Private Information Acquisition and Preemption: a Strategic Wald Problem

Guo Bai *

November 1, 2022

Abstract

This paper studies a dynamic information acquisition model with payoff externalities. Two players can acquire costly information about an unknown state before taking a safe or risky action. Both information and the action taken are private. The first player to take the risky action has an advantage but whether the risky action is profitable depends on the state. The players face the tradeoff between being first and being right. In equilibrium, for different priors, there exist three kinds of randomisation: when the players are pessimistic, they enter the competition randomly; when the players are less pessimistic, they acquire information and then randomly stop; when the players are relatively optimistic, they randomly take an action without acquiring information.

1 Introduction

Popular research questions often attract attentions from multiple research groups. The first research group that publishes their novel discoveries receives the most credit. This is known as a priority race (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 smaller probabilities and receive less citations, the negative effect in general is moderate instead of detrimental. There is a *first-mover advantage* but it is still better to be the second than remaining silent.

Scientific discoveries are often made independently (Merton (1957)). The opponents of a research group in a priority race is sometimes invisible. A research group may work on a project *privately* for decades. Before a result is found, researchers conduct their research in secrecy and do not share information about their research progress (Akcigit & Liu (2015)). Even after a result is found, it still does not become public in an instant. If the result is profound, it takes time for the research group to report and publish their result. If the result is insignificant, it is often not reported (Akcigit & Liu (2015) and Bobtcheff et al. (2021)). As a result, it is difficult for a research group to closely observe other groups' findings or their progress. Because of the delay in observing the opponents' report, or the absence of some results, each research group's research progress and their decisions on reporting a result are essentially private. I refer to this as private priority race.

The private priority race is common in the field of structural biology. Hill & Stein (2020) presents that in a survey of structural biologists, 80 percent of the respondents claim that before journal publication, "they rarely if ever circulate their findings in a working paper or pre-print". In addition, Hill & Stein (2020) establishes that after a research group uploads their findings to a database of biological research community ('deposit date'), there is a two to six months of delay before publication ('release date'). Between deposit and release, the data is hidden from the public and hence also hidden from the competing research groups if there are any. As a result, the research progress and the findings of a research group are essentially private before the release date. At the moment the research groups report their findings, that is, on the deposit date, they know either their competitors' progress nor findings. When a research group does observe the findings of their competitors, it has already been too late. The fact is, Hill & Stein (2020) finds that structural biology projects are involved in priority races where the research groups have finished their projects and reported their findings but later see that their competitors published the findings first and won the priority race.

This paper studies *private* priority race between two players. An unknown payoff-relevant state is either high or low and is constant over time. Before taking an irreversible action, at each time, players can acquire costly information about the state. A safe action guarantees

the player a payoff of zero in both states. A risky action gives the player a positive payoff in high state and a negative payoff in low state. The first player to take the risky action wins the priority race and gets a first-mover premium. The second player to take the risky action does not get the first-mover premium but still gets a positive payoff in high state. Costly information is modelled as stochastic conclusive evidence such that a breakdown reveals a low state and breakthrough reveals a high state. I assume that both information acquisition and the final action taken are private.

To understand the model setup, consider the priority race between two groups of structural biologists.[1] The objective of the research is to understand and solve the structure of a protein. High state represents that given the current technology, the structure of this protein is solvable. Low state represents otherwise. The fact that the state is unknown to the biologists describes that given the existing knowledge, it is not known whether the structure of this protein is solvable or not. The biologists can do costly experiments in the laboratory to better understand this protein and hence solve its structure. By making repeated and varied attempts, experiment results can exhibit fatal flaws that disprove the feasibility of solving the structure of this protein. Or, the results can exhibit successes that shows its feasibility. By conducting experiments, if no fatal flaws or successes have been observed so far, the biologists do not know for certain whether it is solvable, but understand this protein better. At each time, the biologists can choose to stop further experiments and publish what they have understood so far. By publishing their result, they infer that the structure of this protein can be solved by using their method. If the structure of the protein is indeed solvable, publishing their result gives them positive credit and if it is not, they experience some credit losses. the biologists can also choose to stop further experiments and abandon this project, which gives them zero credit.

Except for the priority race in scientific research, first-mover advantage exists in other economic and social activities. Research and development (R&D) of new technology is often competitive. The first firm to verify the feasibility of the technology and conduct mass production generates more sales than any late competitor. A firm's R&D hence affects its competitor's profit and vice versa. In addition, firms may not be able to see their competitor's breakthrough or their start of mass production. A firm may fall behind without even noticing it. Similarly, the firm may not see its competitor's breakdown or their silent exit from the competition. Then, the firm's further R&D will lead to either a 'win' in a doomed-to-fail project or a belated exit.

Another example is market entry. If a firm is the first to enter a new market, they may capture a higher market share. But before entry, they need to investigate whether the market is of high or low demand. This investigation takes time and the preparation of entering a new market is not immediately observable. The firm may decide to enter a new market with

---

[1]This example is borrowed from Hill & Stein (2020) with minor changes.

the belief that they will be the first, but later realise that their competitor has started the whole process earlier than themselves. Their competitor will then be the pioneer and the firm can only follow. At the time the firm observes the competing firm's decision, it is too late to reverse their own decision. Because of the delay of observing the competitor's action, entering a new market is essentially a private action.

In this paper, I use the private priority race model to understand the tradeoff between being first and being right when information acquisition takes time and opponent's action taken is not observable. I focus on understanding the players' equilibrium strategy in the presence of payoff externalities while shutting down the information externalities. The payoff externalities stem from the first-mover advantage. One's early success reduces the opponent's payoff from late success. The absence of information externalities is because of the unobservable actions. There is no information spillover because they do not get any direct or indirect information from each other.

The setup is related to the bandit problem, 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 as 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. The R&D is modelled as a costly activity that contains information about the feasibility of the technology, but does not give the firm direct return. The firm gets return from the R&D only if they take actions using the information generated.

The main result in this paper shows that in equilibrium, players use random stopping strategy if they acquire information. In addition, multiple equilibria can exist in equilibrium. This is significantly different from the single DM case. In the single DM case, the deterministic cutoff strategy is the optimal one: to acquire information if the belief is in the intermediate range and to take an action if the belief is extreme. In my model with competition and first-mover advantage, for relatively pessimistic priors, the players use the random stopping strategy. That is, they acquire information up to some point and then randomly stop and take the risky action. This is a result of the interaction of the learning motive and the preemption motive. At the early stage of information acquisition, players are uncertain about the state and hence have stronger learning motives. In addition, since they are rela-

3

tively pessimistic about the state, they also believe that the probability of being preempted is low. The value associated with taking the risky action is large if the state happens to be high. At this point, acquiring information is a strategic complement. If the opponent has not stopped and taken the risky action, then, the gain from acquiring information is high and the player is willing to acquire more information. The more information the opponent is expected to acquire, the more information the player is willing to acquire. This gives rise to multiplicity of equilibria. After the player has been acquiring information for some time, she becomes more optimistic about the state (because no news is good news). However, the value associated with taking the risky action conditional on the high state becomes lower. This is because the probability of the opponent taking the risky action increases. The player then has a stronger preemption motive to stop acquiring information and to preempt the opponent.

For relatively optimistic priors, there exists an equilibrium where the players randomise between the safe and risky action immediately without acquiring information. The existence of this equilibrium is a result of the strong preemption motive and the weak learning motive. Players undercut the time at which the other takes 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, the opponent's learning is deterred. An interesting result is that even when the information cost is zero, this equilibrium still exists. This is because in my model, the total information cost consists of the exogenous information cost c and an endogenous information cost from being preempted. The endogenous information cost is determined by the opponent's strategy. When the exogenous information cost vanishes, the endogenous information cost does not vanish and hence the total information cost is still positive. This positive total information cost gives rise to this equilibrium where the players take immediate action at time zero without acquiring information.

For sufficiently (but not extremely) pessimistic priors, there exists an asymmetric equilibrium where one player takes the safe action immediately and the opponent acquires information. Information acquisition here has strategic substitutes' property where one player acquires information only if the opponent does not. The intuition is when the prior is sufficiently pessimistic, it takes longer for the belief to drift to the random stopping point. The expected information cost is hence higher. The value associated with taking the risky action must be high enough to compensate for the higher information cost so that the player is willing to acquire information. When the opponent drops out, the value associated with the risky action is the highest and when the opponent participates in the competition, this value decreases. In symmetric equilibrium, the player randomises between these two roles where she mixes between the random stopping strategy and the immediate safe action at time zero.

I discuss how vanishing information cost affects players' equilibrium strategy. I show that

even when information cost vanishes, the players do not acquire information for all the priors in equilibrium. This is different from the single DM case. In the single DM case, since the exogenous information cost c is the total cost of acquiring information, when c vanishes, the DM always acquires information until she is certain. In contrast, in the two-player game, since information cost consists of both exogenous and endogenous components, the total information cost does not vanish with the exogenous information 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 from being preempted is high 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. In addition, 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.

I also investigate how the intensity of competition affect players' equilibrium strategy. When the priority race becomes more intense, the range of priors where the players acquire information shrinks. In addition, if the priority race becomes 'loser gets punished' type of competition, there exists a situation such that players acquire information but later drop out if no revealing information is received.

Related literature I study *strategic* dynamic information acquisition in an optimal stopping framework introduced by Wald (1945, 1947). The single decision maker's optimal stopping problem has been studied in both drift-diffusion models and Poisson models. Fudenberg et al. (2018) study the relationship between decision time and accuracy of the action in a modified drift-diffusion model. Ke & Villas-Boas (2019) and Nikandrova & Pancs (2018) investigate how decision maker optimally acquire information about two different alternatives. The former considers a drift-diffusion model and the latter considers a Poisson model. Che & Mierendorff (2019) and Mayskaya (2020) study how a decision maker optimally choose the bias of the information source in models with Poisson signals. My model adopts the Poisson signal structure and more importantly, considers the *strategic* interaction between players. One way of thinking about it is that instead of having the cost of information as an exogenous parameter, it is now endogenous which depends on the opponent's strategy.

This paper relates to the literature studying preemption races originated from the seminal papers 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). The main difference between this paper and the literature listed above is the assumption on the observability of the actions. The literature listed above assume observable actions with an exception such that Bobtcheff et al. (2021) assumes safe action being unobservable with a positive

probability. In my paper, I shut down the information externalities generated by observable actions in order to focus on the payoff externalities. Bobtcheff et al. (2021) investigates how publicity of actions affects the players' equilibrium strategy and payoff in a preemption game. In their model, similar to mine, the players have both the learning motive and the preemption motive. The difference is that the risky action ('investment' in their model) is always observable while the safe action ('exit' in their model) can be private. They argue that private signals and potential private actions create a winner's curse where the first risky action taker believes the opponent may have received a private breakdown and hence exited. To compensate, the player acquires more information to be more certain about the project before taking the risky action. The main difference between this paper and mine is the unobservability of the irreversible action, the presence of breakthrough and the positive payoff received by the second risky action taker. In my model, both the risky action, safe action and the signals are private. Not being able to observe the risky action creates a stronger preemption motive. The players in my model acquire less information because of this. Ozdenoren et al. (2021) also studies unobservable actions in a discrete-time setting. It shows that preemption motives caused by payoff externalities lead to less experimentation. With a discrete-time setting, they do not have an equilibrium in random stopping.

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.

This paper is also related to strategic experimentation literature that model similar R&D races as a bandit problem (see Bolton & Harris (1999), Keller et al. (2005), and Keller & Rady (2010)). Besides the differences in the model setup introduced earlier, there are three other differences. First, strategic experimentation literature investigates the free-ride 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 free ride. 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. Third, strategic experimentation literature studies the intensity of the experimentation. In my model, I assume the players choose between to acquire information or not, but not the intensity at each time instant. Since bang-bang solution is normally the optimal strategy in bandit problems, this simplification appears to be reasonable.

A distinct but related group of literature is about the equivalence between static and

dynamic information acquisition. Hebert & 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.

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

Assumption 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 2. $u_L < 0 < u_H - \bar{\triangle}_H$.

Assumption 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 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

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

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 an H-state (L-state, resp) revealing signal, it is optimal to stop acquiring the signal and take R (S, resp) 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. *Player i's pure strategy s^i is defined as $(T^i, x^i) \in \mathbb{R}_+ \times \{R, S\}$.* [2]

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

Definition 2. *Player i's mixed strategy γ^i is defined as two non-decreasing measurable functions (ρ^i, σ^i) where $\rho^i : \mathbb{R}_+ \to [0, 1]$ and $\sigma^i : \mathbb{R}_+ \to [0, 1]$ satisfy $\rho^i(t) + \sigma^i(t) \leq 1$ for $\forall t \in \mathbb{R}_+$. The first element ρ^i is the probability that player i stops and takes R before or at time t conditional on no revealing signal. The second element σ^i is the 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 the player's expected payoff. Player i's expected payoff from taking action x at time t in state ω, $\mathbb{E}^{\gamma^j}[u_\omega^x \mid \omega, t]$, 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

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

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

---

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

8

from taking action x^i at time T^i is

$$\int_0^{T^i} \pi_t \left(p_t^i a \mathbb{E}^{\gamma^j} \left[u_H^R \mid H, t \right] - c \right) dt + \pi_{T^i} U_{x^i}^{i,\gamma^j}(t) \tag{2}$$

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

1. $x^i \in \arg\max_{\tilde{x}^i} U_{\tilde{x}^i}^{i,s^{-i}}(T^i)$;

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).

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

The player's strategy and the random arrival of the signal jointly determine the distribution over the action taken and timing. Before proceeding, I define the following two conditional probabilities that characterise the distribution over action taken and timing. Let $F_\omega(t)$ ($G_\omega(t)$, resp) be the probability that the player takes R (S, resp) before or at time t in state ω. Then, $1 - F_\omega(t) - G_\omega(t) \geq 0$ is the probability that player i continues acquiring information at time t in state ω.

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
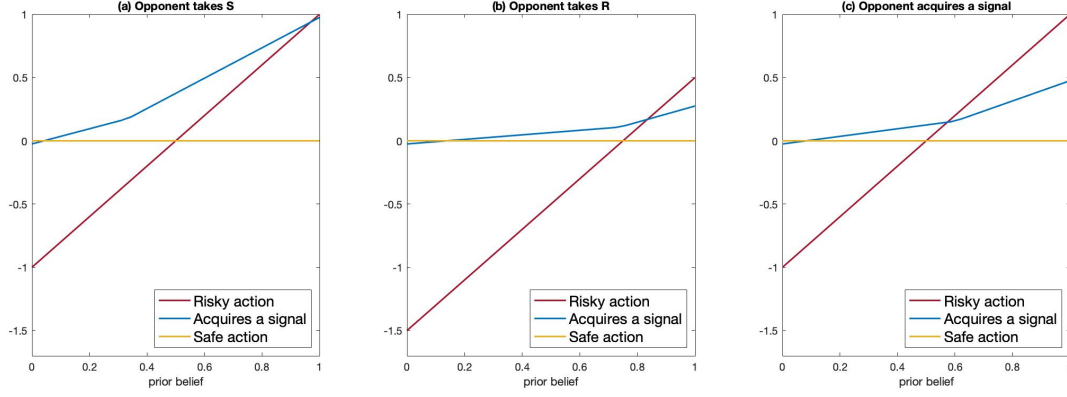
Figure 1: The player's payoffs given opponent's different strategies

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.

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 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 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. [3] When 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

---

[3]The player taking R at time 0 with a positive probability reduces her opponent's expected payoff from taking R at time 1.

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. *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 because she is convinced that the state is more likely to be H. When the DM uses the optimal strategy, if she acquires information, 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 stop 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. *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 Mix 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 Mix 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 [4], 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 Mixed Learning 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 *Mixed Learning Strategy* is a mixed strategy (ρ^{ML}, σ^{ML}) 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 $\bar{T}^{ML} > 0$.

4.4 Main results

The theorem below lists the symmetric equilibria given different priors. Let $\frac{p^M}{1-p^M} := \frac{-(u_L - \triangle_L)}{u_H - \triangle_H}$, $\frac{p^L}{1-p^L} := \frac{-u_L}{u_H}$ 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. *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 Mixed Learning 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 Mix 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, 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 Appendix D. 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

---

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

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 mix strategy, consider a static game where the players do not have access to information [5]. Consider the sufficiently high prior such that there exists an equilibrium where the players take R with a positive probability in this static game [6]. 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 2. *When c is sufficiently small, $p^M < \bar{p}$.*

The cutoff $\bar{p}$ is the upperbound in Proposition 1: the belief above which the player takes R immediately. Lemma 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* can be 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). 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

---

[5]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.

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

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 6 in the appendix. Here in the main text, I give expressions of the marginal cost and marginal benefit from acquiring information and show how they are determined by the opponent's strategy. Suppose player j uses the mixed strategy (ρ, σ) defined in Definition 2. Suppose ρ and σ are differentiable. Then,

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

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

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

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

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

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

$$c + p_t \frac{dF_H(t)}{dt} \bar{\triangle}_H + (1 - p_t) \frac{dF_L(t)}{dt} \bar{\triangle}_L = (1 - p_t) b \left[-u_L + F_L(t) \bar{\triangle}_L \right]. \qquad (3)$$

The left-hand side of (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 6 and Lemma 14 in Appendix D characterises the conditions that the equilibrium strategy satisfies.

Corollary 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 $p_0 \in [p^L, p^M)$. For this range of priors, according to Theorem 1, there exist both kinds of equilibria where the players use the Random Stopping Strategy and where the players use the Immediate Mix Strategy. In other words, there simultaneously exist the 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

17

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 the competition

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

5.1 Vanishing information cost c

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, since the exogenous information cost c is the total cost of acquiring information, when c vanishes, the DM always acquires information until she is certain. 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 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 (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 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 2 and impose Assumption 3.

Assumption 3. $u_L < u_H - \bar{\triangle}_H < 0$.

Assumption 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 [7]. Given Assumption 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 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 3. *Suppose the payoff satisfies Assumption 1 and Assumption 3. There exists a prior cutoff p^{NR} and a non-negative upperbound $T^{PS} = \frac{1}{a} \log \frac{\bar{\triangle}_H}{-(u_H - \bar{\triangle}_H)}$ such that for the prior $p_0 \in (p^{NR}, \tilde{p})$, there exists a symmetric equilibrium where the players use the pure strategy (T^{PS}, S).*

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

---

[7]Note that given Assumption 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.

20

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 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 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 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 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 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 with a 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 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 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.

References

Akcigit, U. & Liu, Q. (2015), 'The Role of Information in Innovation and Competition', *Journal of the European Economic Association* **14**(4), 828–870.

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.

Che, Y.-K. & Mierendorff, K. (2019), 'Optimal Dynamic Allocation of Attention', *American Economic Review* **109**(8), 2993–3029.

Denti, T. (2019), Unrestricted Information Acquisition.

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.

Han, J. & Sangiorgi, F. (2018), 'Searching for information', *Journal of Economic Theory* **175**, 342–373.

Hebert, B. & 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.

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.

Mayskaya, T. (2020), Dynamic Choice of Information Sources.

Merton, R. K. (1957), 'Priorities in Scientific Discovery: A Chapter in the Sociology of Science', *American Sociological Review* **22**(6), 635–659.

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.

Ozdenoren, E., Hoppe-Wewetzer, H. C. & Katsenos, G. (2021), Experimentation, Learning, and Preemption.

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.

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**(PB), 721–738.

A   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

$$\hat{V}^{i,\gamma^j}(0) := \max_{T^i} \int_0^{T^i} \pi_t \left[p_t^i a \mathbb{E}^{\gamma^j} \left[u_H^R \mid H, t \right] - c \right] dt + \pi_{T^i} U^{i,\gamma^j}\left(T^i\right) \tag{4}$$

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

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,γ^j} : $\mathbb{R}_+ \to \mathbb{R}$, where

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

$$\left. U^{i,\gamma^j}(t) - V^{i,\gamma^j}(t) \right\} = 0. \tag{5}$$

The interpretation of (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 (5)) plus the rate of change of the value (labelled as B in (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 (5). The player's best response can then be characterised correspondingly. However, in our problem, (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.

B   Proof of Proposition 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

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

$$\left. U(p) - V(p) \right\} = 0 \qquad (6)$$

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

$$pa \left[u_H - V(p) \right] + (1 - p) b \left[u^S - V(p) \right] + \frac{dV(p)}{dp} (b - a) p (1 - p) = c. \qquad (7)$$

The free boundary solution to (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

$$\left[u_H - u^S - \frac{c}{a} \right] \underline{L} + K \underline{L}^{\frac{b}{b - a}} = \frac{c}{b}. \qquad (8)$$

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 (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, $H\left(p, V(p), V'(p)\right) \leq 0$; (2) if $p \in \left(\underline{p}, \bar{p}\right]$, then, $V(p) \geq U(p)$; (3) if $p < \underline{p}$, then $H\left(p, V(p), V'(p)\right) \leq 0$. At the point where $V(p)$ is not differentiable, that is, at $p = \underline{p}$, I show that $H\left(p, V(p), z\right) \geq 0$ for $z \in D^+$ where $D^+ = \emptyset$ (ignore) and (4) $H\left(p, V(p), z\right) \leq 0$ 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 $H\left(p, U_R(p), U_R'(p)\right) < 0$ 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 $H\left(p, V(p), V'(p)\right) \leq 0$.

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

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.

C Proof of Lemma 1

Consider a strategy $\gamma = (\rho, \sigma)$ as defined in Definition 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, $F_\omega^\gamma(t)$, [8] 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 $\gamma^D = (\rho^D, \sigma)$ 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

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

where

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

When $\triangle > 0$ approaches zero, this gain approaches

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

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 $\tau - \triangle$.

D   Proof of Theorem 1

Let $\frac{p^M}{1-p^M} := \frac{-u_L + \underline{\triangle}_L}{u_H - \bar{\triangle}_H}$, $\frac{p^L}{1-p^L} := \frac{-u_L}{u_H}$ and $\frac{\tilde{p}}{1-\tilde{p}} := \frac{-bu_L - c}{a\triangle_H + c}$ be three prior cutoffs. As $u_L < 0$, when $c < -bu_L$, all of the three cutoffs are positive. The cutoff $\underline{p}$ is defined in Proposition 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 $p^L < p_0 < p^M$.

---

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

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 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 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 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

$$p_0 a dt \left(u_H - \bar{\triangle}_H \right) + \left[1 - p_0 a dt - (1 - p_0) b dt \right] U_R^{IR}(0 + dt) - U_R^{IR}(0). \tag{9}$$

When $dt \to 0$, (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. $\square$

Step 3 I show the following lemma.

Lemma 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,

$$
\frac{p_0}{1 - p_0} = -\frac{\left[m u_L + (1 - m)\left(u_L - \underline{\triangle}_L\right)\right]}{\left[m u_H + (1 - m)\left(u_H - \underline{\triangle}_H\right)\right]}. \tag{10}
$$

For any $p_0 \in \left(p^L, p^M\right)$, there exists $m \in (0, 1)$ such that (10) holds. $\qquad\square$

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 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

$$\frac{d\rho(t)}{dt} = \frac{b(-u_L) - c - L_t c + b\bar{\triangle}_L \int_0^t e^{-bs} \frac{d\rho(s)}{ds} ds - L_t \bar{\triangle}_H a e^{-at}(1 - \rho(t))}{L_t \bar{\triangle}_H e^{-at} + \bar{\triangle}_L e^{-bt}} \tag{11}$$

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 (11). Then Lemma 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

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

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

and

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

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

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,

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

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 7. *The differential equation (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 $L_t = L_0 e^{(b-a)t}$, (11) can be rewritten as

$$\left(\frac{dA}{dt} + bA \right) e^{bt} = \frac{b(-u_L) - c - cL_0 e^{(b-a)t} + b\bar{\triangle}_L \left(A + b \int_0^t A(s) ds \right) - L_0 \bar{\triangle}_H a e^{(b-2a)t}(1 - e^{bt}A)}{L_0 \bar{\triangle}_H e^{(b-2a)t} + \bar{\triangle}_L e^{-bt}}. \tag{12}$$

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

$$z'' + g_1(t) z' + g_2(t) z = g_3(t) \tag{13}$$

where

$$g_0(t) = \frac{1}{L_0 \bar{\triangle}_H e^{2(b-a)t} + \bar{\triangle}_H},$$

$$g_1(t) = b - \frac{b\bar{\triangle}_L + L_0 \bar{\triangle}_H a e^{2(b-a)t}}{g_0(t)},$$

$$g_2(t) = -\frac{b^2 \bar{\triangle}_L}{g_0(t)},$$

$$g_3(t) = \frac{b(-u_L) - c - cL_0 e^{(b-a)t} - L_0 \bar{\triangle}_H a e^{(b-2a)t}}{g_0(t)}.$$

Since $g_0(t)$, $g_1(t)$, $g_2(t)$ and $g_3(t)$ are continuous on $[0, \infty)$, (13) with initial conditions $z(0) = 0$ and $z'(0) = 0$ has a unique solution defined for all t on $[0, \infty)$. The existence of $z(\cdot)$ implies the existence of $\rho(\cdot)$. $\qquad\square$

Lemma 8. $\frac{d\rho}{dt}|_{t=0} > 0$ *iff* $p_0 < \tilde{p}$.

Proof.

$$\frac{d\rho}{dt}\Big|_{t=0} = \frac{b(-u_L) - c - L_0 c - L_0 \bar{\triangle}_H a}{L_0 \bar{\triangle}_H + \bar{\triangle}_L}$$

$$> 0$$

if and only if

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

$\square$

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 $\rho(\cdot)$ increases at the neighbourhood of 0 and the value of ρ is smaller than 1, then, it continues increasing.

Lemma 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 $t = \tau$. $\square$

Next I show that given Lemma 7 and Lemma 9, $\rho(t) = 1$ will be reached in a finite time.

Lemma 10. *Given Lemma 7 and Lemma 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

$$\Phi(t) < \frac{b(-u_L + \bar{\triangle}_L \rho(t)) - c - L_0 e^{(b-a)t} c}{\bar{\triangle}_H a e^{(b-2a)t}(1 - \rho(t))}$$

$$< \frac{b(-u_L + \bar{\triangle}_L) - c - L_0 e^{(b-a)t} c}{\bar{\triangle}_H a e^{(b-2a)t}(1 - \rho(t))}.$$

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 $\rho(T) = 1$. $\square$

Lemma 11. *If $p^L < p_0$ and $\rho(t)$ satisfies the conditions in Lemma 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 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 $\forall t \in [0, T]$. $\qquad\square$

This completes the proof of Lemma 6.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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) $\hat{T} \in [\hat{T}_l, \hat{T}_r]$ 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

$$T_r := \min\left\{ t : \frac{p_0}{1 - p_0} e^{(b-a)t} = \frac{b(-u_L) - c}{c + ae^{-at}\bar{\triangle}_H} \right\} \qquad (14)$$

and

$$T_l := \min\left\{ t : \frac{p_0}{1 - p_0} e^{(b-a)t} = \frac{-u_L}{u_H - (1 - e^{-at})\bar{\triangle}_H} \right\}. \qquad (15)$$

Lemma 12. *If $p_0 < \tilde{p}$, then, $T_r > 0$ exists. If $p_0 < p^L$, then, $T_l > 0$ exists.*

Let

$$\frac{p^*(t)}{1 - p^*(t)} := \frac{\frac{c}{b}}{u_H - \frac{1}{2}\bar{\triangle}_H - \frac{c}{a} + J_2(t)} \qquad (16)$$

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 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 $\underline{p}^* (p_0) < p_0$. $\qquad\square$

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,

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

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

$$\frac{d\lambda (\tau)}{d\tau} = \frac{b (-u_L) - c - \hat{L}_\tau c + b\bar{\triangle}_L \int_0^\tau e^{-bs} \frac{d\lambda(s)}{ds} ds - \hat{L}_\tau \bar{\triangle}_H a e^{-a\tau} (1 - \lambda (t))}{\hat{L}_\tau \left(\eta - \left(u_H - \bar{\triangle}_H\right)\right) e^{-a\tau} + \bar{\triangle}_L e^{-b\tau}} \tag{17}$$

36

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

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

That is,

$$L_0 e^{(b-a)t} < \frac{b(-u_L) - c + b\bar{\triangle}_L \int_0^t e^{-bs}\frac{d\hat{\rho}(s)}{ds}ds}{c + \bar{\triangle}_H ae^{-at}(1 - \hat{\rho}(t))} \tag{18}$$

for all $t > 0$. Equation (18) is derived by substituting in $\tau = t - \hat{T}$, $\lambda(\tau) = \hat{\rho}(t)$ and $\eta = u_H - \left(1 - e^{-a\hat{T}}\right)\bar{\triangle}_H$. The existence of an increasing function $\hat{\rho}(\cdot)$ has been shown in Lemma 7. that satisfies (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,

$$L_0 e^{(b-a)t} < \frac{b(-u_L) - c}{c + ae^{-at}\bar{\triangle}_H} \tag{19}$$

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 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

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

That is,

$$L_0 e^{(b-a)\hat{T}} \geq \frac{-u_L}{u_H - \left(1 - e^{-a\hat{T}}\right)\bar{\triangle}_H}. \tag{20}$$

The lowerbound of $\hat{T}$ is the smallest $\hat{T}$ such that inequality (20) holds. Given Lemma 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 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 $t \le \hat{T}$*

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 (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 $p_0 \ge p^L$.*

Proof. I have shown the existence of an increasing $\hat{\rho}$ function, and (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 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

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

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 $W(0) \ge 0$, which requires

$$L_0 \ge \frac{\frac{c}{b}}{u_H - \frac{1}{2}\bar{\triangle}_H - \frac{c}{a} + J_2\left(\hat{T}\right)} := \underline{L}\left(\hat{T}\right) := \frac{p^*\left(\hat{T}\right)}{1 - p^*\left(\hat{T}\right)}. \tag{21}$$

38

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 14 exists.

Lemma 15. $\underline{p}^* (p_0) \leq p^* \left(\hat{T} \right)$ *for* $\hat{T} \in \left[\hat{T}_l, \hat{T}_r \right]$.

Lemma 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

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

$\square$

Lemma 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 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

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

and

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

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

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

(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 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 $0 \le t < \hat{T}^\beta$, 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

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

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

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

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 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 $\forall t \ge 0$.

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 (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 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

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

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 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 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 $L_0 = \underline{L}^{\beta}\left(\hat{T}_r^{\beta}\right)$. $\square$

Given Lemma 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 $1 - \beta$. $\square$

E Proof of Lemma 2 and Proposition 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 $\frac{c}{b} < \frac{-u_L\left(u_H - \triangle_H\right)}{u_H - \underline{\triangle}_H - \left(u_L - \underline{\triangle}_L\right)}$, we have $p^M < \bar{p}$. Lemma 2 is shown.

The cutoff $\underline{p}$ satisfies

$$\left[u_H - \frac{c}{a}\right]\frac{\underline{p}}{1-\underline{p}} + K\left(\frac{\underline{p}}{1-\underline{p}}\right)^{\frac{b}{b-a}} = \frac{c}{b} \tag{23}$$

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 $\frac{\tilde{p}}{1-\tilde{p}} = \frac{-bu_L - c}{a\triangle_H + c} \to \frac{-bu_L}{a\triangle_H}$ and hence $\tilde{p} \to \frac{\frac{b}{a}\frac{-u_L}{\triangle_H}}{1 + \frac{b}{a}\frac{-u_L}{\triangle_H}} \in (0,1)$.

F Proof of Proposition 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 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

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

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

$$\max\left\{p_t a \left[\max\left\{u_H - \left(1 - e^{-at}\right) \bar{\triangle}_H, 0\right\} - W(t)\right] - (1 - p_t) bW(t) + W'(t) - c,\right.$$
$$\left. U(t) - W(t)\right\} = 0. \tag{24}$$

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 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

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

Proof. Let

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

I show that $W(t)$ is a viscosity solution of the HJB equation (24). For all the points where

43

$W(t)$ 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 $D^+ = \left[W'_L(T^{PS}), 0\right]$ and $H(t, W(t), z) \leq 0$ for $z \in D^-$ where $D^- = \emptyset$.

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

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

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

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

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

$$W'_L\left(T^{PS}\right) = \frac{dp_t}{dt}\mid_{t=T^{PS}}\left[\frac{c}{b} - \left(-u_L - \frac{c}{b}\right)\frac{1-p_{T^{PS}}}{p_{T^{PS}}}\right] + p_{T^{PS}}a\left[\frac{c}{a} - \left(u^S - u_L - \frac{c}{b}\right)\frac{1-p_{T^{PS}}}{p_{T^{PS}}}\right]$$
$$< 0.$$

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 $z \in D^+$.

$\square$

G   Multi-player extension: proof of Proposition 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 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

$$Q_\omega(t) = (1 - F_\omega(t))^{N-1} \tag{25}$$

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

$$F_H(t) = 1 - e^{-at}(1 - \rho(t)) \tag{26}$$

and

$$F_L(t) = \int_0^t \left[e^{-bs} \frac{d\rho(s)}{ds} \right] ds. \tag{27}$$

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

$$c = (1 - p_t) b \left[- \left(u_L - (1 - Q_L(t)) \bar{\triangle}_L \right) \right] + (1 - p_t) \frac{dQ_L}{dt}(t) \bar{\triangle}_L + p_t \frac{dQ_H}{dt}(t) \bar{\triangle}_H. \tag{28}$$

Given (25), (26) and (27), (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 4.

H   Observable actions: proof of Proposition 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. *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

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

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 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.