_i(t))dt + \widetilde{b}(x_i(t))dW_i(t)] \quad (i = 1, 2).$$

The particular case where $\widetilde{a}_1 = \widetilde{a}_2$, and $\widetilde{b}$ are constants is considered in Fernholz et al. (2013), which presents pathwise unique solutions for $x_1(t)$ and $x_2(t)$.

To see the stochastic behavior of $\min\{x_1(t), x_2(t)\}$, we follow the presentation in Fernholz et al. (2013), p.357, with the adjustments that $\widetilde{a}_1(x_1(t)) \neq \widetilde{a}_2(x_1(t))$, and $\widetilde{a}_i(x_i(t))$ for $i = 1, 2$, and $\widetilde{b}(x_i(t))$ are bounded and continuous functions of $x_i(t)$, and where $x_1(t)$ and $x_2(t)$ are unique-in-distribution solutions to (xiii) for $i = 1, 2$. Let us define $Y(t) \equiv x_1(t) - x_2(t)$. Then we can obtain

$$Y(t) = Y(0) + \int_0^t \widetilde{a}_1(x_1(s))\mathbb{1}_{[Y(s)\geq 0]}ds - \int_0^t \widetilde{a}_2(x_2(s))\mathbb{1}_{[Y(s)<0]}ds + \widetilde{W}(t), \tag{xiv}$$

where $\widetilde{W}(t) \equiv \int_0^t \widetilde{b}(x_1(s))\mathbb{1}_{[Y(s)\geq 0]}dW_1(s) - \int_0^t \widetilde{b}(x_2(s))\mathbb{1}_{[Y(s)<0]}dW_2(s)$. We can also obtain

$$x_1(t) + x_2(t) = x_1(0) + x_2(0) + \int_0^t \widetilde{a}_1(x_1(s))\mathbb{1}_{[Y(s)\geq 0]}ds + \int_0^t \widetilde{a}_2(x_2(s))\mathbb{1}_{[Y(s)<0]}ds + \widehat{W}(t), \tag{xv}$$

where $\widehat{W}(t) \equiv \int_0^t \widetilde{b}(x_1(s))\mathbb{1}_{[Y(s)\geq 0]}dW_1(s) + \int_0^t \widetilde{b}(x_2(s))\mathbb{1}_{[Y(s)<0]}dW_2(s)$. The processes $\widetilde{W}(t)$ and $\widehat{W}(t)$ are continuous martingales, and then $Y(t)$ is a semimartingale. Using the definitions $\widetilde{W}(t)$ and $\widehat{W}(t)$

[17] Note that, given the optimal learning strategy, the dispersion coefficient in (xiii) is always strictly positive with probability one if both beliefs start in $(0, 1)$ when a belief is being updated.

we can also obtain

$$\widehat{W}(t) = \int_0^t \operatorname{sgn}(Y(s))d\widetilde{W}(s), \tag{xvi}$$

where the signum function is defined as $\operatorname{sgn}(y) \equiv \mathbb{1}_{[0,\infty)}(y) - \mathbb{1}_{(-\infty,0)}(y)$.

Note now that

$$|Y(t)| = |Y(0)| + \int_0^t \operatorname{sgn}(Y(s))dY(s) + 2L^Y(t) \tag{xvii}$$

by Tanaka's formula (Karatzas and Shreve 1991, p. 205) applied to semimartingales, and where $L^Y(t)$ is the local time of $Y(t)$ at $Y(t) = 0$, which is defined as $L^Y(t) \equiv \lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon}\int_0^t \mathbb{1}_{\{0\le Y(s)<\varepsilon\}}ds$. Substituting the derivative of equation (xiv) and equation (xvi) into equation (xvii) we can obtain

$$|Y(t)| = |Y(0)| + \int_0^t \widetilde{a}_1(x_1(s))\mathbb{1}_{[Y(s)\ge 0]}ds + \int_0^t \widetilde{a}_2(x_2(s))\mathbb{1}_{[Y(s)<0]}ds + \widehat{W}(t) + 2L^Y(t). \tag{xviii}$$

Noting that $\min\{x_1(t), x_2(t)\} = 1/2[x_1(t) + x_2(t) - |x_1(t) - x_2(t)|]$, we can then use equation (xv) and equation (xviii) to obtain

$$\min\{x_1(t), x_2(t)\} = \min\{x_1(0), x_2(0)\} - L^Y(t), \tag{xix}$$

which shows that $\min\{x_1(t), x_2(t)\}$, conditional on the DM being in the region where it is optimal to learn more information, falls through time under the optimal learning strategy.

By the same argument we can also look at the evolution of the stochastic process in the regions L_i^* when the outside option has a relatively low or medium value (see Figures 3 and 4). In those cases, the evolution close to $x_1(t) = x_2(t)$ has to do with the $\max\{x_1(t), x_2(t)\}$ and we can obtain that, given the optimal learning in those regions, $\max\{x_1(t), x_2(t)\} = \max\{x_1(0), x_2(0)\} + L^Y(t)$, increases through time. In the case of medium value of the outside option, there are also two negatively sloped frontiers surrounding the region $L_1^* \cup L_2^*$. In these frontiers the stochastic process would evolve towards $x_1(t) = x_2(t)$ in the upper frontier, and away from $x_1(t) = x_2(t)$ for the lower frontier.

Proof of Proposition 1: Comparative Statics:

Proof. We first show the comparative statics with respect to c. Let us initially show that $\underline{x}$ increases with c. In fact, by taking derivative with respect to c on both sides of equations (17) and (18) and combining the two resulting equations to cancel $\partial\overline{x}/\partial c$, we have

$$\frac{\partial \underline{x}}{\partial c} = \frac{\underline{x}^2(1-\underline{x})^2}{\overline{x}-\underline{x}}\frac{(\Delta\pi)^2\,(\Delta\pi\cdot\overline{x}+\underline{\pi})}{2\sigma^2c^2}.$$

To show $\partial\underline{x}/\partial c > 0$, we only need $\Delta\pi \cdot \overline{x} + \underline{\pi} > 0$. In fact, let us consider any $x \in (\underline{x}, \overline{x})$. First, we know $U(x) \geq 0$, because the DM can always take the outside option and get zero. Also, by following the optimal strategy, the DM ends up either adopting the alternative, or taking the outside option, after paying the learning cost, so we have

$$U(x) = P(x)\left(\Delta\pi \cdot \overline{x} + \underline{\pi}\right) + \left(1 - P(x)\right)\left(\Delta\pi \cdot \underline{x} + \underline{\pi}\right) - cT(x) \geq 0,$$

where $P(x)$ is the probability that the DM ends up adopting the alternative given his current posterior belief as x, and $T(x)$ is the expected time spent on learning. By the inequality above and $\overline{x} > \underline{x}$, we know that $\Delta\pi \cdot \overline{x} + \underline{\pi} > 0$. Therefore, $\partial\underline{x}/\partial c > 0$.

Then, let us show that $V(x_1, x_2)$ decreases with c. This is straightforward by a simple argument. Consider $c' > c''$. Under learning cost c'', the DM can always replicate the optimal strategy under c'' but now incurs less learning cost.

Lastly, we prove that $\overline{X}(x)$ decreases with c. In fact, let us consider any x_1 and x_2 satisfying $x_2 \leq x_1 \leq \overline{X}(x_2)$ and $x_1 \geq \underline{x}$. By equation (31), we have,

$$\frac{\partial V(x_1, x_2)}{\partial c} = \frac{V(x_1, x_2) - (\Delta\pi \cdot x_1 + \underline{\pi})}{c} + \frac{2\sigma^2 c}{(\Delta\pi)^2} \frac{\overline{X}(x_2) - x_1}{\overline{X}(x_2)^2\left(1 - \overline{X}(x_2)\right)^2} \frac{\overline{X}(x_2)}{\partial c} \leq 0,$$

Since $V(x_1, x_2) \geq \Delta\pi \cdot x_1 + \underline{\pi}$, the inequality above implies that $\overline{X}(x_2)/\partial c \leq 0$ for $\forall x_2 \in [0, 1]$.

Now, notice that the HJB equation (9) can be equivalently written as,

$$\max\left\{\max_{1 \leq i \leq n}\left\{\frac{(\Delta\pi_i)^2}{2\sigma_i^2 c_i} x_i^2 (1 - x_i)^2 \widehat{V}_{x_i x_i}(\mathbf{x}) - 1\right\}, g(\mathbf{x}) - \widehat{V}(\mathbf{x})\right\} = 0,$$

so only $\sigma^2 c$ is identified in the model. Therefore, the comparative statics with respect to σ will be the same with that with c.

Next, we prove the comparative statics with respect to $\underline{\pi}$. Let us first show that $\partial V(x_1, x_2)/\partial\underline{\pi} \leq 1$. In fact, given $\Delta\pi$ fixed, consider an increase of $\underline{\pi}$ by $\Delta\underline{\pi} > 0$. If the outside option π_0 also increased $\Delta\underline{\pi}$, then obviously, $V(x_1, x_2)$ would be shifted upward by exactly $\Delta\underline{\pi}$. Since the outside option is kept unchanged, $V(x_1, x_2)$ must increase no more than $\Delta\underline{\pi}$. This implies that $\partial V(x_1, x_2)/\partial\underline{\pi} \leq 1$. Consider any x_1 and x_2 satisfying $x_2 \leq x_1 \leq \overline{X}(x_2)$ and $x_1 \geq \underline{x}$. By equation (31), we have,

$$\frac{\partial V(x_1, x_2)}{\partial\underline{\pi}} = 1 + \frac{2\sigma^2 c}{(\Delta\pi)^2} \frac{\overline{X}(x_2) - x_1}{\overline{X}(x_2)^2\left(1 - \overline{X}(x_2)\right)^2} \frac{\overline{X}(x_2)}{\partial\underline{\pi}} \leq 1.$$

This implies that $\partial\overline{X}(x_2)/\partial\underline{\pi} \leq 0$ for $x_2 \in [0, 1]$. Because $\overline{X}(x_2) = \overline{x}$ for $x_2 \in [0, \underline{x}]$, we have

$\partial\overline{x}/\partial\underline{\pi} \leq 0$. By taking derivative with respect to $\underline{\pi}$ on both sides of equation (18), we have

$$\frac{\partial \underline{x}}{\partial \underline{\pi}} = \frac{\underline{x}^2(1-\underline{x})^2}{\overline{x}^2(1-\overline{x})^2}\frac{\partial \overline{x}}{\partial \underline{\pi}} \leq 0.$$

Finally, we prove the comparative statics with respect to π_0. Obviously, $V(x_1, x_2)$ increases with the outside option π_0, so we have,

$$\frac{\partial V(x_1,x_2)}{\partial \pi_0} = \frac{2\sigma^2 c}{(\Delta\pi)^2}\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)^2\left(1-\overline{X}(x_2)\right)^2}\frac{\overline{X}(x_2)}{\partial \pi_0} \geq 0.$$

This implies that $\partial\overline{X}(x)/\partial\pi_0 \geq 0$ for $x \in [0,1]$, which in turn implies that $\partial\overline{x}/\partial\pi_0 \geq 0$. So we also have

$$\frac{\partial \underline{x}}{\partial \pi_0} = \frac{\underline{x}^2(1-\underline{x})^2}{\overline{x}^2(1-\overline{x})^2}\frac{\partial \overline{x}}{\partial \pi_0} \geq 0.$$

■

Proof of Theorem 3

Proof. We state first the value function $V(x_1, x_2)$ for this case and then prove how this value function can be obtained, which leads to the optimal learning strategies stated in the theorem. The value function is

$$V(x_1,x_2)=\begin{cases}\frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right)-\ln\left(\frac{1-\overline{Y}(x_2)}{\overline{Y}(x_2)}\right)\right] & \\ -\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Y}(x_2)}{1-\overline{Y}(x_2)}\frac{\overline{Y}(x_2)-x_1}{\overline{Y}(x_2)}+\Delta\pi\cdot x_1+\underline{\pi}, & \text{if } 1/2\le x_2\le x_1\le\overline{Y}(x_2)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right)-\ln\left(\frac{1-\overline{Z}(x_2)}{\overline{Z}(x_2)}\right)\right] & \\ -\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Z}(x_2)}{1-\overline{Z}(x_2)}\frac{\overline{Z}(x_2)-x_1}{\overline{Z}(x_2)}+\Delta\pi\cdot x_1+\underline{\pi}, & \text{if } 1/2\le 1-x_2\le x_1\le\overline{Z}(x_2)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_2)\left[\ln\left(\frac{1-x_2}{x_2}\right)+\ln\left(\frac{1-\overline{Y}(1-x_1)}{\overline{Y}(1-x_1)}\right)\right] & \\ +\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Y}(1-x_1)}{1-\overline{Y}(1-x_1)}\frac{1-\overline{Y}(1-x_1)-x_2}{\overline{Y}(1-x_1)}+\Delta\pi\cdot x_1+\underline{\pi}, & \text{if } 1/2\le 1-x_1\le 1-x_2\le\overline{Y}(1-x_1)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_2)\left[\ln\left(\frac{1-x_2}{x_2}\right)+\ln\left(\frac{1-\overline{Z}(1-x_1)}{\overline{Z}(1-x_1)}\right)\right] & \\ +\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Z}(1-x_1)}{1-\overline{Z}(1-x_1)}\frac{1-\overline{Z}(1-x_1)-x_2}{\overline{Z}(1-x_1)}+\Delta\pi\cdot x_1+\underline{\pi}, & \text{if } 1/2\le x_1\le 1-x_2\le\overline{Z}(1-x_1)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_2)\left[\ln\left(\frac{1-x_2}{x_2}\right)-\ln\left(\frac{1-\overline{Y}(x_1)}{\overline{Y}(x_1)}\right)\right] & \\ -\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Y}(x_1)}{1-\overline{Y}(x_1)}\frac{\overline{Y}(x_1)-x_2}{\overline{Y}(x_1)}+\Delta\pi\cdot x_2+\underline{\pi}, & \text{if } 1/2\le x_1\le x_2\le\overline{Y}(x_1)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_2)\left[\ln\left(\frac{1-x_2}{x_2}\right)-\ln\left(\frac{1-\overline{Z}(x_1)}{\overline{Z}(x_1)}\right)\right] & \\ -\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Z}(x_1)}{1-\overline{Z}(x_1)}\frac{\overline{Z}(x_1)-x_2}{\overline{Z}(x_1)}+\Delta\pi\cdot x_2+\underline{\pi}, & \text{if } 1/2\le 1-x_1\le x_2\le\overline{Z}(x_1)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right)+\ln\left(\frac{1-\overline{Y}(1-x_2)}{\overline{Y}(1-x_2)}\right)\right] & \\ +\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Y}(1-x_2)}{1-\overline{Y}(1-x_2)}\frac{1-\overline{Y}(1-x_2)-x_1}{\overline{Y}(1-x_2)}+\Delta\pi\cdot x_2+\underline{\pi}, & \text{if } 1/2\le 1-x_2\le 1-x_1\le\overline{Y}(1-x_2)\\ \frac{2\sigma^2c}{(\Delta\pi)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right)+\ln\left(\frac{1-\overline{Z}(1-x_2)}{\overline{Z}(1-x_2)}\right)\right] & \\ +\frac{2\sigma^2c}{(\Delta\pi)^2}\frac{1-2\overline{Z}(1-x_2)}{1-\overline{Z}(1-x_2)}\frac{1-\overline{Z}(1-x_2)-x_1}{\overline{Z}(1-x_2)}+\Delta\pi\cdot x_2+\underline{\pi}, & \text{if } 1/2\le x_2\le 1-x_1\le\overline{Z}(1-x_2)\\ \Delta\pi\cdot\max\{x_1,x_2\}+\underline{\pi}, & \text{otherwise,}\end{cases} \tag{xx}$$

where $\overline{Y}(x)$ is determined by the following boundary value problem of an ordinary differential equation,

$$\overline{Y}'(x) = -\frac{\Phi\left(\overline{Y}(x)\right)-\Phi(x)+(\Delta\pi)^3/(2\sigma^2c)}{\Phi'\left(\overline{Y}(x)\right)\left(\overline{Y}(x)-x\right)}, \tag{xxi}$$

$$\overline{Y}\left(\frac{1}{2}\right) = \Phi^{-1}\left(-\frac{(\Delta\pi)^3}{4\sigma^2c}\right), \tag{xxii}$$

and $\overline{Z}(x)$ is determined by

$$\overline{Z}'(x) = -\frac{\Phi\left(\overline{Z}(x)\right)+\Phi(x)}{\Phi'\left(\overline{Z}(x)\right)\left(1-x-\overline{Z}(x)\right)}, \tag{xxiii}$$

$$\overline{Z}\left(\frac{1}{2}\right) = \Phi^{-1}\left(-\frac{(\Delta\pi)^3}{4\sigma^2c}\right). \tag{xxiv}$$

Define x^* as the solution to the equation $x^* + \overline{Z}(x^*) = 1$. $\overline{X}(x)$ is then given by

$$\overline{X}(x) = \begin{cases} 1 - \overline{Y}^{-1}(1-x), & \text{if } 0 \le x < 1 - \Phi^{-1}\left(-\frac{(\Delta\pi)^3}{4\sigma^2 c}\right) \\ 1 - \overline{Z}^{-1}(1-x), & \text{if } 1 - \Phi^{-1}\left(-\frac{(\Delta\pi)^3}{4\sigma^2 c}\right) \le x < x^* \\ \overline{Z}(x), & \text{if } x^* \le x < \frac{1}{2} \\ \overline{Y}(x), & \text{otherwise.} \end{cases} \tag{xxv}$$

The proof follows almost the exactly same steps as the proof of Theorem 1 in the main context. First, it is easy to verify all the boundary conditions have been satisfied—actually this is how we come up with the expression of equation (xx). Therefore, the only thing we need to verify is the second inequality of equation (9). By symmetry, we only need to verify this inequality when $1/2 \le x_2 \le x_1 \le \overline{Y}(x_2)$ and when $1/2 \le 1 - x_2 \le x_1 \le \overline{Z}(x_2)$. The proofs are very similar for the two cases. Below, we provide the proof for the case that $1/2 \le x_2 \le x_1 \le \overline{Y}(x_2)$. In fact, similar to the proof of Lemma 3, we can show that this inequality will be satisfied as long as

$$\overline{Y}'(x_2) \le \frac{\overline{Y}(x_2)\left(1 - \overline{Y}(x_2)\right)}{x_2\,(1-x_2)}, \text{ for } \forall x_2 \in [1/2, 1).$$

By equation (xxi), we can write the inequality above equivalently as

$$J(x_2) \equiv x_2(1-x_2)\left(\Phi\left(\overline{Y}(x_2)\right) - \Phi(x_2) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right) - \frac{\overline{Y}(x_2) - x_2}{\overline{Y}(x_2)\left(1-\overline{Y}(x_2)\right)} \le 0, \text{ for } \forall x_2 \in [1/2, 1). \tag{xxvi}$$

Let us define $m = (\Delta\pi)^3/(4\sigma^2 c)$. By equation (xxii), we can calculate that $J(1/2) = m/4 - (\Phi^{-1}(-m) - 1/2)/\left[\Phi^{-1}(-m)(1-\Phi^{-1}(-m))\right]$. Using the definition of $\Phi(\cdot)$, it is not difficult to show that $J(1/2) < 0$. Let us prove by contradiction. Suppose the inequality (xxv) does not hold. Since $J(x)$ is continuous, this implies that there exists $x_2^* \equiv \min_{1/2 \le x_2 \le 1}\{x_2 | J(x_2) = 0\}$ such that $J'(x_2^*) > 0$. By using the equation (xxv) and (xxi) and the condition that $J(x_2^*) = 0$, we can simplify and get the following expression:

$$J'(x_2^*) = -2\left(\Phi\left(\overline{Y}(x_2^*)\right) - \Phi(x_2^*) + \frac{(\Delta\pi)^3}{2\sigma^2 c}\right)\left(x_2^* + \overline{Y}(x_2^*) - 1\right) > 0.$$

However, we know that $\Phi\left(\overline{Y}(x_2^*)\right) - \Phi(x_2^*) + \frac{(\Delta\pi)^3}{2\sigma^2 c} \ge 0$, and $x_2^* + \overline{Y}(x_2^*) - 1 \ge 2x_2^* - 1 \ge 0$. This is a contradiction to the inequality above, which implies that inequality (xxv) will always hold. ∎

Proof of Theorem 4: Adoption Likelihood:

Proof. When $x_1 \geq \overline{X}(x_2)$, the DM adopts alternative 1 immediately, so $P_1(x_1, x_2) = 1$. When $x_1 \leq \underline{x}$ or $x_2 \geq \overline{X}(x_1)$, the DM will never adopt alternative 1, therefore $P_1(x_1, x_2) = 0$. Otherwise when $\underline{x} < x_1 < \overline{X}(x_2)$ and $x_2 < \overline{X}(x_1)$, there are two cases, depending on the value of x_2.

In the first case with $x_2 \leq \underline{x}$, the DM considers alternative 1 only. Given his current belief x_1, he will continue learning until his belief equals either $\underline{x}$ or $\overline{x}$. According to the Optional Stopping Theorem, we have $x_1 = P_1(x_1, x_2)\overline{x} + [1 - P_1(x_1, x_2)]\,\underline{x}$, therefore

$$P_1(x_1, x_2) = \frac{x_1 - \underline{x}}{\overline{x} - \underline{x}}, \quad \underline{x} < x_1 \leq \overline{x}, x_2 \leq \underline{x}.$$

In the second case, $x_2 > \underline{x}$. When $x_1 \geq x_2$, the DM keeps learning alternative 1 until his belief reaches either $\overline{X}(x_2)$ or x_2. Let us define the probability of reaching $\overline{X}(x_2)$ as $q_1(x_1, x_2)$. Then by invoking the Optional Stopping Theorem, we similarly get

$$q_1(x_1, x_2) = \frac{x_1 - x_2}{\overline{X}(x_2) - x_2}.$$

According to symmetry, the probability of reaching $\overline{X}(x_1)$, starting from (x_1, x_2) with $x_1 < x_2$, would be $q_1(x_2, x_1)$. Let us further define $P_0(x)$ as the probability of taking the outside option, given the DM's current beliefs as (x, x). Let us consider an infinitesimal learning on alternative 1 at (x, x), with belief update as dx. By conditioning on the sign of dx, we have

$$\begin{aligned}
P_0(x) &\equiv \Pr[\text{outside option}|(x,x)] \\
&= \frac{1}{2}\Pr[\text{outside option}|(x,x), dx \geq 0] + \frac{1}{2}\Pr[\text{outside option}|(x,x), dx < 0] \\
&= \frac{1}{2}\left[1 - q_1(x+dx, x)\right] P_0(x) + \frac{1}{2}\left[1 - q_1(x, x-dx)\right] P_0(x-dx) \\
&= P_0(x) - \frac{dx}{2}\left[P_0'(x) - \left(\frac{\partial q_1(x_1,x_2)}{\partial x_2} - \frac{\partial q_1(x_1,x_2)}{\partial x_1}\right)\bigg|_{x_1=x_2=x} P_0(x)\right] + o(dx),
\end{aligned}$$

where the last equality is obtained by doing a Taylor expansion of $q_1(x+dx, x)$, $q_1(x, x-dx)$, and $P_0(x-dx)$. By canceling out $P_0(x)$, dividing by dx and taking limit of dx going to zero for the equation above, we have

$$\frac{P_0'(x)}{P_0(x)} = \left(\frac{\partial q_1(x_1,x_2)}{\partial x_2} - \frac{\partial q_1(x_1,x_2)}{\partial x_1}\right)\bigg|_{x_1=x_2=x} = -\frac{2}{\overline{X}(x) - x}.$$

Combining the differential equation above with the initial condition $P_0(\underline{x}) = 1$, we can solve $P_0(x)$

as

$$P_0(x) = e^{-\int_{\underline{x}}^{x} \frac{2}{\overline{X}(\xi)-\xi} d\xi}.$$

Starting from (x_1, x_2) with $x_1 \geq x_2$, the DM learns information on alternative 1. With probability $q_1(x_1, x_2)$, he reaches the adoption boundary $\overline{X}(x_2)$, and immediately adopts alternative 1. With probability $1 - q_1(x_1, x_2)$, he reaches x_2. Then starting from (x_2, x_2), he eventually adopts alternative 1 with probability $1/2[1 - P_0(x_2)]$. Therefore, we have

$$P_1(x_1, x_2) = q_1(x_1, x_2) + [1 - q_1(x_1, x_2)] \frac{1}{2}[1 - P_0(x_2)], \quad \underline{x} < x_2 < x_1 < \overline{X}(x_2).$$

Similarly, starting from (x_1, x_2) with $x_1 < x_2$, the DM learns information on alternative 2. With probability $1 - q_1(x_2, x_1)$, he reaches x_1, upon which, he eventually adopts alternative 1 with probability $1/2[1 - P_0(x_1)]$.

$$P_1(x_1, x_2) = [1 - q_1(x_2, x_1)] \frac{1}{2}[1 - P_0(x_1)], \quad \underline{x} < x_1 < x_2 < \overline{X}(x_1).$$

By combining all the scenarios above, we proved the theorem. ∎

Proof of Theorem 5: Probability of Being Correct:

Proof. By symmetry, we only need to consider the case that $x_1 \geq x_2$. There are four cases to consider.

In the first case, when $x_1 \geq \overline{X}(x_2)$, the DM adopts alternative 1 right away. In this case, he will be correct if and only if the alternative 1 turns out to be of high value. Therefore, $Q(x_1, x_2) = x_1$.

In the second case, when $x_2 \leq x_1 \leq \underline{x}$, the DM will adopt the outside option, which is ex-post correct if and only if both alternatives are of low value, i.e., $Q(x_1, x_2) = (1 - x_1)(1 - x_2)$.

In the third case, when $x_2 \leq \underline{x} \leq x_1 \leq \overline{x}$, the DM will continue learning until his belief equals either $\underline{x}$ or $\overline{x}$. Therefore,

$$\begin{aligned} Q(x_1, x_2) &= \frac{x_1 - \underline{x}}{\overline{x} - \underline{x}} \overline{x} + \left[1 - \frac{x_1 - \underline{x}}{\overline{x} - \underline{x}}\right] (1 - \underline{x})(1 - x_2), \\ &= \frac{\overline{x}(x_1 - \underline{x}) + (1 - \underline{x})(\overline{x} - x_1)(1 - x_2)}{\overline{x} - \underline{x}}. \end{aligned}$$

In the last case, when $\underline{x} < x_2 < x_1 < \overline{X}(x_2)$, the DM keeps learning alternative 1 until his belief reaches either $\overline{X}(x_2)$ or x_2. Let us define $\widehat{Q}_0(x)$ as the probability of being correct, given the DM's beliefs as (x,x). Let us consider infinitesimal learning on alternative 1 at (x,x), with belief update as dx. By conditioning on the sign of dx, we have

$$\begin{aligned}
\widehat{Q}_0(x) &\equiv \Pr[\text{correct}|(x,x)] \\
&= \frac{1}{2}\Pr[\text{correct}|(x,x), dx \geq 0] + \frac{1}{2}\Pr[\text{correct}|(x,x), dx < 0] \\
&= \frac{1}{2}q_1(x+dx,x)\overline{X}(x) + \frac{1}{2}\left[1 - q_1(x+dx,x)\right]\widehat{Q}_0(x) \\
&+ \frac{1}{2}q_1(x,x-dx)\overline{X}(x-dx) + \frac{1}{2}\left[1 - q_1(x,x-dx)\right]\widehat{Q}_0(x-dx).
\end{aligned}$$

Similar to the proof of Theorem 4, by doing a Taylor expansion of $q_1(x+dx,x)$, $q_1(x,x-dx)$,, $\widehat{Q}_0(x-dx)$, and $\overline{X}(x-dx)$, and simplifying, we have

$$\widehat{Q}_0'(x) = \left[\overline{X}(x) - \widehat{Q}_0(x)\right]\left(\frac{\partial q_1(x_1,x_2)}{\partial x_1} - \frac{\partial q_1(x_1,x_2)}{\partial x_2}\right)\bigg|_{x_1=x_2=x} = \frac{2\left[\overline{X}(x) - \widehat{Q}_0(x)\right]}{\overline{X}(x) - x}.$$

Combining the differential equation above with the initial condition $\widehat{Q}_0(\underline{x}) = (1-\underline{x})^2$, we can solve $\widehat{Q}_0(x)$ as

$$\widehat{Q}_0(x) = (1-\underline{x})^2 e^{-\int_{\underline{x}}^{x}\frac{2d\xi}{\overline{X}(\xi)-\xi}} + \int_{\underline{x}}^{x} e^{-\int_{\eta}^{x}\frac{2d\xi}{\overline{X}(\xi)-\xi}}\frac{2\overline{X}(\eta)}{\overline{X}(\eta)-\eta}d\eta.$$

Given $\widehat{Q}_0(x)$, we have,

$$Q(x_1,x_2) = q_1(x_1,x_2)\overline{X}(x_2) + \left[1 - q_1(x_1,x_2)\right]\widehat{Q}_0(x_2).$$

By combining all the scenarios above, we proved the theorem. ■

Solution to the Problem with Two Heterogeneous Alternatives:

The following theorem characterizes the optimal learning strategy in the case of two heterogeneous alternatives under the assumption that alternative 1 is "better" than alternative 2, and (x_1^*, x_2^*) exists. Similar to the proof of Theorem 1, it is straightforward but tedious to show that the variational inequalities in equation (9) are satisfied. The details are thus omitted here.

Theorem A1: *Consider the Decision Maker's optimal learning problem in equation (6) with two heterogeneous alternatives, $\Delta\pi_1 > \Delta\pi_2$ (or $c_1 < c_2$). Suppose that (x_1^*, x_2^*) exits. When the value of* *the outside option is relatively high, the value function $V(x_1, x_2)$ is given by*

$$V(x_1,x_2) = \begin{cases} \frac{2\sigma_1^2 c_1}{(\Delta\pi_1)^2}(1-2x_1)\left[\ln\left(\frac{1-x_1}{x_1}\right) - \ln\left(\frac{1-\overline{X}(x_2)}{\overline{X}(x_2)}\right)\right] & \\ -\frac{2\sigma_1^2 c_1}{(\Delta\pi_1)^2}\frac{1-2\overline{X}(x_2)}{1-\overline{X}(x_2)}\frac{\overline{X}(x_2)-x_1}{\overline{X}(x_2)} + \Delta\pi_1 \cdot x_1 + \underline{\pi}_1, & \underline{x}_1, \underline{X}_1(x_2) \le x_1 \le \overline{X}_1(x_2), x_2 \le y(x_1) \\ \frac{2\sigma_2^2 c_2}{(\Delta\pi_2)^2}(1-2x_2)\left[\ln\left(\frac{1-x_2}{x_2}\right) - \ln\left(\frac{1-\overline{X}(x_1)}{\overline{X}(x_1)}\right)\right] & \\ -\frac{2\sigma_2^2 c_2}{(\Delta\pi_2)^2}\frac{1-2\overline{X}(x_1)}{1-\overline{X}(x_1)}\frac{\overline{X}(x_1)-x_2}{\overline{X}(x_1)} + (\Delta\pi_2)x_2 + \underline{\pi}_2, & \underline{x}_2 \le x_2 \le \overline{X}_2(x_1), x_2 > y(x_1) \\ \Delta\pi_1 \cdot x_1 + \underline{\pi}_1, & x_1 > \overline{X}_1(x_2) \\ \Delta\pi_2 \cdot x_2 + \underline{\pi}_2, & x_2 > \overline{X}_2(x_1) \text{ or } x_1 < \underline{X}_1(x_2) \\ \pi_0, & \text{otherwise.} \end{cases}$$

The adoption boundary of alternative 1, $\overline{X}_1(x_2)$ is given by

$$\overline{X}_1(x_2) = \begin{cases} \overline{\chi}_1(x_2) & x_2 \ge x_2^* \\ \widetilde{\chi}_1\left(y^{-1}(x_2)\right), & x_2^* > x_2 \ge \underline{x}_2 \\ \overline{x}_1, & \text{otherwise.} \end{cases}$$

The adoption boundary of alternative 2, $\overline{X}_2(x_1)$ supported in $[0, x_1^*]$ is given by

$$\overline{X}_2(x_1) = \begin{cases} \overline{\chi}_2(x_1), & x_1^* \ge x_1 \ge \underline{x}_1 \\ \overline{x}_2, & x_1 < \underline{x}_1. \end{cases}$$

The adoption boundary of alternative 2, $\underline{X}_1(x_2)$, and $\overline{\chi}_1(x_2)$, both supported in $[x_2^*, 1]$, are given by the following equations:

$$\begin{aligned} F(\overline{\chi}_1(x_2)) - F(\underline{X}_1(x_2)) &= -\frac{(\Delta\pi_1)^3}{2\sigma_1^2 c_1} \\ \left(\frac{1}{\overline{\chi}_1(x_2)} + \frac{1}{1-\overline{\chi}_1(x_2)}\right) - \left(\frac{1}{\underline{X}_1(x_2)} + \frac{1}{1-\underline{X}_1(x_2)}\right) &= \frac{(\Delta\pi_1)^2\left[2\left(\Delta\pi_2 \cdot x_2 + \underline{\pi}_2 - \underline{\pi}_1\right) - \Delta\pi_1\right]}{2\sigma_1^2 c_1}. \end{aligned}$$

$\overline{\chi}_2(x_1)$, $\widetilde{\chi}_1(x_1)$, and $y(x_1)$, all supported in $[\underline{x}_1, x_1^*]$, are determined by the following boundary value

problem:

$$
\begin{aligned}
\frac{dF\left(\overline{\chi}_2(x_1)\right)}{dx_1} &= -\frac{\frac{2\sigma_1^2 c_1}{(\Delta\pi_1)^2}\left(F\left(\widetilde{\chi}_1(x_1)\right) - F(x_1)\right) + \Delta\pi_1}{\frac{2\sigma_2^2 c_2}{(\Delta\pi_2)^2}\left(\overline{\chi}_2(x_1) - y(x_1)\right)}, \\
\frac{dF\left(\widetilde{\chi}_1(x_1)\right)}{dx_1} &= -y'(x_1)\frac{\frac{2\sigma_2^2 c_2}{(\Delta\pi_2)^2}\left(F\left(\overline{\chi}_2(x_1)\right) - F(y(x_1))\right) + \Delta\pi_2}{\frac{2\sigma_1^2 c_1}{(\Delta\pi_1)^2}\left(\widetilde{\chi}_1(x_1) - x_1\right)}, \\
& \left(\widetilde{\chi}_1(x_1) - x_1\right)\left[\frac{2\sigma_1^2 c_1}{(\Delta\pi_1)^2}\left(F\left(\widetilde{\chi}_1(x_1)\right) - F(x_1)\right) + \Delta\pi_1\right] \\
&= \left(\overline{\chi}_2(x_1) - y(x_1)\right)\left[\frac{2\sigma_2^2 c_2}{(\Delta\pi_2)^2}\left(F\left(\overline{\chi}_2(x_1)\right) - F(y(x_1))\right) + \Delta\pi_2\right], \\
\overline{\chi}_2(\underline{x}_1) &= \overline{x}_2, \\
\widetilde{\chi}_1(\underline{x}_1) &= \tilde{x}_1, \\
y(\underline{x}_1) &= \underline{x}_2.
\end{aligned}
$$

Case of Heterogeneous Alternatives with $\Delta\pi_1 > \Delta\pi_2$ and $c_1 > c_2$:

Now, let us have a look at the more complicated case that alternative 1 may have a higher information-to-noise ratio, while at the same time a higher learning cost. Figure A1 shows the optimal learning strategy under some parameter settings. As we can see, the solution structure remains mostly unchanged. We still have a simple cutoff policy to construct consideration set, but the boundary separating "Learn 1" and "Learn 2" can become more complicated now, and can be sensitive to the parameter setting.

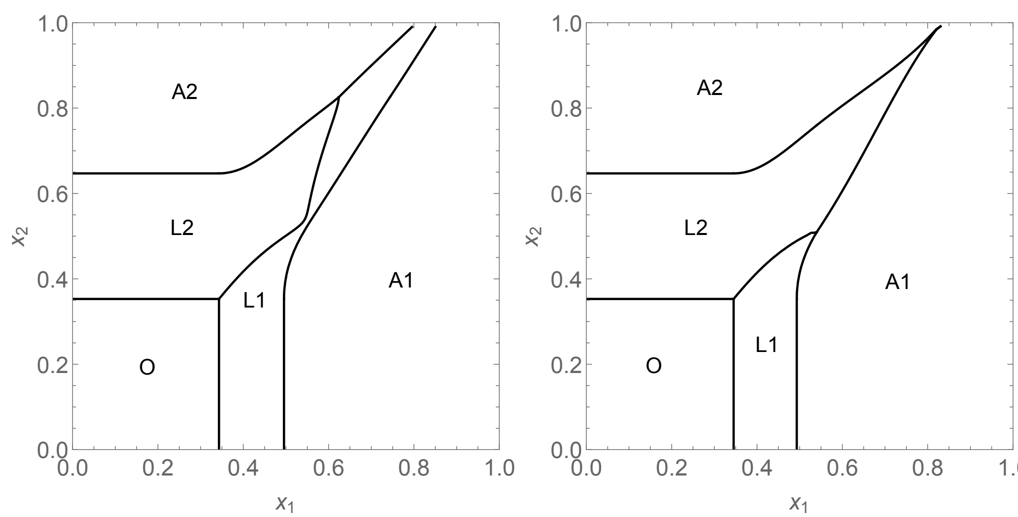

Figure A1: Optimal learning strategy with heterogeneous alternatives. We have $c_1\sigma_1^2/\Delta\pi = 0.33$ for the left panel and $c_1\sigma_1^2/\Delta\pi = 0.34$ for the right panel. For both panels: $\Delta\pi_1 = 1.2\Delta\pi_2 = 1.2\Delta\pi$, $(\pi_0 - \underline{\pi}_1)/\Delta\pi = (\pi_0 - \underline{\pi}_2)//\Delta\pi = 0.5$, and $c_2\sigma_2^2/\Delta\pi = c_2\sigma_2^2/\Delta\pi = 1$.

Solution to the Problem with Time Discounting:

Theorem A2: *There exists a unique solution $V(x_1,x_2)$ to equation (43):*

$$V(x_1,x_2)=\begin{cases}\left[\frac{\alpha-1}{2\alpha}\underline{\pi}Z(x_2)^{\frac{\alpha+1}{2}}+\frac{\alpha+1}{2\alpha}\overline{\pi}Z(x_2)^{\frac{\alpha-1}{2}}\right]x_1^{\frac{\alpha+1}{2}}(1-x_1)^{-\frac{\alpha-1}{2}} & \\ +\left[\frac{\alpha-1}{2\alpha}\overline{\pi}Z(x_2)^{-\frac{\alpha+1}{2}}+\frac{\alpha+1}{2\alpha}\underline{\pi}Z(x_2)^{-\frac{\alpha-1}{2}}\right]x_1^{-\frac{\alpha-1}{2}}(1-x_1)^{\frac{\alpha+1}{2}}, & x_2\le x_1\le \widetilde{X}(x_2), x_1\ge \underset{\sim}{x} \\ \left[\frac{\alpha-1}{2\alpha}\underline{\pi}Z(x_1)^{\frac{\alpha+1}{2}}+\frac{\alpha+1}{2\alpha}\overline{\pi}Z(x_1)^{\frac{\alpha-1}{2}}\right]x_2^{\frac{\alpha+1}{2}}(1-x_2)^{-\frac{\alpha-1}{2}} & \\ +\left[\frac{\alpha-1}{2\alpha}\overline{\pi}Z(x_1)^{-\frac{\alpha+1}{2}}+\frac{\alpha+1}{2\alpha}\underline{\pi}Z(x_1)^{-\frac{\alpha-1}{2}}\right]x_2^{-\frac{\alpha-1}{2}}(1-x_2)^{\frac{\alpha+1}{2}}, & x_1\le x_2\le \widetilde{X}(x_1), x_2\ge \underset{\sim}{x} \\ \Delta\pi\cdot x_1+\underline{\pi}, & x_1>\widetilde{X}(x_2) \\ \Delta\pi\cdot x_2+\underline{\pi}, & x_2>\widetilde{X}(x_1) \\ \pi_0, & \textit{otherwise,}\end{cases}$$

where $\alpha\equiv\sqrt{1+\frac{8r\sigma^2}{(\Delta\pi)^2}}>1$, *and* $\widetilde{X}(x)$ *is given by*

$$\widetilde{X}(x)=\begin{cases}\frac{1}{Z(x)+1}, & x\ge \underset{\sim}{x} \\ \tilde{x}, & \textit{otherwise.}\end{cases}$$

$Z(x)$ *is determined by the following boundary value problem:*

$$\begin{aligned}\left(\alpha^2-1\right)x(1-x)\frac{Z'(x)}{Z(x)} &= \alpha^2+\alpha(1-2x)\frac{Z(x)^\alpha+\left(\frac{1-x}{x}\right)^\alpha}{Z(x)^\alpha-\left(\frac{1-x}{x}\right)^\alpha}-(1-2x)\frac{Z(x)-\frac{\overline{\pi}}{\underline{\pi}}}{Z(x)+\frac{\overline{\pi}}{\underline{\pi}}} \\ &- \alpha\frac{\left[Z(x)^\alpha+\left(\frac{1-x}{x}\right)^\alpha\right]\left[Z(x)-\frac{\overline{\pi}}{\underline{\pi}}\right]}{\left[Z(x)^\alpha-\left(\frac{1-x}{x}\right)^\alpha\right]\left[Z(x)+\frac{\overline{\pi}}{\underline{\pi}}\right]}, \\ Z(\underset{\sim}{x}) &= \frac{1-\tilde{x}}{\tilde{x}}.\end{aligned}$$

$\tilde{x}$ *and* $\underset{\sim}{x}$ *are determined by the following equations:*

$$\begin{aligned}\left(\frac{\tilde{x}}{\underset{\sim}{x}}\right)^{-\frac{\alpha-1}{2}}\left(\frac{1-\tilde{x}}{1-\underset{\sim}{x}}\right)^{\frac{\alpha+1}{2}} &= \frac{\Delta\pi(\alpha-1)\tilde{x}+\underline{\pi}(\alpha+1-2\tilde{x})}{\pi_0(\alpha+1-2\underset{\sim}{x})}, \\ \left(\frac{\tilde{x}}{\underset{\sim}{x}}\right)^{\frac{\alpha+1}{2}}\left(\frac{1-\tilde{x}}{1-\underset{\sim}{x}}\right)^{-\frac{\alpha-1}{2}} &= \frac{\Delta\pi(\alpha+1)\tilde{x}+\underline{\pi}(\alpha-1+2\tilde{x})}{\pi_0(\alpha-1+2\underset{\sim}{x})}.\end{aligned}$$

The proof to Theorem A2 follows similar steps to the proof of Theorem 1, and thus is omitted.

References

Klaus Adam. Learning while searching for the best alternative. *Journal of Economic Theory*, 101 (1):252–280, 2001.

R.F. Bass and É Pardoux. Uniqueness for diffusions with piecewise constant coefficients. *Probability Theory and Related Fields*, 76(4):557–572, 1987.

Dirk Bergemann and Juuso Välimäki. Learning and strategic pricing. *Econometrica*, 64:773–705, 1996.

Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

Fernando Branco, Monic Sun, and J. Miguel Villas-Boas. Optimal search for product information. *Management Science*, 58(11):2037–2056, 2012.

Luis A. Caffarelli. The regularity of free boundaries in higher dimensions. *Acta Mathematica*, 139: 155–184, 1977.

Yeon-Koo Che and Konrad Mierendorff. Optimal sequential decision with limited attention. Working paper, Columbia University, 2016.

Herman Chernoff. Optimal stochastic control. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 221–252, 1968.

Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

Jean-Paul Décamps, Thomas Mariotti, and Stéphane Villeneuve. Irreversible investment in alternative projects. *Economic Theory*, 28(2):425–448, 2006.

Peter A. Diamond. A model of price adjustment. *Journal of Economic Theory*, 3(2):156–168, 1971.

A. Dixit. *Art of smooth pasting*, volume 55. Harwood Academic, Chur, Switzerland, 1993.

Avinash K. Dixit and Robert S. Pindyck. *Investment under uncertainty.* Princeton University Press, 1994.

Laura Doval. Whether or not to open pandora's box. *Working paper, Northwestern University*, 2014.

Bernard Dumas. Super contact and related optimality conditions. *Journal of Economic Dynamics and Control*, 15(4):675–685, 1991.

Leonardo Felli and Christopher Harris. Learning, wage dynamics, and firm-specific human capital. *Journal of Political Economy*, 104(4):838–868, 1996.

E. Roberts Fernholz, Tomoyuki Ichiba, Ioannis Karatzas, and Prokaj Vilmos. Planar diffusions with rank-based characteristics and perturbed tanaka equations. *Probability Theory and Related Fields*, 156(1):343–374, 2013.

Jean G. Forand. Keeping your options open. *Journal of Economic Dynamics and Control*, 53:47–68, 2015.

Drew Fudenberg and Tomasz Strzalecki. Dynamic logit with choice aversion. *Econometrica*, 83(2): 651–691, 2015.

Drew Fudenberg, Philipp Strack, and Tomasz Strzalecki. Stochastic choice and optimal sequential sampling. *Working paper, Harvard University*, 2015.

John C. Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society B*, 41:148–164, 1979.

John C. Gittins, Kevin D. Glazebrook, and Richard Weber. *Multi-armed bandit allocation indices, 2nd Edition.* Wiley Online Library, 2011.

Kevin D. Glazebrook. Stoppable families of alternative bandit processes. *Journal of Applied Probability*, pages 843–854, 1979.

Kiyosi Itô. On stochastic differential equations. *Memoirs of the American Mathematical Society*, 4: 1–51, 1951.

Ioannis Karatzas. Gittins indices in the dynamic allocation problem for diffusion processes. *The Annals of Probability*, pages 173–192, 1984.

Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus.* Springer, New York, 1991.

T.Tony Ke, Zuo-Jun Max Shen, and J.Miguel Villas-Boas. Search for information on multiple products. *Management Science (forthcoming)*, 62(12):3576–3603, 2016.

Dmitri Kuksov and J Miguel Villas-Boas. When more alternatives lead to less choice. *Marketing Science*, 29(3):507–524, 2010.

Pierre-Louis Lions. Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. I. The dynamic programming principle and applications. *Comm. Partial Differential Equations*, 8(10):1101–1174, 1983a.

Pierre-Louis Lions. Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. II. Viscosity solutions and uniqueness. *Comm. Partial Differential Equations*, 8(11):1229–1276, 1983b.

Robert S. Liptser and Albert N. Shiryaev. *Statistics of Random Processes: General Theory*, volume 1. Springer, 2001.

Avi Mandelbaum, Larry A. Shepp, and Robert J. Vanderbei. Optimal switching between a pair of brownian motions. *The Annals of Probability*, 18(3):1010–1033, 1990.

John Joseph McCall. Economics of information and job search. *The Quarterly Journal of Economics*, pages 113–126, 1970.

Giuseppe Moscarini and Lones Smith. The optimal level of experimentation. *Econometrica*, 69(6): 1629–1644, 2001.

Goran Peskir and Albert Shiryaev. *Optimal stopping and free-boundary problems*. Birkhäuser Verlag, Basel, Switzerland, 2006.

Kevin Roberts and Martin L. Weitzman. Funding criteria for research, development, and exploration projects. *Econometrica*, 49(5):1261–1288, 1981.

Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9 (2):185–202, 1974a.

Michael Rothschild. Searching for the lowest price when the distribution of prices is unknown. *The Journal of Political Economy*, 82(4):689–711, 1974b.

Benjamin Scheibehenne, Rainer Greifeneder, and Peter M. Todd. Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, 37(3):409–425, 2010.

Daniel W. Stroock and S.R.S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag, Berlin, 1979.

Bruno Strulovici and Martin Szydlowski. On the smoothness of value functions and the existence of optimal strategies in diffusion models. *Journal of Economic Theory*, 159:1016–1055, 2015.

Helen Weeds. Strategic delay in a real options model of R&D competition. *The Review of Economic Studies*, 69(3):729–747, 2002.

Martin L. Weitzman. Optimal search for the best alternative. *Econometrica*, 47(3):641–654, 1979.

P. Whittle. Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society*, 42 (2):143–149, 1980.