Essays in Learning and Information

Guo Bai

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Economics
University College London

August 6, 2023

Declaration

I, Guo Bai, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Acknowledgements

I would like to express my deepest gratitude to Martin Cripps, Konrad Mierendorff and Deniz Kattwinkel for their patience and encouragement. I am also grateful to Wei Cui, Duarte Gonçalves, Antonio Guarino, Philippe Jehiel, Fabien Postel-Vinay, Franck Portier, Ian Preston, Nikita Roketskiy, Vasiliki Skreta, Ran Spiegler, Michael Thaler, Ming Yang, Ji Hee Yoon and other faculty at UCL for helpful discussions and conversations. I would also like to thank Daniella Harper for her administrative support. Many thanks to all my office mates, and in particular Amir Habibi, Yiming Xia, Jihyun Bae, Gherardo Caracciolo, Morgane Richard, Anusha Guha, Dajana Xhani, Uta Bolt, Riccardo D'Adamo, Enrico Miglino, Nathan Hancart, Alexander Clyde, Wenhao Cheng, Guanyi Wang, Dongxian Guo, Ruimin Ao, Rachel Tan, Xinyu Qing, Jiacheng Fan, Yanziyi William Zhang, Ruolei Zhu and Yilun Allen Shen for their feedback and moral support. Special thanks to my friend, Yingrui, for her support in every aspect of my life. Lastly, I'd like to acknowledge my family, especially my parents for all the financial and emotional support.

Abstract

This thesis consists of three chapters that investigate how economic agents acquire information and learn from the information. I use theoretical models to understand information acquisition and learning. Chapter 1 studies how decision-makers acquire information when they compete for a first-mover advantage. I show that the first-mover advantage gives the decision-makers incentives to preempt. In equilibrium, decision-makers create strategic uncertainties on when they stop acquiring information and taking an action. Chapter 2 studies a single decision-maker's search and learning problem in which the signals that can be obtained are binary. I show that the decision maker's optimal strategy depends on her time risk attitude and her patience level. Chapter 3 investigates the question of whether a designer can learn how a decision-maker learns from information by observing the signals she receives and the actions she takes. It studies the probability of the designer learning the decision-maker's prior and characterises the optimal payoff structure that maximises this probability.

Impact statement

The work presented in this thesis can have an impact both inside and outside academia. Economic agents face uncertainties and they often need to make decisions in this uncertain environment. Acquiring information reduces these uncertainties and help economic agents make better decisions. This thesis aims to contribute to the understanding of information acquisition and learning in economic activities.

All three chapters in this thesis ask questions that have not been extensively investigated yet in the literature. As the first step in contributing to the academic community, Chapter 1 was presented at an academic conference. All three chapters will be submitted to academic journals in order to contribute to the area of economic theory.

Outside academia, the theories in this thesis may bring interests to researchers, firms, corporations and public sectors who make decisions in the uncertain environment. In addition, the work presented in this thesis can have policy implications in terms of competition and how information should be shared among different parties.

Contents

List of Figures

Introduction

This thesis consists of three chapters. The theme of this thesis is how decision-makers acquire information and learn from the information. I use theoretical models to understand the information acquisition problem facing economic agents, including firms, corporations and individuals.

In Chapter 1, I study how decision-makers acquire information when they compete for a first-mover advantage. This chapter uses a dynamic information acquisition model with two players to understand the tradeoff between taking the correct action and preemption. The model applies to a range of settings. These include market entry where the decision to enter a new market is not immediately observable, research and development (R&D) races where the decision to develop new technology is private, and priority races in research where publishing new findings takes time. The strategic component in this model comes from the payoff externality generated by the first-mover advantage. The player's payoff from taking the irreversible action depends on both her belief about the state and her belief about her opponent's action. The equilibria have different qualitative features depending on the prior beliefs. The main result shows for different priors what types of equilibria exist. When players have sufficiently uncertain priors, both players acquire information in equilibrium. The equilibrium strategies consist of two stages: an information acquisition stage and a randomisation stage. At the information acquisition stage, the player only stops and takes an action when a breakthrough or breakdown occurs. If the player observes a breakthrough, she takes the risky action, and if she observes a breakdown, she takes the safe action. When there is no breakthrough or breakdown, the player continues to acquire information. After a fixed period of time with no breakthrough or breakdown, the player transitions to the randomisation stage where she not only stops after observing a breakthrough or breakdown but also stops and takes the risky action at each time at a positive rate. The existence of the random stopping equilibrium is significantly different from the single decision-maker case.

In Chapter 2, I study a single decision-maker's search and learning problem. I ask the following question: when an agent has to learn a realisation of the random variable, when she can only ask questions with yes or no answers, and if she can ask one question each time, what is the optimal learning strategy? I show that the optimal learning strategy depends on the agent's time risk attitude and how patient the agent is. Only two learning strategies can be optimal: the linear search or the binary search. Linear search is a sequence of questions that allows the agent to learn immediately but only with a positive probability. She may learn it today, tomorrow, or the day after tomorrow. It is risky in terms of the timing of learning the realisation of the random variable. Binary search is a sequence of questions that allows the agent to learn on a predetermined day, but not earlier than that. It is safe in terms of the timing of learning the realisation of the random variable. All other learning strategies are sub-optimal. If the agent is time risk averse, or sufficiently patient, she prefers binary search that allows her to learn safely on a predetermined day. If the agent is time risk seeking and impatient, she prefers linear search that allows her to learn immediately.

In Chapter 3, I investigate the following question: by observing what a decision-maker sees and what they do, is it possible that a third party could learn how this decision-maker learns from the information they received? An example is the experimental economists designing experiments to detect how participants process information and form beliefs. In this chapter, I investigate this question in the following situation. Suppose it is common knowledge that the decision-maker updates her beliefs using the Bayes rule but the prior is unknown, I investigate whether it is possible for the designer to learn the decision-maker's prior belief by observing the signals she received and the actions she took. I show in this chapter that it is not always the case that the designer can learn the decision-maker's prior belief.

Chapter 1

Private Information Acquisition and Preemption: a Strategic Wald Problem

1 Introduction

Information helps decision-makers take better actions but information acquisition often takes time. When decision-makers compete for a first-mover advantage, there is tension between acquiring information and preemption. This paper uses a dynamic information acquisition model with two players to understand the tradeoff between taking the correct action and preemption.

The model of dynamic information acquisition studied has an unknown payoff-relevant state that is either high or low and is constant over time. At each time, players can choose among three things: to take an irreversible risky action, to take an irreversible safe action, or to delay the action and pay a cost to acquire information about the state. The safe action guarantees the player a payoff of zero in both states. The risky action is better than the safe action only in the high state. Furthermore, there is always a first-mover premium associated with the risky action. The second player to take the risky action does not get the first-mover premium but still gets a positive payoff in the high state. The player only observes her own signals but not the opponent's signal or actions. The contribution of this paper is to investigate the equilibria of the dynamic information acquisition model where the *actions taken* and the *information acquisition* are both *private.*

The model applies to a range of settings. These include: market entry where the decision to enter a new market is not immediately observable, research and development (R&D) races where the decision to develop new technology is private, and priority races in research where publishing new findings takes time.

In the case of market entry, the first firm to enter a new market captures a higher market share, but before entry, they need to investigate whether the demand in the market is high or low. This investigation takes time and the preparation for entering a new market is not immediately observable. If the firm observes a competing firm's entry, it is too late to reverse their own entry. The delay in observing the competitor's action makes entering a new market essentially a private action.

The R&D of new technologies is often competitive as the first firm to develop the new technology gets more exposure and a higher market share. Whether and *when* to invest in developing a technology is risky. Before investing in developing this technology, firms can collect information about its feasibility. The research progress, the research results and the decision on investing in developing this technology may not be observed by competitors.

The frontier of academic research often attracts attention from multiple research groups. The first research group to publish their discoveries receives most credit. Scientific discoveries are often made independently [1] and their competitors are often invisible. A research group may work on a project privately for decades. It also takes time for the research group to report and publish their result. If the result is insignificant, it is often not reported.[2] As a result, it is difficult for a research group to closely observe other groups' findings or their progress. A research group's research progress and their decisions on reporting a result are frequently private.

Information in this paper is modelled as stochastic conclusive evidence. If the state is high and the player acquires information, a breakthrough occurs at a positive Poisson rate. If the state is low, a breakdown occurs at a different Poisson rate. The evidence is conclusive and revealing. I also assume no news is good news so players become more optimistic if no arrival occurs. In the market entry example, information acquisition can take the form of market research. A breakdown can be a discovery of an incumbent who has gained a large amount of market share. A breakthrough can be a market survey result that confirms the high demand in the market. In the R&D example, a breakdown can take the form

[1] See Merton (1957). Hill & Stein (2020) finds that the scientific priority race is not a 'winner-takes-all' type of competition. Although the losers of a priority race get published in the top journals with a smaller probability and receive fewer citations, the negative effect of losing is in general moderate. There is a first-mover advantage but it is still better to be the second than to remain silent.

[2] See Akcigit & Liu (2015) and Bobtcheff et al. (2021).

of a fatal flaw that disproves the feasibility of the technology, and a breakthrough can take the form of an experimental success that proves its feasibility. In the research priority race example, a breakdown can be a negative test result that disproves the hypothesis the researcher wants to test and a breakthrough can be a positive result that verifies the hypothesis.

Results The strategic component in this model comes from the payoff externality generated by the first-mover advantage. The player's payoff from taking the irreversible action depends on both her belief about the state and the belief about her opponent's action. The equilibria have different qualitative features depending on the prior beliefs. There exist multiple equilibria when the players have sufficiently uncertain prior beliefs. In this case, information acquisition is a strategic complement. In contrast, when players have pessimistic prior beliefs, information acquisition can also be a strategic substitute. I focus on discussing symmetric equilibria.

When players have sufficiently uncertain priors, both players acquire information in equilibrium. The equilibrium strategies consist of two stages: an information acquisition stage and a randomisation stage. At the information acquisition stage, the player only stops and takes an action when a breakthrough or breakdown occurs. If the player observes a breakthrough, she takes the risky action, and if she observes a breakdown, she takes the safe action. When there is no breakthrough or breakdown, the player continues to acquire information. After a fixed period of time with no breakthrough or breakdown, the player transitions to the randomisation stage where she not only stops after observing a breakthrough or breakdown but also stops and takes the risky action at each time at a positive rate.[3] This strategy is referred to as a random stopping strategy. The time at which the player transitions from the information acquisition stage to the randomisation stage is not unique. The strategic complementarity of information acquisition can explain the multiplicity of the equilibria: When the player is sufficiently uncertain about the state, the value associated with acquiring information is high. If the player was not facing an opponent, the gain from acquiring information is strictly higher than the cost. If the player is facing an opponent, there is an endogenous cost of being preempted. When the opponent transitions to the randomisation stage, the cost associated with acquiring information becomes higher. The overall information cost jumps up.

[3] Consider the case that the player acquires information for a short time interval. At the end of this time interval, she takes the risky action if she observes a breakthrough and takes the safe action if she observes a breakdown. Otherwise, she stops and takes the risky action with some positive probability.

In equilibrium, this jump motivates the player to match this transition. This strategic complementarity of information acquisition gives rise to the multiplicity of equilibria for sufficiently uncertain priors.

In contrast, when the prior belief is more pessimistic, information acquisition can also act as a strategic substitute. For these priors, in equilibrium, the players mix between two strategies: they use the random stopping strategy or take the safe action immediately with positive probabilities. Compared to the previous case, the players are more reluctant to acquire information at time zero. In this case, the probability of observing a breakthrough is lower. In addition, when there is no breakthrough or breakdown, the player acquires information for a longer duration before stopping and taking the risky action. For these reasons, the expected information cost is higher. Therefore, the value associated with taking the risky action must be high enough to make it beneficial to acquire information. When the opponent, with positive probability, takes the safe action immediately, the player expects a higher payoff from the risky action because the first-mover premium is secured. In this way, the players are incentivised to acquire info for pessimistic priors. Information acquisition here is a strategic substitute. The player is more willing to acquire information if their opponent does not.

The existence of the random stopping equilibrium is significantly different from the single decision maker case. In the single decision maker case, there cannot exist a random stopping strategy. The only optimal strategy is a deterministic cutoff strategy: to acquire information until the belief drifts to a cutoff and then take the safe or risky action depending on the belief. In the case analysed here, due to the first-mover advantage, any deterministic stopping strategy creates a preemption motive and hence is never optimal.

This paper connects the literature on preemption games and the dynamic information acquisition model. For a large range of prior beliefs, including both extreme prior beliefs and intermediate prior beliefs, there exists a no-learning equilibrium where players take an immediate action without acquiring information. For sufficiently pessimistic prior beliefs, the players take an immediate safe action and for sufficiently optimistic priors, the players take an immediate risky action. Furthermore, for intermediate prior beliefs, the players randomise between the safe action and risky action immediately without acquiring information. The existence of this no-learning equilibrium for intermediate prior beliefs is a result of the strong preemption motive. Players undercut the time at which the other takes the risky action until there is no room for further preemption. However, the players are not optimistic enough to take the risky action with probability one. With a positive probability of taking the safe action, the player secures a zero payoff. At the same time, since taking the risky action immediately increases the opponent's cost from being preempted, information acquisition is deterred.

I discuss how vanishing information cost and the intensity of competition affect the players' equilibrium strategies. In the single decision-maker case, when the information cost decreases, she acquires information for a larger set of beliefs. When information cost vanishes, she always acquires information.[4] This is not the case when the player faces an opponent. A smaller information cost leads to more information acquisition, but does not completely suppress the no-learning equilibrium. The interval of prior beliefs where the players acquire information expands, but this interval is bounded away from one. Hence, for sufficiently optimistic prior beliefs, even for zero information cost, the no-learning equilibrium still exists. This is because when the players acquire information, in addition to the exogenous cost, they also face the endogenous information cost from being preempted. For sufficiently optimistic prior beliefs, the player believes the state is more likely to be high and hence it is also very likely that the opponent may take the risky action early. This endogenous information cost from potentially being preempted does not vanish with the exogenous cost being zero and prevents information acquisition in equilibrium.

I model the intensity of competition as the size of the first-mover premium. When the first-mover premium increases, competition becomes more intense, and the range of prior beliefs where the players acquire information shrinks. This implies that competition for the first-mover premium suppresses information acquisition. In addition, in the equilibrium where the players acquire information, at the randomisation stage, the players stop acquiring information and take the risky action at a higher rate.

There are two technical difficulties when solving for the equilibrium. First, a player's strategy is not Markov in their belief about the state. A player's belief about the opponent's action is also important because of the payoff externalities. Her decision at each time depends on both her current belief about the state and her belief about whether the opponent has taken the risky action. Second, the player's expected payoff from taking the risky action at each time may not be continuous and differentiable everywhere. It depends on their opponent's strategy. I find the equilibrium given each prior by fixing the opponent's strategy and then solving for the player's best responses. Given the opponent's strategy, the player's problem can be written as a Hamilton–Jacobi–Bellman (HJB) equa-

[4] This is because of the no time discounting assumption.

tion. I construct a candidate function that satisfies the HJB equation for all points of differentiability. Because of the potential discontinuity in the player's payoff from taking the risky action, the candidate function that solves the HJB equation may not be continuous and differentiable everywhere. To solve these problems, I show that the player's strategies induce continuous and differentiable expected utility from taking the risky action in any equilibrium where the players acquire information. Therefore, the potential discontinuity is not a problem on the equilibrium path. Since the candidate value function may have a kink,[5] I show that the candidate function is a viscosity solution that satisfies the HJB equation.

Related literature This paper is related to the literature studying strategic experimentation. The model setup is related to the bandit problem (see Bolton & Harris (1999), Keller et al. (2005), Keller & Rady (2010), Hörner & Skrzypacz (2017) and Ozdenoren et al. (2021)), the classical framework studying R&D races, but is different in the following aspects. In the bandit problem with the safe and the risky arms, pulling the safe arm is similar to taking the *safe action* in this model, while pulling the risky arm is similar to *acquiring information.* The payoff generated by the risky arm depends on the nature of the risky arm, which is similar to the *state* in this model. The difference is that in the bandit problem, the player pulling the risky arm gets *information* about the arm from the *payoff* generated by the arm. A high payoff from the risky arm indicates a good arm. In other words, the payoff *is* the information. In contrast, in my model, I separate the information and the payoff from an action. The breakthroughs, breakdowns, or the lack thereof, only contain *information* about the state. If the players want to use the information, they need to take an action. Acquiring information itself does not give the player any payoff. Instead, it incurs a positive information cost. Acquiring information about the technology is modelled as a costly activity that contains information about the feasibility of the technology, but does not give the firm a direct return. The firm gets a return only if they take action using the information generated.

Besides the differences in the model setup, there are two other differences. First, strategic experimentation literature investigates the free-rider problem when information is a public good. This is mainly due to the observability of the actions as well as sometimes the signals themselves. In my model, information is private and information externality does not exist. There is no freeriding.

[5] The player chooses between the risky and safe action when she stops acquiring information. The optimal action switching from the safe to the risky action induces a kink.

Second, there is no exploration-exploitation tradeoff in my model because the players only take the action once and it is irreversible. The players must stop acquiring information and take the risky action in order to exploit the outcome generated by the risky arm.

This paper is related to the literature on preemption games. R&D races and market entry are often studied and modelled as preemption games with complete information. This originated from Reinganum (1981) and Fudenberg & Tirole (1985). There is a growing literature studying preemption games with learning aspects including Hopenhayn & Squintani (2011), Akcigit & Liu (2015), Bobtcheff et al. (2021) and Shahanaghi (2022). In preemption games, there is a first-mover advantage in the final payoff and players make strategic decisions on when to end the game. In the present paper, I explore the interplay between preemption and learning in a model with private information and private actions. By preempting the opponent and taking the risky action early, the player can get the first-mover premium, but she gives up the chance to learn more about the state. The private action assumption distinguishes this paper from the aforementioned papers by shutting down the information spillover generated by observing the opponent's action. The player is only aware of their opponent's existence but does not have any additional information regarding their opponent's action and information.

This paper is related to the literature investigating dynamic information acquisition. A single DM's dynamic information acquisition problem has been studied intensively since Wald (1945, 1947). The recent literature studies both drift-diffusion models (see Fudenberg et al. (2018) and Ke & Villas-Boas (2019)) and Poisson models (see Nikandrova & Pancs (2018), Che & Mierendorff (2019), and Mayskaya (2020)). My model adopts the Poisson signal structure and more importantly, considers the *strategic* interaction between players. The main difference is that instead of having the cost of information completely exogenous, it is now endogenous which depends on the opponent's strategy.

Another relevant strand of literature studies static information acquisition before a game (see Hellwig & Veldkamp (2009), Yang (2015), Han & Sangiorgi (2018) and Denti (2019)). Those papers discuss information acquisition before a coordination game. When the players play a game with strategic complements, information choices exhibit strategic complementarity as well. In my model, information has the features of a strategic complement but could also be a strategic substitute.

A distinct but related group of literature is about the equivalence between static and dynamic information acquisition. Hébert & Woodford (2019) studies a dynamic rational inattention model and shows that the belief dynamics generated can resemble either diffusion processes or processes with large jumps. Morris & Strack (2019) studies what kind of static models with costly information acquisition has a sequential sampling foundation. In my model, because of the preemption feature, dynamic information acquisition is intrinsic as a player's payoff depends on the order that they act.

The remainder of the paper is organised as follows. In Section 2, I introduce the model. Section 3 uses a simple two-period model to illustrate the main tradeoff. In Section 4, I analyse the equilibrium and present the main result. In Section 5, I discuss how information cost and the intensity of competition affect the equilibrium strategy and in Section 6, I discuss two extensions: the case with more than two players and the case with observable actions.

2 The Model

2.1 Model Setup

There are two players $i \in \{1, 2\}$. At the beginning of the game, an unknown, fixed, payoff-relevant state $\omega \in \{H, L\}$ is drawn. At any time $t \in [0, \infty)$, each player can take an irreversible action $x \in \{S(afe), R(isky)\}$, or delay and acquire information about the state. The irreversible action gives the player an one-off payoff at the moment she takes the action. Action S yields a payoff which is normalised to be 0. Action R's payoff depends on the state and whether the player is the first or second to take R. In state ω, the first player to take R gets $u_\omega \in \mathbb{R}$ and the second gets $u_\omega - \bar{\triangle}_\omega$. If the players take R simultaneously, the payoff is $u_\omega - \underline{\triangle}_\omega$. At each time, if the player delays her action and acquires information, she incurs a positive information flow cost c per unit of time. I assume no time discounting. Payoffs satisfy the following two assumptions.

Assumption 1.1. $\bar{\triangle}_\omega > \underline{\triangle}_\omega > 0$ *for* $\omega \in \{H, L\}$.

Assumption 1.1 says that first R taker gets a higher payoff than the second R taker in both states. If the players take R simultaneously, the payoff is in between.

Assumption 1.2. $u_L < 0 < u_H - \bar{\triangle}_H$.

Assumption 1.2 says that R yields a higher payoff than S in state H and a lower payoff in state L. Furthermore, being the second to take R in state H is still better than taking S. Players' incentive to be the first is increasing in the value of $\bar{\triangle}_H$. The difference between the first and second R taker payoff describes the intensity of the competition. Assumption 1.2 describes a gentle competition in the sense that being the second to take R in state H is not too bad. This assumption is dropped in Section 5.2, where the competition is more intense as the second R taker gets a lower payoff than taking S.

Before taking the irreversible action, players have access to costly information about the state. Information is modelled using Poisson signals. If a player acquires information for a short time period $dt > 0$, then, in state H (L, resp), she receives an H-state (L-state, resp) revealing signal with rate adt (bdt, resp). Player i's belief p_t^i is the probability that the state is H. I assume the players have a common prior p_0 and that they observe neither the opponent's action nor their signals. At each time t, players update their beliefs using Bayes' rule. When a player acquires information, her belief jumps to 0 after receiving an L-state revealing signal and jumps to 1 after receiving an H-state revealing signal. In the absence of the revealing signal, player i's belief evolves according to

$$\frac{dp_t^i}{dt} = (b-a)p_t^i(1-p_t^i). \tag{1.1}$$

In the following part of the paper, I assume $b > a > 0$. In the absence of a revealing signal, the player's belief drifts up. This is the 'no news is good news' environment.

2.2 Strategies and Equilibrium

If a player receives a revealing signal, it is optimal to stop acquiring the signal and take R or S regardless of the opponent's behaviour. Therefore, it is sufficient to describe players' strategy conditional on no arrival of a revealing signal. A pure strategy (defined below) specifies the time, T^i, at which the player stops and which action, x^i, they take in the absence of the revealing signals.

Definition 1.1. *Player i's pure strategy s^i is defined as $\left(T^i, x^i\right) \in \mathbb{R}_+ \times \{R, S\}$.* [6]

A mixed strategy (defined below) specifies the probability that the player stops before time t *conditional on no revealing signal.*

Definition 1.2. *Player i's mixed strategy γ^i is defined as two non-decreasing measurable functions $\left(\rho^i, \sigma^i\right)$ where $\rho^i : \mathbb{R}_+ \to [0,1]$ and $\sigma^i : \mathbb{R}_+ \to [0,1]$ satisfying $\rho^i(t) + \sigma^i(t) \le 1$ for $\forall t \in \mathbb{R}_+$. ρ^i is the probability that player i stops and takes R before or at time t conditional on no revealing signal. σ^i is the*

[6] The notation $\mathbb{R}_+$ denotes the set of non-negative real numbers.

probability that player i stops and takes S before or at time t conditional on no revealing signal.

Before defining the equilibrium, I first write down a player's expected payoff. Player i's expected payoff from taking action x at time t in state ω, $\mathbb{E}^{\gamma^j}\left[u_\omega^x \mid \omega, t\right]$, depends on player j's strategy γ^j. The randomness of the payoff from taking R comes from both player j's strategy and the randomness of the signal. For example, the opponent using a pure strategy $(0, R)$ induces degenerate conditional distributions at each time t such that player i's action-R payoff in state ω is $u_\omega - \bar{\triangle}_\omega$ with probability one. Let

$$U_x^{i,\gamma^j}(t) := \mathbb{E}_{\omega|t}\mathbb{E}^{\gamma^j}\left[u_\omega^x \mid \omega, t\right]$$

be player i's expected payoff from taking action x at time t, where the expectation $\mathbb{E}_{\omega|t}$ is taken over the distribution of the state given player i's time t belief. Then, player i's payoff from taking action x^i at time T^i is

$$\int_0^{T^i} \pi_t \left(p_t^i a \mathbb{E}^{\gamma^j}\left[u_H^R \mid H, t\right] - c\right) dt + \pi_{T^i} U_{x^i}^{i,\gamma^j}(t) \tag{1.2}$$

where $\pi_t = p_0 e^{-at} + (1 - p_0) e^{-bt}$ is the probability of no revealing signal up to time t. Let $U^{i,\gamma^j}(t) := \max_{x^i} U_{x^i}^{i,\gamma^j}(t)$ be player i's payoff associated with taking the optimal irreversible action x^i at time t. The following defines the perfect Bayesian equilibrium in pure strategies. The perfect Bayesian equilibrium in mixed strategies can be defined in a similar manner.

Definition 1.3. *A perfect Bayesian equilibrium in pure strategies is a strategy profile (s^i, s^{-i}) and beliefs $\left(\left(p_t^i\right)_{t\in[0,T^i]}, \left(p_t^{-i}\right)_{t\in[0,T^{-i}]}\right)$ such that for $i \in \{1, 2\}$,*

1. $x^i \in \arg\max_{\tilde{x}^i} U_{\tilde{x}^i}^{i,s^{-i}}\left(T^i\right)$;
2. $T^i \in \arg\max_{\tilde{T}^i} \int_0^{\tilde{T}^i} \pi_t \left(p_t^i a \mathbb{E}^{s^{-i}}\left[u_H^R \mid H, t\right] - c\right) dt + \pi_{\tilde{T}^i} U^{i,s^{-i}}\left(\tilde{T}^i\right)$;
3. *The belief p_t^i evolves according to (1.1).*

I seek symmetric equilibria in both pure and mixed strategies. In the following discussion, the superscript i representing the player is ignored.

3 An Illustrative Example

In this section, I use a simple two-period model to illustrate the tradeoff between information acquisition and preemption. I show how players' *learning motives* and *preemption motives* depend on the opponent's strategy and the prior. In equilibrium, mixed strategies create endogenous randomnesses that either deter learning or prevent the opponent from preemption.

At time $t = 0$, players can choose to acquire a signal at cost $c > 0$ or to take one of the actions. At time $t = 1$, players have to take one of the actions. If a player acquires a signal, in state H (L, resp), a revealing signal arrives with probability a (b, resp), and the belief p_1 jumps to 1 (0, resp). If she does not receive the revealing signal, then, her belief is updated to $\frac{p_1}{1-p_1} = \frac{1-a}{1-b}\frac{p_0}{1-p_0}$. Acquiring a signal at time 0 allows the player to learn the state and hence take the 'correct' action (R in state H and S in state L). This gives the player the 'learning motive'. Not acquiring a signal, however, secures the player the first prize. This gives the player the 'preemption motive'.

The *learning motive* is stronger when the prior is in the intermediate range and when the opponent takes S at time 0. When the prior is in the intermediate range, the player is uncertain and hence has stronger incentives to learn. When the opponent takes S at time 0, the player is the single decision maker in this game. The payoff associated with taking the correct action is the highest and hence the value associated with information is higher at each prior. The *preemption motive* is stronger when the opponent acquires a signal at time 0. This is because when the opponent acquires a signal, by taking R at time 0, the player can secure herself the first R taker payoff, while acquiring a signal at time 0 decreases the probability of being the first R taker and hence decreases her expected payoff from taking R. The information becomes less valuable. When the opponent takes R at time 0, the player's incentive to preempt is less strong. It is then optimal for the player to acquire a signal for a larger range of priors. This is because she cannot preempt her opponent only to match their action. This matching reduces the payoff from acting at time 0 and correspondingly increases the payoff of waiting.

Figure 1.1 plots her payoffs associated with taking immediate actions and acquiring a signal: Panel (a) is the player's payoffs when the opponent takes S at time 0, (b) is when the opponent takes R at time 0 and (c) is when the opponent acquires a signal at time 0. For the opponent's different strategies, qualitatively, the player's best responses have similar properties: to take an action at time 0 when the prior belief is extreme and to acquire a signal when the prior belief is in the intermediate range. The difference is the range of the priors at which her best response is to acquire a signal. This range of priors is largest when the opponent takes S at time 0 and is smallest when the opponent acquires a signal.

Next, I briefly discuss the possible equilibria: the equilibrium where players

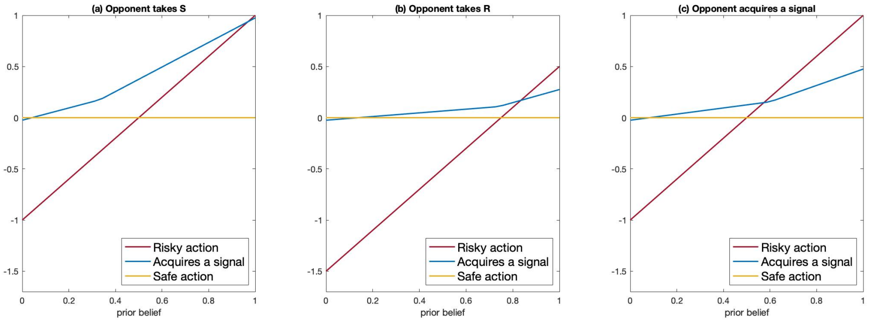


Figure 1.1: The player's payoffs given opponent's different strategies
Notes: This graph is drawn given the following parameter values: $u_H = 1$,$u_L = -1$, $\bar{\triangle}_\omega = 0.7$, $\underline{\triangle}_\omega = 0.5$, $a = 0.6$, $b = 0.8$, $c = 0.025$. In all three panels, the red line represents the player's payoff from taking R at time 0, the blue line represents the player's payoff from acquiring a signal at time 0, and the yellow line represents the player's payoff from taking S at time 0.

preempt R, the equilibrium where players acquire a signal with positive probability, and the equilibrium where players randomise between two actions without acquiring a signal.

When the prior is sufficiently high, the preemption motive dominates and hence there is unravelling. The symmetric equilibrium is such that both players take immediate R at time 0 without acquiring any signal. The weak learning motive and the strong preemption motive work in the same direction which pushes the players to take R immediately. For intermediate priors, there exists an asymmetric equilibrium where one player acquires a signal and the opponent takes R at time 0. In symmetric equilibrium, the players randomise over these two roles. This is referred to as a *random stopping strategy*: the player randomises between stopping (to take R) at time 0 and time 1. By using this strategy, the player creates an endogenous uncertainty that prevents the opponent from preempting. When a player acquires a signal with probability one, her opponent has incentives to preempt and it is easy for them to do so. But when a player randomises, not only the value associated with preemption is reduced, it is also harder for her opponent to preempt because of the endogenous randomness. For some intermediate priors, there exists another symmetric equilibrium where players randomise between two immediate actions without acquiring any signal. This kind of randomisation deters learning. For those intermediate priors, players are uncertain about the state. By simply randomising between R and S at time 0, the player 'hedges' against the uncertainty without paying extra information cost and at the same time, reduces her opponent's value associated with learning. [7] When

[7] The player taking R at time 0 with a positive probability reduces her opponent's expected

this value is sufficiently low, the opponent's learning is deterred.

This two-period example shows the most important tradeoff in the model: the incentive to learn and the fear of being preempted. However, it can only discuss *whether or not* the players acquire information, but not *how much* information they get. To understand the optimal quantity of information the players acquire before taking an action, the dynamic model with multiple periods is of interest. Next, I analyse the dynamic model introduced in Section 2.

4 Equilibrium Analysis

4.1 Single Decision-maker Benchmark

Before getting into the detail of the game, as a benchmark, I first consider the model with one single decision maker (DM). This single DM model is a well-studied sequential sampling model due to Wald (1945, 1947) where the DM chooses an optimal stopping time based on the samples she has observed.

The optimal stopping rule depends on the cost of information and the belief. If the cost of information is too high, then, it is optimal to take an immediate action based on her prior. In this case, learning about the unknown state does not give the DM sufficiently high benefit to compensate for the high cost. When the cost is sufficiently small, it is optimal to acquire information for intermediate beliefs and to take an immediate action for extreme beliefs. The DM's optimal policy at each time t only depends on the current belief p_t but not time t itself. This is because all the past information is summarised by the belief at time t and the information cost in the past is sunk. The DM acquires information if the marginal cost c is smaller than the marginal benefit. The marginal benefit is higher when the DM is uncertain about the state. Hence, it is optimal to take S if the belief is sufficiently small, to acquire information if the belief is in an intermediate range, and to take R if the belief is sufficiently big. This is summarised in the following proposition.

Proposition 1.1. *There exist cutoffs $\underline{p} < \bar{p}$ and $\bar{c}$ such that when $c \leq \bar{c}$, given belief p, the DM's optimal policy is to take S if $p \leq \underline{p}$; to acquire information if $\underline{p} < p < \bar{p}$; and to take R if $p \geq \bar{p}$.*

In the single DM case the optimal strategy is deterministic: the cutoff where the DM stops acquiring information is a constant and is uniquely pinned down by the parameter values. When she stops at the cutoff, the optimal action is R

payoff from taking R at time 1.

because she is convinced that the state is more likely to be H. When the DM uses the optimal strategy and acquires information, then she only makes mistakes in the low state. She correctly takes R in high state with probability one and incorrectly takes R in the low state with a positive probability. This is because when acquiring information, she either receives a revealing signal and then stops or she acquires information until the belief drifts up to the upperbound $\bar{p}$. In the high state the DM takes the correct action R in both events. In the low state, she takes the incorrect risky action if she stops at the upperbound $\bar{p}$.

4.2 Properties of Equilibrium

Now I turn to the model with two players. I establish the properties of the equilibrium strategy. I explain why there are *no jumps* in the player's equilibrium strategy at all $t > 0$. If there *is* a jump in the player's strategy, it only happens at $t = 0$. This property is driven by preemption and the unobservability of actions.

Lemma 1.1. *In equilibrium, $\rho(t)$, the probability that the player stops and takes R before or at time t conditional on no revealing signal, is continuous at all $t > 0$.*

In the absence of the revealing signal, the player does not take R with positive probability mass at any time $t > 0$. This implies that the single DM's *deterministic* optimal strategy is not part of the equilibrium in the two-player game. The reason is twofold. First, the jump at any $t > 0$ in the opponent's strategy gives the player an opportunity to preempt and hence a profitable deviation. If the opponent takes R at some $\tau > 0$ with a positive probability, then, the player is betteroff stopping at $\tau - dt$ where $dt > 0$ is infinitesimal. At time $\tau - dt$, if the player continues acquiring information for dt longer, she gains nothing but loses the probability of being the first R taker. Second, the private action and the private breakdown give rise to *winner's curse.* It is more likely for the player to be the first R taker in state L than in state H because the opponent might have received the breakdown and dropped out. In other words, it is more likely to 'win' in state L. If the opponent takes R at some $\tau > 0$ with a positive probability, then, conditional on no revealing signal and 'winning', the value associated with taking R at time τ has a downward jump. It is then not a best response for the player to take R at time τ.

In equilibrium, ρ can have a jump at $t = 0$. This means that in the absence of the revealing signal, the player takes R with positive probability mass only at time zero. This is a result of preemption. The unravelling comes from the player undercutting the time of the opponent taking R until time zero at which there is no more room for preemption.

4.3 Equilibrium Strategies

This section introduces five strategies that appear in equilibrium.

The Immediate Action Strategy The *Immediate* R (*Immediate* S, resp) *Strategy* is the pure strategy $(0, R)$ ($(0, S)$, resp) where the player does not acquire information and takes R (S, resp) at time 0.

The Immediate Coin Flip Strategy This is a strategy where the player does not acquire information. She randomly takes R and S at time zero. The following is the formal definition. The *Immediate Coin Flip Strategy* is a mixed strategy $\left(\rho^{IM}, \sigma^{IM}\right)$ where $\rho^{IM}(t) = \rho^{IM}(0) \in (0,1)$ for $\forall t > 0$, $\sigma^{IM}(t) = \sigma^{IM}(0) \in (0,1)$ for $\forall t > 0$, and $\rho^{IM}(0) + \sigma^{IM}(0) = 1$.

The Random Stopping Strategy This is a strategy where the player acquires information for a period of time and then stops with a positive rate at each time. The action associated with stopping is R. The following is the definition. The *Random Stopping Strategy* is a mixed strategy $\left(\rho^{RS}, \sigma^{RS}\right)$ such that $\rho^{RS}(\cdot)$, the probability of taking R in the absence of the revealing signal, weakly increases and σ^{RS}, the probability of taking S in the absence of the revealing signal, equals zero for all $t \geq 0$. The first element $\rho^{RS}(\cdot)$ satisfies $\rho^{RS}(0) = 0$ and $\rho^{RS}(t) = 1$ for $t \geq \bar{T}^{RS}$ where $\bar{T}^{RS} > 0$. When using this strategy, conditional on arriving at time t [8] , the player stops and takes action R with rate $\frac{\frac{d\rho^j(t)}{dt}}{1-\rho^j(t)}$. She only stops and takes action S after observing a breakdown.

The Random Existing Strategy This is a strategy where the player mixes between the *Immediate S Strategy* and the *Random Stopping Strategy* at time 0. The following is the definition. The *Random Existing Strategy* is a mixed strategy $\left(\rho^{ML}, \sigma^{ML}\right)$ such that $\rho^{ML}(\cdot)$ weakly increases and $\sigma^{ML}(t) = \beta \in (0,1)$ for $\forall t$. The probability that the player stops and takes R before time t (i.e. $\rho^{ML}(\cdot)$) satisfies $\rho^{ML}(0) = 0$ and $\rho^{ML}(t) = 1 - \beta$ for $t \geq \bar{T}^{ML}$ where $\bar{T}^{ML} > 0$.

[8] That is, the player receives no revealing signal and the player's strategy does not prescribe stopping.

4.4 Main Results

The theorem shows for different priors what types of equilibria exist. Let $\frac{p^M}{1-p^M} := \frac{-(u_L-\underline{\triangle}_L)}{u_H-\underline{\triangle}_H}$, $\frac{p^L}{1-p^L} := \frac{-u_L}{u_H}$ and $\frac{\tilde{p}}{1-\tilde{p}} := \frac{-bu_L-c}{a\triangle_H+c}$ be three prior cutoffs. Both p^M and p^L are positive as u_L is negative. A sufficiently small information cost c guarantees that $p^M < \tilde{p}$

Theorem 1.1. *When $b < 2a$ and c is sufficiently small, there exist cutoffs $\underline{p} < p^* < p^L < p^M < \tilde{p}$ such that:*

If $p_0 \leq \underline{p}$, there exists an equilibrium where both players use the Immediate S Strategy;

If $\underline{p} < p_0 < p^*$, there exists an equilibrium where both players use the Random Existing Strategy;

If $p^* \leq p_0 < \tilde{p}$, there exists an equilibrium where both players use the Random Stopping Strategy;

If $p^L < p_0 < p^M$, there exists an equilibrium where both players use the Immediate Coin Flip Strategy;

If $p_0 \geq p^M$, there exists an equilibrium where both players use the Immediate R Strategy.

The lowest cutoff $\underline{p}$ is the lowerbound in Proposition 1.1, below which the single DM's optimal strategy is to take S immediately. The cutoff p^* is a fixed point and the detail can be found in Section 4. The cutoff p^L (p^M, resp) is the belief at which the player is indifferent between taking S and taking immediate R if she is the first R taker (if she and the opponent take R simultaneously, resp).

When the prior is extreme, both players use the *Immediate R/S Strategy* in equilibrium. It seems to be intuitive as the players' learning motive is weak due to the relatively small value of information. But this is not the whole story. The subtlety is the role of the preemption motive. When the prior is sufficiently small, the preemption motive is absent because S is the optimal action that gives the player the same payoff. The weak learning motive is indeed the only reason why the players take immediate S. However, when the prior is high, the preemption motive is the main reason why the players take immediate R with a positive probability. To understand the existence of the equilibria where the players use the immediate R or the immediate coin flip strategy, consider a static game where the players do not have access to information [9]. Consider the

[9] This is a game where both players have two actions: action R and action S. In state ω, if both players take action R (action S, resp), the payoff is $u_\omega - \underline{\triangle}_\omega$ (0, resp) for each of them. If one player takes action R and the other player takes action S, the player who takes action R gets a payoff of u_ω and the player who takes action S gets a payoff of 0.

sufficiently high prior such that there exists an equilibrium where the players take R with a positive probability in this static game [10] . Then, even if the players are now given access to costly information, it is still optimal to take an immediate action. This is because when the prior is sufficiently high, the preemption motive is strong and the learning motive is weak. In the event that the opponent takes immediate R, if the player also takes immediate R, she then gets the simultaneous-move payoff. If the player acquires information for an infinitesimal time period dt, then, the loss from becoming a second R taker is strictly positive but the gain from acquiring information is negligible. The positive probability of the opponent taking immediate R induces a downward jump in the player's expected payoff from taking R. This deters learning.

Lemma 1.2. *When c is sufficiently small, $p^M < \bar{p}$.*

The cutoff $\bar{p}$ is the upperbound in Proposition 1.1: the belief above which the player takes R immediately. Lemma 1.2 says that the range of priors where the players take immediate R is larger when the players have the incentive to preempt. The preemption motive leads to less information acquisition and more immediate risky action for relatively optimistic prior beliefs.

The intuition of the equilibrium where the players use the *Random Stopping Strategy* is explained by the tradeoff between learning and the preemption motive. When the prior is in the lower intermediate range, the player has the learning motive because the value of information is high. If a player were playing this game alone, the optimal strategy is to acquire information until the belief drifts up to an upperbound and then takes R (see Proposition 1.1). Before she stops acquiring information, the distribution of the time at which she takes R is the distribution of the Poisson breakthrough, which is continuous. At the time the belief drifts up to the upperbound, according to the strategy, the player stops and takes R. Therefore, the induced distribution of the time at which she takes R has an atom at the time when the belief drifts up to the upperbound. If player i were playing against player j who uses this strategy, player i would preempt player j. To avoid being preempted by the opponent, what the player could do is to randomise at which time she stops and takes action R. This randomisation will make the opponent's marginal cost and marginal benefit (explain later) from acquiring information the same and hence eliminate the incentive to preempt. The indifference condition gives rise to an ODE that the equilibrium strategy satisfies which is concluded in Lemma 1.6 in the appendix. Here in the main text, I give expressions of the marginal cost and marginal benefit from acquiring

[10] That is, when the belief is higher than p^L.

information and show how they are determined by the opponent's strategy. Suppose player j uses the mixed strategy (ρ, σ) defined in Definition 1.2. Suppose ρ and σ are differentiable. Then,

$$\begin{aligned} F_H(t) &= \int_0^t \left[e^{-as}(1-\rho(s)) \left(a + \frac{\frac{d\rho(s)}{ds}}{1-\rho(s)} \right) \right] ds \\ &= 1 - e^{-at}(1-\rho(t)) \end{aligned}$$

is the probability that player j takes R before time t in state H and

$$\begin{aligned} F_L(t) &= \int_0^t \left[e^{-bs}(1-\rho(s)) \frac{\frac{d\rho(s)}{ds}}{1-\rho(s)} \right] ds \\ &= \int_0^t \left[e^{-bs} \frac{d\rho(s)}{ds} \right] ds. \end{aligned}$$

is the probability that player j takes R before time t in state L. At time t, player i is indifferent between taking R now and dt later if

$$c + p_t \frac{dF_H(t)}{dt} \bar{\triangle}_H + (1-p_t) \frac{dF_L(t)}{dt} \bar{\triangle}_L = (1-p_t)\, b \left[-u_L + F_L(t) \bar{\triangle}_L \right]. \quad (1.3)$$

The left-hand side of (1.3) is the marginal cost of acquiring information for dt longer at time t and the right-hand side is the marginal benefit. The cost consists of the *direct* information cost c and the *indirect* cost from being preempted. Intuitively, in state ω, the opponent preempts the player in this short time period dt with probability $\frac{dF_\omega(t)}{dt}$. The benefits of acquiring information comes from the breakdown because it corrects the player's action from R to S in state L. The player gets the payoff of zero instead of the negative payoff associated with R in state L. This indifference condition gives an ODE that the equilibrium strategy satisfies. Lemma 1.6 and Lemma 1.14 in Section 4 characterises the conditions that the equilibrium strategy satisfies.

Corollary 1.1. *When c is sufficiently small, there exist multiple equilibria if $\underline{p} < p_0 < p^M$.*

When $\underline{p} < p_0 < \tilde{p}$, in the equilibrium where the players use the Random Stopping Strategy, the players both acquire information up to some time $\hat{T}$ and then randomly stop at each time greater than $\hat{T}$. The time $\hat{T}$ at which the player starts the randomisation is not unique and this leads to multiple equilibria. To get the intuition, suppose player j uses the Random Stopping Strategy such that she starts the random stopping at some time $\hat{T} > 0$. To have both players using this strategy as an equilibrium, before time $\hat{T}$, player i must strictly prefer to acquire information given that the opponent acquires information. That is, given the opponent is still acquiring information, the marginal cost from acquiring information for dt longer must be smaller than the marginal benefit. As a result, the latest time instant at which the player starts randomising is the first time instant at which the marginal cost of acquiring information equals the marginal benefit given that the opponent acquires information. Before this time point, if the opponent acquires the information, the marginal benefit of acquiring information exceeds the cost. The player acquires information. After this time point, even though the player knows the opponent is still acquiring information, the value associated with information is too small to keep the player engaged. After time $\hat{T}$, player i is indifferent between taking R at each time $t > \hat{T}$, and she must prefer to randomise instead of taking S. Therefore, the earliest time instant at which the player is willing to start the random stopping is the earliest time instant such that taking R gives her a payoff of at least zero. Before this time instant, the player is not optimistic enough about the state to take R. After this time instant, the player matches the opponent's action. If the opponent's strategy is to randomise, the player also randomises.

The intuition behind the multiplicity is the strategic complementarity of information. Information is a strategic complement at time $t < \hat{T}$ because when the opponent is still acquiring information, the value of information is relatively high. The player hence is also willing to acquire more information. If the opponent starts randomising, then, the value associated with acquiring information becomes lower. The player then matches the time at which the opponent starts randomising. This strategic complementarity also explains the existence of multiple equilibria when $p_0 \in \left[p^L, p^M\right)$. For this range of priors, according to Theorem 1.1, there exist both kinds of equilibria where the players use the Random Stopping Strategy and where the players use the Immediate Coin Flip Strategy. In other words, there simultaneously exist equilibria where the players acquire information and where the players do not acquire information.

The existence of the equilibrium where both players use the *Random Stopping Strategy* requires the player to start acquiring information at time 0. If the prior is relatively low, then, the expected duration of acquiring information before stopping is long and the cost of delay is high. In addition, the players are pessimistic about the state being high. Therefore, the players do not have strong incentives to start acquiring information and hence this symmetric equilibrium where both players acquire information do not exist. However, there exists asymmetric equilibrium where one player acquires information and the other takes the immediate safe action. This is because when the player faces a longer potential delay of action, the expected benefit associated with taking R must be higher to compensate for the expected information cost. The benefit associated with taking R depends on the opponent's strategy. If the opponent's takes S at time 0, then, the player's gain from information is higher. This higher gain can incentivise the player to start learning at time 0. However, if the opponent acquires information, then, the player gains less from taking R because there is a positive probability that she can only get the lower payoff. This gain may not be sufficient to incentivise the player to start acquiring information at time 0. In this case, information is a strategic substitute. In the symmetric equilibrium, the players mix between two strategies at time 0: the Immediate R Strategy and the Random Stopping Strategy. Compared to the case that the opponent uses Random Stopping Strategy, the opponent takes R with a lower probability. This in turn increases the player's gain from taking R and hence increases the gain from acquiring information. Then, the player is incentivised to start learning at a lower prior.

5 Information Cost and Competition

In this section, I discuss the effects of vanishing information cost, c, and the effects of the intensity of competition.

5.1 Vanishing Information Cost

When the information cost vanishes, the players acquire information for a larger range of priors but not for all the priors. In the equilibrium where the players use the random stopping strategy, they acquire information for a longer duration before randomisation. When they randomise, they stop and take R at a higher rate.

The cost of acquiring information consists of two parts: the exogenous information cost c and the endogenous cost from being preempted. This is the main difference between the two-player case and the single DM case. In the single DM case, when c vanishes, the DM always acquires information until she is certain since the exogenous information cost c is the total cost of acquiring information. In contrast, in the two-player game, the total information cost does not vanish with the exogenous cost c. When the prior is high, the player believes that the state is very likely to be H and infers that the probability of being preempted is high. Thus, the endogenous cost is high for optimistic priors even when c

vanishes. As a result, when c vanishes, the equilibrium such that both players acquire information exists for a larger range of priors, but does not exist for extremely high priors. However, when the prior is low, the player is pessimistic about the state and hence she believes that it is not likely that the opponent takes R. In addition, for low priors, if the player does not acquire information, the optimal action is to take S, the payoff of which does not depend on the opponent's action. Thus, the endogenous information cost is low when the prior is close to zero. As a result, for low priors, the total information cost vanishes with c and the player is willing to acquire information for low priors. This property is summarised in the following proposition.

Proposition 1.2. *The cutoff $\tilde{p}$ decreases in c. When $c \to 0$, $\underline{p} \to 0$ and $\tilde{p} \to \frac{\frac{b}{a}\frac{-u_L}{\triangle_H}}{1+\frac{b}{a}\frac{-u_L}{\triangle_H}} \in (0,1)$.*

For the equilibrium where the players use the random stopping strategy, when c vanishes, the effects are twofold. First, the players acquire information for a longer duration before they start the randomisation stage. Second, when the players just enter the randomisation stage, they stop and take R at a higher rate. The first is intuitive as when c decreases, information becomes cheaper and hence players are willing to acquire information for a longer period. The intuition for the second effect is related to the indifference condition when the players randomise. In equilibrium, the players randomise between acquiring information and taking R because the marginal cost and marginal benefit from acquiring information are the same. The marginal cost from acquiring information consists of the exogenous and the endogenous information cost. When the player just enters the randomisation stage, as shown in (1.3), the marginal benefit is not affected by the exogenous information cost c and the players' stopping rate. The marginal benefit depends only on the cumulative density of the opponent taking R. According the equilibrium strategy, since the opponent has not started the random stopping yet, the marginal benefit only depends on the arrival of the revealing signals but not player's stopping rate. The endogenous information cost however, increases in the opponent's stopping rate. If the players stop and take R at a higher rate, the endogenous information cost is higher. As a result, when the players just enters the randomisation stage, if the exogenous information cost c vanishes, the marginal benefit is unchanged while the marginal cost decreases because of the vanishing information cost c. To have the indifference condition hold, the players stop and take R at a higher rate to increase the endogenous information cost.

5.2 Competition

The competition in this game comes from the payoff difference between the first and second R taker. The intensity of the competition increases in the value of the payoff difference. In all the previous discussion, the payoff difference satisfies Assumption 1.2. That is, the second R taker gets a higher payoff than taking S in state H. With this assumption, after the player learns the state is high, taking R is a dominant strategy. To understand the effect of an intense competition, I drop Assumption 1.2 and impose Assumption 1.3.

Assumption 1.3. $u_L < u_H - \bar{\triangle}_H < 0$.

Assumption 1.3 says that in state H, the payoff of the second R taker is smaller than the safe action payoff. This assumption explicitly imposes a 'loser gets punished' condition [11]. Given Assumption 1.3, whether the player takes R after learning the state is H depends on her belief on the opponent's strategy. Take an extreme case as an example. If the player believes that the opponent takes R immediately, then, taking S is the optimal action regardless of the state. Information becomes worthless because learning the state does not change the player's action and S always gives the player a constant payoff. Then it is never optimal for the player to acquire information. This deters information acquisition completely.

In a less extreme case, information acquisition is not necessarily deterred at the beginning. Instead, there exists an upperbound on the duration of information acquisition. Suppose the opponent uses some well-behaved strategy such that she acquires information at time 0 with probability one. Then, the player's expected payoff from taking R conditional on state H is the highest at time 0 and then decreases. This is because the player gets the first R taker payoff for certain at time 0 and as time passes, the probability of the opponent taking R increases. Since the second R prize is negative, there exists a time instant at which the player's expected payoff from taking R conditional on state H decreases to zero. Then, despite learning the state or not, at and after this time instant, the player's optimal action is S. The player hence gains nothing from acquiring information after this time instant. However, this fact that information becomes worthless after certain time instant does not deter information acquisition at the beginning. When the player just starts acquiring information, the probability of being preempted is low. She puts a higher weight on the first

[11] Note that given Assumption 1.2, the player may eventually be punished due to the total information cost. The difference here is that the player explicitly knows that the second prize from taking R is worse than taking S.

prize conditional on state H. The negative second prize does not matter too much. Information acquisition is not deterred as long as the expected payoff from taking R conditional on state H is positive. Once at the time instant such that R gives the player an expected payoff of zero in state H, information acquisition stops and S will be taken. This puts an upperbound on the duration of information acquisition. Since the player takes the safe action that creates no preemption motive when stopping, there is no random stopping in equilibrium. As a result, there exists a pure strategy equilibrium where in the absence of the revealing signal, the players acquire information up to some time and then take the safe action. This result is summarised in the following proposition.

Proposition 1.3. *Suppose the payoff satisfies Assumption 1.1 and Assumption 1.3. There exists a prior cutoff* p^{NR} *and a non-negative upperbound* $T^{PS} = \frac{1}{a}\log\frac{\bar{\triangle}_H}{-\left(u_H-\bar{\triangle}_H\right)}$ *such that for the prior* $p_0 \in \left(p^{NR}, \tilde{p}\right)$, *there exists a symmetric equilibrium where the players use the pure strategy* $\left(T^{PS}, S\right)$.

This proposition says that when the second prize for R is negative, for intermediate priors, there exists an equilibrium where the players acquires information for a maximum of T^{PS} time. Before time T^{PS}, the players stop and take an action only after receiving the revealing signal. The maximum time T^{PS} is independent of the prior.

The existence of such pure strategy equilibrium is the main difference between the cases with the positive and negative second prize in state H. The intuition is that when the second R prize in state H is negative, the longer the player acquires information, the less attractive R becomes. When the player becomes sufficiently certain that the state is H, she is also convinced that R has been taken. At the end of the information acquisition stage, being sufficiently optimistic is associated with taking S. However, in the positive second prize case, being sufficiently optimistic is associated with taking R, which generates the preemption motive and hence the equilibrium in the random stopping strategy. In equilibrium, the maximum duration T^{PS} is independent of the prior. Given that the opponent takes R after receiving the H-state revealing signal, T^{PS} is the time at which the expected payoff from taking R *conditional* on state H decreases to zero. Since T^{PS} is pinned down by the payoff in state H, it is independent of the prior.

Next I discuss the effect of a more intense competition while Assumption 1.2 holds. In this case, when the difference between the first and second prize from taking R increases, the equilibrium where both players acquire information exists for a smaller range of priors. This is intuitive because the endogenous cost associated with acquiring information increases as the game becomes more competitive. Then, the equilibrium where the players acquire information is harder to be sustained for the lower priors.

The discussion above is about the case when competition becomes more intense. The other limiting case is when there is no competition. That is, when the payoff difference between the first and second R taker is zero. The payoff from taking R does not depend on the opponent's action. Then, there is no strategic interaction between the two players. The equilibrium in this limiting case is such that both players use the single DM optimal strategy.

6 Extensions

6.1 More Than Two players

In this section, I generalise the original model to $N > 2$ players. I assume that the first player to take R gets the first prize and all other R takers get the second prize. I use this generalisation to show that as the number of players increases, the range of priors where the learning equilibrium can exist shrinks.

Increasing the number of players intensifies the competition in the game. In Section 5.2, I discussed the competition in terms of the payoff difference between the first and second prize associated with R. This section further discusses the effect of competition in terms of number of players in the game. I show that the existence of the learning equilibrium requires the competition to be not too intense. The payoff difference between the first and the second prize must decrease as fast as $\frac{1}{N-1}$ to guarantee the existence of the learning equilibrium.

Since I assume that the first R taker gets the first prize and all other R takers get the second prize, the player only cares about whether the first R prize has been taken or not. Because of this, from the player's point of view, the remaining $N-1$ players essentially act as one big opponent. Let $\frac{\tilde{p}_N}{1-\tilde{p}_N} := \frac{-bu_L - c}{(N-1)\triangle_H a + c}$ be a prior cutoff. Since $u_L < 0$, when c is sufficiently small, $\tilde{p}_N$ is positive. The following proposition characterises a necessary condition for the existence of the learning equilibrium where the players use the random stopping strategy.

Proposition 1.4. *For $N > 2$, there exists an equilibrium where the players use the random stopping strategy only if $p_0 < \tilde{p}_N$. The cutoff $\tilde{p}_N$ decreases in N. When $N \to \infty$, $\tilde{p}_N \to 0$.*

The learning equilibrium can exist only if the prior is sufficiently low and this prior cutoff decreases in the number of players. When the number of players goes to infinity, the cutoff $\tilde{p}_N$ approaches zero and the potential learning region vanishes.

Corollary 1.2. *When $N \to \infty$, if $\bar{\triangle}_N$ decreases as fast as $\frac{1}{N-1}$, then, $\tilde{p}_N$ is finite.*

When the number of players increases to infinity, if the prize difference $\bar{\triangle}_H$ decreases as fast as $\frac{1}{N-1}$, then, the potential learning region still exists. The intuition is that the learning equilibrium can exist only if the competition is not too intense. If there are a lot of players competing, then, the payoff difference cannot be too big.

6.2 Observable Actions

The irreversible risky and safe actions are assumed to be private in this paper. This allows me to focus on the role of payoff externalities. In this extension, I assume that after a player takes an irreversible action, it is immediately observed by the opponent. The public actions generate information externalities as observing no action taken is itself informative. The purpose of this extension is to show that the existence of the learning equilibrium found in Theorem 1.1 is robust to some exposure to information externalities. To be more specific, I define *the mimicking and random stopping strategy (MRSS)* and show that there exists a symmetric equilibrium where the players use this strategy.

Definition 1.4. *The mimicking and random stopping strategy (MRSS) is a strategy such that*

1. *After receiving the H-state revealing signal, the player stops and takes R immediately;*

2. *After receiving the L-state revealing signal, the player stops and takes S immediately;*

3. *After receiving no revealing signal and observing no action taken, the player stops and takes R at each time t at rate $h(t) > 0$;*

4. *After receiving no revealing signal and observing the opponent taking S, the player stops and takes S;*

5. *After receiving no revealing signal and observing the opponent taking R, the player uses the single DM optimal strategy.*

The MRSS and the random stopping strategy are similar in the sense that after observing no private revealing signal, the player randomly stops and takes R. The conditions for the existence of the symmetric equilibrium where the players use MRSS is presented in the following proposition.

Proposition 1.5. *Suppose $\bar{\triangle}_H = \bar{\triangle}_L$ and c sufficiently small such that $p^L < \tilde{p}$. If $p^L < p_0 < \tilde{p}$, then, there exists an equilibrium where the players use MRSS.*

Proposition 1.5 suggests that when there are information externalities, there still exists the equilibrium where the players randomly stop and take R after no revealing signal. The reason is twofold. First, the observability of the actions does not eliminate the preemption motive. After the history of no action taken, if the player's strategy prescribes taking R with positive probability mass, then, the opponent has an incentive to preempt. Second, random stopping reduces the information involved in the player's action. The action taken (or no action taken) contains the player's private information. It is essentially an additional signal and the informativeness of it depends on the player's strategy. For example, if the player only stops acquiring information after receiving a revealing signal, then, her action perfectly reveals her private information. In this case, this additional signal is very informative for the opponent. However, if the player stops acquiring information and takes an action randomly, then, her action contains less private information. Information externalities enhance the player's incentive to stop randomly.

7 Conclusion

In this paper, I study a model in which the players can acquire costly private information before taking an irreversible private action. Acquiring information takes time which allows the player to take the 'correct' action but increases the probability of being preempted. With the assumption that there is a first-mover advantage associated with the risky action, I find that in the equilibrium where the players acquire information, they become more optimistic that the state is high but at the same time, the conditional expected payoff from taking the risky action decreases. Because of the interaction between the learning motive and the preemption motive, the players randomly stop and take the risky action. This result is significantly different from the single decision maker case where the optimal strategy is deterministic.

In my model, depending on the prior, the players' decisions on acquiring information can be both strategic substitutes or strategic complements. The strategic substitutability prevails when players have pessimistic priors and it induces players' initial mix between acquiring information and immediate exit. The strategic complementarity prevails when players are relatively optimistic and it gives rise to the multiple equilibria where players use the random stopping strategy. In addition, I find that the equilibrium where the players do not acquire information can exist for not only extreme priors but also a large range of intermediate priors. This stems from the presence of the endogenous information cost. Such equilibrium exists for arbitrarily small exogenous information cost.

Appendix to Chapter 1

1 Formulation of the Player's Problem

This section formulates one player's problem given opponent's strategies. Given player j's strategy γ^j, player i's best-reply problem is to choose a time T^i with value function $\hat{V}^{i,\gamma^j} : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\hat{V}^{i,\gamma^j}(0) := \max_{T^i} \int_0^{T^i} \pi_t \left[p_t^i a \mathbb{E}^{\gamma^j}\left[u_H^R \mid H, t \right] - c \right] dt + \pi_{T^i} U^{i,\gamma^j}\left(T^i\right) \tag{1.4}$$

$$\text{s.t. } \frac{dp_t^i}{dt} = (b-a)p_t^i(1-p_t^i),$$

where $\mathbb{E}^{\gamma^j}\left[u_H^R \mid H, t\right]$ and $U^{i,\gamma^j}\left(T^i\right)$ are as defined in Section 2.2. The Hamilton-Jacobi-Bellman (HJB) equation for player i's problem is the following differential equation in $V^{i,\gamma^j} : \mathbb{R}_+ \to \mathbb{R}$, where

$$\max \left\{ \overbrace{p_t^i a \left[\mathbb{E}^{\gamma^j}\left[u_H^R \mid H, t\right] - V^{i,\gamma^j}(t) \right] + \left(1 - p_t^i\right) b \left[-V^{i,\gamma^j}(t) \right]}^{\text{A}} + \overbrace{\frac{dV^{i,\gamma^j}(t)}{dt}}^{\text{B}} - c, \right. \\ \left. U^{i,\gamma^j}(t) - V^{i,\gamma^j}(t) \right\} = 0. \tag{1.5}$$

The interpretation of (1.5) is that at time t, player i chooses between to continue acquiring information or stopping. She acquires information if the marginal gain is greater than the marginal cost. Otherwise, she stops and gets the payoff $U^{i,\gamma^j}(t)$. The marginal gain consists of the expected gain from receiving the revealing signal (labelled as A in (1.5)) plus the rate of change of the value (labelled as B in (1.5)).

Given opponent's strategy γ^j, if the player's problem is well-behaved, then, the value function $V^{i,\gamma^j}(t)$ is a classical solution to the HJB equation (1.5). The player's best response can then be characterised correspondingly. However, in our problem, (1.5) is not well-behaved because given the opponent's strategy γ^j, $U^{i,\gamma^j}(t)$ has a kink and may not be continuous.

2 Proof of Proposition 1.1

When player i is a single DM, her problem is

$$\hat{V}(p_0) := \max_T \int_0^T \pi_t \left[p_t a u_H + (1-p_t) b u^S - c\right] dt + \pi_T U(T).$$

The HJB equation is

$$\max\left\{p_t a\left[u_H - V(t)\right] + (1-p_t) b\left[u^S - V(t)\right] + \frac{dV(t)}{dt} - c, \right. \\ \left. U(t) - V(t)\right\} = 0$$

Since the argument t only enters the equation via p_t, I use p instead of t as the state variable. Then HJB equation becomes

$$\max\left\{pa\left[u_H - V(p)\right] + (1-p) b\left[u^S - V(p)\right] + \frac{dV(p)}{dp}\frac{dp}{dt} - c, \right. \\ \left. U(p) - V(p)\right\} = 0 \qquad (1.6)$$

To find player i's value function, I construct a candidate value function and show it is a viscosity solution of (1.6).

If the learning region (the range of beliefs at which the DM acquires the signal) exists, the value function is a solution of the ordinary differential equation

$$pa\left[u_H - V(p)\right] + (1-p) b\left[u^S - V(p)\right] + \frac{dV(p)}{dp}(b-a) p (1-p) = c. \qquad (1.7)$$

The free boundary solution to (1.7) is

$$V^L(p) = p\left[u_H - \frac{c}{a}\right] + (1-p)\left[u_L - \frac{c}{b}\right] + K\left(\frac{p}{1-p}\right)^{\frac{b}{b-a}}(1-p)$$

where K is a constant. Suppose the learning region is $(\underline{p}, \bar{p})$. Value matching and smooth pasting pin down the value of $\bar{p}$ and K. Then, value matching pins down $\underline{p}$. We have

$$\frac{\bar{p}}{1-\bar{p}} = \frac{u^S - u_L - \frac{c}{b}}{\frac{c}{b}} := \bar{L},$$

$$K = \frac{c}{b}\left(\frac{b}{a} - 1\right)\bar{L}^{1-\frac{b}{b-a}}$$

and $\underline{L} := \frac{\underline{p}}{1-\underline{p}}$ satisfies

$$\left[u_H - u^S - \frac{c}{a}\right]\underline{L} + K\underline{L}^{\frac{b}{b-a}} = \frac{c}{b}. \tag{1.8}$$

Let $\hat{p}$ be a belief cutoff at which the player is indifferent between R and S. The existence of the learning requires $\bar{p} > \hat{p}$. That is,

$$c < b\frac{(-u_L)\,u_H}{u_H - u_L} := \bar{c}.$$

Outside the learning region, the value function satisfies $V(p) = U(p)$. When $c < \bar{c}$, the candidate value function is

$$V(p) = \begin{cases} p\left[u_H - \frac{c}{a}\right] + (1-p)\left[u_L - \frac{c}{b}\right] + K\left(\frac{p}{1-p}\right)^{\frac{b}{b-a}}(1-p) & p \in \left(\underline{p}, \bar{p}\right) \\ U(p) & \text{o.w.} \end{cases}.$$

This candidate has a kink at $\underline{p}$ and is differentiable everywhere else. I next show that it is a viscosity solution of (1.6). Let

$$H\left(p, V(p), V'(p)\right) := pa\left[u_H - V(p)\right] + (1-p)\,b\left[u^S - V(p)\right] + V'(p)\,(b-a)\,p\,(1-p) - c.$$

For the points where $V(p)$ is differentiable, I show that (1) if $p > \bar{p}$, then, $H(p, V(p), V'(p)) \leq 0$; (2) if $p \in \left(\underline{p}, \bar{p}\right]$, then, $V(p) \geq U(p)$; (3) if $p < \underline{p}$, then $H(p, V(p), V'(p)) \leq 0$. At the point where $V(p)$ is not differentiable, that is, at $p = \underline{p}$, I show that $H(p, V(p), z) \geq 0$ for $z \in D^+$ where $D^+ = \emptyset$ (ignore) and (4) $H(p, V(p), z) \leq 0$ for $z \in D^-$ where $D^- = \left[\frac{dU_S(p)}{dp}\,|_{p=\underline{p}}, \frac{dV^L(p)}{dp}\,|_{p=\underline{p}}\right]$.

Step (1): When $p > \bar{p}$, we have $V(p) = U_R(p)$ and $H(p, U_R(p), U_R'(p)) < 0$ if and only if $p > \bar{p}$.

Step (2): It can be shown that $V^L(p)$ is convex. At $p = \bar{p}$, we have $V^L(p) = U_R(p)$. If we decrease p by a little bit, $U_R(\cdot)$ decreases faster than $V^L(\cdot)$. Therefore, we have $V^L(p) \geq U_R(p)$ for $p \leq \bar{p}$. At $p = \underline{p}$, we have $V^L\left(\underline{p}\right) = U_S\left(\underline{p}\right)$ and $\frac{dV^L(p)}{dp} > \frac{dU_S(p)}{dp}$. As a result, we have $V^L(p) \geq U_S(p)$ for $p > \underline{p}$.

Step (3): If $p < \underline{p}$, then, $V(p) = U_S(p)$. To have $H\left(p, U_S(p), \frac{dU_S(p)}{dp}\right) \leq 0$, we need $p \leq \frac{\frac{c}{a}}{u_H - u^S}$. It can be shown that $\underline{p} < \frac{\frac{c}{a}}{u_H - u^S}$. As a result, if $p < \underline{p}$, then $H(p, V(p), V'(p)) \leq 0$.

Step (4): Since $H(p, V(p), z)$ is increasing in z and we have $H\left(\underline{p}, V\left(\underline{p}\right), \frac{dV^L(p)}{dp} \mid_{p=\underline{p}}\right) = 0$, it is true that $H(p, V(p), z) \leq 0$ for $\forall z \in \left[\frac{dU_S(p)}{dp} \mid_{p=\underline{p}}, \frac{dV^L(p)}{dp} \mid_{p=\underline{p}}\right]$.

To conclude, if $c < \bar{c}$, the DM's optimal strategy is to acquire the signal if the belief if in the range $\left(\underline{p}, \bar{p}\right)$, to take action R if $p \geq \bar{p}$ and to take action S if $p \leq \underline{p}$. If $c \geq \bar{c}$, the DM's optimal strategy is to take an action without acquiring the signal.

3 Proof of Lemma 1.1

Consider a strategy $\gamma = (\rho, \sigma)$ as defined in Definition 1.2. Suppose ρ is discontinuous at some $\tau > 0$ and continuous everywhere else. Since by definition ρ is right-continuous and weakly increasing, this implies that $\rho(\tau) > \lim_{t \to \tau_-} \rho(t)$. Then, the induced probability that the player takes R before or at time t in state ω, $F_\omega^\gamma(t)$, [12] is right continuous such that $F_\omega^\gamma(\tau) > \lim_{t \to \tau_-} F_\omega^\gamma(t)$ for $\forall \omega \in \{H, L\}$. Let M_ω denote the mass that $F_\omega^\gamma(\tau)$ places on τ. Consider a deviation $\gamma^D = \left(\rho^D, \sigma\right)$ such that it is identical to γ except that the mass that ρ places on τ is shifted to $\tau - \triangle$. There exists a $\triangle > 0$ such that given the opponent uses the γ strategy, using γ^D gives the player a higher payoff than γ.

Suppose the opponent uses the γ strategy described above. Consider the history that the player receives no revealing signal until time $\tau - \triangle$ where $\triangle > 0$. The total gain from acquiring information for $\triangle$ time longer is

$$p_{\tau-\triangle} a \triangle \left[u_H - F_H^\gamma(\tau) \bar{\triangle}_H\right] + \left[1 - p_{\tau-\triangle} a \triangle - \left(1 - p_{\tau-\triangle}\right) b \triangle\right] U_R(\tau) - U_R(\tau - \triangle)$$

where

$$U_R(t) = p_t \left[u_H - F_H^\gamma(t) \bar{\triangle}_H\right] + (1 - p_t) \left[u_L - F_L^\gamma(t) \bar{\triangle}_L\right]$$

if $t < \tau$ and

$$U_R(t) = p_t \left[u_H - F_H^\gamma(t) \bar{\triangle}_H - \left(1 - F_H^\gamma(t)\right) \underline{\triangle}_H\right] + (1 - p_t) \left[u_L - F_L^\gamma(t) \bar{\triangle}_L - \left(1 - F_L^\gamma(t)\right) \underline{\triangle}_L\right]$$

[12] The superscript indicates the strategy that this function is induced from.

if $t = \tau$. When $\triangle > 0$ approaches zero, this gain approaches

$$p_\tau \left[-M_H \bar{\triangle}_H - \left(1 - F_H^\gamma(\tau)\right) \underline{\triangle}_H\right] + (1 - p_\tau) \left[-M_L \bar{\triangle}_L - \left(1 - F_L^\gamma(\tau)\right) \underline{\triangle}_L\right],$$

which is negative due to the mass point M_ω and the tie-breaking at time τ. Therefore, there is always a deviation to put the mass places on τ to $\tau - \triangle$.

4 Proof of Theorem 1.1

Let $\frac{p^M}{1-p^M} := \frac{-u_L+\underline{\triangle}_L}{u_H-\underline{\triangle}_H}$, $\frac{p^L}{1-p^L} := \frac{-u_L}{u_H}$ and $\frac{\tilde{p}}{1-\tilde{p}} := \frac{-bu_L-c}{a\bar{\triangle}_H+c}$ be three prior cutoffs. As $u_L < 0$, when $c < -bu_L$, all of the three cutoffs are positive. The cutoff $\underline{p}$ is defined in Proposition 1.1 and p^* is a fixed point that will be defined later in step 5.

The method to find the symmetric equilibrium is 'guess and verify'. The following outlines the steps of the proof.

1. I show that there exists an equilibrium where both players use the pure strategy $(0, S)$ when $p_0 \leq \underline{p}$.

2. I show that there exists an equilibrium where both players use the pure strategy $(0, R)$ when $p_0 \geq p^M$.

3. I show that there exists an equilibrium where both players use the Immediate Mix with No Learning Strategy when $p^L < p_0 < p^M$.

4. I show that there exists an equilibrium where both players use a Randomised Stopping Time Strategy when $p^L < p_0 < \tilde{p}$.

5. I show that there exists an equilibrium where both players use a Randomised Stopping Time Strategy when $p^* < p_0 < p^L$.

6. I show that there exists an equilibrium where both players use the Mixed Learning Strategy when $\underline{p} < p_0 < p^*$.

Step 1 When $p_0 \leq \underline{p}$, Proposition 1.1 implies that a single DM takes S immediately. If a player takes S immediately, the other player is the single DM in the game. Given the player takes S immediately, the opponent's best response is to take S. Therefore, when $p_0 \leq \underline{p}$, both players taking S immediately is an equilibrium. This is summarised in the following lemma.

Lemma 1.3. *If $p_0 \leq \underline{p}$, then, there exists an equilibrium where both players use the strategy $(0, S)$.*

Step 2 I prove the following lemma.

Lemma 1.4. *If $p_0 \geq p^M$, there exists an equilibrium where both players use the pure strategy* $(0, R)$

Proof. Suppose player j uses the strategy $(0, R)$. I check whether player i wants to stop at time 0 and take action R or to acquire the signal for dt longer. If player i takes the immediate R action at time 0, the payoff is

$$U_R^{IR}(0) = p_0 \left(u_H - \underline{\triangle}_H\right) + (1 - p_0) \left(u_L - \underline{\triangle}_L\right).$$

If the player takes action R at time $t > 0$, the payoff is

$$U_R^{IR}(t) = p_t \left(u_H - \bar{\triangle}_H\right) + (1 - p_t) \left(u_L - \bar{\triangle}_L\right)$$

At time 0, the gain from acquiring the signal for dt longer is

$$p_0 a dt \left(u_H - \bar{\triangle}_H\right) + [1 - p_0 a dt - (1 - p_0) b dt] U_R^{IR}(0 + dt) - U_R^{IR}(0). \qquad (1.9)$$

When $dt \to 0$, (1.9) tends to something negative, which is smaller than the cost of information. Therefore, if the opponent stops at time 0, player i prefers taking action R at time 0 to acquiring the signal for dt longer. At time 0, player i prefers action R to action S at time 0 if $U_R^{IR}(0) \geq 0$. Since $p_0 \geq p^M$, the inequality $U_R^{IR}(0) \geq 0$ holds. □

Step 3 I show the following lemma.

Lemma 1.5. *If $p^L < p_0 < p^M$, then, there exists an equilibrium where both players use the Immediate Mix Strategy.*

Proof. Suppose player j uses the strategy $\left(\rho^{IM}, \sigma^{IM}\right)$ such that $\rho^{IM}(t) = m$ for $\forall t \in [0, \infty)$ and $\sigma^{IM}(t) = 1 - m$ for $\forall t \in [0, \infty)$. Then, player i's payoff from taking action R at time 0 is

$$\begin{aligned} U_R^{IM}(0) &= m \left[p_0 u_H + (1 - p_0) u_L\right] + (1 - m) \left[p_0 \left(u_H - \underline{\triangle}_H\right) + (1 - p_0) \left(u_L - \underline{\triangle}_L\right)\right] \\ &= p_0 \left[m u_H + (1 - m) \left(u_H - \underline{\triangle}_H\right)\right] + (1 - p_0) \left[m u_L + (1 - m) \left(u_L - \underline{\triangle}_L\right)\right]. \end{aligned}$$

Player i's payoff from taking action R at time $t > 0$ is

$$\begin{aligned} U_R^{IM}(t) &= m \left[p_t u_H + (1 - p_t) u_L\right] + (1 - m) \left[p_t \left(u_H - \bar{\triangle}_H\right) + (1 - p_t) \left(u_L - \bar{\triangle}_L\right)\right] \\ &= p_t \left[u_H - (1 - m) \bar{\triangle}_H\right] + (1 - p_t) \left[u_L - (1 - m) \bar{\triangle}_L\right]. \end{aligned}$$

At time $t = 0$, player i does not want to acquire the information due to the same reasoning as in step 2.

Next I show that for the prior $p_0 \in \left(p^L, p^M\right)$, there exists $m \in (0,1)$ such that player i is indifferent between taking action R and action S at time $t = 0$. The indifference requires $0 = U_R(0)$. That is,

$$\frac{p_0}{1 - p_0} = -\frac{\left[mu_L + (1-m)\left(u_L - \underline{\triangle}_L\right)\right]}{\left[mu_H + (1-m)\left(u_H - \underline{\triangle}_H\right)\right]}. \tag{1.10}$$

For any $p_0 \in \left(p^L, p^M\right)$, there exists $m \in (0,1)$ such that (1.10) holds. □

Step 4 This step shows the existence of the equilibrium where players use the randomised stopping time strategy. I show that there exists an equilibrium where both players take action R at each time $t \geq 0$ with positive rate. Suppose cost c is sufficiently small such that $p^L < \tilde{p}$. Let $L_t = \frac{p_t}{1-p_t}$ be the likelihood ratio. I prove the following lemma.

Lemma 1.6. *Suppose c is sufficiently small and $b < 2a$. If $p^L < p_0 < \tilde{p}$, then, there exists an equilibrium in mixed strategies (ρ, σ) such that $\sigma(t) = 0$ for $\forall t \in \mathbb{R}_+$ and $\rho : \mathbb{R}_+ \to [0,1]$ satisfies the following conditions:*

1. ***(Initial condition)*** *$\rho(0) = 0$.*

2. ***(Increasing condition)*** *There exists a $T > 0$ such that for $t \in [0,T]$, ρ is a solution to the following differential equation*

$$\frac{d\rho(t)}{dt} = \frac{b(-u_L) - c - L_t c + b\bar{\triangle}_L \int_0^t e^{-bs}\frac{d\rho(s)}{ds}ds - L_t\bar{\triangle}_H a e^{-at}(1-\rho(t))}{L_t\bar{\triangle}_H e^{-at} + \bar{\triangle}_L e^{-bt}} \tag{1.11}$$

   with the initial condition $\rho(0) = 0$ where $\frac{d\rho}{dt} > 0$ for $\forall t \in [0,T]$ and $\rho(T) = 1$.

3. ***(Terminal condition)*** *$\rho(t) = 1$ for $\forall t > T$.*

In words, the equilibrium strategy is that the players stop and take action R at each time with positive rate.

Proof. Fix player j's strategy (ρ, σ), I first show that at each time t, player i is indifferent between taking action R now or dt later if ρ satisfies (1.11). Then Lemma 1.7 shows that when c is sufficiently small and $b < 2a$, if $p_0 < \tilde{p}$, we can find a ρ function such that $\frac{d\rho}{dt} > 0$. Last, I show that if $p_0 > p^L$, player i prefers acquiring information to taking action S.

Suppose player j uses the mixed strategy (ρ, σ). This strategy induces

$$\begin{aligned} F_H(t) &= \int_0^t \left[e^{-as}(1-\rho(s)) \left(a + \frac{\frac{d\rho(s)}{ds}}{1-\rho(s)} \right) \right] ds \\ &= 1 - e^{-at}(1-\rho(t)) \end{aligned}$$

and

$$\begin{aligned} F_L(t) &= \int_0^t \left[e^{-bs}(1-\rho(s)) \frac{\frac{d\rho(s)}{ds}}{1-\rho(s)} \right] ds \\ &= \int_0^t \left[e^{-bs} \frac{d\rho(s)}{ds} \right] ds. \end{aligned}$$

Player i's payoff from taking action R at time t is

$$U_R(t) := p_t \left[u_H - F_H(t) \bar{\triangle}_H \right] + (1-p_t) \left[u_L - F_L(t) \bar{\triangle}_L \right].$$

The equilibrium condition is that given player j's strategy, player i is indifferent between taking action R and acquiring the signal at each time instant. That is,

$$p_t a dt \left[u_H - F_H(t+dt) \bar{\triangle}_H \right] + \left[1 - p_t a dt - (1-p_t) b dt \right] U_R(t+dt) - U_R(t) = cdt.$$

The interpretation is that the marginal cost and marginal benefit associated with acquiring information for dt longer are the same. When $dt \to 0$, we have the following differential equation

$$c = (1-p_t) b \left[-u_L + F_L(t) \bar{\triangle}_L \right] - p_t \frac{dF_H(t)}{dt} \bar{\triangle}_H - (1-p_t) \frac{dF_L(t)}{dt} \bar{\triangle}_L$$

After plugging in the expressions of $F_H(t)$ and $F_L(t)$, we have

$$\frac{d\rho(t)}{dt} = \frac{b(-u_L) - c - L_t c + b\bar{\triangle}_L \int_0^t e^{-bs} \frac{d\rho(s)}{ds} ds - L_t \bar{\triangle}_H a e^{-at}(1-\rho(t))}{L_t \bar{\triangle}_H e^{-at} + \bar{\triangle}_L e^{-bt}}.$$

Lemma 1.7. *The differential equation (1.11) with initial condition $\rho(0) = 0$ has a unique solution defined for all $t \in [0, \infty)$.*

Proof. Let $A(t) := e^{-bt}\rho(t)$. Then, we have $\frac{dA}{dt} = -bA + e^{-bt}\frac{d\rho}{dt}$. Use the formula that $L_t = L_0 e^{(b-a)t}$, (1.11) can be rewritten as

$$\left(\frac{dA}{dt} + bA\right) e^{bt} =$$

$$\frac{b\left(-u_L\right) - c - cL_0 e^{(b-a)t} + b\bar{\triangle}_L \left(A + b\int_0^t A\left(s\right) ds\right) - L_0 \bar{\triangle}_H a e^{(b-2a)t}\left(1 - e^{bt}A\right)}{L_0 \bar{\triangle}_H e^{(b-2a)t} + \bar{\triangle}_L e^{-bt}}. \tag{1.12}$$

Let $z\left(t\right) := \int_0^t A\left(s\right) ds$. Then, the differential equation can be written as

$$z'' + g_1\left(t\right) z' + g_2\left(t\right) z = g_3\left(t\right) \tag{1.13}$$

where

$$\begin{aligned} g_0\left(t\right) &= \frac{1}{L_0 \bar{\triangle}_H e^{2(b-a)t} + \bar{\triangle}_H}, \\ g_1\left(t\right) &= b - \frac{b\bar{\triangle}_L + L_0 \bar{\triangle}_H a e^{2(b-a)t}}{g_0\left(t\right)}, \\ g_2\left(t\right) &= -\frac{b^2 \bar{\triangle}_L}{g_0\left(t\right)}, \\ g_3\left(t\right) &= \frac{b\left(-u_L\right) - c - cL_0 e^{(b-a)t} - L_0 \bar{\triangle}_H a e^{(b-2a)t}}{g_0\left(t\right)}. \end{aligned}$$

Since $g_0\left(t\right)$, $g_1\left(t\right)$, $g_2\left(t\right)$ and $g_3\left(t\right)$ are continuous on $[0, \infty)$, (1.13) with initial conditions $z\left(0\right) = 0$ and $z'\left(0\right) = 0$ has a unique solution defined for all t on $[0, \infty)$. The existence of $z\left(\cdot\right)$ implies the existence of $\rho\left(\cdot\right)$. □

Lemma 1.8. *$\frac{d\rho}{dt} |_{t=0} > 0$ iff $p_0 < \tilde{p}$.*

Proof.

$$\begin{aligned} \frac{d\rho}{dt} |_{t=0} &= \frac{b\left(-u_L\right) - c - L_0 c - L_0 \bar{\triangle}_H a}{L_0 \bar{\triangle}_H + \bar{\triangle}_L} \\ &> 0 \end{aligned}$$

if and only if

$$L_0 = \frac{p_0}{1 - p_0} < \frac{b\left(-u_L\right) - c}{a\bar{\triangle}_H + c} = \frac{\tilde{p}}{1 - \tilde{p}}.$$

□

This lemma implies that if $p_0 < \tilde{p}$, then, $\frac{d\rho}{dt} > 0$ at the neighbourhood of 0. The following lemma and its proof shows that under some conditions, if $\rho\left(\cdot\right)$

increases at the neighbourhood of 0 and the value of ρ is smaller than 1, then, it continues increasing.

Lemma 1.9. *Suppose $b < 2a$ and c sufficiently small. If $\rho(t) \in [0,1)$ for $\forall t \in [0,\tau]$ and $\frac{d\rho}{dt} > 0$ $\forall t \in [0,\tau)$, then, $\frac{d\rho}{dt} > 0$ at $t = \tau$.*

Proof. If $b < 2a$, c sufficiently small, and $\frac{d\rho}{dt} > 0$ $\forall t \in [0,\tau)$, then,

$$\Phi(t) := \frac{b\left(-u_L + \bar{\triangle}_L \int_0^t e^{-bs}\frac{d\rho}{ds}ds\right) - c - L_0 e^{(b-a)t}c}{\bar{\triangle}_H a e^{(b-2a)t}(1-\rho(t))}$$

increases in t for $\forall t \in [0,\tau)$. Since $\rho(t) \in [0,1)$ for $\forall t \in [0,\tau)$, then, $\Phi(t) > L_0$ for $\forall t \in [0,\tau)$ iff $\frac{d\rho}{dt} > 0$ for $\forall t \in [0,\tau)$. Next, I show by contradiction that if $\frac{d\rho}{dt} > 0$ for $\forall t \in [0,\tau)$ and $\rho(t) \in [0,1)$ for $\forall t \in [0,\tau]$, then, it must be that $\frac{d\rho}{dt} > 0$ at $t = \tau$. Suppose $\frac{d\rho}{dt} = 0$ at $t = \tau$. Then, it must be that $\Phi(\tau) = L_0$. However, we know that $\Phi(t)$ increases in t for $\forall t \in [0,\tau)$ and $\Phi(t) > L_0$ for $\forall t \in [0,\tau)$. Then, it cannot be that $\Phi(\tau) = L_0$. There is a contradiction and hence $\frac{d\rho}{dt} \neq 0$ at $t = \tau$. Since ρ is continuous, it cannot be that $\frac{d\rho}{dt} > 0$ at $t = \tau$. □

Next I show that given Lemma 1.7 and Lemma 1.9, $\rho(t) = 1$ will be reached in a finite time.

Lemma 1.10. *Given Lemma 1.7 and Lemma 1.9, there exists a $T < \infty$ such that $\rho(T) = 1$.*

Proof. If $\frac{d\rho}{dt} > 0$ and $\rho(t) \in [0,1)$, then

$$\begin{aligned}\Phi(t) &< \frac{b\left(-u_L + \bar{\triangle}_L \rho(t)\right) - c - L_0 e^{(b-a)t}c}{\bar{\triangle}_H a e^{(b-2a)t}(1-\rho(t))} \\ &< \frac{b\left(-u_L + \bar{\triangle}_L\right) - c - L_0 e^{(b-a)t}c}{\bar{\triangle}_H a e^{(b-2a)t}(1-\rho(t))}.\end{aligned}$$

As $t \to \infty$, we have $\Phi(t) < 0$. If $\Phi(t) < 0$ and $\rho(t) < 1$ when $t \to \infty$, then, $\frac{d\rho}{dt} < 0$ as $t \to \infty$. There is a contradiction. Therefore, it must be that $\rho(t) > 1$ for some t. Because of continuity, there exists a T such that $\rho(T) = 1$. □

Lemma 1.11. *If $p^L < p_0$ and $\rho(t)$ satisfies the conditions in Lemma 1.6, then, $\sigma(t) = 0$ for $\forall t \in [0,T]$.*

Proof. Let $V(t)$ be the value associated with the mixed strategy characterised in Lemma 1.6. Then,

$$V(t) = U_R(t) - ct = U_R(0)$$

because the player's payoffs are the same at each time when she is randomising between continuing and stopping (to take action R). Since $p^L < p_0$, we have $U_R(0) > 0$. Hence, at each $t < T$, we have $U_R(t) > 0$. As a result, $\sigma(t) = 0$ for $\forall t \in [0, T]$. □

This completes the proof of Lemma 1.6.

□

Step 5 In the previous step, I consider the situation that the player starts randomisation at time $t = 0$. However, it is possible that the player strictly prefers to acquire information for a certain time period and then starts randomisation. This step shows the existence of the equilibrium where players use a mixed strategy $(\hat{\rho}, \hat{\sigma})$ such that (1) $\hat{\rho}(t) = 0$ for $t \leq \hat{T}$, (2) $\frac{d\hat{\rho}(t)}{dt} \mid_{t \geq \hat{T}} > 0$, (3) $\hat{T} \in \left[\hat{T}_l, \hat{T}_r\right]$ and (4) $\hat{\sigma}(t) = 0$ for $\forall t \geq 0$. That is, a randomised stopping time strategy such that the players acquire information with probability one until some time $\hat{T} > 0$ and then start randomising between stopping and acquiring information. I first define three parameters T_l, T_r, p^* and show their existence. Then, I show the conditions that $\hat{\rho}$ satisfies in equilibrium and under what conditions such equilibrium exists.

Let

$$T_r := \min\left\{t : \frac{p_0}{1-p_0} e^{(b-a)t} = \frac{b(-u_L) - c}{c + ae^{-at}\bar{\triangle}_H}\right\} \tag{1.14}$$

and

$$T_l := \min\left\{t : \frac{p_0}{1-p_0} e^{(b-a)t} = \frac{-u_L}{u_H - (1 - e^{-at})\bar{\triangle}_H}\right\}. \tag{1.15}$$

Lemma 1.12. *If $p_0 < \tilde{p}$, then, $T_r > 0$ exists. If $p_0 < p^L$, then, $T_l > 0$ exists.*

Let

$$\frac{p^*(t)}{1 - p^*(t)} := \frac{\frac{c}{b}}{u_H - \frac{1}{2}\bar{\triangle}_H - \frac{c}{a} + J_2(t)} \tag{1.16}$$

where

$$J_2(t) = \left[\frac{1}{2}e^{-at}\bar{\triangle}_H + \frac{c}{a} - \left(-u_L - \frac{c}{b}\right)\frac{1}{L_t}\right]e^{-at}.$$

Let

$$\underline{p}^*(p_0) := p^*(T_r)$$

be $p^*(t)$ evaluated at $t = T_r$. I denote it as $\underline{p}^*(p_0)$ because T_r depends on p_0. Let p^* be the fixed point such that $p^* = \underline{p}^*(p^*)$.

Lemma 1.13. *Suppose c is sufficiently small. There exists a $p^* < p^L$ such that $p^* = \underline{p}^* (p^*)$. We have $p^* < p_0$ if and only if $\underline{p}^* (p_0) < p_0$.*

Proof. When $0 < p_0 < p^L$, $\underline{p}^* (\cdot)$ decreases in p_0. When $p_0 \to 0$, $p_0 < \underline{p}^* (p_0) < p^L$. When $p_0 \to p^L$ and c sufficiently small, we have $\underline{p}^* (p_L) < \underline{p}^* (0) < p^L$. Therefore, there exists $p^* < p^L$ such that $p^* = \underline{p}^* (p^*)$. Since $\underline{p}^* (\cdot)$ decreases in p_0, $p^* < p_0$ if and only if $\underline{p}^* (p_0) < p_0$. □

Next, I show the existence of the equilibrium and the conditions $\hat{\rho}$ satisfies in equilibrium. Suppose player j uses the mixed strategy $(\hat{\rho}, \hat{\sigma})$ such that (1) $\hat{\rho}(t) = 0$ for $t \leq \hat{T}$, (2) $\frac{d\hat{\rho}(t)}{dt} \mid_{t \geq \hat{T}} > 0$, (3) $\hat{T} \in \left[\hat{T}_l, \hat{T}_r\right]$ and (4) $\hat{\sigma}(t) = 0$ for $\forall t \geq 0$. Given the assumption that $\hat{\rho}(t) = 0$ for $t \leq \hat{T}$ and $\hat{\rho}(t) > 0$ for $t > \hat{T}$, consider time $\hat{T}$ as a new time 0 and denote the time line starting from $\hat{T}$ as τ. Let $\tau := t - \hat{T}$ and $\lambda(\tau) := \hat{\rho}\left(\tau + \hat{T}\right)$. Let $q_\tau = p_{\tau + \hat{T}}$ be the belief and let $\hat{U}_R(\tau)$ be player i's payoff from taking action R at time τ, then,

$$\hat{U}_R(\tau) = q_\tau \left[\left(1 - \hat{F}_H(\tau)\right)\eta + \hat{F}_H(\tau)\left(u_H - \bar{\triangle}_H\right)\right] + (1 - q_\tau)\left[u_L - \hat{F}_L(\tau)\bar{\triangle}_L\right]$$

where

$$\hat{F}_H(\tau) = 1 - e^{-a\tau - a\hat{T}} + e^{-a\tau - a\hat{T}}\lambda(\tau),$$

$$\hat{F}_L(\tau) = \int_0^{\tau + \hat{T}} e^{-bs}\lambda(s)\,ds$$

and $\eta := u_H - \left(1 - e^{-a\hat{T}}\right)\bar{\triangle}_H$. The intuition is that when considering time $\hat{T}$ as a new time 0, player i's problem is essentially the same as in step 4 with η being the payoff for the first action R taker instead of u_H.

After time $\hat{T}$, the equilibrium condition requires player i being indifferent between acquiring the information and taking action R at each time instant $\tau > 0$. That is,

$$\begin{aligned} c = &(1 - q_\tau)\, b\left[-\left(u_L - \hat{F}_L(\tau)\bar{\triangle}_L\right)\right] \\ &- q_t \frac{d\hat{F}_L(\tau)}{d\tau}\left(\eta - \left(u_H - \bar{\triangle}_H\right)\right) - (1 - q_\tau)\frac{d\hat{F}_L(\tau)}{d\tau}\bar{\triangle}_L. \end{aligned}$$

After plugging in the expressions of $\hat{F}_H(t)$ and $\hat{F}_H(t)$, we have

$$\frac{d\lambda(\tau)}{d\tau} = \frac{b(-u_L) - c - \hat{L}_\tau c + b\bar{\triangle}_L \int_0^\tau e^{-bs}\frac{d\lambda(s)}{ds}ds - \hat{L}_\tau \bar{\triangle}_H a e^{-a\tau}(1 - \lambda(t))}{\hat{L}_\tau\left(\eta - \left(u_H - \bar{\triangle}_H\right)\right)e^{-a\tau} + \bar{\triangle}_L e^{-b\tau}} \tag{1.17}$$

where $\hat{L}_\tau = \frac{q_\tau}{1-q_\tau}$. To have $\frac{d\lambda(s)}{ds} > 0$, we need

$$\hat{L}_\tau < \frac{b\left(-u_L\right) - c + b\bar{\triangle}_L \int_0^\tau e^{-bs}\frac{d\lambda(s)}{ds}ds}{c + \left(\eta - \left(u_H - \bar{\triangle}_H\right)\right) ae^{-a\tau}\left(1 - \lambda\left(\tau\right)\right)}.$$

That is,

$$L_0 e^{(b-a)t} < \frac{b\left(-u_L\right) - c + b\bar{\triangle}_L \int_0^t e^{-bs}\frac{d\hat{\rho}(s)}{ds}ds}{c + \bar{\triangle}_H ae^{-at}\left(1 - \hat{\rho}\left(t\right)\right)} \tag{1.18}$$

for all $t > 0$. Equation (1.18) is derived by substituting in $\tau = t - \hat{T}$, $\lambda\left(\tau\right) = \hat{\rho}\left(t\right)$ and $\eta = u_H - \left(1 - e^{-a\hat{T}}\right)\bar{\triangle}_H$. The existence of an increasing function $\hat{\rho}\left(\cdot\right)$ has been shown in Lemma 1.7. that satisfies (1.18).

Next, I characterise the condition that $\hat{T}$ satisfies in equilibrium. I am going to show that there exists an interval $\left[\hat{T}_l, \hat{T}_r\right]$ such that an equilibrium exists when $\hat{T} \in \left[\hat{T}_l, \hat{T}_r\right]$. The idea is that before time $\hat{T}$, player i must strictly prefer to acquire the signal and after time $\hat{T}$, player i is indifferent between acquiring the signal and taking action R at each time instant. To have the player strictly prefer to acquire the signal before time $\hat{T}$, we need the marginal cost smaller than the marginal benefit associated with acquiring the signal. That is,

$$L_0 e^{(b-a)t} < \frac{b\left(-u_L\right) - c}{c + ae^{-at}\bar{\triangle}_H} \tag{1.19}$$

for all $t \leq \hat{T}$. The upperbound $\hat{T}_r$ is the first time the marginal cost of acquiring the signal exceeds the marginal benefit. Given Lemma 1.12, if $p_0 < \tilde{p}$, then, $\hat{T}_r = T_r > 0$ exists.

The lowerbound of $\hat{T}$ is the earliest time point at which the player is willing to start randomising. That is, if player j starts randomising at time $\hat{T}$, player i must prefer to start randomising at time $\hat{T}$ instead of taking action S. At time $\hat{T}$, the value associated with randomisation is the same as the value associated with taking action R because of the opponent's randomisation. Therefore, in order to have player i prefer randomisation to taking action S at time $\hat{T}$, we need

$$p_{\hat{T}}\left[u_H - \left(1 - e^{-a\hat{T}}\right)\bar{\triangle}_H\right] + \left(1 - p_{\hat{T}}\right)u_L \geq 0.$$

That is,

$$L_0 e^{(b-a)\hat{T}} \geq \frac{-u_L}{u_H - \left(1 - e^{-a\hat{T}}\right)\bar{\triangle}_H}. \tag{1.20}$$

The lowerbound of $\hat{T}$ is the smallest $\hat{T}$ such that inequality (1.20) holds. Given Lemma 1.12, if $p_0 < p^L$, then, $\hat{T}_l = T_l > 0$. If $p_0 \geq p^L$, then, $\hat{T}_l = 0$.

The following lemma characterises the equilibrium when $p^* < p_0 < \tilde{p}$.

Lemma 1.14. *Suppose $b < 2a$ and c is sufficiently small. If $p^* < p_0 < \tilde{p}$, there exists an equilibrium in mixed strategies $(\hat{\rho}, \hat{\sigma})$ such that $\hat{\sigma}(t) = 0$ for $\forall t \in \mathbb{R}_+$ and $\hat{\rho}(\cdot)$ satisfies the following conditions:*

1. *$\hat{\rho}(t) = 0$ for $t \leq \hat{T}$*
2. *$\hat{\rho}(t) \in (0,1]$ for $t \in \left(\hat{T}, \bar{T}\right]$ and $\hat{\lambda}(\tau) = \hat{\rho}\left(\tau + \hat{T}\right)$ is a solution to the differential equation (1.17) with initial condition $\hat{\lambda}(0) = 0$ where $\frac{d\hat{\rho}}{dt} > 0$ for $\forall t \in \left[\hat{T}, \bar{T}\right]$*
3. *$\hat{\rho}(t) = 1$ for $\forall t > \bar{T}$*
4. *$\hat{T} \in \left[\hat{T}_l, \hat{T}_r\right]$ where $\hat{T}_r = T_r$, $\hat{T}_l = T_l$ if $p_0 < p^L$ and $\hat{T}_l = 0$ if $p_0 \geq p^L$.*

Proof. I have shown the existence of an increasing $\hat{\rho}$ function, and (1.17) guarantees that the players stop and take action R with a positive rate at each time $t \in \left[\hat{T}, \bar{T}\right]$. I have also derived the conditions for $\hat{T}$. What left to show is the equilibrium exists when $p_0 > p^*$. Given Lemma 1.13, $p^* < p_0$ if and only if $\underline{p}^*(p_0) < p_0$. What left to show is the equilibrium exists when $\underline{p}^*(p_0) < p_0$.. Suppose player j uses $(\hat{\rho}, \hat{\sigma})$ strategy described in the lemma. Since I have discussed what happens after time $\hat{T}$, I will characterise player i's value $W(\cdot)$ at time $t < \hat{T}$. The HJB equation is

$$\begin{aligned} \max \{ & p_t a \left[u_H - \left(1 - e^{-at}\right)\bar{\triangle}_H - W(t)\right] \\ & + (1 - p_t) b [-W(t)] - c + W'(t), \\ & U(t) - W(t)\} = 0 \end{aligned}$$

If the learning region exists, in the learning region, we have

$$W(t) = p_t \left(\left(u_H - \bar{\triangle}_H\right) - \frac{c}{a} + \frac{1}{2}\bar{\triangle}_H e^{-at}\right) + (1 - p_t)\left(-\frac{c}{b}\right) + p_t e^{at} J_2$$

where J_2 is a constant. At time $\hat{T}$, we have $W\left(\hat{T}\right) = U\left(\hat{T}\right) = U_R\left(\hat{T}\right)$. This pins down J_2 as a function of $\hat{T}$, where

$$J_2\left(\hat{T}\right)=\left[\frac{1}{2}e^{-a\hat{T}}\bar{\triangle}_H+\frac{c}{a}-\left(-u_L-\frac{c}{b}\right)\frac{1}{L_{\hat{T}}}\right]e^{-a\hat{T}}.$$

Player 1 acquires the signal at time $t=0$ if $W(0)\geq 0$, which requires

$$L_0\geq\frac{\frac{c}{b}}{u_H-\frac{1}{2}\bar{\triangle}_H-\frac{c}{a}+J_2\left(\hat{T}\right)}:=\underline{L}\left(\hat{T}\right):=\frac{p^*\left(\hat{T}\right)}{1-p^*\left(\hat{T}\right)}. \tag{1.21}$$

The discussion shows that given the opponent uses the strategy described in the lemma, player i's best response is to use the Randomised Stopping Time Strategy if $p_0\geq p^*\left(\hat{T}\right)$.

Since $\hat{T}$ can be any value in the interval $\left[\hat{T}_l,\hat{T}_r\right]$, I use this to characterise the lowerbound of the prior p_0 such that the equilibrium described in Lemma 1.14 exists.

Lemma 1.15. *$\underline{p}^*(p_0)\leq p^*\left(\hat{T}\right)$ for $\hat{T}\in\left[\hat{T}_l,\hat{T}_r\right]$.*

Lemma 1.15 is true because when $\hat{T}\leq\hat{T}_r$, $J_2(\cdot)$ increases in its argument. Since $\frac{p^*(\hat{T})}{1-p^*(\hat{T})}$ decreases in J_2, given $\hat{T}\leq\hat{T}_r$, we have

$$\underline{p}^*(p_0)\leq p^*\left(\hat{T}\right).$$

□

Lemma 1.15 implies that for any $p_0>\underline{p}^*(p_0)$, there exists an equilibrium where the players use the Randomised Stopping Time Strategy as described in Lemma 1.14.

Step 6 Suppose player j uses the Mixed Learning Strategy $\left(\rho^{ML},\sigma^{ML}\right)$ such that (1) $\sigma^{ML}(t)=\beta>0$ for $\forall t\geq 0$, (2) $\rho^{ML}(t)\in(0,1]$ for $t\in\left(\hat{T}^\beta,\bar{T}^\beta\right]$, (3) $\rho^{ML}(t)=1-\beta$ for $\forall t>\bar{T}^\beta$, and (4) $\hat{T}^\beta\in\left[\hat{T}_l^\beta,\hat{T}_r^\beta\right]$ where

$$\hat{T}_r^\beta:=\min\left\{t:L_0e^{(b-a)t}=\frac{b(-u_L)-c}{c+(1-\beta)ae^{-at}\bar{\triangle}_H}\right\}.$$

and

$$\hat{T}_l^\beta:=\min\left\{t:L_0e^{(b-a)t}=\frac{-u_L}{e^{-aT}u_H-(1-\beta)\left(e^{-at}\right)\bar{\triangle}_H}\right\}.$$

That is, she uses the Immediate action S Strategy and the Randomised Stopping Time Strategy with probability $\beta \in (0,1)$ and $1-\beta$ such that $\hat{T}^{\beta}$ is the time at which player j starts randomising conditional on she uses the Randomised Stopping Time Strategy.

Given the assumption that $\rho^{ML}(t) = 0$ for $t \leq \hat{T}^{\beta}$ and $\rho^{ML}(t) > 0$ for $t > \hat{T}^{\beta}$, consider time $\hat{T}^{\beta}$ as a new time 0 and denote the time line starting from $\hat{T}^{\beta}$ as τ. Let $\tau := t - \hat{T}^{\beta}$ and $\lambda^{ML}(\tau) := \rho^{ML}\left(\tau + \hat{T}\right)$. Let $q_{\tau}^{ML} := p_{\tau+\hat{T}}$ be the belief and let $L_{\tau}^{ML} := \frac{q_{\tau}^{ML}}{1-q_{\tau}^{ML}}$ be the likelihood ratio. Following the same discussion as in step 5, at time $t \geq \hat{T}^{\beta}$, that is, $\tau \geq 0$, player i is indifferent between acquiring information and taking action R if

$$\begin{aligned}&\frac{d\lambda^{ML}(\tau)}{d\tau} = \\ &\frac{b(-u_L) - c - L_{\tau}^{ML}c + b\bar{\triangle}_L \int_0^{\tau} e^{-bs}\frac{d\lambda^{ML}(s)}{ds}ds - L_{\tau}^{ML}\bar{\triangle}_H a e^{-a\tau}\left(1-\lambda^{ML}(t)\right)}{L_{\tau}^{ML}\left(\eta^{ML} - \left(u_H - \bar{\triangle}_H\right)\right)e^{-a\tau} + \bar{\triangle}_L e^{-b\tau}}\end{aligned} \tag{1.22}$$

where $\eta^{ML} := u_H - (1-\beta)\left(1-e^{-a\hat{T}}\right)\bar{\triangle}_H$. The intuition is that when considering time $\hat{T}^{\beta}$ as a new time 0, player i's problem is essentially the same as in step 5 with η^{ML} being the payoff for the first action R taker instead of η. The existence of an increasing function $\rho^{ML}(\cdot)$ can be shown following the same logic as in the proof of Lemma 1.7.

Next, I characterise the upperbound and lowerbound of $\hat{T}^{\beta}$. Following similar argument as in step 5, the upperbound $\hat{T}_r^{\beta}$ is the first time the marginal cost of acquiring information exceeds the marginal benefit. At time $0 \leq t < \hat{T}^{\beta}$, given player j's strategy, player i's payoff associated with taking action R at time t is

$$\begin{aligned}U_R^{\beta}(t) =& \beta\left[p_t u_H + (1-p_t)u_L\right] \\ &+ (1-\beta)\left[p_t\left(u_H - \left(1-e^{-at}\right)\bar{\triangle}_H\right) + (1-p_t)u_L\right].\end{aligned}$$

That is,

$$U_R^{\beta}(t) = p_t\left[\beta u_H + (1-\beta)\left(u_H - \left(1-e^{-at}\right)\bar{\triangle}_H\right)\right] + (1-p_t)u_L.$$

The marginal cost of acquiring information is smaller than the marginal benefit if

$$\frac{b\left(-u_L\right)-c}{c+(1-\beta)\, ae^{-at}\bar{\triangle}_H} > L_0 e^{(b-a)t}$$

Then, if $L_0 < \frac{b(-u_L)-c}{c+(1-\beta)a\bar{\triangle}_H}$, there exists a $\bar{T}_r^\beta > 0$ such that

$$\hat{T}_r^\beta := \min\left\{t : L_0 e^{(b-a)t} = \frac{b\left(-u_L\right)-c}{c+(1-\beta)\, ae^{-at}\bar{\triangle}_H}\right\}.$$

The lowerbound of $\hat{T}_l^\beta$ is the earliest time point at which the player is willing to start randomising instead of taking action S given player j's strategy. Following similar argument as in step 5, if $p_0 < p^L$, then, there exists a $\bar{T}_l^\beta > 0$ such that

$$\hat{T}_l^\beta := \min\left\{t : L_0 e^{(b-a)t} = \frac{-u_L}{e^{-aT}u_H - (1-\beta)\left(e^{-at}\right)\bar{\triangle}_H}\right\}.$$

The following lemma characterises the equilibrium when $\underline{p} < p_0 < p^*$.
tbc

Lemma 1.16. *Suppose $b < 2a$ and c is sufficiently small. If $\underline{p} < p_0 < p^*$, then, there exists an equilibrium where both players use the Mixed Learning Strategy $\left(\rho^{ML}, \sigma^{ML}\right)$ such that*

1. *$\sigma^{ML}(t) = \beta > 0$ for $\forall t \geq 0$.*

2. *$\rho^{ML}(t) = 0$ for $t < \hat{T}^\beta$.*

3. *$\rho^{ML}(t) \in (0, 1-\beta]$ for $t \in \left(\hat{T}^\beta, \bar{T}^\beta\right]$ and $\lambda^{ML}(\tau) = \rho^{ML}\left(\tau + \hat{T}\right)$ is a solution to the differential equation (1.22) with initial condition $\lambda^{ML}(0) = 0$ where $\frac{d\rho^{ML}}{dt} > 0$ for $\forall t \in \left[\hat{T}^\beta, \bar{T}^\beta\right]$.*

4. *$\rho^{ML}(t) = 1-\beta$ for $\forall t > \bar{T}^\beta$.*

5. *$\hat{T}^\beta \in \left[\hat{T}_l^\beta, \hat{T}_r^\beta\right]$.*

Proof. Suppose player j uses the Mixed Learning Strategy $\left(\rho^{ML}, \sigma^{ML}\right)$.

I have shown that at time $t \geq \hat{T}^\beta$, player i is indifferent between taking action R and acquiring information for dt longer and that at time $0 < t < \hat{T}^\beta$, player i prefers acquiring information to taking action R. What left is to show that if $\underline{p} < p_0 < \underline{p}^*$, there exists a $\beta \in (0,1)$ such that player i is indifferent between the Immediate action S Strategy and the Randomised Stopping Time Strategy at time 0. Given Lemma 1.13, I show that such $\beta \in (0,1)$ exists if $\underline{p} < p_0 < \underline{p}^*(p_0)$.

At time $t < \hat{T}^{\beta}$, player i's value associated with acquiring information is

$$\begin{aligned} W^{\beta}(t) =& p_t \left[u_H - (1-\beta)\bar{\triangle}_H + \frac{1}{2}(1-\beta)\bar{\triangle}_H e^{-at} - \frac{c}{a} \right] \\ & - (1-p_t)\frac{c}{b} + p_t e^{at} H\left(\hat{T}^{\beta}\right) \end{aligned}$$

where

$$H\left(\hat{T}^{\beta}\right) = \left[\frac{1}{2}(1-\beta)\bar{\triangle}_H e^{-a\hat{T}^{\beta}} + \frac{c}{a} - \frac{1}{L_{\hat{T}^{\beta}}}\left(-u_L - \frac{c}{b}\right) \right] e^{-a\hat{T}^{\beta}}$$

is a constant. The value of acquiring the information at time $t = 0$ equals the value associated with taking action S without acquiring the signal if $W^{\beta}(0) = 0$. That is,

$$L_0 = \frac{\frac{c}{b}}{u_H - \frac{1}{2}\bar{\triangle}_H + \frac{1}{2}\beta\bar{\triangle}_H - \frac{c}{a} + H\left(\hat{T}^{\beta}\right)} := \underline{L}^{\beta}\left(\hat{T}^{\beta}\right).$$

Then, β can be pinned down by $L_0 = \underline{L}^{\beta}\left(\hat{T}^{\beta}\right)$.

Lemma 1.17. *When $\underline{p} < p_0 < \underline{p}^*(p_0)$, there exists a $\beta \in (0,1)$ such that $L_0 = \underline{L}^{\beta}\left(\hat{T}_r^{\beta}\right)$.*

Proof. The proof uses the intermediate value theorem. When $\beta = 0$, we have $\underline{L}^{\beta=0}\left(\hat{T}_r^{\beta=0}\right) = \underline{L}\left(\hat{T}_r\right)$. Since $p_0 < \underline{p}^*(p_0)$, we have $\underline{L}^{\beta=0}\left(\hat{T}_r^{\beta=0}\right) = \underline{L}\left(\hat{T}_r\right) > L_0$.

Next, I show that when $\beta = 1$, we have $\underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right) < \underline{L}$. When $\beta = 1$, $\hat{T}_r^{\beta=1}$ satisfies

$$L_0 e^{(b-a)\hat{T}_r^{\beta=1}} = \frac{b(-u_L) - c}{c} = \bar{L}.$$

Then,

$$\underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right) = \frac{\frac{c}{b}}{u_H - \frac{c}{a} + H\left(\hat{T}_r^{\beta=1}\right)}$$

where $H\left(\hat{T}_r^{\beta=1}\right) = \left(\frac{c}{a} - \frac{c}{b}\right) e^{-a\hat{T}_r^{\beta}}$. Then, we have

$$\underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right)\left[u_H - \frac{c}{a}\right] + \left(\frac{c}{a} - \frac{c}{b}\right) e^{-a\hat{T}_r^{\beta}} \underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right) = \frac{c}{b}.$$

Since $L_0 e^{(b-a)\hat{T}_r^{\beta=1}} = \bar{L}$, from the proof of proposition 1.1 we know that

$$\underline{L}\left[u_H - \frac{c}{a}\right] + \left(\frac{c}{a} - \frac{c}{b}\right) e^{-a\hat{T}_r^{\beta}} \left(\frac{\underline{L}}{L_0}\right)^{\frac{a}{b-a}} \underline{L} = \frac{c}{b}.$$

Since $\underline{p} < p_0$, we have $\underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right) < \underline{L}$ and thus $\underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right) < L_0$.

Since $\underline{L}^{\beta=1}\left(\hat{T}_r^{\beta=1}\right) < L_0 < \underline{L}^{\beta=0}\left(\hat{T}_r^{\beta=0}\right)$ there exists a $\beta \in (0,1)$ such that $L_0 = \underline{L}^{\beta}\left(\hat{T}_r^{\beta}\right)$. □

Given Lemma 1.17, we know that when $\underline{p} < p_0 < \underline{p}^*(p_0)$, there exists at least a pair $(\beta \in (0,1), \hat{T}_r^{\beta})$ such that the players play the Immediate action S strategy with probability β and player the Randomised Stopping Time Strategy with probability $1-\beta$. □

5 Proof of Lemma 1.2 and Proposition 1.2

By definition, $\frac{p^M}{1-p^M} := \frac{-\left(u_L - \underline{\triangle}_L\right)}{u_H - \underline{\triangle}_H}$ and $\frac{\bar{p}}{1-\bar{p}} = \frac{-u_L - \frac{c}{b}}{\frac{c}{b}}$. When $\frac{c}{b} < \frac{-u_L\left(u_H - \underline{\triangle}_H\right)}{u_H - \underline{\triangle}_H - \left(u_L - \underline{\triangle}_L\right)}$, we have $p^M < \bar{p}$. Lemma 1.2 is shown.

The cutoff $\underline{p}$ satisfies

$$\left[u_H - \frac{c}{a}\right]\frac{\underline{p}}{1-\underline{p}} + K\left(\frac{\underline{p}}{1-\underline{p}}\right)^{\frac{b}{b-a}} = \frac{c}{b} \tag{1.23}$$

where $K = \frac{c}{b}\left(\frac{b}{a} - 1\right)\left(\frac{\bar{p}}{1-\bar{p}}\right)^{1-\frac{b}{b-a}}$. When $c \to 0$, $\bar{p} \to 1$ and $\underline{p} \to 0$.

When $c \to 0$, the cutoff $\frac{\tilde{p}}{1-\tilde{p}} = \frac{-bu_L - c}{a\bar{\triangle}_H + c} \to \frac{-bu_L}{a\bar{\triangle}_H}$ and hence $\tilde{p} \to \frac{\frac{b}{a}\frac{-u_L}{\bar{\triangle}_H}}{1 + \frac{b}{a}\frac{-u_L}{\bar{\triangle}_H}} \in (0,1)$.

6 Proof of Proposition 1.3

Suppose player j uses the pure strategy $\left(T^{PS}, S\right)$ where $T^{PS} = \frac{1}{a}\log\frac{\bar{\triangle}_H}{-\left(u_H - \bar{\triangle}_H\right)}$. I show that when the conditions in Proposition 1.3 are satisfied, player i's best respond is to use the same pure strategy. The method is 'guess and verify'. Given that player j uses the pure strategy $\left(T^{PS}, S\right)$, player i's time t payoff from taking R is

$$U_R(t) := \begin{cases} p_t\left[u_H - \left(1 - e^{-at}\right)\bar{\triangle}_H\right] + (1 - p_t)\,u_L & t < T^{PS} \\ p_t\left[u_H - \left(1 - e^{-aT^{PS}}\right)\bar{\triangle}_H\right] + (1 - p_t)\,u_L & t \geq T^{PS} \end{cases}.$$

Let

$$U(t) := \max\{U_R(t), 0\}$$

be player i's time t payoff if she stops and takes an irreversible action. Let $W(\cdot)$ be the value function associated with player i's best response. Then, the value function satisfies the following HJB equation

$$\max\left\{p_t a\left[\max\left\{u_H-\left(1-e^{-at}\right)\bar{\triangle}_H,0\right\}-W(t)\right]-(1-p_t)bW(t)+W'(t)-c,\right.$$
$$\left.U(t)-W(t)\right\}=0. \tag{1.24}$$

Let

$$\frac{p^{NR}}{1-p^{NR}}:=\frac{\frac{c}{b}}{\frac{1}{2}\left(2u_H+\bar{\triangle}_H\right)-\frac{c}{a}-u^S+\psi\left(T^{PS}\right)}$$

where

$$\psi(t):=\left[\frac{1}{2}e^{-at}\bar{\triangle}_H+\frac{c}{a}-\left(-u_L-\frac{c}{b}\right)\frac{1-p_t}{p_t}\right]e^{-at}$$

I prove the following lemma.

Lemma 1.18. *When c sufficiently small, if* $p_0\in\left(p^{NR},\tilde{p}\right)$*, the value function* $W(\cdot)$ *is*

$$W(t)=\begin{cases}W_L(t) & t<T^{PS}\\ 0 & t\geq T^{PS}\end{cases}$$

where

$$W_L(t)=p_t\left(\left(u_H^R-\bar{\triangle}_H\right)-\frac{c}{a}+\frac{1}{2}\bar{\triangle}_He^{-at}\right)-(1-p_t)\frac{c}{b}+p_te^{at}\psi\left(T^{PS}\right).$$

Proof. Let

$$H\left(t,W(t),W'(t)\right):=p_ta\left[\max\left\{u_H-\left(1-e^{-at}\right)\bar{\triangle}_H,0\right\}-W(t)\right]$$
$$-(1-p_t)bW(t)+W'(t)-c.$$

I show that $W(t)$ is a viscosity solution of the HJB equation (1.24). For all the points where $W(t)$ is differentiable: if $t<T^{PS}$, then $W(t)\geq U(t)$; if $t>T^{PS}$, then $H(t,W(t),W'(t))\leq 0$. At the point where $W(t)$ is not differentiable, that is, $t=T^{PS}$, I show that $H(t,W(t),z)\geq 0$ for $z\in D^+$ where $D^+=\left[W_L'\left(T^{PS}\right),0\right]$ and $H(t,W(t),z)\leq 0$ for $z\in D^-$ where $D^-=\emptyset$.

Step 1 For $t < T^{PS}$,

$$\begin{aligned} W(t) - U(t) &= W_L(t) - U_R(t) \\ &= p_t \left[-\frac{1}{2} \left(\bar{\triangle}_H \right) e^{-at} + e^{at} \psi \left(T^{PS} \right) - \frac{c}{a} \right] + (1 - p_t) \left(-u_L - \frac{c}{b} \right). \end{aligned}$$

When $p_0 < \tilde{p}$, $\psi(t)$ increases in t. Then,

$$W(t) - U(t) > p_t \left[-\frac{1}{2} \left(\bar{\triangle}_H \right) e^{-at} + e^{at} \psi(t) - \frac{c}{a} \right] + (1 - p_t) \left(-u_L - \frac{c}{b} \right) = 0$$

Step 2 When $t > T^{PS}$, $W(t) = 0$ and $W'(t) = 0$. Therefore, $H(t, W(t), W'(t)) = -c < 0$.

Step 3 I first show that for sufficiently small c, $W_L' \left(T^{PS} \right) < 0$. We have

$$\begin{aligned} W_L'(t) = &\frac{dp_t}{dt} \left[u_H - \bar{\triangle}_H - \frac{c}{a} + \frac{1}{2} \bar{\triangle}_H e^{-at} \right] - p_t \frac{1}{2} a \bar{\triangle}_H e^{-at} \\ &+ \frac{dp_t}{dt} \frac{c}{b} + \frac{dp_t}{dt} e^{at} \psi \left(T^{PS} \right) + p_t a e^{at} \psi \left(T^{PS} \right) \end{aligned}$$

and if $\frac{c}{a} < \left(-u_L - \frac{c}{b} \right) \frac{1 - p_{T^{PS}}}{p_{T^{PS}}}$,

$$\begin{aligned} &W_L' \left(T^{PS} \right) = \\ &\frac{dp_t}{dt} |_{t = T^{PS}} \left[\frac{c}{b} - \left(-u_L - \frac{c}{b} \right) \frac{1 - p_{T^{PS}}}{p_{T^{PS}}} \right] + p_{T^{PS}} a \left[\frac{c}{a} - \left(u^S - u_L - \frac{c}{b} \right) \frac{1 - p_{T^{PS}}}{p_{T^{PS}}} \right] \\ &< 0. \end{aligned}$$

When $t = T^{PS}$, $H(t, W(t), W'(t)) = 0$. Since $H(t, W(t), z)$ increases in z, we have $H(t, W(t), z) \geq 0$ for $z \in D^+$.

□

7 Multi-player Extension: Proof of Proposition 1.4

In this extension, I generalise the two-player model to a multi-player model. There are $N > 2$ players in this model. The first R taker gets the first prize u_ω in state $\omega \in \{H, L\}$ and all other R takers get the second prize $u_\omega - \bar{\triangle}_\omega$. In case of the simultaneous move, the payoff is assumed to be the convex combination of the first and second prizes. I focus on discussing the conditions for the existence of the learning equilibrium where the players use the random stopping strategy.

Suppose the players use the mixed strategy (ρ, σ) defined as in Definition 1.2, where ρ is the probability that the player stops and takes R before or at time t conditional on no revealing signal. By definition, ρ must be weakly increasing. I derive a necessary condition for $\frac{d\rho}{dt} \geq 0$. Let $Q_\omega(t)$ be the probability that no one has taken R before or at time t in state ω. Since I focus on symmetric equilibrium, we have

$$Q_\omega(t) = (1 - F_\omega(t))^{N-1} \tag{1.25}$$

where $F_\omega(t)$ is the probability that a player takes R before or at time t in state ω. We have

$$F_H(t) = 1 - e^{-at}(1 - \rho(t)) \tag{1.26}$$

and

$$F_L(t) = \int_0^t \left[e^{-bs} \frac{d\rho(s)}{ds} \right] ds. \tag{1.27}$$

Let

$$U_R(t) := p_t \left[u_H - (1 - Q_H(t)) \bar{\triangle}_H \right] + (1 - p_t) \left[u_L - (1 - Q_L(t)) \bar{\triangle}_L \right]$$

be the time t payoff from taking R. When the players randomise between taking R and acquiring information for dt longer, the indifference condition in equilibrium is

$$p_t a dt \left[u_H - (1 - Q_H(t + dt)) \bar{\triangle}_H \right] + (1 - p_t a dt - (1 - p_t) b dt) U_R(t + dt) - c dt = U_R(t).$$

When $dt \to 0$, we have the following indifference condition

$$c = (1 - p_t) b \left[- \left(u_L - (1 - Q_L(t)) \bar{\triangle}_L \right) \right] + (1 - p_t) \frac{dQ_L}{dt}(t) \bar{\triangle}_L + p_t \frac{dQ_H}{dt}(t) \bar{\triangle}_H. \tag{1.28}$$

Given (1.25), (1.26) and (1.27), (1.28) can be written as a differential equation that involves ρ and its derivatives. That is,

$$\frac{d\rho}{dt}(t) = \frac{b(-u_L) - c - L_t c - b\bar{\triangle}_L (1 - F_L(t))^{N-1} - L_t (N-1) \bar{\triangle}_H (1 - F_H(t))^{N-2} a e^{-at} (1 - \rho(t))}{(N-1) \bar{\triangle}_L (1 - F_L(t))^{N-2} e^{-bt} + L_t (N-1) \bar{\triangle}_H (1 - F_H(t))^{N-2} e^{-at}}.$$

To have $\frac{d\rho}{dt} \geq 0$, we need the numerator to be positive (as the denominator is positive). A necessary condition for $\frac{d\rho}{dt} \geq 0$ is

$$\frac{p_0}{1-p_0} \leq \frac{b\left(-u_L\right)-c}{\left(N-1\right)\bar{\triangle}_H a+c} =: \frac{\tilde{p}_N}{1-\tilde{p}_N}.$$

This shows Proposition 1.4.

8 Observable Actions: Proof of Proposition 1.5

When actions are observable, the history contains both the public component and the private component. The public component is the action taken or not taken by the opponent and the private component is the signal received by the player herself. The following observation eliminates the histories that are not interesting.

Observation 1.1. *After receiving an H-state (L-state, resp) revealing signal, the player takes R (S, resp) immediately.*

This observation says that after receiving a revealing signal, the player does not have incentives to postpone the action. Specifically, she does not have incentive to conceal the fact that she has learned the state even though she knows that her action is informative. This is because there is the first-mover advantage. The player gains nothing from postponing an action after she has learned the state. Given this observation, the interesting histories are the ones associated with no revealing signal. It is thus sufficient to check the player's strategy conditional on no arrival of a revealing signal.

Suppose the opponent uses MRSS. I first consider the history after observing the opponent taking S. Given the opponent's strategy, she only stops acquiring information and takes S after observing the L-state revealing signal. Therefore, after observing the opponent taking S, the player infers that the state is L. Her best response is hence to take S immediately.

Then consider the history after observing no action taken. First notice that the player's belief evolve in a different way. Given the opponent's strategy, not observing any action taken indicates that the opponent has not received any revealing signal. As a result, the belief after observing no revealing signal and no action taken up to time t is

$$\frac{p_t}{1-p_t} = e^{2(b-a)t}\frac{p_0}{1-p_0}.$$

Let

$$\bar{U}_R(t) := p_t u_H + (1-p_t)\, u_L$$

be the player's time t payoff from taking R if she is the first R taker and let

$$\underline{U}_R(t) := p_t \left(u_H - \bar{\triangle}_H\right) + (1-p_t)\left(u_L - \bar{\triangle}_L\right)$$

be the player's time t payoff from taking R if she is the second R taker. Since $p_0 > p^L$, we have $p_t > p^L$ and hence $\bar{U}_R(t) > 0$. The player is indifferent between taking R and acquiring information for dt longer if

$$\begin{aligned}
&p_t \left\{adt \left[u_H + (1-(1-adt)(1-h(t)\,dt))\,\bar{\triangle}_H\right]\right.\\
&\left.+ (1-adt)\left[(1-adt)(1-h(t)\,dt)\,\bar{U}_R(t+dt) + (1-(1-adt)(1-h(t)\,dt))\,\underline{U}_R(t+dt)\right]\right\}\\
&+ (1-p_t)\left\{(1-bdt)\left[(bdt + (1-bdt)(1-h(t)\,dt))\,\bar{U}_R(t+dt) + (1-bdt)\,h(t)\,dt\underline{U}_R(t+dt)\right]\right\}\\
&- cdt\\
&= \bar{U}_R(t)\,.
\end{aligned}$$

When $dt \to 0$, the equation above is equivalent to

$$\begin{aligned}
&p_t a\left(u_H - \bar{U}_R(t)\right) + (1-p_t)\, b\left(-\bar{U}_R(t)\right) - p_t a\left(p_t\bar{\triangle}_H + (1-p_t)\,\bar{\triangle}_L\right)\\
&- h(t)\left(p_t\bar{\triangle}_H + (1-p_t)\,\bar{\triangle}_L\right) + \frac{d\bar{U}_R(t)}{dt} = c.
\end{aligned}$$

The hazard rate

$$h(t) = \frac{b(-u_L) - c - \frac{p_t}{1-p_t}\left[a\left(p_t\left(\bar{\triangle}_H - \bar{\triangle}_L\right) + \bar{\triangle}_L\right) + c\right]}{\frac{p_t}{1-p_t}\bar{\triangle}_H + \bar{\triangle}_L}$$

is positive if and only if

$$\frac{p_t}{1-p_t} < \frac{b(-u_L) - c}{a\left[p_t\left(\bar{\triangle}_H - \bar{\triangle}_L\right) + \bar{\triangle}_L\right] + c}.$$

Since Proposition 1.5 assumes $\bar{\triangle}_H = \bar{\triangle}_L$ and $p_0 < \tilde{p}$, the hazard rate $h(t) > 0$.

Last, consider the history after the opponent taking R. Let p_t^R be the player's belief after observing the opponent taking R at time t. Since the opponent takes R with positive rate after no revealing signal, the player's belief p_t^R is smaller than one. After the opponent has taken R, the player becomes the only player in this game. The payoff associated with R now is $u_\omega - \bar{\triangle}_\omega$ in state $\omega \in \{H, L\}$. The player's best response is to use the single DM optimal strategy.

Chapter 2

Linear Search or Binary Search

1 Introduction

Pooled testing is a method combining the same type of specimen from several people and conducting one laboratory test on the combined pool of specimens to detect viruses. This method is used to screen SARS-CoV-2 (the virus that causes COVID-19) infections in the community. If a pooled test result is negative then all the specimens can be presumed negative with the single test. If it is positive, each of the specimens in the pool will need to be tested individually to determine which specimens are positive. Suppose there are sixteen specimens waiting to be tested and one of them is positive. There are multiple ways to conduct the tests in order to search for the positive specimen. I will call it *Binary Search* if the pooled testing with a pool size of eight (that is, half of the specimens) is used and *Linear Search* if the specimens are tested individually.[1]

When people only have access to binary signals, a single round of search normally is not sufficient. Learning the truth takes time. Learning immediately is possible but relies on luck. In the example, the laboratory tests only deliver positive or negative results. To find the positive specimen, after a pooled test with a pool size larger than one, follow-up tests are certainly needed. A test with the pool size of one could find the positive specimen immediately only if it coincidentally tests the positive specimen. In most circumstances, a sequence of tests is needed. I refer to the sequence of tests the *test protocol.*

Any sensible test protocol allows the learner to find the positive specimen. The main difference between test protocols is in terms of the riskiness involved. That is, how and when the positive specimen is found. The test protocol con-

[1] Linear Search and Binary Search are standard terminologies in Computer Science literature.

sisting of a sequence of linear searches allows the learner to learn immediately, but it is risky and can be slow. When this linear search protocol is used, the specimens are tested individually and sequentially. The positive specimen can be found with a minimum of one and a maximum of fifteen tests. The test protocol consisting of a sequence of binary searches is faster and safer, but it is impossible to learn immediately. When this binary search protocol is used, the positive specimen can be found for certain after four tests.[2] The possibility of learning immediately induced by the linear search protocol could be useful in an inpatient clinical setting where isolating the positive case immediately can minimise the risk of hospital transmission. The guaranteed learning after a fixed number of tests can be beneficial when the test kits have to be pre-ordered or when the result release date has to be announced in advance.

The main purpose of this paper is to understand the factors that affect the learner's choice of the test protocols. The learner will learn the truth eventually and the question of interest is how the learner acquires pieces of information to learn the truth over time. Given the structure of the tests, some test protocols allow the learner to acquire information and learn the truth in a safe and steady way, while some test protocols surprise the learner with each piece of information which allows the learner to learn immediately. The learner faces the tradeoff between *possibly learning today but more likely nothing* and *no learning today but learning for certain in the near future.*

The learner's preferences, in particular, the patience level and her *time* risk attitude are the key factors that determine the choices of the test protocols. If a learner is impatient, then, she appreciates the test protocols that allow her to learn today. The *time* risk attitude is the learner's risk attitude towards the potentially uncertain learning time. To understand the time risk attitude, suppose the learner can choose between two 'lotteries'. Lottery one is a test protocol that allows her to learn today or the day after tomorrow with equal probabilities and lottery two allows her to learn tomorrow for certain.[3] If learning the truth gives the learner a monetary payment of unity, then, these two lotteries generate the same expected payment. The difference between the two lotteries is in terms of the risk involved. Lottery one is riskier as the learner may learn

[2] The first pooled testing has a pool size of eight. Eight random specimens will be tested together in one pooled test. If the result is positive, four random specimens from these eight specimens will be tested again in the second pooled test. If the result is negative, the second pooled testing will be conducted on the untested specimens in the first round. The first pooled test eliminates eight negative specimens, the second pooled test eliminates four negative specimens, the third test eliminates two and the fourth test pins down the positive specimen.

[3] In my model, the learner never chooses between these two lotteries because of the structure of the tests. These two lotteries are just for illustration purposes

the truth today or the day after tomorrow while lottery two is safe. Given that both lotteries generate the same expected payment, a learner is *time risk averse* if she prefers lottery two and is *time risk seeking* if she prefers lottery one. A test protocol with different pool sizes induces a ‘lottery’ of learning time. A time risk averse learner would appreciate a test protocol that allows her to learn at some time for certain.

To model both time risk attitude and the patience level, the standard lifetime utility with exponential discounting is not sufficient. It automatically describes a time risk seeking preference whenever the discount factor is smaller than one. This can be seen by evaluating the lifetime utility associated with the two lotteries introduced earlier. Suppose learning the truth gives the learner a monetary payment of unity and the learner has a discount factor $\delta \in (0, 1)$. Then, the utility associated with lottery one is $.5 + .5\delta^2$ and the utility associated with lottery two is δ. The utility from lottery one is always greater than that from lottery two if δ is between zero and one. In other words, the lifetime utility with exponential discounting implicitly describes a time risk loving preference. This time risk loving attitude is a side product of assuming exponential discounting. Because of this, the standard utility function with exponential discounting does not allow us to isolate the effect of the time risk attitude and the patience level on the learner’s optimal choice.

To disentangle the effects of the time risk attitude and the patience level, I discuss the main tradeoff between learning safely and possibly learning today with a single-agent model where the agent’s utility is modelled by a generalised discounted utility function. This utility function introduced in DeJarnette et al. (2020) allows for different time risk attitudes as well as patience levels. The agent’s goal is to learn the true value of a parameter, which is drawn from a finite parameter set. The agent can learn the true value of a parameter by doing a sequence of tests. At each time, the agent can choose any number of elements from the parameter set and then conduct a single test to check if the true value of the parameter is among the elements she has chosen. The test result is binary and truthful. Given the true value of the parameter and the chosen elements, the agent receives a positive or negative test result. The positive result indicates that the true value of the parameter is among one of the elements she has chosen while the negative result indicates the opposite. The test result cannot indicate which element is the true value of the parameter unless the agent only picks up one element.

The overall informativeness of a test protocol is independent of the use of Linear or Binary Search. Any sensible test protocol allows the learner to learn the truth as the positive specimen will be found eventually. However, the informativeness of one round of Linear Search and Binary Search are not the same. In the example, the test result is a binary signal. The informativeness of a negative or positive signal, and the probability of receiving a negative or positive signal, varies with the pool size. A positive signal of the individual test, that is, a pooled test with the pool size of one, is very informative as it pins down the positive specimen immediately. But the probability of receiving this informative signal is low. It is more likely to get a negative signal that does not contain much information. For a pooled test with a pool size of eight, both the negative and positive signals have the same informativeness level and are equally likely to be received. They are equally informative in the sense that they both eliminate half of the negative specimens, but neither test result is informative enough to pin down the positive specimen. At each time the learner chooses the pool size of a test, she essentially chooses both the informativeness of the signals and the ex-ante probability of receiving each signal.

This paper shows that given the uniform prior, only two test protocols can be optimal: the Linear Search test protocol or the Binary Search protocol. All other test protocols are suboptimal. Whether the Linear Search or the Binary Search is optimal depends on the agent's time risk attitude and patience level. If the agent is time risk averse, then, the Binary search protocol is always optimal. If the agent is time risk seeking and sufficiently impatient, then, the Linear Search protocol is optimal.

This paper is closely related to Zhong (2022). Zhong (2022) discusses a general dynamic information acquisition framework where the agent can choose any information structure subject to a flow cost constraint at each time. Zhong (2022) shows that when the agent is an expected utility maximiser who discounts the future payoffs exponentially, the Poisson signal is the optimal signal structure. This is because the Poisson signal generates the learning time with the highest variance compared to other signal structures. Since the agent discounts the future payoffs exponentially, she has the risk-loving time preference, and hence prefers the signal structure that induces a high variance of learning time. The other paper Zhong (2017) explicitly discusses the relationship between the optimal information acquisition and the agent's time preference. Zhong (2017) shows that subject to a flow cost constraint, any information structure induces the same expected learning time, and the Poisson signal induces the largest variance. The agent with risk-loving time preference hence prefers the Poisson signal. I focus on comparing how an agent obtains information when she has different preferences given a specific class of information structures, while Zhong's papers focus on discussing the agent's optimal information choices in a general framework.

The signal structure in this paper tells the agent the learning direction. It can be imagined as the indicator at the intersection of two roads that tells the agent which direction to go in order to find the target. This shares some similar features as in Callander (2011). In Callander (2011), the agent wants to learn about a mapping from the choices to the outcomes. The mapping is modelled as a realised Brownian motion, where the agent knows the parameters that characterise the Brownian motion, but does not know its realisation. In order to learn about the realised Brownian motion, the agent can observe the alternative and outcome pairs that has been chosen by his predecessors and then choose an alternative to learn its outcome. Because of the property of the Brownian motion, when the agent observes the alternative and outcome pairs of his predecessors, if the previous outcomes are not of satisfaction, the agent learns the direction to search for the next alternative. This has the similar features as the signal structure in my paper, but it is different. In Callander (2011), when the agent chooses an alternative, the agent learns the outcome associated with that alternative. However, in my paper, the signal itself does not allow the agent to learn the unknown parameter directly. The signal only serves as an indicator that tells the agent the learning direction. Learning happens when there is only one candidate left towards that direction. Learning can be considered as indirect in this sense.

This paper is related to the literature that discusses the preferences on time lotteries. This is because the final payoff associated with learning can be considered as a time lottery. Given a search method, the agent faces a time lottery that gives him a payoff of one at a random time. Dillenberger et al. (2018) and DeJarnette et al. (2020) discuss the preferences on time lotteries. One of the ideas in those papers is that the commonly used expected utility with exponential discounting describes a risk-loving time preference. In order to characterise other time risk attitude, a generalised expected utility should be considered. My paper can be considered as an application of the generalised expected utility to the learning and searching problem.

This paper is also related to the literature that discusses the timing of resolution of uncertainty. If the agent always uses Binary Search, then the uncertainty about the timing of learning is resolved, and hence it can be regarded as an early resolution of uncertainty. This can be shown from the example in the introduction. If the employer always uses Binary Search, she knows that she will learn the type of the employee at the second time period. However, if the agent always uses Linear Search, then the timing of learning the unknown remains uncertain, and it thus can be regarded as a late resolution of uncertainty. There is a group of literature discussing the preferences on early and late resolution of uncertainty, including Epstein & Zin (1989), Kreps & Porteus (1978), Dillenberger (2010) and Palacios-Huerta (1999). This paper can be considered as an application of the preferences on the resolution of uncertainty to the learning and searching problem.

The last group of the literature is the computer science literature on Linear and Binary Search. In computer science, Linear Search and Binary Search are algorithms to find the position of a target value. The details of these algorithms can be found in Knuth (1998). The recent computer science papers including Kumari (2012), Mehta et al. (2015), and Rahim et al. (2017) compare the Linear Search and Binary Search algorithms in different situations. While the computer science literature focuses on comparing the speed and the complexity of the search algorithms, this paper focuses on discussing how people's time preferences affect their optimal choices of the search algorithms. Without discussing which search algorithm allows the agent to learn faster, this paper discusses how agent's preferences, in particular, agent's patient level and the risk attitude, affect their choice of the search algorithms.

2 The Model

This section first introduces the model setup and then discusses the assumptions of the model.

2.1 Setup

There is a single decision-maker in this model. I call her the *agent*. The agent wants to learn the true value of a parameter $\theta \in \Theta$, where Θ is finite with the cardinality $\bar{N}$. I assume that the set Θ is sorted such that $\theta_1 < \theta_2 < \cdots < \theta_{\bar{N}}$. The parameter θ is drawn from a distribution F with the probability mass function f at the beginning and it is fixed over time. Time $t = 0, 1, \ldots, T$ is discrete and finite. The final period T is greater than $\bar{N} - 2$. The agent's action at each time t is to choose an element $r_t \in \Theta$ to test.

The test at time t has two outcomes: pass or fail. The element chosen fails the test if it is greater than the true value, and it passes otherwise. By choosing the test at time t, the agent effectively chooses the information structure as described below. The information structure at each time t consists of a binary signal $s_t \in \{0, 1\}$ and a probability distribution over the signals. The 0 signal is the fail signal, while the 1 signal is the pass signal. Conditional on the true value of the parameter and the test chosen, the probabilities of receiving the signals are

$$\Pr(s_t = 0 \mid \theta < r_t) = \Pr(s_t = 1 \mid \theta \geq r_t) = 1.$$

A pair (r_t, s_t) is the agent's action and the signal received at time t. The agent remembers her past actions and the past signals received. Let $r^t = \{r_0, r_1, \dots, r_{t-1}\}$ be the sequence of the past actions, and let $s^t = \{s_0, s_1, \dots, s_{t-1}\}$ be the sequence of the past signals. At the beginning of time t, $I_t = \{r^t, s^t\}$ denotes the history up to that point. The set of histories is denoted by $\mathcal{I} = \bigcup_{t=1,2,\dots,T} I_t \cup \emptyset$. The strategy of the agent is given by a mapping $\mathcal{R} : \mathcal{I} \to \Theta$ from the histories to the test choices.

The agent has the prior belief $f_0 = \big(f_0(\theta)\big)_{\theta\in\Theta}$ with $f_0(\theta) \in (0,1)$ for $\forall\theta \in \Theta$. At the start of time t, the agent has the belief f_t. After choosing the action r_t, the agent receives the signal s_t and updates the belief to f_{t+1} using Bayes rule. I assume the agent has a uniform prior belief. That is $f_0(\theta) = \frac{1}{N}$ for $\forall\theta \in \Theta$.

At each time t, the agent either learns or fails to learn the true value of the parameter. If she learns the true value of the parameter at time t, then she gets a reward $x_t = x > 0$, and the game ends. If she fails to learn the true value of the parameter at time t, she gets $x_t = 0$ and the game enters time $t+1$. Let $u : \mathbb{R} \to \mathbb{R}_+ \cup \{0\}$ be the utility function that maps from the ex-post reward to the set of positive real numbers. The agent evaluates the ex-post reward x_t at time t by $u(x_t)$, where $u(\cdot)$ is increasing and $u(0) = 0$. The objective of the agent is to choose the strategy that maximises her lifetime utility

$$\mathbb{E}\left[\phi\left(\delta^\tau u(x)\right)\right], \tag{2.1}$$

where ϕ is strictly increases, τ is the time that the agent learns the value of the parameter, and the expectation is taken over the distribution of learning time τ. This lifetime utility is the generalised expected discounted utility introduced in DeJarnette et al. (2020). It allows for different time risk attitude. When $\phi(z) = z$, we have the standard exponential discounting.

2.2 Discussion

This model is a single agent's learning problem. The unknown parameter is predetermined and fixed overtime. The agent can costlessly do tests in order to obtain information about the unknown parameter. The results of the tests are purely informational. That is, the agent's flow payoff at time t does not depend on the results of the tests. Therefore, the agent does not have incentive to pass the test. These assumptions are similar to Meyer (1994), where the learner ('the principal' in their model) can costlessly design tasks to learn the type of the worker and the principal does not gain any payoffs related to the task completion. Deb & Stewart (2018) also has the similar learning feature, but the object the learner wants to learn in their model is a strategic player, and hence not a fixed parameter.

The agent in this paper just wants to learn the true value of the parameter. There is no exploitation and exploration tradeoffs. Since the agent only gets a reward when she learns the unknown parameter, there is no exploitation. The agent does not settle before learning the true value of the parameter. Therefore, the model is not about when the agent stop learning. It is about how the agent acquires information to learn the true parameter when she has different intertemporal preferences.

The parameter set is assumed to be finite and sorted. It is assumed to be finite so that the agent can eventually learn the true parameter. This assumption is equivalent to assuming the unknown parameter is drawn from a compact set, and the agent learns the unknown parameter when she is ϵ-close to the true value where ϵ is exogenous. The parameter set is assumed to be sorted to simplify the expression of the agent's action and the signal structure. If this assumption is dropped, the following specification of the agent's action and the signal structure is equivalent to what is in the model setup. The agent's action is to partition the parameter set into two subsets. The signal tells the agent which subset the true parameter belongs to. Then Linear Search and Binary Search should also be redefined. The partition is considered to be equivalent to Linear Search if one of the subsets only contains one element. The partition is considered to be equivalent to Binary Search if the two subsets have the same number of elements. In the computer science literature, Binary Search always requires the parameter set to be sorted. But for the purpose of this paper, this assumption is not needed. This assumption is made to simplify the expression of the agent's action and the signal structure.

Time is finite and the duration of the game is long enough such that the agent always has enough time to learn the unknown parameter. [4] Linear Search

[4] This requires that the agent does not choose fully uninformative signals. An alternative situation is when there exists a deadline such that the agent may not have enough time to learn. That is, the final period T is smaller than $\bar{N}-2$. If there were a deadline, it may result in different optimal search behaviour. An extreme case is that if the agent only has one period to learn the value of the parameter, then Linear Search is always optimal, as it allows the agent to learn with positive probability, while other search strategies do not. When a deadline exists, the agent might want to maximise the probability of learning the unknown parameter. The

requires the longest time to learn. If the cardinality of the parameter set is $\bar{N}$, it requires at most $\bar{N}-1$ periods to learn the unknown parameter. Therefore, with the assumption that $T \geq \bar{N}-2$, the agent always has enough time to learn.

The agent is assumed to be Bayesian. However, because of the signal structure, where the the signal is binary and noise free, the Bayesian assumption is not as demanding as in the literature. The agent just needs to be able to partition the parameter set and then re-scale the prior belief after receiving the signal. Bayes rule is not needed for the agent to revise her belief.

Finally, the agent's prior belief is assumed to be uniform. Uniform distribution is symmetric. It describes that all the parameters in the parameter set are equally likely. Following the Principle of Insufficient Reason, if there are N indistinguishable parameters, each of them should be assigned a probability $\frac{1}{N}$. In Bayesian probability, the uniform prior is the simplest diffuse prior. It basically describes that the agent only has vague information about the unknown parameter. The important property I exploit in this paper is the symmetry of the uniform prior. The result that the agent does not switch between Binary Search and Linear Search relies on this symmetry property. In Section 1.2, I discuss a simple relaxation of the uniform prior assumption.

3 The Benchmark Case

I consider the case that $\phi(z) = z$ and $u(z) = z$ as the benchmark case. The agent discounts the future payoff exponentially, and the time-t utility is risk neutral. The lifetime utility now becomes

$$\max \mathbb{E}\left[\delta^{\tau} x\right], \tag{2.2}$$

where $\delta \in (0,1)$ is the discount factor. I normalise the reward x to 1 in the following discussion to simplify the calculation.

I first rewrite the agent's problem as a dynamic programming problem and then characterise the optimal strategy. The main results Proposition 2.1 characterise the optimal strategy.

3.1 Dynamic Programming Setup

I first define the state variable and the choice at time t, and then formally define Binary Search and Linear Search. Then I write down the Bellman equation.

optimal search behaviours might be different from the optimal ones in this paper.

At each time t, by choosing the test r_t, the agent partitions the parameter set into two subsets. The signals tell the agent which set the true parameter belongs to. The agent then can eliminate the other subset. Therefore, from the agent's point of view, the parameter set shrinks to one of the subsets after receiving the signal. Since the agent remembers the past actions and the past signals, the optimal action at next time must belong to the subset that is not eliminated. It is then sufficient to keep track of the evolution of the parameter set. Let $\Theta_0 = \Theta$ be the initial parameter set. Let Θ_t be the parameter set at the beginning of time t. Since the agent's action can be considered as partitioning the parameter set, and the agent has uniform prior, the cardinality of the parameter set can be modelled as the state variable at time t. The action of the agent at time t is to choose how to partition the parameter set into two subsets. Let N_t be the cardinality of the parameter set Θ_t. Let m_t and n_t be the cardinalities of the two subsets the agent chooses. Thus, at time t, the state variable is N_t, and the choice is $(m_t, n_t) \in \mathcal{F}_t = \{(m_t, n_t) | m_t + n_t = N_t, (m_t, n_t) \in ((\mathbb{Z}_+)^2 \cap [1, N_t])^2\}$, where $\mathcal{F}_t$ is the feasible set of the choice at time t. Due to symmetry, assume without loss of generality that $m_t \leq n_t$. The evolution of the state is as follows. Given the state at time t and the choice (m_t, n_t), the state at time $t+1$ is

$$N_{t+1} = \begin{cases} m_t & \text{with probability } \frac{m_t}{N_t} \\ n_t & \text{with probability } \frac{n_t}{N_t} \end{cases}.$$

Next, I formally define *Binary Search*, *Linear Search* and *the agent learns the true parameter at time t* using the terminologies defined above.

Definition 2.1. *The Binary Search Policy in state N_t is the choice $(m_t, n_t) \in \mathcal{F}_t$ such that $m_t = \lfloor \frac{N_t}{2} \rfloor$ and $n_t = \lceil \frac{N_t}{2} \rceil$* [5] *. The Binary Search Strategy is the strategy that prescribes the agent the Binary Search Policy in all the states.*

Definition 2.2. *The Linear Search Policy in state N_t is the choice $(m_t, n_t) \in \mathcal{F}_t$ such that $m_t = 1$ and $n_t = N_t - 1$. The Linear Search Strategy is the strategy that prescribes the agent the Linear Search Policy in all the states.*

Definition 2.3. *The agent learns the true parameter at time t if $N_{t+1} = 1$ (N_{t+1} is the state at the end of the period t).*

To simplify the expression, I ignore the t subscript in the following discussion. If the agent uses the Linear Search Policy in state N, the agent can learn the true parameter with probability $\frac{1}{N}$. The state evolves to $N-1$ in the next period

[5] The notation $\lfloor x \rfloor$ rounds $x \in \mathbb{R}$ to the nearest integer less than or equal to x, and the notation $\lceil x \rceil$ rounds $x \in \mathbb{R}$ to the nearest integer greater than or equal to x.

with probability $\frac{N-1}{N}$. If the agent uses the choices other than the Linear Search Policy in state N, the agent cannot learn the true parameter today if $N > 2$, and the state evolves according to his choice. Define another set $\mathcal{F}^\dagger = \mathcal{F} \backslash \{(m = 1, n = N-1)\}$ to be the set of the choices excluding Linear Search in state N. The Bellman equation is

$$V(N) = \max\left\{\frac{1}{N} + \frac{N-1}{N}\delta V(N-1), \max_{(m,n)\in\mathcal{F}^\dagger} \delta\Big\{\frac{m}{N}V(m) + \frac{n}{N}V(n)\Big\}\right\}.$$

The first term is the value of the Linear Search Policy in state N and the second term is the value of other choices in state N. If $N = 2$, the agent learns the value of the parameter at the end of this period. That is, $V(2) = 1$. From this, the initial condition of the dynamic programming problem is $V(1) = \frac{1}{\delta}$. If there is only on element in the parameter set, it means that the agent learns the true parameter in the last time period. The Bellman equation can be simplified to

$$V(N) = \max_{(m,n)\in\mathcal{F}} \delta\Big\{\frac{m}{N}V(m) + \frac{n}{N}V(n)\Big\} \tag{2.3}$$

with the initial condition $V(1) = \frac{1}{\delta}$. The closed form of the value function when $N \leq 3$ are easy to compute, where $V(2) = 1$ and $V(3) = \frac{1}{3} + \frac{2}{3}\delta$. Let $W(N) := NV(N)$ be the product of N and the value function $V(N)$. The Bellman equation (2.3) can be rewritten as

$$W(N) = \max_{(m,n)\in\mathcal{F}} \delta\Big\{W(m) + W(n)\Big\}.$$

3.2 The Optimal Strategy

In this section, I will show that the Linear Search Strategy is optimal if the agent is sufficiently impatient, and the Binary Search Strategy is optimal otherwise. All other strategies are weakly sub-optimal.

Using the Linear Search Strategy allows the agent to test one element at each time in any state $N > 2$. The best-case scenario is that the agent learns the unknown parameter immediately at $t = 0$, but most likely the agent learns the unknown parameter at some other time $0 < t < \bar{N} - 2$. By using the Linear Search Strategy, the agent effectively check one element in the parameter set at a time. But, when there are only two parameters in the parameter set, learning that one parameter is not the true value of parameter is equivalent to learning that the other parameter left is. It is as if the agent is able to check both parameters at once. Since the agent's prior belief is uniform, the expected value associated with the Linear Search Strategy is thus the sum of the geometric series $\mathcal{L} = (\frac{1}{N}, \frac{1}{N}\delta, \ldots, \frac{1}{N}\delta^{N-3})$ plus $\frac{2}{N}\delta^{N-2}$.

Using the Binary Search Strategy, the agent can learn the value of the parameter within two consecutive periods. When the state is a power of two, say $N = 2^K$, where K is a positive integer, the agent learns the unknown parameter at time $\tau_2^N = K - 1$ with probability one. If the state is greater than 2^K and smaller than 2^{K+1}, the agent learns the unknown parameter at time $\tau_1^N = K$ or $\tau_2^N = K - 1$ with positive probabilities $\frac{\pi^N}{N}$ and $1 - \frac{\pi^N}{N}$ respectively. The uncertainties associated with the timing of learning is small compared to that of the Linear Search Strategy. The agent does face uncertainties in each state that is not a power of two. In an odd state, the uncertainty arises because the following state is not deterministic. In an even state, the uncertainty can still arise if the following state is odd. When starting from a large initial state $\bar{N}$, a big number of states can be visited before learning the unknown parameter. Since there can be uncertainties in each of the odd state visited, one may think that there could be a high aggregate uncertainty associated with the Binary Search Strategy. However, this is not true. This is because the agent only learns the unknown parameter when the state evolves to one. By using the Binary Search Strategy, in any state $N > 3$, the state never evolves to one directly. Instead, state one only occurs after states $N = 2$ and $N = 3$. When the state $N = 3$ occurs, with probability of a third, the agent learns the unknown parameter. When the state $N = 2$ occurs, the agent learns the parameter for sure today. If the state evolves to $N = 2$ directly without reaching state three, the agent learns the unknown for certain. If the state evolves to $N = 3$ before evolving to $N = 2$, the agent might learn the unknown parameter today or tomorrow. By using the Binary Search Strategy, the learning procedure ends in visiting either state $N = 2$ or $N = 3$. As a consequence, the agent always learn the unknown parameter within two consecutive periods.

Let $V^L(\cdot)$ and $V^B(\cdot)$ be the values associated with the Linear Search Strategy and the Binary Search Strategy. Let $\pi^N = 2N - 2^{\lfloor \log_2 N \rfloor + 1}$, $\tau_1^N = \lceil \log_2 N \rceil - 1$, and $\tau_2^N = \lfloor \log_2 N \rfloor - 1$, where τ_1^N and τ_2^N are the two consecutive time at which the agent learns the unknown parameter, and $\frac{\pi^N}{N} \in [0, 1)$ is the probability that the agent learns the unknown parameter at time τ_1^N.

Lemma 2.1. *The value associated with the Linear Search Strategy is*

$$V^L(N) = \frac{1}{N}\left(\frac{1 - \delta^{N-1}}{1 - \delta} + \delta^{N-2}\right).$$

The value associated with the Linear Search Strategy is

$$V^B(N) = \frac{1}{N}\big[\pi^N \delta^{\tau_1^N} + (N - \pi^N)\delta^{\tau_2^N}\big].$$

Given the reward x, the function $V^L(\cdot)$ and $V^B(\cdot)$ decrease in N. Let $W^L(N) := NV^L(N)$ and $W^B(N) := NV^B(N)$.

Lemma 2.2. *When $\delta > 0.5$ ($\delta < 0.5$), the first-order difference of $W^L(\cdot)$ is positive (negative), and the second-order difference of $W^L(\cdot)$ is negative (positive). The first-order difference of $W^B(\cdot)$ is positive (negative), and the second-order difference of $W^B(\cdot)$ is non-positive (non-negative). When $\delta = 0.5$, both $W^L(N)$ and $W^B(N)$ are independent of N.*

Lemma 2.1 and Lemma 2.2 can be used to show the optimality of the Linear and Binary Search Strategy.

Proposition 2.1. *There exists a unique threshold $\bar{\delta} = 0.5$ such that the Linear Search Strategy is weakly optimal if the agent has a discount factor $\delta \leq \bar{\delta}$, and the Binary Search Strategy is weakly optimal if the agent has a discount factor $\delta \geq \bar{\delta}$.*

To check the optimality of the Linear (Binary) Search Strategy, I define a *Linear (Binary) Search Deviating Strategy* and then show that the *Linear (Binary) Search Deviating Strategy* that gives the agent a higher payoff than the Linear (Binary) Search Strategy does not exist. A *Linear (Binary) Search Deviating Strategy* is a one-step deviation strategy from the Linear Search Strategy, such that the agent follows Linear (Binary) Search Strategy in all the states $n \neq N$, and deviates from the Linear (Binary) Search Policy in state N.

When the agent deviates from the Linear Search Policy to some other policy $(m, n) \in \mathcal{F}^\dagger$ in state N, the payoff today will be zero. The most profitable one-step deviation strategy hence must maximise the continuation value. If the agent uses the $(m, n) \in \mathcal{F}^\dagger$ policy in state N, then, the (undiscounted) continuation value is the convex combination of $V^L(m)$ and $V^L(n)$, with the weights being $\frac{m}{N}$ and $\frac{n}{N}$. Since $m \leq n$ and $V^L(\cdot)$ is decreasing, the value of $V^L(m)$ is greater than the value of $V^L(n)$, but the weight attaches to $V^L(m)$ is smaller than the weight attaches to $V^L(n)$. Increasing m (i.e. decreasing n) decreases the value of $V^L(m)$, but puts a higher weight to that value. At the same time, it increases the value of $V^L(n)$, but decreases the weight attaches to $V^L(n)$. As a consequence, it is ambiguous whether increasing m and decreasing n increases the continuation value. Whether increasing m is optimal depends on the value of the function

$W^L(\cdot)$. According to Lemma 2.2, the function $W^L(\cdot)$ is decreasing and convex when $\delta \leq \bar{\delta}$. Therefore, the increase of the continuation value from decreasing n is smaller than the decrease of the continuation value from increasing m. Thus, if the agent were to deviate from the Linear Search Strategy in state N, the most profitable Linear Search Deviating Strategy is to choose $(m, n) = (2, N-2)$ in state N.

In state N, When the agent deviates from the Linear Search Policy in state N to $(m, n) = (2, N-2)$, the agent gives up the flow payoff of $\frac{1}{N}$ today, and increases the (discounted) continuation value from $\delta \frac{N-1}{N} V^L(N-1)$ to $\delta[\frac{N-2}{N} V^L(N-2) + \frac{2}{N} V^L(2)]$. To check whether the one-step deviation is optimal, the agent compares the value she gives up today (the loss) with the increase of the discounted continuation value (the gain). If the loss and the gain are scaled up by N, then the rescaled loss is a constant and the rescaled gain is decreasing in N when $\delta < \bar{\delta}$. That is, the rescaled gain is maximised in the smallest state $N = 4$. Therefore, if it is not optimal for the agent to deviate from the Linear Search Policy in state $N = 4$, then, the agent would not want to deviate from the Linear Search Policy in any other state that is greater than four. In state $N = 4$, when $\delta \leq \bar{\delta}$, the discounted increase in the continuation payoff by deviating from the Linear Search Policy is smaller than the payoff given up today. The agent hence does not want to deviate in state $N = 4$, and does not want to deviate in any other states greater than four.

When the agent deviates from the Binary Search Strategy to the policy $(m, n) \in \mathcal{F}^\dagger$ in state N, it does not change the payoff today, which is still zero. but changes the continuation values. If the continuation value in state N is scaled up by N, then, given any policy $(m, n) \in \mathcal{F}^\dagger$ in state N, the rescaled continuation value is the sum of the two elements $mV^B(m)$ and $nV^B(n)$. By definition, the Binary Search Policy in state N maximises the value of m. Whether decreasing m increases the continuation value depends on the property of the function $W^B(\cdot)$. According to Lemma 2.2, $W^B(\cdot)$ function is increasing and concave when $\delta > \bar{\delta}$. Since m is assumed to be smaller than n, decreasing m by one does not offset the effect of increasing n by one. Therefore, deviating to the policy $(m, n) \in \mathcal{F}^\dagger$ in state N is not optimal when δ is sufficiently big.

If instead of deviating to a policy $(m, n) \in \mathcal{F}^\dagger$, the agent deviates to the Linear Search Policy in state N, effectively, the agent postpones Binary Search until tomorrow and takes the risk of learning the unknown parameter today. If the agent fails to learn the unknown parameter today, she gets a small benefit of decreasing the state by one in next period. The cost and benefit of deviating to the choice Linear Search in state N are different in different states. It is more beneficial to deviate to the Linear Search Policy in a small state, say in state $N = 4$, rather than in a big state, say in state $N = 400$. First, it is because the probability of learning the unknown parameter, which is $\frac{1}{N}$, is higher when N is small. Second, it is because Binary Search is more efficient in terms of eliminating the number of impossible candidates when the state is big. For example, in state $N = 400$, the Binary Search Policy today will eliminate 200 impossible candidates, while in state $N = 4$, the Binary Search Policy will only eliminate 2 impossible candidates. Because of the two reasons above, if the agent would like to deviate to the Linear Search Policy in some state, it is most profitable for him to deviate in state $N = 4$, rather than in state $N = 400$. If deviating to the Linear Search Policy in state $N = 4$ is not optimal, then deviating to Linear Search Policy in other states is also not optimal. The gain from deviating to the Linear Search Policy in state $N = 4$ is $\frac{1}{4}x$, and the loss is $V^B(4) - \delta\frac{3}{4}V^(3)$. When $\delta > \bar{\delta}$, the gain is always smaller than the loss. Deviating to the Linear Search Policy in state $N = 4$ is thus not optimal. Therefore, when $\delta > \bar{\delta}$, deviating to the Linear Search Policy in any state N is not optimal.

3.3 Discussion

The thresholds $\bar{\delta}$ in Proposition 2.1 is a constant $\frac{1}{2}$. This means that for any value $\delta \in (0, 1)$, either the Binary Search Strategy or the Linear Search Strategy are *weakly* optimal. Some of the alternative strategies are weakly sub-optimal but not *strictly* sub-optimal because they induce the same distribution of the learning time as is induced by the Linear or Binary Search Strategy. For example, when $N = 6$, the following two strategies induce the same distribution of the learning time. That is, the agent learns the parameter at $t = 1$ with probability $\frac{1}{3}$ and learns the parameter at $t = 2$ with probability $\frac{2}{3}$.

Strategy One: The Binary Search Strategy.

Strategy Two: The agent chooses $(m, n) = (2, 4)$ *in state* $N = 6$*. If the state* 4 *occurs, the agent chooses* $(m, n) = (2, 2)$*. and if the state* 2 *occurs, the agent chooses* $(m, n) = (1, 1)$*.*

Since these two strategies induce the same distribution of the learning time, these two strategies induce the same expected utility. The *Strategy Two* above is only weakly dominated by the Binary Search Strategy.

The threshold $\bar{\delta}$ is a constant, and is independent of the state N. For example, if an agent has the discount parameter $\delta = 0.45$, then it is optimal for him to use the Linear Search Policy in state $N = 4$, and it is also optimal for the agent to use the Linear Search Policy in state $N = 400$. If the agent has the discount parameter $\delta = 0.9$, then it is optimal for him to use the Binary Search

Policy in state $N = 4$ and in state $N = 400$. Thus the agent does not need to commit to these policies, they are the consequences of the optimal choices at every state. The simple form of the policy is due to the simple form of the state variable. The state variable is just the cardinality of the parameter set because of the uniform prior assumption. Since the agent's prior belief is assumed to be uniform, the agent's posterior belief is still uniform. The agent always believes that the remaining candidates are equally likely, that is, the likelihood ratio of any two candidates is one. The number of candidates in the parameter set only scales up or down the absolute value of the probability attached to each candidate, but does not change the likelihood ratio of any two candidates. As a result, with the uniform prior assumption, in different states, the agent is facing the problem with the same properties: all the elements in the parameter set are equally likely to be the true value of the parameter. If the likelihood ratio of any two elements changes as the state changes, the agent might use different policies in different states. For example, let N be the cardinality of the initial parameter set, which is the initial state. Consider the agent's prior belief to be such that the first element θ_1 in the parameter set is the most likely one with probability a half, and all other parameters are equally likely. The likelihood ratio of θ_1 to other elements is $N - 1$. In this case, since θ_1 is very likely compared to other elements in the parameter set, the agent has a strong incentive to choose the Linear Search Policy in state N regardless the value of δ. If the agent does not learn the unknown parameter, the state $N - 1$ occurs and the agent's belief is revised to a uniform distribution. In state $N - 1$, the agent hence is facing the problem as in the benchmark case. The likelihood ratio of any two elements is one. The agent's strategy then depends on δ. In Section 1.2, I consider another prior belief to check how the it can affect the agent's optimal strategy.

The argument above appears to say that being patient or impatient explains why the Linear Search Strategy or Binary Search Strategy is optimal. But, there are other models of time preference where this apparent relationship fails. Consider the expected utility with a linear discounting, that is, the agent pays a fixed cost c at each time t if she searches and does not learn the unknown parameter (see Section 1.1 for details). My main result in this case is that the Binary Search Strategy is *always* optimal as long as the cost c is positive and bounded. The agent is also impatient in this case, but this impatience is modelled by a linear discounting. Even though the linear discounting and the exponential discounting both describe the agent being impatient, the optimal strategies in these case are different from each other. In fact, the expected utility with the exponential discounting does not just describe the agent being impatient, it also describes the agent being "time risk loving". To understand *time risk*, consider the following two options one can choose from. These are just illustrative examples to show what time risk refers to [6].

Option One: Learn the unknown parameter today ($t = 0$) or the day after tomorrow ($t = 2$) with equal probabilities.

Option Two: Learn the unknown parameter tomorrow ($t = 1$) for certain.

If the agent discounts future utilities exponentially, Option One gives the agent an ex-ante utility of $\frac{1}{2}(1 + \delta^2)x$, and Option Two gives the agent an ex-ante utility of δx. The agent then *always* prefers Option One, as it gives the agent a higher expected utility with any $\delta < 1$. Since these two options yield the same expected learning time: tomorrow ($t = 1$), and the agent prefers the one that is risky in terms of learning time, the agent is hence time risk loving. If the agent discounts future utilities with linear discounting, then the agent is indifferent between these two. The agent is hence time risk neutral.

The comparison between the two different preferences shows that the time risk preference affects the optimal strategy. In fact, the fixed search cost utility can be expressed as the utility function (2.1) with $\phi(z) = \log(z)$. The $\phi(\cdot)$ function determines the agent's time risk attitude (DeJarnette et al. (2020)) and affects the agent's optimal strategy.

4 Time Risk Attitude

In this section, I consider the lifetime utility as in (2.1). The lifetime utility is identified with a triple (ϕ, δ, u). As discussed in DeJarnette et al. (2020), the agent's time risk attitude is determined by the concavity of ϕ, the intertemporal substitution is determined by δ and ϕ, and the atemporal risk attitude towards the lotteries with only immediate payment is determined by $\phi \circ u$. I show that if ϕ is a power function with power $1 - \gamma$ and u is the inverse of ϕ, then, the optimal strategy of the agent with utility (ϕ, δ, u) is the same as the benchmark agent (the agent with utility function as in (2.2)) with a discount parameter ρ where $\rho = \delta^{1-\gamma}$.

If ϕ is a power function, say $\phi(z) = \frac{1}{1-\gamma} z^{1-\gamma}$ where $\gamma \in (0, 1) \cup (1, +\infty)$, then, eq. (2.1) can describe both time risk loving and time risk averse utility. The following lemma presents this result.

Lemma 2.3. *If $\phi(z) = sign(1 - \gamma) a z^{1-\gamma}$ where $a > 0$ is a constant, then, for any increasing u function, when $\gamma \in (0, 1)$, (ϕ, δ, u) describes time risk loving*

[6] This example and the discussion of the time risk are borrowed from DeJarnette et al. (2020).

preference, and when $\gamma > 1$, (ϕ, δ, u) describes time risk averse preference.

This lemma follows directly from DeJarnette et al. (2020) Prop 2. To understand why this is true, reconsider the two options introduced in Section 3.

Option One: Learn the unknown parameter and get a reward of one today ($t = 0$) or the day after tomorrow ($t = 2$) with equal probabilities.

Option Two: Learn the unknown parameter and get a reward of one tomorrow ($t = 1$) for certain.

Suppose $\phi(z) = \frac{1}{1-\gamma} z^{1-\gamma}$. Given the preference (ϕ, δ, u), the utility from Option One is $\frac{1}{1-\gamma}(\frac{1}{2}u(1)^{1-\gamma} + \frac{1}{2}[\delta^2 u(1)]^{1-\gamma})$, and the utility from Option Two is $\frac{1}{1-\gamma}[\delta u(1)]^{1-\gamma}$. When $\gamma \in (0, 1)$, the utility from Option One is greater than the utility from Option Two. When $\gamma > 1$, the utility from Option One is smaller than the utility from Option Two, and The special cases discussed in Section 3, the exponential discounting and the search with a fixed cost, are limiting cases. When $\gamma \to 0$, $\phi(z)$ converges to $\phi(z) = z$, which is the exponential discounting case. When $\gamma \to 1$, $\phi(z)$ converges to the natural log function, which describes the linear discounting preference.

Next, I show that there exists a (ϕ, δ, u) such that given any strategy, the utility derived from (ϕ, δ, u) equals the utility derived from the exponential discounting with a different discount parameter ρ. In addition, such functional forms of ϕ and u are unique.

Lemma 2.4. *Let $a \neq 0$ be a constant. Then, $\phi\big(\delta^\tau u(x)\big) = \rho^\tau x$ if and only if $\hat{\phi}(z) = a z^{\frac{\log \rho}{\log \delta}}$ and $\hat{u}(x) = \left(\frac{x}{a}\right)^{\frac{\log \delta}{\log \rho}}$. If in addition $\rho = \delta^c$ where $c \neq 0$, then, $\hat{\phi}(z) = a z^c$ and $\hat{u}(x) = \left(\frac{x}{a}\right)^{\frac{1}{c}}$.*

Each strategy pins down a distribution of the learning time. According to Lemma 2.4, given a strategy, the utility derived from $(\hat{\phi}, \delta, \hat{u})$, which is $\mathbb{E}\big[\phi\big(\delta^\tau u(x)\big)\big]$, equals the utility derived from the exponential discounting with a different discount parameter ρ, which is $\mathbb{E}\big[\rho^\tau x\big]$. This shows that the agent with preference $(\hat{\phi}, \delta, \hat{u})$ behaves as if she discounts future payoffs exponentially with a discount parameter $\rho = \delta^c$. Because of this equivalence of the behaviour, this lemma together with Proposition 2.1 can be used to find the optimal strategy of the agent with preference $(\hat{\phi}, \delta, \hat{u})$.

In the following discussion, let $\hat{\phi}(z) = \frac{1}{1-\gamma} z^{1-\gamma}$ and $\hat{u}(x) = [(1-\gamma)x]^{\frac{1}{1-\gamma}}$ where $\gamma \in (0, 1) \cup (1, \infty)$. This specification is not necessary for the following results, but it gives a clearer intuition. I refer to $\rho = \delta^{1-\gamma}$ as a *virtual discount parameter.*

The following proposition characterises the optimal strategy for preference $(\hat{\phi}, \rho, \hat{u})$.

Proposition 2.2. *For each $\gamma \in (0,1)$, there exists a unique threshold $\breve{\delta}_\gamma := (\bar{\delta})^{\frac{1}{1-\gamma}} < \bar{\delta} = \frac{1}{2}$, such that the Linear Search Strategy is optimal if $\delta \leq \breve{\delta}_\gamma$ and the Binary Search Strategy is optimal if $\delta \geq \breve{\delta}_\gamma$. The threshold $\breve{\delta}_\gamma$ is decreasing in γ, and it approaches $\bar{\delta}$ when γ approaches zero.*

For $\gamma > 1$, the Binary Search Strategy is optimal given any discount parameter $\delta \in (0,1)$.

When $\gamma \in (0,1)$, as shown in Lemma 2.4, the agent with preference $(\hat{\phi}, \delta, \hat{u})$ behaves as if she discounts the future payoffs exponential with a discount parameter $\rho = \delta^{1-\gamma}$. Given Proposition 2.1, the Linear Search Strategy is optimal if the discount parameter ρ is weakly smaller than a half. That is, the Linear Search Strategy is optimal if δ is weakly smaller than $\left(\frac{1}{2}\right)^{\frac{1}{1-\gamma}} := \breve{\delta}_\gamma$. The similar argument holds for the Binary Search Strategy. Since $\gamma \in (0,1)$, $\breve{\delta}_\gamma$ decreases in γ.

When $\gamma \in (0,1)$, the agent's virtual discount parameter ρ is greater than his real discount parameter δ. Due to the effect of the time risk aversion, the agent behaves as if she is more patient. For example, consider an agent A who has the benchmark utility and discounts future payoffs exponentially with the discount parameter $\tilde{\delta} < \frac{1}{2}$. The optimal strategy of this agent is then the Linear Search Strategy. Consider another agent B with preference $(\hat{\phi}, \tilde{\delta}, \hat{u})$ where $\gamma \in (0,1)$. Suppose γ takes the value such that the optimal strategy of the agent B is the Binary Search Strategy. Imagine there is an outside observer observing these two agents' strategies. If the outside observer incorrectly believes that the agent B has the preference as in the benchmark case, then the outside observer would conclude that the agent B is more patient than the agent A. However, the fact is that the agent B is as patient as the agent A, but less time risk loving. The decreasing $\breve{\delta}_\gamma$ shows that when γ increases, the agent becomes less time risk loving, and has a stronger incentive to use the Binary Search Strategy. Being less time risk loving, the less patient agent still prefers to use the Binary Search Strategy.

When $\gamma > 1$, the agent is time risk averse. Since we know from Section 3 that the Binary Search Strategy is optimal for the time risk neutral agent as it induces a distribution of learning time with a smaller mean, when the agent becomes time risk averse, the agent prefers the Binary Search Strategy more. Since the Binary Search Strategy induces a distribution of the learning time with a smaller variance, it is hence also optimal for the risk averse agent.

Figure 2.1 demonstrates the optimal strategy given the value of the discount parameter δ and the time risk aversion parameter γ. From Proposition 2.2, we know that the functional form of the red curve in Figure 2.1 is $\delta^{1-\gamma} = \frac{1}{2}$. When

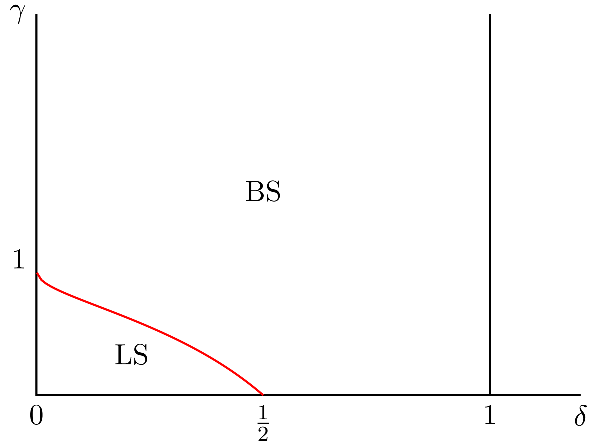


Figure 2.1: The optimal strategy for each value of the discount parameter δ and the risk-aversion parameter γ. The functional form of the red curve is $\delta^{1-\gamma} = \frac{1}{2}$.

the discount parameter δ is greater than a half, the optimal strategy is the Binary Search Strategy given any value of the time risk aversion parameter γ. That is, when the agent is sufficiently patient, the effect of being patient dominates the effect of time risk attitude. It is optimal for the agent with $\delta > \frac{1}{2}$ to use the Binary Search Strategy regardless the time risk attitude. When the time risk aversion parameter γ is greater than one, the optimal strategy is the Binary Search Strategy given any value of δ. That is, when the agent is time risk averse, the effect of the time risk attitude dominates the effect of patience. It is optimal for the time risk averse agent to use the Binary Search Strategy regardless the patience level. When the agent is time risk loving and not sufficiently patient, that is, when $\delta < \frac{1}{2}$ and $\gamma < 1$, there is a trade-off between learning in a risky way and learning faster in expectation. Being time risk loving incentivises the agent to take the risk of learning the parameter today, while being patient incentivises the agent to learn faster in expectation. As a consequence, the optimal strategy depends on the effect that dominates.

Given the value of γ, the threshold $\breve{\delta}_\gamma$ in Proposition 2.2 is a constant, and it does not depend on N. This means that the agent's optimal strategies have the same time-consistent property as in Section 3. That is, if the agent has the discount parameter δ and the risk-aversion parameter γ such that $\rho = \delta^{1-\gamma} = 0.45$, then it is optimal for him to use the Linear Search Policy in state $N = 4$ and the state $N = 400$. If the agent has the discount parameter δ and the risk-aversion parameter γ such that $\rho = \delta^{1-\gamma} = 0.9$, then it is optimal for him to use the Binary Search Policy in state $N = 4$ and the state $N = 400$. Therefore, the agent does not need to commit to a certain policy, the commitment to a certain policy along the path of learning is a result of the optimal strategy. This is true because of the equivalence condition specified in Lemma 2.4. The optimal strategies in this section thus preserve the same properties as in the benchmark case.

The Linear Search Strategy is not only risky in terms of the timing of learning the unknown parameter, it is also risky in terms of whether the agent can learn at a certain time t. That is, if N is the state at time t, by using the Linear Search Strategy, the agent learns the parameter with the probability $\frac{1}{N}$, and does not learn the parameter with the probability $\frac{N-1}{N}$. There are two different types of risks associated with the Linear Search Strategy. One is the time risk that has been discussed in the previous part of this section, and the other one is the temporal risk at each time t. Because $\hat{u}$ is the inverse of $\hat{\phi}$, the agent is assumed to be risk-neutral towards time-t lotteries. To discuss the effect of the temporal risk attitude, I consider EZ preferences in Section 5.

5 Epstein and Zin Preferences

In order to discuss the EZ preferences, I first rewrite the one time reward at time t as a stream of reward over time. Since the agent essentially gets a reward of zero at the time when the parameter is not learned, learning the parameter at time τ is equivalent to getting a reward stream $(x_1, x_2, \ldots, x_\tau, x_{\tau+1} \ldots, x_T) = (0, 0, \ldots, x, 0, \ldots, 0)$. After introducing this reward stream, the EZ recursive setup can be used to evaluate the agent's lifetime utility.

The recursive formulation of EZ preferences is developed in Epstein & Zin (1989), which is originated in Kreps & Porteus (1978) in a finite time setting. The EZ recursive utility function consists of two components: a CES time aggregator that characterises the preference over the deterministic payoff vector, and a risk aggregator that aggregates the risk associated with future uncertain payoffs. Consider a deterministic payoff vector $(z_0, z_1, \ldots, z_T)$, with z_t denoting the payoff at time t. The utility from time t onwards is

$$U(z_t, z_{t+1}, \ldots, z_T) = \left(z_t^\rho + \delta U(z_{t+1}, \ldots, z_T)^\rho\right)^{\frac{1}{\rho}} \tag{2.4}$$

where $\delta \in (0, 1)$ is the discount factor and $\frac{1}{1-\rho} > 0$ is the elasticity of intertemporal substitution (EIS). The greater the value of EIS, the greater willingness the agent has (or, the easier it is) to substitute today's payoff to tomorrow's payoff. In case of the uncertain payoffs, the utility is evaluated by a certainty equivalent operator that is introduced in Kreps & Porteus (1978). Let z be a set of future payoff vectors. Let p be a probability measure on the set z. Then the utility from the uncertain payoffs is

$$\big(\mathbb{E}_p U(z)^{\alpha}\big)^{\frac{1}{\alpha}} \tag{2.5}$$

where $1 - \alpha > 0$ is the relative risk aversion (RRA). A smaller value of α corresponds to a greater risk aversion.

How does this utility specification affect the agent's evaluation of Binary Search and Linear Search? The Binary Search Policy ensures the agent a certain payoff today, either zero or one, but the Linear Search Policy gives the agent an uncertain payoff both today and in the future. To be more specific, if the agent chooses the Binary Search Policy at time t, then the payoff at time t is deterministic and the future payoffs are uncertain. The agent then evaluates the uncertain future payoffs by the certainty equivalent operator, and the total utility at time t aggregates the certain payoff at time t and the certainty equivalent of the future uncertain payoffs using the CES aggregator. If the agent chooses the Linear Search Policy at time t, the payoff today and the payoffs in the future are both uncertain. To evaluate the total utility associated with the uncertain payoffs, the agent first evaluates the utility of each payoff stream using the CES aggregator. The total utility is then calculated using the certainty equivalent operator. In other words, when the agent chooses Binary Search, the CES aggregator is the 'outside operator'. The certainty equivalent operator is only used to evaluate the future payoffs, and hence is the 'inside operator' When the agent chooses the Linear Search Policy, the CES aggregator becomes the inside aggregator, and the certainty equivalent operator is the outside operator.

Given the recursive formulation of the EZ preference, the Bellman equation in state N is

$$\begin{aligned} V^{EZ}(N) = \max\Bigg\{ &\Big(\frac{1}{N}(1^{\rho} + \delta 0^{\rho})^{\frac{\alpha}{\rho}} + \frac{N-1}{N}(0^{\rho} + \delta V^{EZ}(N-1)^{\rho})^{\frac{\alpha}{\rho}}\Big)^{\frac{1}{\alpha}}, \\ &\max_{(m,n)\in\mathcal{F}^{\dagger}} \Big(0^{\rho} + \delta\big(\frac{m}{N}V^{EZ}(m)^{\alpha} + \frac{n}{N}V^{EZ}(n)^{\alpha}\big)^{\frac{\rho}{\alpha}}\Big)^{\frac{1}{\rho}}\Bigg\} \end{aligned}$$

with the initial condition $V^{EZ}(1) = \delta^{\frac{\rho}{\alpha^2}}$. The initial condition is computed from the fact that $V^{EZ}(2) = 1$, that is, when the state $N = 2$, the agent learns the true value of the parameter immediately after one search. The first element in the Bellman equation is the value associated with the Linear Search Policy in state N. By choosing the Linear Search Policy today, the agent either gets a payoff of one today and zero tomorrow, or a payoff of zero today and the continuation value $V^{EZ}(N-1)$ tomorrow. It is then as if the agent is facing two possible payoff vectors: $(1,0)$ and $(0, V^{EZ}(N-1))$, with probability $\frac{1}{N}$ and $\frac{N-1}{N}$ respectively. The utility is evaluated by the certainty equivalent of the utilities associated with the two payoff vectors. The second element in the Bellman equation is the value associated with the policy $(m,n) \in \mathcal{F}^\dagger$ in state N. By choosing the policy $(m,n) \in \mathcal{F}^\dagger$, the payoff of zero today is certain. The continuation value however, is $V^{EZ}(m)$ or $V^{EZ}(n)$ with probability $\frac{m}{N}$ and $\frac{n}{N}$. The certainty equivalent operator is used to calculate the value associated with the future payoff, and the CES aggregator aggregates the payoff today and the certainty equivalent of the payoff tomorrow. The agent chooses the policy (m,n) that gives the agent the highest value. With the initial condition $V^{EZ}(1) = \delta^{\frac{\rho}{\alpha^2}}$, the Bellman equation can be simplified to

$$V^{EZ}(N) = \max_{(m,n)\in\mathcal{F}} \left(\zeta \Big(\frac{m}{N} V^{EZ}(m)^\alpha + \frac{n}{N} V^{EZ}(n)^\alpha \Big) \right)^{\frac{1}{\alpha}}, \tag{2.6}$$

where $\zeta \equiv \delta^{\frac{\alpha}{\rho}}$ is a virtual discount parameter.

In general, the parameters α and ρ can take any value smaller than one, that is, both of the parameters can be negative. But, in the following discussion, I restrict the values of α and ρ to be positive. First, to make sure the parameter ζ is indeed a virtual discount parameter, I restrict the parameters α and ρ to have the same sign. Second, Epstein & Zin (1989) points out that decreasing α can change the agent's preference in terms of the timing of the resolution of uncertainty. To shut down this possible effect, I restrict the value of α to be positive. That is, I only consider the case that the agent is not 'too risk averse', and it is not 'too hard' to substitute the consumption intertemporally.

5.1 The Optimal Strategy

This section discusses the agent's optimal strategy and compares it with the benchmark case.

To find the agent's optimal strategy in this section, I show that the Bellman equation (2.6) is closely related to the Bellman equation (2.3) in the benchmark case. Let $\mathcal{W}(N) = V^{EZ}(N)^\alpha$. Equation (2.6) can be rewritten as

$$\mathcal{W}(N) = \max_{(m,n)\in\mathcal{F}} \zeta \Big(\frac{m}{N} \mathcal{W}(m) + \frac{n}{N} \mathcal{W}(n) \Big) \tag{2.7}$$

with the initial condition $\mathcal{W}(1) = \frac{1}{\zeta}$. Equation (2.7) is essentially the same Bellman equation as Equation (2.3) with the same initial condition. The only difference is that the discount parameter in Equation (2.7) is the virtual discount parameter ζ, and the discount parameter in Equation (2.3) is the real discount parameter δ. Since the Bellman equation in this section and the Bellman equation in the benchmark case are closely related in the way described above, the agent with EZ preferences behaves the same as the agent in the benchmark case with a discount parameter ζ. In the benchmark case, the value of the real discount parameter determines the agent's optimal strategy. In this section, the virtual discount parameter ζ hence determines the optimal strategy of the agent with EZ preferences. That is, if ζ is greater than a half, the Binary Search Strategy is optimal. If ζ is smaller than a half, the Linear Search Strategy is optimal. Since the result Proposition 2.1 in the benchmark case are presented as how the real discount parameter δ affects the agent's optimal strategy, I characterise the agent's optimal strategy below in the similar way. Let $\bar{\zeta} = \frac{1}{2}$.

Proposition 2.3. *If the agent has EZ preferences, given any (α, ρ) pair, there exists a unique threshold $\tilde{\delta} = (\bar{\zeta})^{\frac{\rho}{\alpha}}$ such that if $\delta > \tilde{\delta}$, the Binary Search Strategy is the optimal strategy. If $\delta < \tilde{\delta}$, the Linear Search Strategy is the optimal strategy. If $\delta = \tilde{\delta}$, the agent is indifferent between the Binary Search Strategy and the Linear Search Strategy.*

To better understand how the benchmark agent and the agent with EZ preferences behave differently, I consider the marginal agent who is just indifferent between the Linear Search Strategy and the Binary Search Strategy. The marginal agent in the benchmark case (the benchmark marginal agent) has a discount parameter $\delta = \bar{\delta}$, and the marginal agent with EZ preference (the EZ marginal agent) has the parameter triple (δ, α, ρ) such that $\delta = \tilde{\delta} = (\bar{\zeta})^{\frac{\rho}{\alpha}}$. Since the two thresholds $\bar{\delta}$ and $\bar{\zeta}$ are indeed the same, the relationship of $\bar{\delta}$ and $\tilde{\delta}$ depends on the value of the ratio of ρ and α. Proposition 2.3 hence can be used to check which marginal agent is more patient. Note that $EIS = \frac{1}{1-\rho}$ and $RRA = 1 - \alpha$.

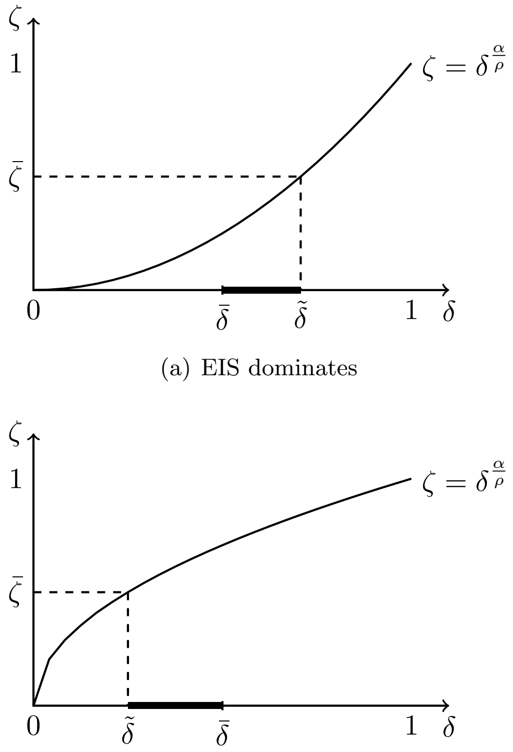
Lemma 2.5. *If $\alpha > \rho$ ($EIS < \frac{1}{RRA}$), then $\tilde{\delta} > \bar{\delta}$. If $\alpha < \rho$ ($RRA > \frac{1}{EIS}$), then $\tilde{\delta} < \bar{\delta}$.*

Lemma 2.5 says that if EIS is smaller than the reciprocal of RRA, then the EZ marginal agent is more patient than the benchmark marginal agent. If RRA is greater than the reciprocal of EIS, then the EZ marginal agent is less patient than the benchmark marginal agent. This can be better understood by considering the agent's incentives and the features of the Linear and Binary Search Strategies. When EIS is small, it is hard to substitute the payoffs intertemporally. The agent hence has stronger incentives to use Linear Search Strategy, as it allows the agent to get immediately payoff. However, this payoff is uncertain. Therefore, there is a tradeoff between getting the payoff today and facing uncertain payoffs. The optimality of Linear Search or Binary Search depends on the value of EIS and RRA. If it is sufficiently hard to substitute intertemporal payoffs such that the gain of getting a payoff today overweighs the agent's aversion towards risky payoff, that is, if $EIS < \frac{1}{RRA}$, then the agent with EZ preference is willing to use Linear Search even when the agent has the discount parameter $\delta > \bar{\delta}$. This case is referred to the case that EIS dominates RRA. If, however, the agent is sufficiently risk averse such that the aversion towards risky overweighs the benefits of getting the payoff today, that is, if $RRA > \frac{1}{EIS}$, then the agent with EZ preference is willing to use Binary Search even when the agent has the discount parameter $\delta < \bar{\delta}$. This case is referred to as the case that RRA dominates EIS. Figure 2.2 plots the virtual discount parameter ζ as a function of the real discount parameter δ. The left panel plots the case that EIS dominates RRA, and the right panel plots the case that RRA dominates EIS. When EIS dominates RRA, it is optimal for the agent with the discount parameter $\delta \in (\bar{\delta}, \tilde{\delta})$ (thick black line on the left panel) to use the Binary Search Strategy in the benchmark case, but to use the Linear Search Strategy in the EZ preference case. This is because Linear Search satisfies the agent's preference on not willing to substitute intertemporal payoffs. When RRA dominates EIS, it is optimal for the agent with the discount parameter $\delta \in (\tilde{\delta}, \bar{\delta})$ (thick black line on the right panel) to use the Linear Search Strategy in the benchmark case, but to use the Binary Search Strategy in the EZ preference case. This is because Binary Search is less risky than Linear Search. Since the agent is very risk averse, she prefers Binary Search.

6 The Present-biased Preference

The optimal strategies in previous sections all have a consistent property. That is, if it is optimal for the agent to use the Linear Search Policy in state $N = 400$, then it is also optimal for the agent to use the Linear Search Policy in state $N = 4$. It is not optimal for the agent to switch between actions. In this section, I show that if the agent has a time-inconsistent preference, there exists the situation such that it is optimal for the agent to use the Binary Search Policy when the state is big, say $N = 400$, and it is optimal for him to use the Linear Search Policy when the state is small, say $N = 4$.

This section discusses the present-biased preference. Consider the case that



(a) EIS dominates

(b) RRA dominates

Figure 2.2: The relationship between $\bar{\delta}$ and $\tilde{\delta}$

an agent is asked which of the two options she prefers: one is to get a payment of one on Wednesday, and the other one is to get a payment of two on Friday. If the agent has the time-consistent preference, she will make the same decision regardless the question is asked on Monday or Wednesday. If the agent has the present-biased preference, she may, however, make different decisions when asked on different days. That is, when asked on *Monday* and *Tuesday*, the agent prefers to get the payment of two on Friday, but when asked on *Wednesday*, she changes his mind and prefers to get the payment of one on Wednesday, which is *today*. The agent changes his mind when the *future* becomes the *present*. The agent is biased towards the payment at the *present* when asked on Wednesday and thus has a stronger incentive to get the payment today.

These present-biased preferences could be described by quasi-hyperbolic discounting (see Phelps & Pollak (1968), Laibson (1997), Fischer (1999) and O'Donoghue & Rabin (1999)). To discuss the agent's present-biased preference, I consider the reward stream as in Section 5. Since the agent's preference changes over time, it is as if the agent has different selves at different times. The agent's time-t self evaluates the payments at and after time t as

$$\mathbb{E}\left[x_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} x_\tau\right], \tag{2.8}$$

where $\beta \in (0,1)$ is the present-biased parameter, and $\delta \in (0,1)$ is the time-consistent discount parameter. The present-biased parameter β describes the magnitude of the present bias. When $\beta = 1$, there is no present bias, and quasi-hyperbolic discounting is identical to exponential discounting. When $\beta < 1$, the smaller the β, the smaller weight the agent attaches to the time that is 'the future', and hence the greater the present bias is. When $\beta = 0$, the agent only cares about the payment today. In addition to the fact that the agent is present-biased, the agent is also assumed to be aware of his present bias. That is, the agent knows that when the *future* becomes the *present*, his preference changes. This type of the agent is referred to as the *sophisticated* agent in the literature.

In this section, I show that the optimal strategy for the present-biased agent with the discount parameter $\delta \leq \bar{\delta}$ and the present-biased parameter $\beta \in (0,1)$ is still the Linear Search Strategy. This is because the present-biased preference gives the agent a stronger incentive to choose the Linear Search Policy in each state. Since it is already optimal for his time-consistent counterpart to choose the Linear Search Policy in each state, it is hence also optimal for the present-biased agent to choose the Linear Search Policy in each state when $\delta \leq \bar{\delta}$.

For the present-biased agent with $\delta > \bar{\delta}$, the optimal strategy depends on the agent's present-biased parameter. When the agent is not very present-biased, the Binary Search Strategy is still the optimal strategy. Suppose the agent's future selves use the Binary Search Policy in all future states. First notice that when $\delta > \bar{\delta}$, any $(m, n) \in \mathcal{F}^{\dagger}$ policy gives a weakly smaller payoff than that of the Binary Search Policy in state N. This is because $(m, n) \in \mathcal{F}^{\dagger}$ policy and Binary Search Policy give the agent zero immediate reward. Since the present-biased agent is assumed to be consistent when evaluating the future payoffs, she behaves essentially the same as the time-consistent agent when there is no immediate payoff getting involved. As a result, I only consider whether it is optimal for the agent to use the Linear Search Policy in the current state. If the agent uses the Linear Search Policy in the current state, the benefit is the immediate expected reward associated with learning. That is $\frac{1}{N}$. The (future) cost is the difference between using the Binary Search today and tomorrow $V^B(N) - \delta \frac{N-1}{N} V^B(N-1)$. If the agent is not very present-biased, that is, the present-biased parameter is greater than the ratio of the benefit to the cost, then, using the Linear Search Policy in current state N is not optimal. Using the Binary Search Policy in state N hence is optimal, given that her future selves use the Binary Search Policy. Using the idea of backward induction, one can check whether using the Binary Search Policy is optimal in state N given that the agent's future selves use the Binary Search Policy. If this is true for all the states, then, the Binary Search Strategy is the optimal strategy. This requires the agent's present-biased parameter to be greater than $\bar{\beta}^N := \frac{\frac{1}{N}}{V^B(N) - \delta \frac{N-1}{N} V^B(N-1)}$ for $N = 4, 5, \ldots, \bar{N}$ where $\bar{N}$ is the cardinality of the parameter set. The benefit associated with the Linear Search Policy is the highest in state $N = 4$, and the cost associated with the Binary Search Policy is the lowest in state $N = 4$. As a result, if the present-biased parameter β is weakly greater than $\check{\beta} := \bar{\beta}^4$, then, the Binary Search Strategy is the optimal strategy.

When $\delta > \bar{\delta}$, for any $\beta < \check{\beta}$, there exists a state $\underline{N}$ such that the Linear Search Policy is optimal in all the states $H < \underline{N}$, and the Binary Search Policy is optimal in state $\underline{N}$. That is, there exists a switch of actions from Binary Search (when the state is big) to Linear Search (when the state becomes small). Suppose the agent's future selves use the Linear Search Policy in all future states. In order to check which policy is optimal in state N, let

$$P(N) \equiv \max_{\{m,n\} \in \mathcal{F}^{\dagger}} \delta \left\{ \frac{m}{N} V^L(m) + \frac{n}{N} V^L(n) \right\}$$

be the highest value associated with the policy $(m, n) \in \mathcal{F}^{\dagger}$ in state N given that the agent's future selves use the Linear Search Policy in all the states smaller than N. If the agent uses the Linear Search Policy in the current state, the benefit is the immediate expected reward associated with learning. That is $\frac{1}{N}$. The cost of using the Linear Search Policy in state N is the forgone payoff from using the $(m, n) \in \mathcal{F}^{\dagger}$ policy, which is $P(N) - \delta \frac{N-1}{N} V^L(N-1)$. Using the idea of backward induction, one can check which Policy is optimal in state N given that the agent's future selves use the Linear Search Policy. If the agent's present-biased parameter is smaller than $\tilde{\beta}^N := \frac{\frac{1}{N}}{P(N)-\delta\frac{N-1}{N}V^L(N-1)}$, the ratio of the benefit to the cost of using the Linear Search Policy in state N, then, it is optimal for the agent to use the Linear Search Policy in state N. Otherwise, it is optimal for her to use the $(m, n) \in \mathcal{F}^{\dagger}$ policy that achieves $P(N)$ in state N. From the discussion in the benchmark case,we know that when $\delta > \bar{\delta}$, the $(m, n) \in \mathcal{F}^{\dagger}$ policy that achieves $P(N)$ in state N is the Binary Search Policy because of the concavity of the function $W^L(N) = NV^L(N)$. Therefore, when the agent's present-biased parameter is smaller than $\tilde{\beta}^N$, it is optimal for the agent to use the Binary Search Policy in state N.

When it is optimal for the agent to use the Binary Search Policy in a state $\underline{N}$ given the belief that her future selves use the Linear Search Policy, we still need to check whether it is indeed optimal for her future selves to use the Linear Search Policy. In the appendix, I show that $\tilde{\beta}^N$ decreases in N. This is because the benefit of using the Linear Search Policy in state N decreases in N, and the cost increases in N. As a result, in order to guarantee that it is optimal for the agent to use the Linear Search Policy in any state smaller than $\underline{N}$, the agent's present-biased parameter should be greater than $\tilde{\beta}^{\underline{N}-1}$.

The following proposition concludes the result discussed above.

Proposition 2.4. *When the present-biased agent has the discount parameter* $\delta \leq \bar{\delta}$ *and the present-biased parameter* $\beta \in (0, 1)$*, the optimal strategy is the Linear Search Strategy.*

When the present-biased agent has the discount parameter $\delta > \bar{\delta}$ *and the present-biased parameter* $\beta \geq \check{\beta}$*, the optimal strategy is the Binary Search Strategy.*

When the present-biased agent has the discount parameter $\delta > \bar{\delta}$ *and the present-biased parameter* $\beta \in [\tilde{\beta}^{\underline{N}}, \tilde{\beta}^{\underline{N}-1})$*, it is optimal for the agent to use the Binary Search Policy in state* $\underline{N}$ *and to use the Linear Search Policy in state* $H < \underline{N}$*.*

Consider an agent with discount parameter $\delta > \bar{\delta}$ and the present-biased parameter $\beta \in [\tilde{\beta}^{\underline{N}}, \tilde{\beta}^{\underline{N}-1})$, the last point of Proposition 2.4 indicates that in states

$N < \underline{N}$, the agent's optimal policy is the Linear Search Policy, and the agent's optimal policy in state $\underline{N}$ is the Binary Search Policy. But in states $N > \underline{N}$, we do not know what the optimal policy of the agent is. Proposition 2.4 does not characterise the complete optimal strategy when $\delta > \bar{\delta}$ and $\beta < \tilde{\beta}^4$ due to the complexity of the calculation. Instead, it only specifies that there exists at least one switch between the Linear Search Policy and the Binary Search Policy. If there exists more than one switch between actions, the last point of Proposition 2.4 characterises the smallest state in which the agent switches actions. This serves the purpose to show that when the agent has time-inconsistent preferences, the optimal strategy does not have the consistent property as in the cases with time-consistent preferences.

To better explain the result, I introduce the *present-biased agent's time-consistent counterpart.* If a present-biased agent has a discount parameter $\hat{\delta}$ and a present-biased parameter $\hat{\beta}$, then his time-consistent counterpart also has a discount parameter $\hat{\delta}$, but a present-biased parameter $\beta = 1$. The result above indicates that in comparison with the present-biased agent's time-consistent counterpart, it is optimal for the present-biased agent to use the Linear Search Policy in small states, while it is *always* optimal for his time-consistent counterpart to use the Binary Search Policy. This is because the value associated with a policy in state N consists of two parts: the expected payment today and the perceived discounted continuation value. Given the future policies remaining the same, consider the cost and the benefit of the Linear Search Policy in the current state N. The cost of the Linear Search Policy comes from the discounted continuation value. The greater the current state N, the more costly to use the Linear Search Policy. The benefit of the Linear Search Policy is from the expected payment today. The smaller the state N, the greater the benefit associated with the Linear Search Policy. Notice that when the agent is present-biased, the cost of the Linear Search is perceived to be smaller because of the present-biased parameter, but she evaluates the expected payment today the same as his time-consistent counterpart. As a result, there are two driving forces that incentivise the agent to use the Linear Search Policy: one is that the state is sufficiently small and thus the benefit of Linear Search becomes higher, and the other one is that the agent is sufficiently present-biased and thus perceives the cost of Linear Search to be smaller. The main tradeoff is that when the agent is very present-biased, that is, she has a smaller present-biased parameter β, she perceives the Linear Search Policy to be attractive in bigger states, while for the agent who is less present-biased, she only perceives the Linear Search Policy to be attractive in smaller states.

7 Winner Takes All

In the previous section, I have discussed a single agent's decision problem. There is only one agent learning the unknown parameter. In this section, I consider a game where the players compete to learn the unknown parameter. At each time t, the player $i \in \{1, 2, \ldots, I\}$ is active with probability p_i where $\sum_i^I p_i = 1$ and $p_i \in (0, 1)$ for all i. There is only one active player at each time t. The active player i can use the Linear Search Policy, the Binary Search Policy, or any $(m, n) \in \mathcal{F}^\dagger$ policy in state N. At time t, if the active player learns the unknown parameter, then, she gets the reward associated with learning and the game ends. All other players get no reward. If the active player does not learn the unknown parameter, the game enters next period. Assume the lifetime utility of the players is as in eq. (2.2) and all the players have the same discount parameter δ.

If all the players use the Linear Search Strategy,then, the outcome of the game, that is, the distribution of the learning time, is the same as the case with one player learning alone using the Linear Search Strategy. The players' payoffs are different because now they only have the chance to get the reward associated with learning when they are active. One could imagine that now the players have less incentives to use the Binary Search Strategy because it gives the player zero probability of getting the reward in state greater than two. In addition, using the Binary Search Strategy increases the other players' probabilities of getting the reward by shrinking the parameter set. It is then interesting to check when it is an equilibrium such that all the players use the Linear Search Strategy I show that this equilibrium exists when the discount parameter δ is not too close to one.

Proposition 2.5. *There exists an upperbound $\mathring{\delta} = \min \frac{1}{1+p_i}$ such that if $\delta \leq \mathring{\delta}$, there exists an equilibrium where all the players use the Linear Search Strategy.*

Suppose all the players use the Linear Search Strategy and then I check whether there exists any deviation in state N that gives the player a higher payoff. First notice that, when $\delta \leq \frac{1}{2}$, the player does not want to deviate to other policies in any state. This is because when players are competing to get the reward associated with learning, the players have stronger incentives to use the Linear Search Strategy. If it is optimal for the player to use the Linear Search Strategy when she is learning alone, it is still optimal for her to use the Linear Search Strategy when competing with other players. When $\delta > \frac{1}{2}$, the player faces a tradeoff where the Linear Search Policy gives the player the probability of getting the immediate payoff, while using other policies reduces the future state and increases her payoff when she is active again in the future. Which effect dominates depends on the player's discount parameter and the probability that she is active in each state. If the player is active with a high probability, say, p_i is close to one, then, the Linear Search Strategy best responses to other players using the Linear Search Strategy only when δ is close to $\frac{1}{2}$. If the player is active with a probability close to zero, then, the Linear Search Strategy best responses to other players using the Linear Search Strategy even when δ is close to one. If the best response of the most active player (i.e. the player with the highest probability of being active) to the Linear Search Strategy is the Linear Search Strategy, then, it is an equilibrium where all the players use the Linear Search Strategy.

When p_i is smaller than one for all i, the upperbound $\mathring{\delta}$ is greater than a half. This shows that the sufficiently patient players (whose optimal strategy is the Binary Search Strategy when learning alone) can switch to the Linear Search Strategy when competing with others. This leads to an inefficient outcome because it takes longer to learn the parameter in expectation when the players use the Linear Search Strategy. This inefficiency prevails for a larger range of the discount parameter when the players are active with equal probabilities. In addition, if the players are active with equal probabilities, increasing the number of players intensifies the inefficiency. That is, all the players using the Linear Search Strategy is an equilibrium for a larger range of the discount parameter. Suppose there are two players who are active in each state with probability 0.9 and 0.1 respectively. Then, the upperbound $\mathring{\delta}$ is 0.53. Thus, both players using the Linear Search Strategy is an equilibrium when the discount parameter δ is weakly smaller than 0.53. If both players are active in each state with probability a half, then, the upperbound $\mathring{\delta}$ is 0.67. In this case, both players using the Linear Search Strategy is an equilibrium when the discount parameter δ is weakly smaller than 0.67. When the players are active with equal probability, both players using the Linear Search Strategy is an equilibrium for a greater range of the discount parameter. When there are I players and all the players are active with equal probability in each state, we have $p_i = \frac{1}{I}$. Since the upperbound $\mathring{\delta} = \frac{I}{I+1}$ increases in I, all the players using the Linear Search Strategy is an equilibrium for a larger discount parameter set when I increases. Increasing the number of players hence intensifies the inefficiency.

Appendix to Chapter 2

1 Extensions

1.1 Search with a Fixed Cost

In this section, the agent does not discount future payoffs with a discount parameter. Instead, she pays a fixed cost $c > 0$ for each search if she does not learn the true parameter. As mentioned in Section 3, the agent's preference is time risk neutral. The agent's optimal strategy is the Binary Search Strategy as long as the fixed cost c is bounded. Otherwise, the agent does not search. The main intuition is that since the agent has to pay a fixed cost c for each search, the agent's only incentive is to learn with the least number of searches. The Binary Search Strategy satisfies this requirement. However, since the agent only gets a reward of one associated with learning, she is thus not willing to search when the fixed cost c is too high. Let $\bar{N}$ denote the initial state, which is the cardinality of the parameter set Θ. Let $\bar{c}(\bar{N}) \equiv \frac{\bar{N}}{(2\bar{N}-2^{\lfloor \log_2 \bar{N} \rfloor+1})(\lceil \log_2 \bar{N} \rceil-1)+(2^{\lfloor \log_2 \bar{N} \rfloor+1}-\bar{N})(\lfloor \log_2 \bar{N} \rfloor-1)}$.

Proposition 2.6. *Given any initial state $\bar{N}$, if the fixed cost $c \leq \bar{c}(\bar{N})$, the agent searches and the optimal strategy is the Binary Search Strategy. Otherwise, the agent does not search. The Linear Search Strategy is always sub-optimal.*

The idea of the proof is the same as in the benchmark case: compare the lifetime utility of the Binary (Linear, resp) Search Strategy and the lifetime utility of the Binary (Linear, resp) Search Deviating Strategies. There is always a Linear Search Deviating Strategy that gives a higher lifetime utility than the Linear Search Strategy, and no Binary Search Deviating Strategy gives a higher lifetime utility than the Binary Search Strategy.

1.2 Relax the Uniform Prior Assumption

Note that the agent's prior belief is $f_0 = (f_0(\theta))_{\theta\in\Theta}$ with $f_0(\theta) \in (0, 1)$ for $\forall\theta \in \Theta$. In the previous discussion, f_0 is assumed to be uniform. That is, $f_0(\theta) = \frac{1}{\bar{N}}$ for $\forall\theta \in \Theta$, where $\bar{N}$ is the cardinality of the parameter set Θ. In this section, I assume that the agent believes that θ_1, the first element in the parameter set, is most likely to be the true parameter, whereas all other parameters in the parameter set are equally likely. That is $f(\theta_1) = f_1$, and $f(\theta) = \frac{1-f_1}{N-1}$ for $\theta \in \Theta \setminus \{\theta_1\}$. I refer to this distribution as a *pseudo-uniform distribution with a peak of* f_1. To keep the model simple, I consider the preference in the benchmark case: expected utility with exponential discounting. In addition, I make the following assumption.

Assumption 2.1. *The probability attached to the element θ_1 is not smaller than a half, which is $f_1 \geq \frac{1}{2}$.*

This assumption guarantees that any revised belief of the agent is either a uniform distribution or a distribution with a *peak*. If this assumption is violated, the revised belief may not describe the fact that the agent believes the first element is *most* likely to be the true parameter, but rather the first element is *least* likely to be the true parameter. This assumption allows me to retain the simplification of the agent's choice in each state. In the previous discussion, the feasible set of the choice in state N is denoted by $\mathcal{F} = \{(m,n) | m+n = N, m, n \in \mathbb{Z}^+ \cap [1,N]\}$. In the benchmark case, due to the symmetry of the uniform distribution, assuming $m \leq n$ is without loss of generality. In this section, with the assumption that $f_1 \geq \frac{1}{2}$, choosing $m > \frac{N}{2}$ is always sub-optimal. For example, if N is even, choosing $m = \frac{N}{2} + 1$ is dominated by choosing $m^{'} = \frac{N}{2} - 1$. The simplification that $m \leq n$ is still without loss of generality.

The main question of interest in this section is to find the conditions under which *the Focal Point Search Strategy* (defined below) is optimal.

Definition 2.4. *The Focal Point Search Strategy is a strategy such that (1) in state N, if the belief of the agent has a peak, then the agent chooses $(m,n) = (1, N-1)$, (2) in state N, if the belief of the agent is uniform, then the agent chooses the Linear Search Policy if $\delta \leq 0.5$ and the Binary Search Policy if $\delta > 0.5$.*

When $\delta \leq 0.5$, the Focal Point Search Strategy coincides with the Linear Search Strategy. When $\delta > 0.5$, the Focal Point Search Strategy can be considered as the agent uses the Linear Search Policy when the belief is not uniform, and then switch to the Binary Search Strategy when the belief becomes uniform. The agent with $\delta > 0.5$ is only willing to use the Linear Search Policy under one situation: when she is quite certain that the first element in the parameter set is the true value of the parameter. Hence, asking the question that under what conditions the Focal Point Search Strategy is optimal, is equivalent to asking how certain the agent has to be so that she is willing to test the first element in the parameter set before doing anything else.

Proposition 2.7. *When $\delta \leq 0.5$, the Focal Point Search Strategy is optimal.*

The intuition is that when the agent uses the Linear Search Strategy, the agent's utility is a convex combination of the payoff of learning the unknown parameter and the discounted continuation value. The continuation value is the same under uniform prior and under the prior with a peak, and it is smaller than the payoff of learning the unknown parameter. When the agent's prior belief is the distribution with a peak, the agent puts higher weight to the payoff of learning the unknown parameter today. Therefore, if it is optimal for the agent to use the Linear Search Strategy when the prior is uniform, it is also optimal for the agent to use the Focal Point Search Strategy when the prior is the distribution with a peak.

Proposition 2.8. *If $\delta \in (0.5, f_1]$, then the Focal Point Search Strategy is optimal.*

This proposition says that given the value of f_1, when the agent is not too patient, that is, when $\delta \leq f_1$, the Focal Point Search Strategy is optimal. The other way to interpret this proposition is that given the value of the agent's discount parameter $\delta > \frac{1}{2}$, if the agent is sufficiently certain that the first element is the true value of the parameter, that is, $f_1 \geq \delta$, then, he/she is willing to use the Linear Search Policy to check the first element in the parameter set, and then uses the Binary Search Strategy if the first element is not the true value of the parameter.

This proposition provides a sufficient condition under which the Focal Point Search Strategy is optimal. It does not characterise the optimal strategy when $\delta > f_1$. Even with this seemingly simple prior assumption, it is still hard to get the characterisation of the optimal strategy for any given value of δ and f_1. However, the optimal strategy under some extreme cases may be helpful to understand the agent's behaviour. When the agent has $\delta = 1$, she is indifferent between the Linear Search Strategy, the Binary Search Strategy and the Focal Point Search Strategy because all the future payoffs are perceived to be the same. Whenever and however the agent learns the unknown parameter, he/she gets a payoff of one. This shows that increasing δ does not monotonically increase the agent's incentive to use the Binary Search Strategy, otherwise, when δ takes the highest value, the optimal strategy must be unique: the Binary Search Strategy.

I may make the *conjecture* that when δ is close to one, it might be optimal for the agent to use the Focal Point Search Strategy. The intuition is that when the agent is very patient, the agent does not mind using one period of Linear Search Policy to test the most likely element. But, this is just a conjecture, there is yet no proof to support this conjecture.

2 Proofs

2.1 Proof of Lemma 2.1

The value associated with the Linear Search Strategy is

$$\begin{aligned} V^L(N) &= \frac{1}{N} + \delta\frac{N-1}{N}V^L(N-1) \\ &\dots \\ &= \frac{1}{N} + \delta\frac{1}{N} + \delta^2\frac{1}{N} + \cdots + \delta^{N-2}\frac{2}{N}V^L(2) \\ &= \frac{1}{N}\Big(\sum_{i=0}^{N-2}\delta^i + \delta^{N-2}\Big) = \frac{1}{N}\Big(\frac{1-\delta^{N-1}}{1-\delta} + \delta^{N-2}\Big) \end{aligned}$$

To derive the value associated with the Binary Search Strategy, I use mathematical induction. To simplify the notation, I rewrite the value in the following way

$$NV^B(N) = \pi^N\delta^{\tau_1^N} + (N-\pi^N)\delta^{\tau_2^N}. \tag{2.9}$$

I also establish the properties of π^N and τ^N.

Remark 2.1. *The following equations hold.*

$$\begin{aligned} \pi^{2N} &= 2\pi^N \\ \tau_i^{2N} &= \tau_i^N + 1 \text{ for } i = 1, 2 \end{aligned}$$

I first show that if eq. (2.9) is the value function in state N associated with the Binary Search Strategy, then we have $2NV^B(2N)$, the value function in state $2N$ associated with the Binary Search Strategy, following the same functional form as eq. (2.9). The calculation is as follows.

$$\begin{aligned}2NV^B(2N) &= \delta\Big\{NV^B(N) + NV^B(N)\Big\}\\ &= 2\pi^N\delta^{\tau_1^N+1} + (2N - 2\pi^N)\delta^{\tau_2^N+1}\\ &= \pi^{2N}\delta^{\tau_1^{2N}} + (2N - \pi^{2N})\delta^{\tau_2^{2N}}\end{aligned}$$

Before showing the next part of the proof, I establish some properties of π^N and τ^N. Let $K \in \mathbb{Z}^+$.

Remark 2.2. *If $N = 2^K$, the following equations hold.*

$$\begin{aligned}\pi^{2N+1} &= \pi^N + \pi^{N+1} = \pi^{N+1}\\ \tau_1^{2N+1} &= \tau_1^{N+1} + 1 = \tau_1^N + 2\\ \tau_2^{2N+1} &= \tau_2^{N+1} + 1 = \tau_2^N + 1\end{aligned}$$

If $N \in [2^K + 1, 2^{K+1} - 2]$, the following equations hold.

$$\begin{aligned}\pi^{2N+1} &= \pi^N + \pi^{N+1}\\ \tau_i^{2N+1} &= \tau_i^N + 1 = \tau_i^{N+1} + 1 \textit{ for } i = 1, 2\end{aligned}$$

If $N = 2^{K+1} - 1$, the following equations hold.

$$\begin{aligned}N - \pi^N &= (2N + 1) - \pi^{2N+1}\\ \pi^N &= N - 1;\ \pi^{2N+1} = 2N;\ \pi^{N+1} = 0\\ \tau_1^N &= \tau_1^{N+1} = \tau_2^{N+1}\\ \tau_1^{2N+1} &= \tau_1^N + 1 = \tau_1^{N+1} + 1\\ \tau_2^{2N+1} &= \tau_2^N + 1 = \tau_2^{N+1}\end{aligned}$$

Next, I show that If the value function in state N associated with the Binary Search Strategy follows the functional form eq. (2.9), and the value function in state $N + 1$ associated with the Binary Search Strategy follows the functional form eq. (2.9), then, the value function in state $2N+1$ associated with the Binary Search Strategy follows the same function form as eq. (2.9). The calculation is as follows.

$$\begin{aligned}(2N + 1)V^B(2N + 1) &= \delta\Big\{NV^B(N) + (N + 1)V^B(N + 1)\Big\}\\ &= \pi^N\delta^{\tau_1^N+1} + (N - \pi^N)\delta^{\tau_2^N+1} + \pi^{N+1}\delta^{\tau_1^{N+1}+1} + (N + 1 - \pi^{N+1})\delta^{\tau_2^{N+1}+1}\end{aligned}$$

If $N = 2^K$, then

$$\begin{aligned}(2N+1)V^B(2N+1) &= N\delta^{\tau_2^N+1} + \pi^{N+1}\delta^{\tau_1^{N+1}+1} + (N+1-\pi^{N+1})\delta^{\tau_2^{N+1}+1} \\ &= N\delta^{\tau_2^{2N+1}} + \pi^{2N+1}\delta^{\tau_1^{2N+1}} + (N+1-\pi^{2N+1})\delta^{\tau_2^{2N+1}} \\ &= \pi^{2N+1}\delta^{\tau_1^{2N+1}} + (2N+1-\pi^{2N+1})\delta^{\tau_2^{2N+1}}\end{aligned}$$

follows the same functional form as eq. (2.9). If $N \in [2^K+1, 2^{K+1}-2]$, then

$$\begin{aligned}(2N+1)V^B(2N+1) &= \pi^N\delta^{\tau_1^{2N+1}} + (N-\pi^N)\delta^{\tau_2^{2N+1}} + \pi^{N+1}\delta^{\tau_1^{2N+1}} + (N+1-\pi^{N+1})\delta^{\tau_2^{2N+1}} \\ &= \pi^{2N+1}\delta^{\tau_1^{2N+1}} + (2N+1-\pi^{2N+1})\delta^{\tau_2^{2N+1}}\end{aligned}$$

follows the same functional form as eq. (2.9). If $N = 2^{K+1}-1$, then

$$\begin{aligned}(2N+1)V^B(2N+1) &= \pi^N\delta^{\tau_1^N+1} + (N-\pi^N)\delta^{\tau_2^N+1} + (N+1)\delta^{\tau_2^{N+1}+1} \\ &= (N-1)\delta^{\tau_1^{2N+1}} + (2N+1-\pi^{2N+1})\delta^{\tau_2^{2N+1}} + (N+1)\delta^{\tau_1^{2N+1}} \\ &= \pi^{2N+1}\delta^{\tau_1^{2N+1}} + (2N+1-\pi^{2N+1})\delta^{\tau_2^{2N+1}}\end{aligned}$$

follows the same functional form as eq. (2.9).

Lastly, I show that when $N \in \{3, 4, 5\}$, the value function in state N associated with Binary Search follows the same functional form as eq. (2.9). If $N = 3$, $3V^B(3) = 1 + 2\delta$ follows eq. (2.9). If $N = 4$, $4V^B(4) = 4\delta$ follows eq. (2.9). If $N = 5$, $5V^B(5) = 3\delta + 2\delta^2$ follows eq. (2.9).

2.2 Proof of Lemma 2.2

To show the properties of $W^L(N)$, It is without loss of generality to treat N as a continuous variable and then compute the first and second order derivatives. The first order derivative of $W^L(N)$ is $(\log\delta)\frac{1-2\delta}{1-\delta}\delta^{N-2}$, and the second order derivative of $W^L(N)$ is $(\log^2\delta)\frac{1-2\delta}{1-\delta}\delta^{N-2}$. When $\delta < \frac{1}{2}$, the first order derivative of $W^L(N)$ is negative and the second order derivative is positive. When $\delta > \frac{1}{2}$, the first order derivative of $W^L(N)$ is positive and the second order derivative is negative. When $\delta = \frac{1}{2}$, $W^L(N)$ is independent of N.

The first and second order difference of $W^B(N)$ is summarised in the following lemma. Let $\triangle W^B(N) := W^B(N) - W^B(N-1)$ and $\triangle^2 W^B(N) := \triangle W^B(N) - \triangle W^B(N-1)$.

Lemma 2.6. *If $N = 2^Y$, then, $\triangle W^B(N) = \delta^{Y-2}(2\delta-1)$ for all $Y \in \mathbb{Z}^+$.*

If $2^Y < N < 2^{Y+1}$, then, $\triangle W^B(N) = \delta^{Y-1}(2\delta-1)$ for all $Y \in \mathbb{Z}^+$.

If $N = 2^Y$ and $2^{Y-1} < N-1 < 2^Y$, then, $\triangle^2 W^B(N) = 0$.

If $2^Y < N < 2^{Y+1}$ *and* $2^Y < N-1 < 2^{Y+1}$*, then,* $\triangle^2 W^B(N) = 0$.
If $2^Y < N < 2^{Y+1}$ *and* $N-1 = 2^Y$*, then,* $\triangle^2 W^B(N) = (2\delta-1)(\delta-1)\delta^{Y-2}$.

Lemma 2.6 is derived directly from calculating the first difference of the function $W^B(\cdot)$. When $\delta > 0.5$ ($\delta < 0.5$), The first-order difference of $W^B(\cdot)$ is positive (negative), and the second-order difference of $W^B(\cdot)$ is non-positive (non-negative). When $\delta = 0.5$, $W^B(N)$ is independent of N.

2.3 Proof of Proposition 2.1

To show that the Linear Search Strategy is optimal, I show that there is no Linear Search Deviating Strategy give the agent a higher payoff than that of the Linear Search Strategy. In state N, the value associated with the Linear Search Deviating Strategy is

$$\begin{aligned}
V^D(N) &= \delta\Big\{\frac{m}{N}V^L(m) + \frac{n}{N}V^L(n)\Big\}\\
&= \delta\Big[\frac{m}{N}\frac{1}{m}\Big(\frac{1-\delta^{m-1}}{1-\delta} + \delta^{m-2}\Big) + \frac{n}{N}\frac{1}{n}\Big(\frac{1-\delta^{n-1}}{1-\delta} + \delta^{n-2}\Big)\Big]\\
&= V^L(N)\frac{2\delta + (1-2\delta)(\delta^{m-1}+\delta^{n-1})}{2\delta + (1-2\delta)(1+\delta^{N-2})}
\end{aligned}$$

Let $M = \frac{2\delta+(1-2\delta)(\delta^{m-1}+\delta^{n-1})}{2\delta+(1-2\delta)(1+\delta^{N-2})}$. I show that $M \leq 1$ when $\delta \leq 0.5$ and $M > 1$ when $\delta > 0.5$. Notice that $2\delta + (1-2\delta)(\delta^{m-1}+\delta^{n-1}) > 0$ and $2\delta + (1-2\delta)(1+\delta^{N-2}) > 0$. I calculate the differences of the denominator and numerator of M.

$$\begin{aligned}
&\Big[2\delta + (1-2\delta)(\delta^{m-1}+\delta^{n-1})\Big] - \Big[2\delta + (1-2\delta)(1+\delta^{N-2})\Big]\\
&= (1-2\delta)\Big[(\delta^{m-1}+\delta^{n-1}) - (1+\delta^{m+n-2})\Big]
\end{aligned}$$

Notice that $(\delta^{m-1}+\delta^{n-1})-(1+\delta^{m+n-2}) = -(1-\delta^{n-1})(1-\delta^{m-1}) < 0$ Therefore, $2\delta+(1-2\delta)(\delta^{m-1}+\delta^{n-1}) > 2\delta+(1-2\delta)(1+\delta^{N-2})$ when $\delta > 0.5$ and $2\delta+(1-2\delta)(\delta^{m-1}+\delta^{n-1}) \leq 2\delta+(1-2\delta)(1+\delta^{N-2})$ when $\delta \leq 0.5$. Therefore, $M \leq 1$ when $\delta \leq 0.5$ and $M > 1$ when $\delta > 0.5$. Thus, $V^L(N) \geq V^D(N)$ if $\delta \leq 0.5$. As a result, a *Linear Search Deviating Strategy* that gives the agent a higher payoff than the Linear Search Strategy does not exist.

To show that the Binary Search Strategy is optimal, I first show that the Binary Search Deviating Strategy such that the $(m,n) \in \mathcal{F}^\dagger$ policy in state N is used gives the agent a lower payoff than the Binary Search Strategy. Since I assume that $m \leq n$, the Binary Search Policy in state N by definition maximises m and minimises n. To show that deviating to other choices $(m,n) \in \mathcal{F}^\dagger$ is not profitable, I show that

$$\delta\Big\{mV^B(m)+nV^B(n)\Big\}\geq\delta\Big\{(m-1)V^B(m-1)+(n+1)V^B(n+1)\Big\}$$

if $\delta\geq 0.5$.

Rearrange the inequality, it is equivalent to show that

$$\triangle W^B(m)\geq\triangle W^B(n+1). \tag{2.10}$$

According to Lemma 2.6, I consider four cases based on the values of m and n.

Case 1: $2^Y<m<2^{Y+1}$ and $2^K-1<n<2^{K+1}-1$. Since $m\leq n$, $\log_2 m\leq\log_2 n$. Then, $\lfloor\log_2 m\rfloor\leq\lfloor\log_2 n\rfloor$, which is $Y\leq K$. Therefore,

$$\delta^{Y-1}(2\delta-1)\geq\delta^{K-1}(2\delta-1).$$

The inequality (2.10) holds if $\delta\geq 0.5$.

Case 2: $m=2^Y$ and $n=2^K-1$. Since $m\leq n$, we have $Y\leq K-1$. Thus $Y-2<K-2$ Therefore,

$$\delta^{Y-2}(2\delta-1)>\delta^{K-2}(2\delta-1).$$

The inequality (2.10) holds if $\delta\geq 0.5$.

Case 3: $2^Y<m<2^{Y+1}$ and $n=2^K-1$. Since $m\leq n$, we have $Y\leq K-1$. Therefore

$$\delta^{Y-1}(2\delta-1)\geq\delta^{K-2}(2\delta-1).$$

The inequality (2.10) holds if $\delta\geq 0.5$.

Case 4: $m=2^Y$ and $2^K-1<n<2^{K+1}-1$. Since $m\leq n$, we have $Y\leq K$. Therefore,

$$\delta^{Y-2}(2\delta-1)>\delta^{K-1}(2\delta-1).$$

The inequality (2.10) holds if $\delta\geq 0.5$.

Next I show that deviating to Linear Search in state N is not profitable. Before showing the result, I introduce the following lemma.

Lemma 2.7. *If $\delta\in[0.5,1)$, $W^B(N)-\delta W^B(N-1)$ weakly increases in N.*

This lemma can be shown given Lemma 2.6

To show that deviating to Linear Search in state N is not profitable, I show that

$$\frac{1}{N} + \frac{N-1}{N}\delta V^B(N-1) \leq V^B(N),$$

which is equivalent to

$$1 \leq NV^B(N) - \delta(N-1)V^B(N-1).$$

Given Lemma 2.7, the minimum of $W(N)$ is 1. Since $W(N)$ is increasing, $W(N) \geq 1$. Therefore, deviating to Linear Search in state N is not profitable if $\delta \geq 0.5$.

To summarise, there is no Binary Search Deviating strategy that gives the agent a higher payoff than the Binary Search Strategy if $\delta \geq 0.5$.

2.4 Proof of Lemma 2.4

In this proof, I find $\phi(\cdot)$ and $u(\cdot)$ such that $\phi\big(\delta^\tau u(x)\big) = \rho^\tau x$. Let $z = \delta^\tau$ and $u(x) = y$. Then,

$$\phi\big(\delta^\tau u(x)\big) = \phi(zy) = z^{\frac{\log\rho}{\log\delta}} u^{-1}(y).$$

Let $t = zy$. Then,

$$\phi(t) = u^{-1}(y)\left(\frac{1}{y}\right)^{\frac{\log\rho}{\log\delta}} t^{\frac{\log\rho}{\log\delta}} := at^{\frac{\log\rho}{\log\delta}}$$

where $a := u^{-1}(y)\left(\frac{1}{y}\right)^{\frac{\log\rho}{\log\delta}}$. Given the functional form of $\phi(\cdot)$, next, I find the functional form of $u(\cdot)$ such that

$$a\left[\delta^\tau u\left(x\right)\right]^{\frac{\log\rho}{\log\delta}} = \rho^\tau x.$$

Then, we have $u(x) = \left(\frac{x}{a}\right)^{\frac{\log\delta}{\log\rho}}$. Let $c := \frac{\log\rho}{\log\delta}$. Then, $\phi(x) = ax^c$ and $u(x) = \left(\frac{x}{a}\right)^{\frac{1}{c}}$.

2.5 Proof of Proposition 2.3

Let $T(\mathcal{W})(N) = \max_{(m,n)\in\mathcal{F}} \zeta\Big(\frac{m}{N}\mathcal{W}(m) + \frac{n}{N}\mathcal{W}(n)\Big)$ and $\mathcal{B}(\mathbb{Z}^+)$ be a space of the bounded functions $\mathcal{W}$: $\mathbb{Z}^+ \to \mathbb{R}$. The operator $T(\mathcal{W})$ is a contraction mapping maps from $\mathcal{B}(\mathbb{Z}^+)$ to $\mathcal{B}(\mathbb{Z}^+)$. The fixed point can be derived based on the discussion in the benchmark case. If $\zeta \in (0, \bar{\zeta})$, the fixed point of the mapping is $\mathcal{W}(N) = V^L(N)$. The corresponding strategy is the Linear Search Strategy. If $\zeta \in (\bar{\zeta}, 1)$, the fixed point of the mapping is $\mathcal{W}(N) = V^B(N)$. The corresponding strategy is the Binary Search Strategy. Since $\alpha \in (0,1)$, the value function $\mathcal{E}(N)$ is maximised when $\mathcal{W}(N)$ is maximised. Therefore, if $\zeta \in (0, \bar{\zeta})$, the optimal strategy to achieve the maximum $\mathcal{E}(N)$ is the Linear Search Strategy. If $\zeta \in (\bar{\zeta}, 1)$, the optimal strategy to achieve the maximum $\mathcal{E}(N)$ is the Binary Search Strategy. Since $\zeta = \delta^{\frac{\alpha}{\rho}}$, given any (α, ρ) pair, there exists a unique threshold $\tilde{\delta} = (\bar{\zeta})^{\frac{\rho}{\alpha}}$ such that if $\delta > \tilde{\delta}$, the Binary Search Strategy is the optimal strategy. If $\delta < \tilde{\delta}$, the Linear Search Strategy is the optimal strategy.

2.6 Proof of Proposition 2.4

Before showing the proofs of the propositions, I first introduce some definitions that will be used in the proofs.

Given a strategy, the agent's state-N self's value function associated with that strategy is related to the value functions of the time-consistent agent. For example, the agent's state-N self's value associated with the Linear Search Strategy $\mathcal{U}^L(N)$ is

$$\mathcal{U}^L(N) = \frac{1}{N} + \beta\delta\frac{N-1}{N}V^L(N-1),$$

where $V^L(\cdot)$ is the value associated with the Linear Search Strategy in the benchmark case (see Lemma 2.1). This is because the present-biased agent perceives all the future payments to be less important than the payment at present. This idea is formally characterised in the following lemma. Let $\mathcal{S}$ be a strategy and let $\mathcal{U}^{\mathcal{S}}(N)$ be the agent's state-N self's value function associated with this strategy. Let $V^{\mathcal{S}}(\cdot)$ be the time-consistent agent's value function associated with the strategy $\mathcal{S}$, and let $(m^{\mathcal{S}}, n^{\mathcal{S}})$ denote the strategy $\mathcal{S}$ induced policy in state N.

Lemma 2.8. *The present-biased agent's state-N self's value function associated with the strategy $\mathcal{S}$ is*

$$\mathcal{U}^{\mathcal{S}}(N) = \begin{cases} \frac{1}{N} + \beta\delta\frac{N-1}{N}V^{\mathcal{S}}(N-1) & \text{if } (m^{\mathcal{S}}, n^{\mathcal{S}}) = (1, N-1), \\ \beta\delta\{\frac{m^{\mathcal{S}}}{N}V^{\mathcal{S}}(m^{\mathcal{S}}) + \frac{n^{\mathcal{S}}}{M}V^{\mathcal{S}}(n^{\mathcal{S}})\} & \text{otherwise.} \end{cases}$$

This lemma can be used to check the optimality of the strategy $\mathcal{S}$ induced policy $(m^{\mathcal{S}}, n^{\mathcal{S}})$ for the agent's state-N self, and hence the optimality of the strategy $\mathcal{S}$ for the present-biased agent.

Definition 2.5. *The strategy $\mathcal{S}$ induced policy $(m^{\mathcal{S}}, n^{\mathcal{S}})$ in state N is optimal for the agent's state-N self if it is not optimal for the agent's state-N self to deviate to other policies given that he will follow the strategy $\mathcal{S}$ in the future.*

Then, the optimality of the strategy $\mathcal{S}$ is defined in the following way.

Definition 2.6. *The strategy $\mathcal{S}$ is optimal for the present-biased agent, if, for all N, the strategy $\mathcal{S}$ induced policy $(m^{\mathcal{S}}, n^{\mathcal{S}})$ in state N is optimal for the agent's state-N self.*

The optimality of a strategy can be shown using the same one-step deviation principle as in the benchmark case. The difference is that in the benchmark case, since the agent is time-consistent, it can be considered as the agent's state-N selves are the same for all N. To show the optimality of the Linear Search Strategy (Binary Search Strategy, resp), it is thus sufficient to show that no Linear Search Deviating Strategy (Binary Search Deviating Strategy, resp) is beneficial in an arbitrary state N. However, when the agent is present-biased, the agent's preferences are different in each state N, to show the optimality of a strategy, we need to check that for all N, the agent's state-N self does not want to deviate. This idea coincides with the backward induction. The agent's state-N self chooses the optimal policy in state N given that for all $m < N$, his state-m selves use the optimal policies in state m.

I first show the first bullet point of Proposition 2.4. I ask the question: if the present-biased agent believes that his future selves will use the Linear Search Policy, what is the smallest state $\underline{N}$ such that the agent's state-$\underline{N}$ self finds it beneficial to use other policies in state $\underline{N}$? If the state $\underline{N}$ does not exist, then I have shown that the first bullet point of Proposition 2.4 is true. This proof uses the idea of backward induction. Since the agent does not make effective decisions in state $N \leq 3$, where the Linear Search Policy the Binary Search Policy coincide, it can be considered as the agent uses Linear Search Policy in state $N \leq 3$. By backward induction, the optimal policy in state 4 gives the agent's state-4 self the highest value function given that the agent's future selves will use the Linear Search Policy. In state 4, given that the agent's future selves will use the Linear Search Policy, the agent's state-4 self's value function associated with Linear Search is $\mathcal{U}^L(4) = \frac{1}{4} + \delta\beta\frac{3}{4}V^L(3)$, where $V^L(\cdot)$ is the value function associated with the Linear Search Strategy when the agent is time consistent (see Lemma 2.1). If the agent's state-4 self uses some other policies, which can only be Binary Search in state 4, given the agent's future selves use the Linear Search Policy, then the value function $\mathcal{U}^D(\cdot)$ associated with this deviation is $\mathcal{U}^D(4) = \beta\delta V^L(2)$. By deviating to the Binary Search Policy in state 4, the agent gives up the payment of $\frac{1}{4}$ at present, and increases the discounted continuation value by $\delta[V^L(2) - \frac{3}{4}V^L(3)]$. Proposition 2.1 implies that when the agent is time consistent and has the discount parameter $\delta < \bar{\delta}$, the increasing of the discounted continuation value is smaller than the payment the agent gives up today, and the Linear Search Policy is hence optimal in state 4. When the agent is present-biased, he perceives the increasing of the discounted continuation value to be even smaller than the real discounted continuation value. The Linear Search Policy is thus also optimal in state 4 for the present-biased agent's state-4 self. This argument can be generalised to the present-biased agent's state-N self for all N. The Linear Search Strategy is thus optimal for the present-biased agent with $\delta < \bar{\delta}$ and $\beta \in (0, 1)$.

To find the optimal strategy of the present-biased agent with $\delta > \bar{\delta}$, I ask the following two questions:

- If the present-biased agent's future selves will use the Binary Search Policy, what is the smallest state $\underline{N}$ such that the agent's state-$\underline{N}$ self finds it beneficial to use other policies in state $\underline{N}$?
- If the present-biased agent's future selves will use the Linear Search Policy, what is the smallest state $\underline{N}$ such that the agent's state-$\underline{N}$ self finds it beneficial to use other policies in state $\underline{N}$?

Answering the first question is useful to show the second bullet point of Proposition 2.4, while answering the second question is useful to show the third bullet point. If the smallest state $\underline{N}$ in the first question does not exist, then the Binary Search Strategy is optimal.

The proof uses the idea of backward induction. Since the agent does not make any effective decisions in state $N \leq 3$, and the Binary Search Policy and the Linear Search Policy coincide, it can be considered as the agent uses Binary Search Policy in state $N \leq 3$. The optimal policy in state 4 gives the agent's state-4 self the highest value function given that the agent's future selves will use the Binary Search Policy. In state 4, given that the agent's future selves will use the Binary Search Policy, the agent's state-4 self's value function associated with Binary Search is $\mathcal{U}^B(4) = \beta\delta V^L(2) = \beta V^B(4)$, where $V^B(\cdot)$ is the value function associated with the Binary Search Strategy when the agent is time consistent (see Lemma 2.1). Proposition 2.1 implies that deviating to any policy that gives the agent zero payment today is not beneficial. Therefore, only the Linear Search Policy should be considered. If the agent's state-4 self believes that his future selves will use the Binary Search Policy, and he uses Linear Search Policy in state 4, the value function is $\mathcal{U}^D(4) = \frac{1}{4} + \beta\delta\frac{3}{4}V^B(3)$. It can be regarded as the agent asks himself this question: in comparison with always using Binary Search Policy, is it beneficial for me to postpone the Binary Search to tomorrow and use Linear Search today? The benefit of using the Linear Search Policy today is the positive expected payment $\frac{1}{4}$, and the cost of using the Linear Search Policy is from the delay of Binary Search, which is $V^B(4) - \delta\frac{3}{4}V^B(3)$. Since this cost is future cost, the present-biased agent perceives the cost as $\beta[V^B(4) - \delta\frac{3}{4}V^B(3)]$. In state 4, delaying the Binary Search to tomorrow is not beneficial when the agent's state-4 self's perceived cost is greater than the benefit, that is, when $\beta > \bar{\beta}^4$, where $\bar{\beta}^4 = \frac{\frac{1}{4}}{V^B(4)-\delta\frac{3}{4}V^B(3)}$ is the ratio of the benefit to the cost of delaying the Binary Search Policy and use the Linear Search Policy in state 4 instead. The discussion above shows that when the present-biased agent has $\delta > \bar{\delta}$ and $\beta > \bar{\beta}^4$, the Binary Search Policy is optimal in state 4 for the agent's state-4 self. Using the idea of backward induction, the Binary Search Policy can be shown to be optimal in state 5 for the agent's state-5 self if $\beta > \bar{\beta}^5$. The Binary Search Policy is thus optimal for the present-biased agent in all the states up to N if $\beta > \max\{\bar{\beta}^m\}_{m=\{4,5,\ldots,N\}}$, where $\bar{\beta}^m := \frac{\frac{1}{m}}{V^B(m)-\delta\frac{m-1}{m}V^B(m-1)}$ is the ratio of the benefit to the cost of using the Linear Search Policy in state m given that the agent's future selves will use the Binary Search Policy. The cost $V^B(m) - \delta\frac{m-1}{m}V^B(m-1)$ is increasing in m because the value function $V^B(\cdot)$ is concave, and the benefit $\frac{1}{m}$ is decreasing in m. The value of $\bar{\beta}^m$ is thus decreasing in m. Therefore, if $\beta > \bar{\beta}^4$, then the value of β is greater than $\bar{\beta}^N$ for all $N > 4$. The intuition is that delaying the Binary Search Policy to tomorrow is the most beneficial for the present-biased agent's state-4 self. If the present-biased agent's state-4 self finds it optimal to use the Binary Search Policy, then all the present-biased agent's selves will find it optimal to use the Binary Search Policy. When the present-biased agent is not too present-biased, that is, when β is big enough, the present-biased agent has the same optimal strategy as the time-consistent agent.

In state N, if the agent's future selves all use the Linear Search Policy, and the agent's state-N self uses the $(m,n) \in \mathcal{F}^\dagger$ policy, the highest payoff from these policies is

$$\beta P(N) \equiv \beta \max_{\{m,n\}\in\mathcal{F}^\dagger} \delta\left\{\frac{m}{N}V^L(m) + \frac{n}{N}V^L(n)\right\}$$

where $V^L(\cdot)$ is value associated with the Linear Search Strategy (see Lemma 2.1). When $\delta > \bar{\delta}$, it has been shown that the maximum value is achieved at the Binary Search Policy in state N. Following the backward induction method, consider the agent's state-4 self. Since Linear Search and Binary Search coincide in states smaller than 4, it can be considered as the agent's future selves use Linear Search Policy in each future states. If the agent uses the Linear Search Policy in state 4, then the expected payoff is $\frac{1}{4} + \beta\delta\frac{3}{4}V^L(3)$. If the agent uses the Binary Search Policy in state 4, then the expected payoff is $\beta P(4)$. Then, given that the agent's future selves all use the Linear Search Policy, the agent's state-4 self uses the Binary Search Policy if $\beta \geq \tilde{\beta}^4$ where $\tilde{\beta}^4 \equiv \frac{\frac{1}{4}}{P(4)-\delta\frac{3}{4}V^L(3)}$ is the ratio of the benefit to the cost of using the Linear Search Policy in state 4. Let $\tilde{\beta}^N \equiv \frac{\frac{1}{N}}{P(N)-\delta\frac{N-1}{N}V^L(N)}$ be the ratio of the benefit to the cost of using the Linear Search Policy in state N. Note that $\tilde{\beta}^4 = \bar{\beta}^4$ because Linear Search and Binary Search coincide in states smaller than 4.

When $\beta < \tilde{\beta}^4$, the agent uses the Linear Search Policy in state 4 given that his future selves also use Linear Search Policy. Then, consider the optimal policy in state 5 given that the agent's all future selves use the Linear Search Policy. Following the same calculation as in state 4, the agent uses the Binary Search Policy if $\beta \geq \tilde{\beta}^5$, and uses the Linear Search Policy if $\beta < \tilde{\beta}^5$.

Lemma 2.9. *The ratio of the benefit to the cost of using the Linear Search Policy in state N given that the agent's future selves all use the Linear Search Policy $\tilde{\beta}^N$ is decreasing in N.*

This is because the benefit of using the Linear Search Policy in state N is decreasing in N, and due to the concavity of the function $V^L(\cdot)$, the cost of using the Linear Search Policy is increasing in N. As a result, $\tilde{\beta}^N$ is decreasing in N.

Given this lemma, in state 5, it can be concluded that if the agent has the present-biased parameter $\beta \in [\tilde{\beta}^5, \tilde{\beta}^4)$, it is optimal for him to use the Binary Search Policy in state 5, and to use the Linear Search Policy in all future states. For $\beta < \tilde{\beta}^5$, I can keep discussing the agent's policy in state 6 using the same approach. Because of the decreasing property of $\tilde{\beta}^N$, it will be the case that if the agent has the present-biased parameter $\beta \in [\tilde{\beta}^N, \tilde{\beta}^{N-1})$, it is optimal for him to use the Binary Search Policy in state N, and to use the Linear Search Policy in all future states.

2.7 Proof of Proposition 2.5

Suppose all the players use the Linear Search Strategy. In state N, prior to knowing whether she is active or not, if player i uses the Linear Search Strategy when she is active, her value is

$$V_i^L(N) = p_i\left\{\frac{1}{N} + \delta\frac{N-1}{N}V_i^L(N-1)\right\} + (1-p_i)\left\{\delta\frac{N-1}{N}V_i^L(N-1)\right\}.$$

Given the initial condition $V_i^L(2) = p_i$, it can be computed that

$$V_i^L(N) = \frac{1}{N} p_i \left[\frac{1-\delta^{N-1}}{1-\delta} + \delta^{N-2} \right].$$

If player i uses the $(m, n) \in \mathcal{F}^\dagger$ policy in state N, and uses the Linear Search Policy in all other states, prior to knowing whether she is active or not, her value is

$$V_i^D(N) = p_i \left\{ \delta \frac{m}{N} V_i^L(m) + \delta \frac{n}{N} V_i^L(n) \right\} + (1-p_i) \left\{ \delta \frac{N-1}{N} V_i^L(N-1) \right\}.$$

Next, I show that when $\delta \leq \frac{1}{1+p_i}$, we have $V_i^D(N) \leq V_i^L(N)$. That is, given that all other player uses the Linear Search Strategy, it is optimal for player i to use the Linear Search Strategy (if she is active). Let $W_i^L(N) := NV_i^L(N)$ and $W_i^D(N) := NV_i^D(N)$. To show that $V_i^D(N) \leq V_i^L(N)$ is equivalent to show that $W_i^D(N) \leq W_i^L(N)$. We have

$$W_i^D(N) - W_i^L(N) = \delta W_i^L(m) + \delta W_i^L(n) - \delta W_i^L(N-1) - 1.$$

If $\delta > \frac{1}{2}$, $W_i^L(m) + W_i^L(n)$ is maximised at $m = \frac{N}{2}$. If $\delta \leq \frac{1}{2}$, $W_i^L(m) + W_i^L(n)$ is maximised at $m = 2$. When $\delta \leq \frac{1}{2}$,

$$\begin{aligned} W_i^D(N) - W_i^L(N) &< \delta W_i^L(1) + \delta W_i^L(N-1) - \delta W_i^L(N-1) - 1 \\ &= 2\delta p_i \leq p_i < 1. \end{aligned}$$

Therefore, when $\delta \leq \frac{1}{2}$, there exists an equilibrium where all the players use the Linear Search Strategy. When $\delta > \frac{1}{2}$,

$$\begin{aligned} W_i^D(N) - W_i^L(N) &\leq \delta 2 W_i^L(\frac{N}{2}) - \delta W_i^L(N-1) - 1 \\ &= \frac{\delta + (2\delta - 1)\left(\delta^{N-2} - 2\delta^{\frac{N}{2}-1}\right)}{1-\delta} p_i - 1 \\ &< \frac{\delta}{1-\delta} p_i - 1. \end{aligned}$$

If $\delta < \frac{1}{1+p_i}$, we have $W_i^D(N) - W_i^L(N) < 0$. As a result, if $\frac{1}{2} < \delta < \frac{1}{1+p_i}$, given that all the players use the Linear Search Strategy, it is optimal for player i to use the Linear Search Strategy.

2.8 Proof of Proposition 2.6

In this proof, I first derive the lifetime utility associated with the Linear Search Strategy and the Binary Search Strategy. Next, I show that there is always a Linear Search Deviating Strategy that gives a higher lifetime utility than the Linear Search Strategy. The Linear Search Strategy is hence always sub-optimal. Then, I show that no Binary Search Deviating Strategy gives a higher lifetime utility than the Binary Search Strategy, and hence the Binary Search Strategy is optimal.

To derive the lifetime utility associated with the Linear Search Strategy, I first write down the Bellman equation

$$\mathcal{S}^L(N) = \frac{1}{N} + \frac{N-1}{N}(\mathcal{S}^L(N-1) - c).$$

Iterate backwards and plug in the initial condition $\mathcal{S}(2) = 1$, we have

$$\begin{aligned}\mathcal{S}^L(N) &= \frac{1}{N} + \frac{N-1}{N}\left(\frac{1}{N-1} + \frac{N-2}{N-1}\big(\mathcal{S}^L(N-2) - c\big) - c\right)\\ &= 1 - \frac{(N+1)(N-2)}{2N}c\end{aligned}$$

By using the Binary Search Strategy, the agent learns the state after $\lceil \log_2(N) \rceil - 1$ periods with probability $\frac{2N - 2^{\lfloor \log_2 N \rfloor + 1}}{N}$ and learns the state after $\lfloor \log_2(N) \rfloor - 1$ periods with probability $\frac{2^{\lfloor \log_2 N \rfloor + 1} - N}{N}$. The lifetime utility associated with the Binary Search Strategy is hence

$$\begin{aligned}\mathcal{S}^B(N) &= \frac{2N - 2^{\lfloor \log_2 N \rfloor + 1}}{N}\big(1 - (\lceil \log_2(N) \rceil - 1)c\big) + \frac{2^{\lfloor \log_2 N \rfloor + 1} - N}{N}\big(1 - (\lfloor \log_2(N) \rfloor - 1)c\big)\\ &= 1 - \frac{(2N - 2^{\lfloor \log_2 N \rfloor + 1})(\lceil \log_2 N \rceil - 1) + (2^{\lfloor \log_2 N \rfloor + 1} - N)(\lfloor \log_2 N \rfloor - 1)}{N}c.\end{aligned}$$

Next, I show that there is always a Linear Search Deviating Strategy that gives a higher lifetime utility than the Linear Search Strategy if $c > 0$. The Linear Search Deviating Strategy is to choose $(m, n) \in \mathcal{F}^\dagger$ in state N, and use the Linear Search Policy in all other states. Let $\mathcal{S}^D(N)$ be the lifetime utility associated with the Linear Search Deviating Strategy in state N, then

$$\begin{aligned}\mathcal{S}^D(N) &= \frac{m}{N}\mathcal{S}^L(m) + \frac{n}{N}\mathcal{S}^L(n) - c\\ &= 1 - \frac{m^2 + n^2 + N - 4}{2N}c\end{aligned}$$

If N is even, $\mathcal{S}^D(N)$ is maximised at $m = \frac{N}{2}$, where

$$\max \mathcal{S}^D(N) = 1 - \frac{N^2 + 2N - 8}{4N} c > \mathcal{S}^L(N)$$

if $N > 2$. If N is odd, $\mathcal{S}^D(N)$ is maximised at $m = \frac{N-1}{2}$, where

$$\max \mathcal{S}^D(N) = 1 - \frac{N^2 + 2N - 7}{4N} c > \mathcal{S}^L(N)$$

if $N > 3$. Therefore, whenever the Binary Search Strategy and the Linear Search Strategy does not coincide, there is always a Linear Search Deviating Strategy that gives a higher lifetime utility than the Linear Search Strategy if $c > 0$. Linear Search is hence sub-optimal.

Then, I show that no Binary Search Deviating Strategy gives a higher lifetime utility than the Binary Search Strategy. I first derive the expression of $m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1)$, which will be useful for the rest of the proof. I will show that one-step deviation to the Linear Search Policy in state N is not profitable, and then I show that one-step deviation to $(m, n) \in \mathcal{F}^\dagger$ in state N is not profitable. Lastly, I show that search happens in state N when the cost c is smaller than a threshold $\bar{c}(N)$.

Let $K \in \mathbb{Z}^+$ and $K \geq 2$ be a constant.

Lemma 2.10. *If* $m = 2^K$,

$$m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) = 1 - Kc.$$

If $m \in [2^K + 1, 2^{K+1} - 1] \cap \mathbb{Z}^+$,

$$m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) = 1 - (K+1)c.$$

Proof. We have

$$m\mathcal{S}^B(m) = m - (2m - 2^{\lfloor \log_2 m \rfloor + 1})(\lceil \log_2 m \rceil - 1) + (2^{\lfloor \log_2 m \rfloor + 1} - m)(\lfloor \log_2 m \rfloor - 1)c.$$

Then, the difference is

$$\begin{aligned} m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) = 1 - \Bigg[& m\big(2(\lceil \log_2 m \rceil - 1) - \lfloor \log_2 m \rfloor - 1\big) \\ & - (m-1)\big(2(\lceil \log_2 (m-1) \rceil - 1) - \lfloor \log_2 (m-1) \rfloor - 1\big) \\ & - 2^{\lfloor \log_2 m \rfloor + 1}\big(\lceil \log_2 m \rceil - \lfloor \log_2 m \rfloor\big) \\ & + 2^{\lfloor \log_2 (m-1) \rfloor + 1}\big(\lceil \log_2 (m-1) \rceil - \lfloor \log_2 (m-1) \rfloor\big)\Bigg] c \end{aligned}$$

Case 1. First consider the case that $m = 2^K$. In this case, $\lceil \log_2 m \rceil = \lfloor \log_2 m \rfloor = \lceil \log_2 (m-1) \rceil = K$, and $\lfloor \log_2 (m-1) \rfloor = K - 1$. Then,

$$m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) = 1 - Kc.$$

Case 2. Then consider the case that $m = 2^K + 1$. In this case, $\lceil \log_2 m \rceil = K+1$, and $\lfloor \log_2 m \rfloor = \lceil \log_2 (m-1) \rceil = \lfloor \log_2 (m-1) \rfloor = K$. Then,

$$m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) = 1 - (K+1)c.$$

Case 3. Finally, consider the case that $m \in [2^K + 2, 2^{K+1} - 1] \cap \mathbb{Z}^+$. In this case, $\lceil \log_2 m \rceil = \lceil \log_2 (m-1) \rceil = K + 1 = K + 1$, and $\lfloor \log_2 m \rfloor = \lfloor \log_2 (m-1) \rfloor = K$. Then,

$$m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) = 1 - (K+1)c.$$

□

Next, I show that one-step deviation to the Linear Search Policy in state N is not profitable. Let $D^L(N)$ be the lifetime utility in state N given that the agent uses the Binary Search Deviating Strategy and chooses the Linear Search Policy in state N, We have

$$D^L(N) = \frac{1}{N} + \frac{N-1}{N}\Big(\mathcal{S}^B(N-1) - c\Big).$$

To show that the one-step deviation is not profitable is equivalent to show $ND^L(N) < N\mathcal{S}^B(N)$, which is equivalent to show $1 - (N-1)c < N\mathcal{S}^B(N) - (N-1)\mathcal{S}^B(N-1)$. According to Lemma 2.10, the inequality holds when $N > 2$.

Then, I show that one-step deviation to $(m, n) \in \mathcal{F}^\dagger$ in state N is not profitable. Let $D^P(N)$ be the lifetime utility in state N if the agent uses the

Binary Search Deviating Strategy and chooses $(m,n) \in \mathcal{F}^\dagger$ in state N, then

$$D^P(N) = \max_{(m,n)\in\mathcal{F}^\dagger} \{\frac{m}{N}\mathcal{S}^B(m) + \frac{n}{N}\mathcal{S}^B(n) - c\}$$

If I can show that $D^P(N) = \mathcal{S}^B(N)$ when the agent chooses Binary Search in state N, then there is no profitable Binary Search Deviating Strategy. To show this, I introduce the following corollary.

Corollary 2.1. *For $(m,n) \in \mathcal{F}^\dagger$, the following inequality holds.*

$$m\mathcal{S}^B(m) - (m-1)\mathcal{S}^B(m-1) \geq (n+1)\mathcal{S}^B(n+1) - n\mathcal{S}^B(n)$$

Proof. Given Lemma 2.10, I consider four cases. Let $K, J \in \mathbb{Z}^+$ and $K, J \geq 2$ be two constants. **Case 1.** First consider the case that $m = 2^K$ and $n+1 = 2^J$. The left-hand side of the inequality is $1 - Kc$ and the right-hand side of the inequality is $1 - Jc$. Since $m \leq n$, we have $K < J$. Therefore, the inequality holds with the strict inequality. **Case 2.** Next, consider the case that $m = 2^K$ and $n+1 \in [2^J+1, 2^J-1]$. The left-hand side of the inequality is $1 - Kc$ and the right-hand side of the inequality is $1-(J+1)c$. Since $m \leq n$, we have $K < J$. Therefore, the inequality holds with the strict inequality. **Case 3.** Next, consider the case that $m \in [2^K+1, 2^K-1]$ and $n+1 = 2^J$. The left-hand side of the inequality is $1-(K+1)c$ and the right-hand side of the inequality is $1 - Jc$. Since $m \leq n$, we have $m < n+1$ and hence $2^{K+1} < 2^J + 1$. Since $m, n \in \mathbb{Z}^+$, it must be that $K+1 \leq J$. Therefore, the inequality holds with the weak inequality. **Case 4.** Lastly, consider the case that $m \in [2^K+1, 2^K-1]$ and $n+1 \in [2^J+1, 2^J-1]$. The left-hand side of the inequality is $1-(K+1)c$ and the right-hand side of the inequality is $1-(J+1)c$. Since $m \leq n$, we have $K \leq J$. Therefore, the inequality holds with the weak inequality. □

Given Corollary 2.1, the following inequality holds

$$m\mathcal{S}^B(m) + n\mathcal{S}^B(n) \geq (m-1)\mathcal{S}^B(m-1) + (n+1)\mathcal{S}^B(n+1).$$

Therefore, $D^P(N)$ is achieved by using the Binary Search Policy in state N. Therefore, there is no profitable Binary Search Deviating Strategy.

Since the agent only searches when learning gives the agent a non-negative payoff. That is $V^B(N) \geq 0$. Therefore, the fixed cost c has to be smaller than the threshold $\bar{c}(N)$ so that the agent starts searching.

2.9 Proof of Proposition 2.8

This proof follows the following steps. I first write down the Bellman equation. Next, I compute the value associated with the Focal Point Search Strategy. Then, I derive a sufficient condition under which there is no one-step deviation strategy that gives the agent a higher payoff than the Focal Point Search Strategy.

The Bellman equation consists of the payoff at the current period and the continuation value. The agent's revised belief determines the continuation value. If the prior belief is a distribution with a peak, the revised belief at the next time can be one of the following three distributions: a degenerated distribution, a uniform distribution, or a distribution with a peak with a different support. The shape of the revised belief depends on the agent's policy at that time. If the agent uses the Linear Search Policy, the revised belief will be a degenerated distribution or a uniform distribution. If the agent chooses $(m, n) \in \mathcal{F}^{\dagger}$, the revised belief will be a uniform distribution or a distribution with a peak with a different support. Since the degenerated revised belief means that the agent learns the unknown parameter, the continuation value is hence zero. The positive continuation value thus takes two different functional forms: one corresponding to the uniform revised belief, and the other corresponding to the belief with a peak.

Let $V_p(N)$ be the value function in state N when the belief of the agent is a distribution with a peak of f_1. Let $V_u(N)$ be the value function in state N when the belief of the agent is a uniform distribution. If the agent's belief in state N is the distribution with a peak f_1, the Bellman equation in state N is

$$V_p(N) = \max \left\{ f_1 + (1 - f_1)\delta V_u(N-1), \max_{(m,n)\in\mathcal{F}^{\dagger}} \delta\big\{\mu V_p(m) + (1-\mu)V_u(n)\big\} \right\},$$

with the initial condition $V_p(1) = \frac{1}{\delta}$ and $\mu = f_1 + \frac{1-f_1}{N-1}(m-1)$. The first element is the value associated with the Linear Search Policy in state N, and the second element without the max operator is the value associated with (m, n) in state N. With the max operator, it is the highest value the agent can get by choosing $(m, n) \in \mathcal{F}^{\dagger}$.

Let $V_p^F(N)$ be the value associated with the Focal Point Search Strategy. Then,

$$V_p^F(N) = f_1 + (1 - f_1)\delta V^B(N-1),$$

where $V^B(\cdot)$ is the value function in the benchmark case associated with the Binary Search Strategy (see Lemma 2.1). This is because if the agent uses the

Focal Point Search Strategy and does not learn the unknown parameter, the revised belief becomes the uniform distribution. Then the problem becomes the one that has been discussed in the benchmark case. Since the discount parameter is greater than a half, the corresponding value function is $V^B(\cdot)$.

To check the optimality of the Focal Search Strategy, I derive a sufficient condition under which there is no one-step deviation strategy that gives the agent a higher payoff than the Focal Point Search Strategy.

Definition 2.7. *The Focal Point Search Deviating Strategy is an one-step deviation strategy such that the agent chooses $(m, N-m) \in \mathcal{F}^\dagger$ in state N and uses the Focal Point Search Strategy in all other states.*

Let $G(m)$ be the value associated with the Focal Point Search Deviating Strategy. Then,

$$G(m) = \delta\Big\{\mu V_p^F(m) + (1-\mu)V^B(N-m)\Big\}.$$

Plug in the value function associated with the Focal Point Search Strategy, we have

$$G(m) = \delta\Big\{\mu\Big(f_1 + (1-f_1)\delta V^B(m-1)\Big) + (1-\mu)V^B(N-m)\Big\}.$$

The functional form of $V^B(\cdot)$ is known (see Lemma 2.1). Since $m \geq 2$ and $m \leq N-m$, the upperbound of $V^B(m-1)$ is $V^B(1)$, which is $\frac{1}{\delta}$, and the upperbound of $V^B(N-m)$ is $V^B(2) = 1$. Therefore,

$$G(m) \leq \delta\Big\{\mu\Big(f_1 + (1-f_1)\delta\frac{1}{\delta}\Big) + (1-\mu)\Big\} = \delta.$$

Then, if $\delta \leq f_1$, it is always true that $G(m) \leq V_p^F(N)$. As a consequence, when $\delta \in (\frac{1}{2}, f_1]$, there is no Focal Point Search Deviating Strategy that gives the agent a higher payoff than the Focal Point Search Strategy. The Focal Point Search Strategy is hence optimal.

Chapter 3

Learning How People Learn

1 Introduction

When a payoff-relevant state is unknown, a decision maker (DM) learns from information about the state before making decisions. For example, before buying a product, the consumers watch the advertisement and the introduction of the product, search the internet for reviews about the product, and then decide whether to buy or not. Investors read the financial report and news of a company to check if the company is worth investing in and then decide whether to invest or not. Chain restaurants do market research and use the data from the market research to see whether it is profitable to open a new restaurant. The DM collects information, processes information and learns from the information.

How does a DM learn from the information? One of the canonical theories posits that the DM holds a prior belief about the unknown state. When the DM receives new signals and information about the unknown state, she adjusts her prior belief based on these new observations. This process is referred to as 'learning from the information' in this chapter. Mainstream economic theories impose Bayesian DM assumptions where the DM updates the belief using Bayes rule given some prior. However, the psychology and behavioural economic literature suggest that people tend to process information in a manner that deviates from Bayesian updating rules (see Tversky & Kahneman (1974), Rabin (1998) and Camerer (1998)). Behavioural economists have explored and developed theories regarding alternative belief updating rules. Epstein et al. (2010) provides a non-Bayesian updating rule to capture the underreaction and overreaction to signals. Rabin (1998) analyses the bias in beliefs when the order of the signals matters in how people infer future signals. Besides the theories about the non-Bayesian updating rules, there are some ad hoc tests used to understand how a DM learn from the information. For example, Angrisani et al. (2017) studies how people process information in a social learning network.

The question of interest is whether it is possible that a third party could learn how a DM learns from the information they received by observing what this DM sees and what they do. For example, consider the experimental economists designing an experiment to learn how the experiment participant processes the information. If the designer sees the information received by the participant and the actions taken by the participant, then, is it possible for the designer to learn about the participant's beliefs? Theoretically speaking, answering this question provides a theoretical foundation for the experiments that are interested in detecting how people learn from the information. In addition, answering this question has applications in industry. Firms can benefit from learning how their consumers react to the information they provide. For example, firms can design advertising and free trials better to attract targeted consumers. Central banks can benefit from knowing how the public process the news they release if they can do better forward guidance.

In this chapter, I investigate this question in the following situation. Suppose it is common knowledge that the DM updates their beliefs using Bayes rule but the prior is unknown, I investigate whether it is possible for the designer to learn the DM's prior belief by observing the signals she received and the actions she took. Learning the DM's prior can be considered a starting point for learning how the DM learns due to its tractability. Given that the DM updates beliefs using Bayes rule, if the state is binary, all the designer needs to learn is one parameter. If the state is finite but not binary, the designer just needs to learn a distribution with a finite number of parameters. This is the least demanding situation in terms of what the designer needs to learn. If the designer cannot learn how the DM learns in this situation, then, it is very likely that the designer is not able to learn in a more demanding environment. It is hence considered a starting point.

Whether the designer can learn the DM's prior depends on the variation in the DM's actions and the informativeness of the signals. The variation in the DM's actions determines how much information is involved in the DM's action. For example, if the DM always takes the same action regardless of the beliefs, then, the designer can learn nothing about the DM from observing her actions. If the DM's action is reporting the beliefs directly, it is likely that the designer can learn how the decision maker learns with sufficiently many observations. The informativeness of the signals about the state determines how quickly the DM's actions stabilise. If the signals are very informative, the decision-maker learns the state very fast. Then, there will not be too many variations in the DM's actions eventually. It is likely that the designer cannot learn anything after a certain point. If the signals are not very informative, the DM's beliefs fluctuate a lot. The fluctuation of the beliefs then may induce variations in the actions. Then it is more likely that the designer can learn how the DM learns.

In this chapter, I consider the case where there are two possible states. It is drawn at the beginning and it is constant over time. At each time, the DM can observe a signal about the state and then take an action. The DM's action at each time is binary. There is a designer who can observe the signals received and the action taken by the DM at each time. The designer wants to learn the DM's prior belief. I assume the designer can design the DM's payoff function in order to achieve her objective. An example of this setting is that a DM has some initial opinions on whether being vegan is a good idea or not. She can read the news about being vegan and observe the purchasing history of plant-based meals and then decide whether she wants to purchase plant-based meals. The seller of plant-based meals can choose the price of the plant-based meals, sees what the DM sees on the news and whether the DM purchased the plant-based meal. The seller then tries to recover the DM's initial opinion on whether being vegan is a good idea or not.

1.1 Related Literature

This chapter is related to the experimental economics papers about detecting the bounded rationality of individuals. Angrisani et al. (2017) design the experiment to detect how individuals to update beliefs in a social learning network. The agents in their model act sequentially after observing statements from the neighbours and a private signal about the state. The agents' actions in their model are continuous. This chapter is different from Angrisani et al. (2017). This chapter focuses on asking the question of whether the experiment designer can learn how the agents update their beliefs by observing their actions and the signals. The focus is not on how the individual interprets the signals from different sources. Augenblick & Rabin (2018) examines the time-inconsistency preference of individuals. Their paper focuses on the experiment designer observing the actions to detect the form of the individual's preferences. The aim of their paper is different from this one. This chapter is mainly about beliefs rather than preferences.

Mathematically, this chapter is using the idea of a sequential probability ratio test by Wald (1945). The sequential probability ratio test considers the odds ratio as a function of the probability of each observation and the number of observations. The sequential probability ratio test is for hypothesis testing. The aim of the test is to decide which hypothesis is correct in the shortest time period, i.e. with the smallest amount of observations. There are two thresholds: a lower one and a higher one. The sequential probability ratio test ends if the odds ratio falls below the lower threshold or jumps above the higher threshold. The stopping time of the test follows Wald's identity. It allows us to calculate the probability that the odds ratio hits the two thresholds. This chapter uses Wald's identity to calculate the probability of the designer learning the prior.

2 The Model

2.1 Model Setup

There are two players: a designer and a decision maker (DM). Time $t = 0, 1, 2, \ldots$ is discrete and potentially infinite. The state is drawn from the set $\Theta = \{\theta, \theta'\}$ at time $t = 0$ and is constant over time. Both the designer and the DM are uninformed about the state. The designer's prior belief about the state is denoted $\mu_0 = \Pr(\theta) > 0$ and the DM's prior belief is denoted $p_0 = \Pr(\theta) > 0$. The DM's prior belief is private information and is referred to as the DM's type. The designer believes that p_0 is drawn from P according to distribution Σ_0. I discuss two cases in this chapter. One is when the designer believes there are two types and the other is when the designer believes there are continuous types. In the two-type case, $P = \{\underline{p_0}, \bar{p_0}\}$ and the probability mass function σ_0 of Σ_0 is such that $\sigma_0\left(p_0\right) > 0$ for $\forall p_0 \in P$. In the continuous-type case, $P = [0, 1]$ and Σ_0 is a uniform distribution with support P.

At time $t = 0$, the designer designs the DM's payoff function $U : \{0, 1\} \times \Theta \to \mathbb{R}^+$. The DM observes this payoff function and takes an action $\alpha_0 \in \{0, 1\}$. At each time $t > 0$, the DM observes a signal $s_t \in \{1, 2, , \ldots, S\} := \mathcal{S}$ about the state and then takes an action $\alpha_t \in \{0, 1\}$. The signal structure is exogenous. If the state is θ, the DM observes signal $s_t = s$ with probability $\Pr(s_t = s|\theta) = \pi_s > 0$ where $\pi := \left(\pi_s\right)_{s\in\mathcal{S}}$ is the distribution of the signals in state θ. If the state is $\theta^{'}$, the DM observes signal $s_t = s$ with probability $\Pr(s_t = s|\pi') = \pi'_s > 0$ where $\pi' := \left(\pi'_s\right)_{s\in\mathcal{S}}$ is the distribution of the signals in state $\theta^{'}$. Let $h_t = (s_0, s_1, \ldots, s_{t-1}) \in s^t$ denote the history of signals. The DM observes h_t and forms a posterior belief $p_t = \Pr(\theta|h_t)$ using Bayes rule. The DM's objective is to maximise her time-t utility by taking action $\alpha_t \in \{0, 1\}$ at each time t. The payoff at time t is not observed by the DM until the experiment ends. The designer's objective is to learn the DM's type. At time $t > 0$, the designer observes the signal s_t and the action α_t, and then updates her belief about the

DM's type. If the designer learns the type of DM, the experiment ends. If the designer does not learn the type of DM, the experiment continues to the next period.

The notion of 'learning the DM's type' When $P = \{\underline{p_0}, \bar{p_0}\}$, the designer learns the DM's type when the designer believes that the DM's type must be $\underline{p_0}$ but not $\bar{p_0}$, or the other way around. That is, the designer's posterior belief is such that $\sigma_t\left(\underline{p_0}\right) = 1$ or $\sigma_t\left(\bar{p_0}\right) = 1$. This notion of learning can easily be extended to the case where P has a finite cardinality and $\sigma_0\left(p_0\right) > 0$ for $\forall p_0 \in P$. However, it is less straightforward in the case when there are continuous types. I introduce the notion of learning in the continuous case in Section 3.

The simplification of the designer's action Given the binary actions and binary states, instead of designing the utility function $U : \{0,1\} \times \Theta \to \mathbb{R}^+$, the designer's action can be simplified to choosing a threshold

$$r := \frac{U\left(0, \theta^{'}\right) - U\left(1, \theta^{'}\right)}{U\left(1, \theta\right) - U\left(0, \theta\right) + U\left(0, \theta'\right) - U\left(1, \theta'\right)} \in (0,1)$$

such that it is optimal for the DM to take action $\alpha_t = 1$ if and only if $p_t \geq r$.

2.2 An Example

The designer has two kinds of biased coins: A and B. The name of the coin is the 'state' in the model. Coin A comes up Heads with probability $\frac{2}{3}$ and Tails with probability $\frac{1}{3}$. Coin B comes up Heads with probability $\frac{1}{3}$ and Tails with probability $\frac{2}{3}$. These two kinds of biased coins are in a non-transparent box. The composition of the coins inside the box is unknown. At the beginning of the game, the DM draws a coin from the box. Then, the DM tosses the coin and guesses which coin it is. Since the draw is random and the composition of the coin is unknown, the state is unknown and the DM's prior belief is private.

The designer designs the DM's payoff. The payoff function is announced before the coin is drawn. An example of the payoff function is as follows. If the coin is A and the DM guesses it correctly, then, the DM gets M. If the coin is A and the DM guesses it wrong, then, she gets m. If the coin is B and the DM guesses it correctly, then, she gets N. If the coin is B and the DM guesses it wrong, then, she gets n. The designer wants to learn the DM's prior by observing the coin toss results and the DM's guesses.

The timing of the payment guarantees that the DM does not learn the state from the payoff. If there is only one period, after the guess, the designer and the DM check the name of the coin, and then the designer pays the DM accordingly. If there is more than one period, the coin is tossed and the DM makes guesses at each time. The payment is made at the end of the experiment. The experiment ends after a predetermined time period T (if the time horizon is finite), after the designer learns the type of the DM (if the time horizon is infinite), or continues forever if there is no learning. Since the state is unknown to both the DM and the designer, the designer records the DM's guess at each time and calculates the payoffs at the end of the experiment when the state is revealed.

2.3 Discussion

I make the assumptions that the designer can design the DM's payoffs and DM's actions are observable. These assumptions make it easier for the designer to learn the DM's type. The purpose of this chapter is to check whether the designer can learn in this environment. If the designer cannot learn in this case, then, it is even harder to learn when actions are noisy or the utility function is exogenous.

In the example, the DM announces the 'guess' each time and the designer observes this guess when it is announced. This is the case that the action is observable. It may not always be the case that the action is observable. Sometimes, only the noisy version of the action is observable. For example, suppose the DM cannot talk but has two coins, A and B, in her pocket. After tossing the coin drawn from the box, the DM cannot announce her guess but she can pick up the corresponding coin in her pocket. Then, the designer can flip the coin picked by the DM, observes a signal of the DM's guess and then infers the DM's guess, The observable action assumption makes it easier for the designer to learn. During the experiment, what the designer observes are a series of actions taken by the DM and a series of signals about the state. There is no extra level of uncertainty compared to the noisy version of the actions. In this section, I consider the case with observable actions to check if the designer can learn the type of the DM. If the designer cannot learn the type of the DM in this case, it is highly possible that the designer cannot learn when the observed actions are noisy.

In the example, the designer can choose the payoffs the DM gets. This makes it easier for the designer to learn. Consider the case that the designer cannot choose the payoff function in the coin example. Suppose that the DM gets 1 if she makes a correct guess and 0 otherwise. Assume that the DM is risk neutral. Then the DM would guess A if she believes that the coin is A with a probability of at least a half. Suppose that the DM has the prior that the coin is A with probability 0.8, and if the coin comes up heads more than tails, it is likely that the DM's belief is always bigger than a half. The designer then keeps observing the A guesses. The designer may not learn whether the DM has a prior 0.8 or the DM has other priors bigger than 0.8. If the DM with prior 0.8 guesses A all the time, the DM who has a prior bigger than 0.8 would guess A as well. The designer then cannot learn. However if the designer can choose the payoff function, the designer has the ability to induce actions that are favourable in terms of learning. Suppose that the designer somehow believes that the DM has the prior that the coin is A with a high probability. The DM believes that the coin is A with a probability of 0.8. Then, the designer may learn the prior of the DM if the designer chooses the payoff function as follows. If the coin is A and the DM guesses it correctly, the DM gets 1; if the coin is B and the DM guesses it correctly, the DM gets 3; if the guess is wrong, the DM gets 0. Given this payoff function, the DM will guess B if she believes that the probability that the coin is A is smaller than $\frac{3}{4}$. The payoff of guessing B correctly is big enough so that it is worth taking the risk. Consider the action of the DM who believes that the coin is A with a probability of 0.8. Before the first signal, the DM guesses A because the prior is 0.8. After the first signal, if the signal is tails, the DM believes that the coin is A with probability $\frac{2}{3}$ and then guesses B. Then, the designer can back up the prior belief of the DM. The designer can back up that the prior of the DM is in the range $(\frac{3}{4}, \frac{6}{7})$.

3 Analysis

This section analyses the model. Section 3.1 gives a positive result where the designer always learns the DM's type. Section 3.2 establishes the importance of observing a switch of action and re-defines the notion of learning. Section 3.3 discusses a special case when there are two signals and Section 3.4 gives a negative result where the designer does not always learn the DM's type and characterises the optimal threshold that maximises the probability of learning.

3.1 Two Types

When there are two types of DM, the designer can learn the type of the DM immediately.

Theorem 3.1. *When $P = \{\underline{p_0}, \bar{p_0}\}$, the designer can learn the type of the DM immediately by choosing $r^* \in \big(\underline{p_0}, \bar{p_0}\big]$.*

By choosing the threshold $r^* \in \big(\underline{p_0}, \bar{p_0}\big]$, the $\underline{p_0}$-type DM takes action $\alpha_0 = 0$

and the $\bar{p}_0$-type DM takes action $\alpha_0 = 1$. Since different types of the DM take different actions at time 0, after observing the action taken at time 0, the type of the DM is fully revealed. The designer then learns the DM's type immediately.

The role of action α_0 The model setup is such that the DM takes an action α_0 at time 0 before receiving any signal. That is, α_0 is taken based on the DM's prior. This makes it easier for the designer to learn the DM's type. If, however, the DM's first action is taken at time $t = 1$, that is, $\alpha_0 \in \emptyset$, then, the designer learns the DM's type after observing α_1 if the two possible types are not too close. This is summarised in Lemma 3.1. Let $\bar{p}_1 \in \left[\bar{p}_1^l, \bar{p}_1^r\right]$ be the $\bar{p}_0$-type DM's posterior at time $t = 1$ where $\bar{p}_1^l$ is the lowest possible posterior after receiving any $s_1 \in \mathcal{S}$ and $\bar{p}_1^r$ is the highest possible posterior after receiving any $s_1 \in \mathcal{S}$. Similarly, let $\underline{p}_1 \in \left[\underline{p}_1^l, \underline{p}_1^r\right]$ be the $\underline{p}_0$-type DM's posterior at time $t = 1$ where $\underline{p}_1^l$ is the lowest possible posterior after receiving any $s_1 \in \mathcal{S}$ and $\underline{p}_1^r$ is the highest possible posterior after receiving any $s_1 \in \mathcal{S}$.

Lemma 3.1. *Suppose $\alpha_0 \in \emptyset$. If $P = \{\underline{p_0}, \bar{p}_0\}$ and the time $t = 1$ posterior beliefs $\bar{p}_1 \in \left[\bar{p}_1^l, \bar{p}_1^r\right]$ and $\underline{p}_1 \in \left[\underline{p}_1^l, \underline{p}_1^r\right]$ satisfy $\underline{p}_1^r < \bar{p}_1^l$, then, the designer can learn the type of the DM at time $t = 1$ by choosing $r^* \in \left(\underline{p}_1^r, \bar{p}_1^l\right]$.*

Lemma 3.1 says that when the two possible types of the DM are significantly different from each other, that is, the $\bar{p}_0$-type DM's lowest possible posterior belief at time $t = 1$ is higher than the $\underline{p}_0$-type DM's highest possible posterior belief at time $t = 1$, the designer can learn the type of the DM by choosing the threshold r such that it is optimal for the $\bar{p}_0$-type DM to choose action $\alpha_1 = 1$ and optimal for the $\underline{p}_0$-type DM to choose action $\alpha_1 = 0$. On the contrary, when $\bar{p}_0$ and $\underline{p}_0$ do not satisfy the condition in Lemma 3.1, that is, when $\bar{p}_0$ and $\underline{p}_0$ are close to each other, then, the designer may not always learn the DM type at time $t = 1$.

To understand this, for simplicity, suppose $\mathcal{S} = \{1, 2\}$ and the distribution of the two signals is $(\pi, 1 - \pi)$ in state θ and $(1 - \pi, \pi)$ in state θ' where $\pi > 0.5$. Then, the posterior belief of both types after observing $s_t = 1$ is greater than the posterior belief after observing $s_t = 2$. After observing $s_t = 1$, the DM of both types believes that it is more likely that the state is θ and then revises the belief upwards. After observing $s_t = 2$, the DM of both types believes that it is more likely that the state is θ' and then revises the belief downwards. To be consistent with the notation in Lemma 3.1, let $\bar{p}_1^l$ and $\underline{p}_1^l$ be the posterior beliefs after observing $s_t = 2$ and let $\bar{p}_1^r$ and $\underline{p}_1^r$ be the posterior beliefs after observing $s_t = 1$.

When $\bar{p}_0$ and $\underline{p}_0$ are far away from each other such that $\underline{p}_1^r < \bar{p}_1^l$, then, if the designer chooses $r^* \in \left(\underline{p}_1^r, \bar{p}_1^l\right]$, after observing any realisation $s_1 \in \mathcal{S}$ at time $t = 1$, the $\bar{p}_0$-type DM maker takes action 1 while the $\underline{p}_0$-type DM maker takes action 0 at time. In this case, the designer always learns the type after observing the action taken at time $t = 1$. On the contrary, when $\bar{p}_0$ and $\underline{p}_0$ are close to each such that $\underline{p}_1^r > \bar{p}_1^l$, the designer may or may not learn the DM's type after observing one action because the DM may choose the same action given any choice of r. Suppose the designer chooses $r \in \left(\bar{p}_1^l, \underline{p}_1^r\right]$, then, both DM types take action 1 after observing $s_1 = 1$ and take action 0 after observing $s_1 = 2$. In this case, the designer cannot learn immediately after observing one action. If the designer chooses $r \in \left(\underline{p}_1^r, \bar{p}_1^r\right]$, then the designer can learn if the signal realisation is $s_1 = 1$ but cannot learn if the signal realisation is $s_1 = 0$. This is because when $s_1 = 1$, the $\bar{p}_0$-type takes action 1 while the $\underline{p}_0$-type takes action 0 at time $t = 1$. The designer can distinguish these two types. When $s_1 = 0$, both types take action 0 and then the designer cannot distinguish. As a result, the designer cannot always learn after observing one action. If the designer chooses $r > \bar{p}_1^r$ or $r < \underline{p}_1^l$, she cannot learn after observing one action. This is because both DM types take the same action at time $t = 1$ after observing any realisation of the signal $s_1 \in \mathcal{S}$.

More than two types The immediate learning relies heavily on the assumption that there are only two types of the DM. When there are three types, since the DM can only choose one cutoff r, there must be at least two types of the DM that take the same action at time 0. Then, whether a DM can learn the type or when the DM can learn the type depends on the sequence of actions and signal realisations. The next section tackles how to approach a more general case when there are more than two types.

3.2 Switch of Actions and Notion of Learning Re-defined

In this section, I argue that observing a switch of action is key for the designer to learn the type of the DM. I then re-introduce the notion of learning as observing a switch of action. Then, I compute the probability of observing a switch of action conditional on the state θ and θ' given a threshold r and a type $p_0 \in P$.

Although the time in this model is infinite, the real time for the designer to learn the type of the DM is actually finite. This is because a Bayesian DM learns the state after observing sufficiently many signals. Therefore, given any type $p_0 \in (0, 1)$, the sequences of the DM's actions eventually converge. The sequence of the DM's actions converges to 1 if the state is θ and converges to 0 if the state is θ' for any $r \in (0,1)$. As a result, if a designer wants to recover the DM's type, the designer must observe some variation in the actions before the DM's action converges.

As discussed in the previous section, the designer can distinguish between two types of the DM if the sequences of actions taken by these two types are *different* after they receive the same sequence of signals. Consider two types of the DM. After receiving the same sequence of signals, if one type of the DM takes a sequence of actions $(1,0,0,0,\dots,0,0,\dots)$ and the other type of the DM takes a sequence of actions $(0,0,0,0,\dots,0,0,\dots)$, the designer will be able to distinguish between these two types of the DM. When there are more than two types of the DM, the timing at which the DM switches from one action to the other allows the designer to recover the DM's prior. Suppose there are three types of the DM. After observing the same sequence of signals, if type-1 takes a sequence of actions $(1,1,0,0,\dots,0,0,\dots)$, type-2 takes a sequence of actions $(1,1,1,0,\dots,0,0,\dots)$, and type-3 takes a sequence of actions $(0,0,0,0,\dots,0,0,\dots)$, then, the designer will be able to distinguish between these three types of the DM. Given the examples above, it appears to be the case that as long as different types take different sequences of actions is enough for the DM to distinguish the type, a switch of actions is not necessary. For example, after observing a sequence of action 0's, it is still possible to eliminate type 1 and type 2 in the three-type example. This is because the types are discrete. It is possible to eliminate the type whose sequence of actions has a switch of actions. If the designer is belief is such that $P = [0,1]$ and Σ_0 is a uniform distribution, then, I define the notion of learning as observing a switch of actions.

Definition 3.1. *When $P = [0,1]$ and Σ_0 is uniform, the designer learns the type of the DM at time t if she observes a switch of action at time t.*

When does a switch of action happen? It depends on the choice of the threshold r and the state. Suppose the designer chooses a threshold r. The DM takes action 1 if the belief about the state being θ is greater than or equal to r. If the DM has a prior p_0 that is smaller than r, then the DM starts by taking action 0 at time 0. After receiving a sequence of signals, the DM first switches the action from 0 to 1 if the belief about the state goes from below the threshold r to above the threshold r. If the state is indeed θ, this switch of actions happens with probability one as the sequence of actions converges. If, however, the DM has the prior p_0 larger than r, then the DM starts by taking action 1. The designer can observe a switch of actions only if the DM's belief about the state being θ falls below the threshold r. Since the DM is Bayesian and the belief converges to 1 if the state is θ, the probability of the DM's belief falling below r is positive but not 1. As a result, the designer may or may not be able to observe such a switch of actions from 1 to 0 if the state is θ

Proposition 3.1 characterises the probability of observing a switch of actions in state θ and $\theta^{'}$ given the type of the DM p_0 and the threshold r. Let $v^* < 0$ satisfy $\sum_{s\in\mathcal{S}} \pi_s (\frac{\pi_s}{\pi'_s})^{v^*} = 1$, let $u^* > 0$ satisfy $\sum_{s\in\mathcal{S}} \pi'_s (\frac{\pi_s}{\pi'_s})^{u^*} = 1$. and let $k := \ln \frac{r}{1-r} - \ln \frac{p_0}{1-p_0}$.

Proposition 3.1. *Assume that θ is the underlying state and p_0 is the type of the DM.*

If the designer chooses $r > p_0$, the switch of actions is observed w.p. 1.

If the designer chooses $r < p_0$, the switch of actions is observed w.p. $\frac{1}{e^{v^* k}}$.

If the designer chooses $r = p_0$, the switch of actions is observed w.p. $\frac{1}{e^{v^* \eta}}$ where $\eta = \lim_{r\to p_0} k$.

Assume that θ' is the underlying state and p_0 is the type of the DM.

If the designer chooses $r < p_0$, the switch of actions is observed w.p. 1.

If the designer chooses $r > p_0$, the switch of actions is observed w.p. $\frac{1}{e^{u^* k}}$.

If the designer chooses $r = p_0$, the switch of actions is observed w.p. $\frac{1}{e^{u^* \eta}}$ where $\eta = \lim_{r\to p_0} k$.

3.3 A Special Case: Binary Signals

The probability of observing a switch of actions requires solving for v^* or u^* numerically. It is hard to compute the exact values when there are many signals. In this section, I consider a special binary-signal case.

There are two signals $\mathcal{S} = \{1, 2\}$ with distributions $(\pi, 1-\pi)$ in state θ and $(1-\pi, \pi)$ in state θ'. I assume $\pi > 0.5$ so that it is more likely to receive signal 1 in state θ and more likely to receive signal 2 in state θ'. When π is very close to a half, i.e. when the signal is very uninformative, the belief of the DM takes very small steps after receiving one more signal. In the limit, it is like the case with continuous times. The probability of learning is easier to compute given this signal structure.

Proposition 3.2 characterises the probability of observing a switch of actions when $\mathcal{S} = \{1, 2\}$. Let $B := \frac{\ln \frac{r}{1-r} - \ln \frac{p_0}{1-p_0}}{\ln \frac{\pi}{1-\pi}}$.

Proposition 3.2. *Assume that θ is the underlying state and p_0 is the type of the DM.*

If the designer chooses $r > p_0$, a switch of actions is observed with probability 1.

If the designer chooses $r < p_0$*, a switch of actions is observed with probability* $(\frac{1-\pi}{\pi})^{-B}$*.*

If the designer chooses $r = p_0$*, a switch of actions is observed with probability* $\frac{1-\pi}{\pi}$*.*

Assume that θ' *is the underlying state and* p_0 *is the type of the DM.*

If the designer chooses $r \leq p_0$*, a switch of actions is observed with probability 1.*

If the designer chooses $r > p_0$*, a switch of actions is observed with probability* $(\frac{1-\pi}{\pi})^{B}$*.*

In the following sections, I discuss the designer's optimal choice of the threshold r and the probability of learning the DM's type when the signals are binary.

3.4 No Learning in Continuous-Type Case

Suppose $P = [0, 1]$ and Σ_0 is uniform. This describes the situation in that the designer knows nothing about the DM's prior. The designer does not know the potential prior the DM may have and the designer does not know which prior is more likely. The continuously uniformly distributed prior can be considered as a conservative assumption the designer has about the DM. In this section, I give a negative result that the designer does not always learn the type of the DM and then I characterise the optimal threshold r that maximises the probability of learning when the signals or binary.

Theorem 3.2. *When* $P = [0, 1]$*,* Σ_0 *is a uniform distribution, the probability of learning the DM's type is smaller than one.*

The main tradeoff facing the designer is to learn for almost certain in one of the states or to learn with some probability in both states. If the designer chooses the threshold r very close to 1, then, if the state is indeed θ, she will observe a switch of actions from 0 to 1 at some point for all the possible types. If, however, the state is indeed θ', then, a switch of actions is almost impossible to be observed. If the designer chooses the threshold $r = 0.5$, then, if the state is θ, the types $p_0 < r$ will switch their action from 0 to 1 at some point and then the designer will be able to learn those types. For the types $p_0 > r$, if the state is θ, a switch of actions from 1 to 0 can only be observed with a positive probability. The greater the distance between p_0 and r, the less likely a switch of actions will be observed. However, the similar logic applies to the case if the state is θ'. As a result, the designer can learn with a positive probability in both states.

In general, the designer's prior belief about the state and the designer's prior belief about the type of the DM affect the choices of the threshold r when the designer wants to maximise the probability of learning. The designer's prior belief about the state determines whether the threshold r is closer to 0 or closer to 1. If the designer believes that the state is more likely to be θ, the designer believes that it is more likely that the beliefs of the DM are going up to 1 regardless of the types of the DM. Choosing the threshold r closer to 1 can maximise the probability of learning the type of the DM in the state θ. However, if the designer believes that the state is more likely to be θ', the designer believes that it is more likely that the beliefs of the DM are going down to 0 regardless of the types of the DM. Choosing the threshold r closer to 0 can maximise the probability of learning the type of the DM in the state θ'. The designer's prior belief about the type determines whether the designer wants to choose the threshold r closer to a certain prior. The closer the threshold r is to a range of certain priors, the higher the probability is for the designer to learn those priors when the state is unknown. In this section, since the designer believes that the Σ_0 is a uniform distribution, the second factor affects the designer's choice of r less significantly.

Next, I characterise the condition that the optimal threshold r^* satisfies. I show that the optimal threshold that maximises the probability of learning the prior is a function of μ_0 and it is increasing in μ_0. If the designer believes that the state is more likely to be θ, the designer will choose r^* closer to 1. It will allow the designer to learn more priors in state θ, but learn less in state θ'. If the designer believes that the state is more likely to be θ',the designer will choose the r^* closer to 0. It will allow the designer to learn more priors in state θ', but learn less in state θ.

Proposition 3.3. *The optimal choice of the threshold r^* that maximises the probability of learning the prior satisfies*

$$\frac{(r^*)^2}{(1-r^*)^2}\frac{\ln r^* + 1 - r^*}{\ln(1-r^*) + r^*} = \frac{1-\mu_0}{\mu_0}.$$

The optimal threshold r^* increases in μ_0.

4 Finite Periods of Learning

Previous sections discuss the cases that the time horizon is infinite. I focus on discussing the probability of learning in the long run. In this section, I consider the case that the time $t = 0, 1, \ldots, T$ is finite. The finite-time assumption is more relevant to real-life applications. For experimental economists who design experiments to detect the belief of the experiment candidates, the experiments cannot last forever. For firms that are doing trials to understand their customers, they cannot do an infinite number of trials

Proposition 3.4 characterises the probability of observing a switch of action in state θ and $\theta^{'}$ given the type of the DM, the threshold r and the final time period of the experiment T. Before presenting the proposition, I define $\Phi_B(s)$, $\Psi_B(s)$, $\Xi_B(s)$ and $\Omega_B(s)$ functions where

$$\Phi_B(s) := (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2(1-\pi)s})^B,$$

$$\Psi_B(s) := (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2\pi s})^{-B},$$

$$\Xi_B(s) := (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2\pi s})^B,$$

and

$$\Omega_B(s) := (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2(1-\pi)s})^{-B}.$$

Let $\phi_B(\tau) = \frac{\Phi_B^{(\tau)}(0)}{\tau!}$, $\psi_B(\tau) = \frac{\Psi_B^{(\tau)}(0)}{\tau!}$, $\xi_B(\tau) = \frac{\Xi_B^{(\tau)}(0)}{\tau!}$, and $\omega_B(\tau) = \frac{\Omega_B^{(\tau)}(0)}{\tau!}$ where $f^{(\tau)}(0)$ is the τ-th derivative of the function f evaluated at 0.

Proposition 3.4. *Assume that θ is the underlying state and p_0 is the type of the DM. If the designer sets $r > p_0$, the designer learns the type of the DM with probability $\sum_{\tau=1}^{T} \phi_B(\tau)$. If the designer sets $r < p_0$, the designer learns the type of the DM with probability $\sum_{\tau=1}^{T} \psi_B(\tau)$.*

Assume that θ' is the underlying state and p_0 is the type of the DM. If the designer sets $r > p_0$, the designer learns the type of the DM with probability $\sum_{\tau=1}^{T} \xi_B(\tau)$. If the designer sets $r < p_0$, the designer learns the type of the DM with probability $\sum_{\tau=1}^{T} \omega_B(\tau)$.

5 Conclusion

This chapter investigates whether a designer can learn a DM's prior belief about an underlying state by observing the public signals and the DM's actions. I show that when the DM's action is binary, it is not always the case that the designer can learn the DM's prior. The key to learning the DM's prior is being able to observe a switch in the DM's actions. If the DM's prior only has two possible values, then, the designer can immediately learn the DM's prior. However, if the DM's prior is uniformly distributed from zero to one, then, the probability of learning is smaller than one.

Appendix to Chapter 3

1 Proof of Proposition 3.1

First, consider the DM's belief. The DM forms the belief using the Bayes rule. Define the log odds ratio $\Lambda_t := \ln(\frac{p_t}{1-p_t})$. The following part shows that Λ_t follows a random walk. The odds ratio is

$$\frac{p_{t+1}}{1-p_{t+1}} = \frac{p_t}{1-p_t}\frac{\Pr(s_t|\theta)}{\Pr(s_t|\theta')} = \frac{p_0}{1-p_0}\prod_{\tau=0}^{t}\frac{\Pr(s_\tau|\theta)}{\Pr(s_\tau|\theta')} = \frac{p_0}{1-p_0}\prod_{s\in\mathcal{S}}(\frac{\pi_s}{\pi'_s})^{\hat{t}_s}$$

where $\hat{t}_s$ is the number of s signals before time $t+1$. Taking the logarithms of the odds ratio, we have

$$\ln\frac{p_{t+1}}{1-p_{t+1}} = \ln\frac{p_t}{1-p_t} + \ln\frac{\pi_{s_t}}{\pi'_{s_t}} = \ln\frac{p_0}{1-p_0} + \sum_{\tau=0}^{t}\ln\frac{\pi_{s_\tau}}{\pi'_{s_\tau}}$$

Thus

$$\Lambda_{t+1} = \Lambda_t + \ln\frac{\pi_{s_t}}{\pi'_{s_t}} \tag{3.1}$$

The log odds ratio follows a random walk with steps $\ln\frac{\pi_{s_t}}{\pi'_{s_t}}$. If the underlying state is θ, the random walk takes steps with the probability determined by the probability distribution a.

Now consider the DM's action. The DM takes action $\alpha_t = 1$ if $p_t \geq r$ and $\alpha_t = 0$ otherwise. Therefore, the DM takes action $\alpha_t = 1$ if $\Lambda_t \geq \ln\frac{r}{1-r}$. If the DM is observed to switch action from 0 to 1 in period t, then we have $\Lambda_t \approx \ln\frac{r}{1-r}$. Thus,

$$\ln\frac{p_0}{1-p_0} + \sum_{\tau=0}^{t}\ln\frac{\pi_{s_\tau}}{\pi'_{s_\tau}} \approx \ln\frac{r}{1-r}.$$

Therefore, observing a switch of action is equivalent to observing the type of the DM.

Next step is to characterise the probability of observing a switch of action given the choice of r. Let $x_t := \ln \frac{\pi_{s_t}}{\pi'_{s_t}}$. Conditional on the state θ, $x_1, x_2, \ldots$ are identically and independently distributed random variables. Let $y_t = x_1 + \cdots + x_t$. The sequence $y = \{y_\tau : \tau > 0\}$ is a random walk starting at the origin. Let $k := \ln \frac{r}{1-r} - \ln \frac{p_0}{1-p_0}$. The probability of observing a switch of action is equal to the probability that the random walk y hits the value k.

If $k > 0$, the probability that y hits the value k is 1. Conditional on θ, we have $\mathbb{E}_\theta[\Lambda_{t+1} \mid \Lambda_t] > \Lambda_t$. Therefore, Λ_t is a submartingale. According to the Martingale convergence theorem, if the state is θ, $\Lambda_t \to \infty$ almost surely. If the state is θ', $\Lambda_t \to -\infty$ almost surely.

Since the random walk y satisfies

$$y_t = \Lambda_{t+1} - \ln \frac{p_0}{1-p_0},$$

if the state is θ, the random walk y hits the value $k > 0$ with probability 1.

Now consider the situation that $k < 0$. Let $A > 0$. Define a stopping time τ to be the first time the random walk y exists the interval $[k, A]$. The stopping time τ is finite with probability 1 as the the random walk y tends to infinity. According to Wald's Identity, the stopping time τ satisfies

$$1 = \mathbb{E}[\frac{e^{vy_\tau}}{(\sum_s \pi_s(\frac{\pi_s}{\pi'_s})^v)^\tau}] \text{ for } \forall v \neq 0 \text{ s.t. } \sum_s \pi_s(\frac{\pi_s}{\pi'_s})^v \geq 1.$$

Choose $v^* < 0$ such that $\sum_s \pi_s(\frac{\pi_s}{\pi'_s})^{v^*} = 1$, then,

$$1 = \mathbb{E}[e^{v^* y_\tau}].$$

Let τ_A and τ_k be the two stopping times that y_τ hits A and y_τ hits k, then

$$1 \approx \Pr(y_\tau = A)e^{v^* A} + \Pr(y_\tau = k)e^{v^* k}.$$

Notice that this is an approximation because when y_t hits k, it does not exactly equal k. The equality holds with an equal sign if either the steps the random walk takes are infinitely small (the continuous time case) or the random walk takes steps up and down of the equal sizes. Next section will discuss the case that the random walk takes steps up and down of the equal sizes. Use the fact that $\Pr(y_\tau = k) = 1 - \Pr(y_\tau = A)$ we can get that

$$\Pr(y_\tau = k) \approx \frac{1 - e^{v^* A}}{e^{v^* k} - e^{v^* A}}.$$

Let $A \to \infty$, we have

$$\Pr(y_\tau = k) \to \frac{1}{e^{v^* k}}.$$

When the state is $\theta^{'}$, the proof follows the same approach.

If $k < 0$, the random walk y hits the value k with probability 1. From the proof for lemma 3.1, conditional on θ', Equation (3.1) implies that $\mathbb{E}_\theta[\Lambda_{t+1} \mid \Lambda_t] < \Lambda_t$. Therefore, $-\Lambda_t$ is a submartingale. If the state is θ', $\Lambda_t \to -\infty$ almost surely. The random walk y hits the value $k < 0$ with probability 1.

Consider the case that $k > 0$. Let $A < 0$. Define the stopping time τ to be the first time the random walk y exits the interval $[A, k]$. The stopping time τ is finite with probability 1. According to Wald's identity, the stopping time satisfies

$$1 = \mathbb{E}[\frac{e^{uy_\tau}}{(\sum_s \pi'_s(\frac{\pi_s}{\pi'_s})^u)^\tau}] \text{ for } \forall u \neq 0 \text{ s.t. } \sum_s \pi'_s(\frac{\pi_s}{\pi'_s})^u \geq 1.$$

Choose $u^* > 0$ such that $\sum_s \pi_s(\frac{\pi_s}{\pi'_s})^{u^*} = 1$. Follow the same steps as in the proof for lemma 3.1. We have

$$\Pr(y_\tau = k) \to \frac{1}{e^{u^* k}}.$$

2 Proof of Proposition 3.2

Consider the random walk y defined previously where $y_t = x_1 + \cdots + x_t$. Given the signal structure characterised above, if $s_t = 1$, we have $x_t = \ln \frac{\pi}{1-\pi}$, and if $s_t = 2$, we have $x_t = \ln \frac{1-\pi}{\pi}$. The random walk y now starts at the origin and takes steps up and down by equal amounts $\ln \frac{\pi}{1-\pi}$. If the threshold is r, let $B = \frac{\ln \frac{r}{1-r} - \ln \frac{p_0}{1-p_0}}{\ln \frac{\pi}{1-\pi}}$ be the corresponding steps the random walk y takes to hit the value $\ln \frac{r}{1-r} - \ln \frac{\pi}{1-\pi}$. The expressions 'the random walk y hits the value $\ln \frac{r}{1-r} - \ln \frac{\pi}{1-\pi}$' and the 'random walk y hits the step B' are the same.

We now have a random walk y starting at the origin taking steps up and down by equal amounts $\ln \frac{\pi}{1-\pi}$. If the state is θ, the random walk y taking steps up with probability a and steps down with probability $1-a$. If $r > p_0$, $B > 0$. The random walk y hits the step B with probability 1. If $r < p_0$, $B < 0$. The random walk y hits the step B with probability $(\frac{1-\pi}{\pi})^{-B}$.

If the state is θ', the random walk y taking steps up with probability $1-a$ and steps down with probability a. If $r > p_0$, $B > 0$. The random walk y hits the step B with probability $(\frac{1-\pi}{\pi})^B$. If $r < p_0$, $B < 0$. The random walk y hits the step B with probability 1.

3 Proof of Theorem 3.2 and Proposition 3.3

Let $B = \frac{\ln\frac{r}{1-r} - \ln\frac{p_0}{1-p_0}}{\ln\frac{\pi}{1-\pi}}$ be the steps. If the state is θ, for $p_0 < r$, the designer learns the type with probability 1. For $p_0 > r$, the designer learns the type with probability $(\frac{1-\pi}{\pi})^{-B}$. If the state is θ', for $p_0 < r$, the designer learns the type with probability $(\frac{1-\pi}{\pi})^{B}$. For $p_0 > r$, the designer learns the type with probability 1. The probability of learning is

$$\begin{aligned}
\Pr(\mathit{Learning}) &= \mu_0(\int_0^r 1dp_0 + \int_r^1 (\frac{1-\pi}{\pi})^{-B} dp_0) + (1-\mu_0)(\int_0^r (\frac{1-\pi}{\pi})^{B} dp_0 + \int_r^1 1dp_0) \\
&= \mu_0(r + \frac{r}{1-r}(-1 - \ln r + r)) + (1-\mu_0)(\frac{1-r}{r}(-r - \ln(1-r)) + 1 - r) \\
&< 1
\end{aligned}$$

Take the first order derivative with respect to r,

$$\begin{aligned}
\frac{d\Pr(\mathit{Learning})}{dr} &= \mu_0(\frac{d}{dr}\int_0^r 1dp_0 + \frac{d}{dr}\int_r^1 (\frac{1-\pi}{\pi})^{-B} dp_0) \\
&+ (1-\mu_0)(\frac{d}{dr}\int_0^r (\frac{1-\pi}{\pi})^{B} dp_0 + \frac{d}{dr}\int_r^1 1dp_0) \\
&= \ln\frac{1-\pi}{\pi}\frac{dB}{dr}(-\mu_0\int_r^1 (\frac{1-\pi}{\pi})^{-B} dp_0 + (1-\mu_0)\int_0^r (\frac{1-\pi}{\pi})^{B} dp_0) \\
&= \ln\frac{1-\pi}{\pi}(\ln\frac{\pi}{1-\pi})^{-1}\frac{1}{r(1-r)}[-\mu_0\frac{r}{1-r}(r - 1 - \ln r) \\
&- (1-\mu_0)\frac{1-r}{r}(r + \ln(1-r))] \\
&= \mu_0\frac{1}{(1-r)^2}(r - 1 - \ln r) + (1-\mu_0)\frac{1}{r^2}(r + \ln(1-r))
\end{aligned}$$

Take the second order derivative with respect to r,

$$\begin{aligned}
\frac{d^2\Pr(\mathit{Learning})}{dr^2} &= -\mu_0\frac{-r^2 + 2r\ln r + 1}{r(1-r)^3} - (1-\mu_0)\frac{2\ln(1-r) - \frac{(r-2)r}{1-r}}{r^3} \\
&< 0
\end{aligned}$$

The probability of learning is concave in $r \in (0, 1)$.

Therefore, there exists a $r^* \in (0, 1)$ such that the probability of learning is maximised. Next, show r^* is increasing in μ_0. Take the first order condition, we have

$$\mu_0\frac{1}{(1-r^*)^2}(r^* - 1 - \ln r^*) + (1-\mu_0)\frac{1}{(r^*)^2}(r^* + \ln(1-r^*)) = 0$$

Then we have

$$(1-\mu_0)\frac{1}{(r^*)^2}(r^*+\ln(1-r^*)) = \mu_0\frac{1}{(1-r^*)^2}(\ln r^*+1-r^*)$$

Therefore,

$$\frac{\mu_0\frac{1}{(1-r^*)^2}(\ln r^*+1-r^*)}{(1-\mu_0)\frac{1}{(r^*)^2}(r^*+\ln(1-r^*))} = 1$$

Thus

$$\frac{(r^*)^2}{(1-r^*)^2}\frac{\ln r^*+1-r^*}{r^*+\ln(1-r^*)} = \frac{1-\mu_0}{\mu_0}$$

We can write r^* as a function of μ_0. The left hand side is decreasing in r^*, the right hand side is decreasing in μ_0. Therefore r^* is increasing in μ_0.

4 Proof of Proposition 3.4

Consider the random walk y defined in Proposition 3.1. Let $B = \frac{\ln\frac{r}{1-r}-\ln\frac{p_0}{1-p_0}}{\ln\frac{\pi}{1-\pi}}$. Let

$$\phi_B(\tau) = \Pr(y_1 \neq B, \ldots, y_{\tau-1} \neq B, y_\tau = B)$$

be the probability that the random walk y first hits the step B at the τ-th step when $B > 0$ with generating function

$$\Phi_B(s) = \sum_{\tau=1}^{\infty}\phi_B(\tau)s^n = (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2(1-\pi)s})^B. \tag{3.2}$$

Let

$$\psi_B(\tau) = \Pr(y_1 \neq B, \ldots, y_{\tau-1} \neq B, y_\tau = B)$$

be the probability that the random walk y first hits the step B at the τ-th step when $B < 0$ with generating function

$$\Psi_B(s) = \sum_{\tau=1}^{\infty}\psi_B(\tau)s^n = (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2\pi s})^{-B}. \tag{3.3}$$

We have $\phi_B(\tau) = \frac{\Phi_B^{(\tau)}(0)}{\tau!}$ and $\psi_B(\tau) = \frac{\Psi_B^{(\tau)}(0)}{\tau!}$ where $f^{(\tau)}(0)$ is the τ-th derivative of the function f evaluated at 0.

Assume that θ is the underlying state. The random walk y takes steps up with probability π and steps down with probability $1-\pi$. The generating functions can be written as (3.2) and (3.3) [1]. Then summing ϕ_B and ψ_B over τ from $\tau = 1$ to $\tau = T$ gives us the probability of hitting the step B within T periods.

Similar results hold for assuming θ' being the underlying state. Let

$$\xi_B(\tau) = \Pr(y_1 \neq B, \ldots, y_{\tau-1} \neq B, y_\tau = B)$$

be the probability that the random walk y first hits the step B at the τ-th step when $B > 0$ with generating function

$$\Xi_B(s) = \sum_{\tau=1}^{\infty} \xi_B(\tau) s^n = (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2\pi s})^B. \tag{3.4}$$

Let

$$\omega_B(\tau) = \Pr(y_1 \neq B, \ldots, y_{\tau-1} \neq B, y_\tau = B)$$

be the probability that the random walk y first hits the step B at the τ-th step when $B < 0$ with generating function

$$\Omega_B(s) = \sum_{\tau=1}^{\infty} \omega_B(\tau) s^n = (\frac{1-(1-4\pi(1-\pi)s^2)^{\frac{1}{2}}}{2(1-\pi)s})^{-B}. \tag{3.5}$$

Assume that θ' is the underlying state. The random walk y takes steps up with probability $1-\pi$ and steps down with probability π. The generating functions can be written as (3.4) and (3.5). Then summing ξ_B and ω_B over τ from $\tau = 1$ to $\tau = T$ gives us the probability of hitting the step B within T periods.

[1] See Grimmett & Stirzaker (2001)

Bibliography

Akcigit, U. & Liu, Q. (2015), 'The Role of Information in Innovation and Competition', Journal of the European Economic Association **14**(4), 828–870.

Angrisani, M., Guarino, A., Jehiel, P., Kitagawa, T. et al. (2017), 'Information redundancy neglect versus overconfidence: a social learning experiment'.

Augenblick, N. & Rabin, M. (2018), 'An experiment on time preference and misprediction in unpleasant tasks'.

Bobtcheff, C., Levy, R. & Mariotti, T. (2021), Negative Results in Science: Blessing or (Winner's) Curse?

Bolton, P. & Harris, C. (1999), 'Strategic Experimentation', Econometrica **67**(2), 349–374.

Callander, S. (2011), 'Searching and Learning by Trial and Error', American Economic Review **101**(6), 2277–2308.

Camerer, C. (1998), 'Bounded rationality in individual decision making', Experimental Economics **1**(2), 163–183.

Che, Y.-K. & Mierendorff, K. (2019), 'Optimal Dynamic Allocation of Attention', American Economic Review **109**(8), 2993–3029.

Deb, R. & Stewart, C. (2018), 'Optimal Adaptive Testing: Informativeness and Incentives', Theoretical Economics **13**(3), 1233–1274.

DeJarnette, P., Dillenberger, D., Gottlieb, D. & Ortoleva, P. (2020), 'Time Lotteries and Stochastic Impatience', Econometrica **88**(2), 619–656.

Denti, T. (2019), Unrestricted Information Acquisition.

Dillenberger, D. (2010), 'Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior', Econometrica **78**(6), 1973–2004.

Dillenberger, D., Gottlieb, D. & Ortoleva, P. (2018), Stochastic Impatience and the Separation of Time and Risk Preferences.

Epstein, L. G., Noor, J. & Sandroni, A. (2010), ‘Non-bayesian learning’, The B.E. Journal of Theoretical Economics **10**(1).

Epstein, L. G. & Zin, S. E. (1989), ‘Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework’, Econometrica **57**(4), 937.

Fischer, C. (1999), Read This Paper Even Later: Procrastination with Time-Inconsistent Preferences.

Fudenberg, D., Strack, P. & Strzalecki, T. (2018), ‘Speed, Accuracy, and the Optimal Timing of Choices’, American Economic Review **108**(12), 3651–3684.

Fudenberg, D. & Tirole, J. (1985), ‘Preemption and Rent Equalization in the Adoption of New Technology’, The Review of Economic Studies **52**(3), 383.

Grimmett, G. & Stirzaker, D. (2001), Probability and Random Processes, Oxford university press.

Han, J. & Sangiorgi, F. (2018), ‘Searching for Information’, Journal of Economic Theory **175**, 342–373.

Hébert, B. M. & Woodford, M. (2019), Rational inattention when decisions take time.

Hellwig, C. & Veldkamp, L. (2009), ‘Knowing What Others Know: Coordination Motives in Information Acquisition’, Review of Economic Studies **76**(1), 223–251.

Hill, R. & Stein, C. (2020), Scooped! Estimating Rewards for Priority in Science.

Hopenhayn, H. A. & Squintani, F. (2011), ‘Preemption Games with Private Information’, Review of Economic Studies **78**(2), 667–692.

Hörner, J. & Skrzypacz, A. (2017), ‘Learning, Experimentation, and Information Design’, Advances in Economics and Econometrics pp. 63–98.

Ke, T. T. & Villas-Boas, J. M. (2019), ‘Optimal Learning Before Choice’, Journal of Economic Theory **180**, 383–437.

Keller, G. & Rady, S. (2010), ‘Strategic Experimentation with Poisson Bandits’, Theoretical Economics **5**(2), 275–311.

Keller, G., Rady, S. & Cripps, M. (2005), ‘Strategic Experimentation with Exponential Bandits’, Econometrica **73**(1), 39–68.

Knuth, D. (1998), Sorting and Searching, in ‘The Art of Computer Programming. 3’, 2nd editio edn.

Kreps, D. M. & Porteus, E. L. (1978), ‘Temporal Resolution of Uncertainty and Dynamic Choice Theory’, Econometrica **46**(1), 185.

Kumari, A. (2012), ‘Linear Search Versus Binary Search: A Statistical Comparison For Binomial Inputs’, International Journal of Computer Science, Engineering and Applications **2**(2), 29–39.

Laibson, D. (1997), ‘Golden Eggs and Hyperbolic Discounting’, The Quarterly Journal of Economics .

Mayskaya, T. (2020), Dynamic Choice of Information Sources.

Mehta, A., Saxena, A., Patel, J. & Thanna, A. (2015), ‘A Review on Comparison of Binary Search AND Linear Search’, International Journal OF Engineering Sciences & Management Research **2**(10).

Merton, R. K. (1957), ‘Priorities in Scientific Discovery: A Chapter in the Sociology of Science’, American Sociological Review **22**(6), 635–659.

Meyer, M. A. (1994), ‘The Dynamics of Learning with Team Production: Implications for Task Assignment’, The Quarterly Journal of Economics **109**(4), 1157–1184.

Morris, S. & Strack, P. (2019), The Wald Problem and the Relation of Sequential Sampling and Ex-Ante Information Costs.

Nikandrova, A. & Pancs, R. (2018), ‘Dynamic Project Selection’, Theoretical Economics **13**(1), 115–143.

O’Donoghue, T. & Rabin, M. (1999), ‘Doing It Now or Later’, American Economic Review **89**(1), 103–124.

Ozdenoren, E., Hoppe-Wewetzer, H. C. & Katsenos, G. (2021), Experimentation, Learning, and Preemption.

Palacios-Huerta, I. (1999), ‘The Aversion to the Sequential Resolution of Uncertainty’, Journal of Risk and Uncertainty **18**(3), 249–269.

Phelps, E. S. & Pollak, R. A. (1968), 'On Second-best National Saving and Game-equilibrium Growth', The Review of Economic Studies **35**(2), 185–199.

Rabin, M. (1998), 'Psychology and economics', Journal of Economic Literature **36**(1), 11–46.

Rahim, R., Nurarif, S., Ramadhan, M., Aisyah, S. & Purba, W. (2017), 'Comparison Searching Process of Linear, Binary and Interpolation Algorithm', Journal of Physics: Conference Series **930**(1), 12007.

Reinganum, J. F. (1981), 'On the Diffusion of New Technology: A Game Theoretic Approach', Review of Economic Studies **48**(3), 395–405.

Shahanaghi, S. (2022), Competition and Errors in Breaking News.

Tversky, A. & Kahneman, D. (1974), 'Judgment under uncertainty: Heuristics and biases', Science **185**(4157), 1124–1131.

Wald, A. (1945), 'Sequential Tests of Statistical Hypotheses', The Annals of Mathematical Statistics **16**(2), 117–186.

Wald, A. (1947), 'Foundations of a General Theory of Sequential Decision Functions', Econometrica **15**(4), 279–313.

Yang, M. (2015), 'Coordination with flexible information acquisition', Journal of Economic Theory **158**, 721–738.

Zhong, W. (2017), Time Preference and Information Acquisition.

Zhong, W. (2022), 'Optimal dynamic information acquisition', Econometrica **90**(4), 1537–1582.