HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Tatiana Mayskaya

DYNAMIC CHOICE OF INFORMATION SOURCES

Working Paper WP9/2019/05
Series WP9
Research of economics and finance

Moscow
2019

УДК 303
ББК 60в7
М 43

Editor of the Series WP9
“Research of economics and finance”
Maxim Nikitin

Mayskaya Tatiana.
M43 Dynamic Choice of Information Sources* [Electronic resource] : Working paper WP9/2019/05 / T. Mayskaya ; National Research University Higher School of Economics. – Electronic text data (500 Kb). – Moscow : Higher School of Economics Publ. House, 2019. – (Series WP9 “Research of economics and finance”). – 63 p.

The state has two (possibly correlated) binary components, (θ_1, θ_2), $\theta_i \in \{0, 1\}$. Before taking an action, an agent can search for conclusive evidence of $\theta_i = 1$. No matter what the actions and the payoffs from these actions are in different states, any optimal strategy consists of two phases. In a special case when $\theta_1 + \theta_2 \leq 1$, the agent searches in the most promising direction during phase 1 (possibly changing the direction as the search progresses) and completely ignores one of the state components during phase 2. Consequently, when the stakes are high, groups with opposed interests agree on the direction of search.

УДК 303
ББК 60в7

Keywords: Optimal learning, sequential optimal experimental design, Poisson process, limited attention

JEL classification: D83

Tatiana Mayskaya
ICEF, National Research University Higher School of Economics, Russian Federation.
Postal address: Office 3116b, 26 Shabolovka, Moscow, 119049, Russian Federation.
E-mail: tmayskaya@gmail.com

* The study has been funded by the Russian Academic Excellence Project ‘5-100’. I am grateful to Johannes Hörner, Federico Echenique, Jakča Cvitanić, Benjamin J. Gillen, Jean-Laurent Rosenthal, Caroline Thomas and Annie Liang for many productive discussions and helpful suggestions. I also thank seminar audience in Caltech, Oxford, Warwick, Cambridge, TSE, CMU, and HSE.

1 Introduction

How important is the goal of learning? Does it matter who brings evidence to the court or contributes to science? When can people with diverse interests agree on what to research? The dependence of the optimal learning strategy on the goal of learning (payoff matrix) is the central question of this paper.

When we talk about learning, we need to define two things: *how* we learn (tools or means) and *why* we learn (goal or end). In general, the tools are a combination of two extremes: gradual learning and breakthrough learning. Gradual learning describes the process of reading a book or making small, incremental innovations. This paper focuses exclusively on breakthrough learning. I assume there are two hypotheses an agent can investigate. Each hypothesis is either true or false. The agent can find conclusive proof that a given hypothesis is correct, but he cannot falsify that hypothesis. If a hypothesis is true, the agent makes a discovery of its proof at a random time (hence, the name "breakthrough" learning). Whenever he decides to invest in learning, the agent chooses how to split his attention between two hypotheses. The more attention he pays to a given hypothesis, the sooner he finds the proof if this hypothesis is true.

The goal of learning is to take an action with a higher payoff, such as to decide on an economic policy, an investment project, a marketing strategy, or a technological design. The agent chooses when to stop learning and take an action. The payoff from each action depends on the correctness of each of the two hypotheses.

For example, a city council has to decide on the fate of some area. This area has a unique environment and at the same time it is a potential recreation area

for people. Objectively, there are many options for this area, such as whether to develop it for housing, public events, or a national park, to restrict the use of this area by various tax regulations, or leave it as it is and do nothing. Different council members might have different views on the best course of action (plus, simply formulating the options often requires a lot of resources). Assume that had it been known that the environment is not unique and / or any construction is dangerous in this area, all members would have agreed on what to do. Intuitively, additional information might help to resolve the disagreement. If the council discovers another place of similar ecological conditions, ecologists would not insist on protecting this area anymore. If experts prove that the construction is dangerous in the area, construction companies would back off. The question is, whether the council will be able to agree on the *type* of information to invest in, to search for a similar environment or to study the construction safety.[1]

The short answer is, yes, under some quite general conditions, they will agree. A formal argument lies in the characterization of the optimal strategy which I describe below.

Once one hypothesis is proven, the agent faces an optimal stopping problem

[1]This example is based on two real cases. The first one happened in June 2018 in Essen, Germany. A popular musician Ed Sheeran had run into trouble when organizing an open air concert: Sheeran's music might disturb the protected birds called skylarks. While the matter was discussed, a hundred unexploded bombs were found at the venue, and the concert had to be moved to a different venue. See more at https://www.telegraph.co.uk/news/2018/06/13/ed-sheerans-open-air-german-concert-may-thwarted-trees/. I thank Natalia Zabelina for this reference.

The second case happened in Southern California where a bird called California Gnatcatcher like a coastal area, but so do humans. By the time the bird was recognized as an endangered species in 1993, its habitat had been taken over almost completely by housing tracts. See more at https://www.audubon.org/field-guide/bird/california-gnatcatcher and https://www.biologicaldiversity.org/species/birds/coastal_California_gnatcatcher/index.html.

with only one hypothesis to investigate. Since the solution to the optimal stopping problem is already known, the only remaining question is, in the absence of any proof, what behavior is optimal? One possibility is when it is optimal to focus on one hypothesis only and ignore the other one, until either the proof is found or the search is abandoned and the decision is made. Suppose this is not true and it is optimal to investigate both hypotheses (either at the same time or sequentially, it does not matter) before any proof is found or the search is stopped. The main result of the paper (Theorem 1) states that in this case the optimal behavior is to test the hypothesis with the highest index, where hypothesis 1 index is

$$\text{probability of } (1,0) + \text{probability of } (1,1) \times \text{payoff in (1,1) from action } a_1$$

and hypothesis 2 index is

$$\text{probability of } (0,1) + \text{probability of } (1,1) \times \text{payoff in (1,1) from action } a_2,$$

where action a_i is optimal to take when hypothesis i is proven, state $(1,0)$ $[(0,1)]$ corresponds to the situation when hypothesis 1 is true [false] and hypothesis 2 is false [true], and both hypotheses are correct under the state $(1,1)$.

Intuitively, hypothesis i should be tested when this hypothesis is true and the other one is false. This corresponds to the terms *"probability of (1,0)"* and *"probability of (0,1)"*. When both hypotheses are false, it does not matter which one is investigated. Hence there is no term with *"probability of (0,0)"*. Finally, when both hypotheses are true, it is better to focus on the one that, when proven, implies the action that leads to a higher payoff. That explains the terms *"probability of (1,1) × payoff in (1,1) from action a_i"*, $i = 1,2$.

To summarize, the optimal strategy consists of two phases. During the first phase, the index policy described above is used. During the second phase, it is optimal to focus on one hypothesis only and ignore the other one, until either the proof is found or the search is abandoned and the decision is made.

Returning to the example, the council will agree on the type of information to invest in whenever all members value *both* types of information (possibly differently) — so that it is optimal to investigate both hypotheses before any proof is found or the search is stopped (the condition for the first phase). That happens when learning about each hypothesis is cheap, or equivalently, when potential payoffs are high (Theorem 2). Due to the assumption that members' payoffs are the same conditional on any discovery, everybody will use the same indices and therefore agree on the hypothesis to investigate despite having possibly opposed interests in the absence of discoveries.

From the descriptive perspective, the two-phase strategy seems to be used in criminal or air crash investigations. When the stakes are high, such as in the airline industry, the focus of investigation is always on the most likely cause of an accident (in line with the index policy described above in the special case when at most one hypothesis is true, so that it is optimal to investigate the most likely hypothesis), no matter whether it is a pilot error (in which case no major changes to the industry are required) or a mechanical failure (which does have major consequences). Closer to home, researchers seem to follow the same pattern in their work. In his Nobel Prize interview, Robert Aumann said: "People but too often want to emphasize the practical importance, practical applications. And

that is not what science is about. Science is simply following your curiosity. And if you are doing interesting things, then eventually it will find its applications."[2]

Other applications are discussed in Section 5.

Related literature

Literature that studies the dynamics of information collection started with the simplest case when only one information source is available. Since in that case the only question is when to stop learning and take an action, this is called the *optimal stopping problem* (Wald (1947), Peskir and Shiryaev (2006)).

When there are two or more information sources, the problem becomes much harder. Rational inattention literature (Sims (2003)) provides a tractable approach by assuming away any restrictions on the type of collected information. This assumption allows a researcher to "skip" the dynamics and work with a posterior-dependent function (such as Shannon entropy) that measures the cost of collected information. Recently, Steiner et al. (2017) and Zhong (2017) present dynamic rational inattention models, where at each moment of time, the agent faces an unrestricted choice of information type.

When the number and nature of information sources is restricted, the problem is called the *optimal sequential information acquisition problem* (Chaloner and Verdinelli (1995)). The difficulty of the problem and the need to solve it in computer science[3] and econometrics[4] drew attention to myopic strategies

[2] Full interview can be found at https://www.nobelprize.org/mediaplayer/?id=1133.

[3] A classical result from computer science literature is asymptotic optimality of myopic strategies (see Chernoff (1959), Naghshvar and Javidi (2013)).

[4] Chapman et al. (2018) demonstrate how myopically optimized sequential experimentation improves an estimate of loss aversion.

(strategies that maximize the next period payoff neglecting dynamic considerations, myopic strategies are easy to calculate) and their relationship with optimal strategies. Assuming that only one state component is payoff-relevant, Liang et al. (2017) show that a generalized myopic strategy is optimal (Theorem 1) and the myopic strategy is eventually optimal (Theorem 3). In their follow-up paper Liang and Mu (2017) show that the myopic strategy might lead to "learning traps" (persistent inefficiency in information gathering). In my model, the myopic strategy is optimal in phase 2, in line with Theorem 3 from Liang et al. (2017). However, the phase 1 rule is inconsistent with myopic behavior because it largely follows belief-based incentives rather than payoff-based incentives (see discussion on page 20).

Ke et al. (2016), Fudenberg et al. (2018), Ke and Villas-Boas (2019), Liang et al. (2019) derived the optimal strategy when there are two or more information sources modeled as Brownian motions. Each source provides information about one state component through state-dependent drift. Ke et al. (2016) assume independent components. Fudenberg et al. (2018) and Ke and Villas-Boas (2019) focused on two information sources, while Liang et al. (2019) allow for multiple sources with general correlation structure. This approach corresponds to gradual learning.

This paper belongs to the stream of literature that models information sources as Poisson processes with state-dependent intensities, which corresponds to breakthrough learning. Since a breakthrough (or "success") often leads to an immediate decision, there is a close relationship with literature that interprets the success as monetary payoff. Multi-armed bandit literature focuses on maximizing the total number of successes (see Section 5.3), search problems aim to mini-

mize the time needed to achieve the first success (Chatterjee and Evans (2004), Klein and Rady (2011), Francetich et al. (2018)). Presman and Sonin (1990) provide a good review of this literature.

Two papers that are closest to mine are Che and Mierendorff (2019) and Nikandrova and Pancs (2018). Che and Mierendorff (2019) consider the case with two states, (1,0) and (0,1), so their benchmark model is a special case of mine (see Section 3.1). However, they go further and generalize to positive discounting, non-conclusive Poisson signals and non-linear return to attention.[5] Nikandrova and Pancs (2018) consider four states, — (0,0), (1,0), (0,1), and (1,1), — but they assume its components are independent (that corresponds to $\xi = 0$ in my model, see (2) for definition of ξ).[6] Most importantly, both Nikandrova and Pancs (2018) and Che and Mierendorff (2019) focus on a given decision problem by making very restrictive assumptions on the payoff matrix. Therefore, they provide a full characterization of the solution, including what action to take at the stopping time in the absence of discoveries. In contrast, my focus is on the general form of the optimal strategy and I do not give the full characterization of the solution in the strict sense.

2 Model

An agent must choose among a finite set of actions, $\mathscr{A}$. His payoff from these actions depends on what the true state of the world is. Before taking an action, the agent can learn about the state.

[5] Damiano et al. (2019) studied a similar setup but with an additional source of learning. In their model the agent can also choose to experiment with a risky arm that can be either good (state (1,0)) or bad (state (0,1)).

[6] Austen-Smith and Martinelli (2018) are studying multiple sources discrete time version of Nikandrova and Pancs (2018) allowing for exogenous deadline.

Timing. Time $t \geq 0$ is continuous. At each moment, the agent chooses either to stop collecting information by taking an action or to wait and gather more information. Once the action is chosen, the game is over.

Learning. The state of the world is a vector with two binary components, $(\theta_1, \theta_2) \in \{0,1\}^2$. Each state component θ_i has an information source attached to it. If at time t the agent decides to learn about the state, he allocates a unit of attention between two information sources. The amount of attention paid to source i at time t is denoted by $x_i(t) \in [0,1]$, with $x_1(t) + x_2(t) = 1$. The attention process $x_i = \{x_i(t) \mid t \geq 0\}$, together with the state component θ_i, define the time-dependent intensity $\theta_i x_i(t)$ for a Poisson process $N_i = \{N_i(t) \mid t \geq 0\}$, $N_i(0) = 0$, observed by the agent. Note that a jump in $N_i(t)$ reveals $\theta_i = 1$, and the probability of the jump is proportional to $x_i(t)$.[7] Naturally, the attention allocation plan $x = (x_1, x_2)$ has to be measurable with respect to information available at time t. In sum, while learning, the agent observes two information sources, $\{N_1(t) \mid t \geq 0\}$ and $\{N_2(t) \mid t \geq 0\}$, and the more attention he pays to a source, the more informative it is about the corresponding component of the state.

Payoff. Denote by $u_{\theta_1\theta_2}(a)$ the payoff the agent gets if he takes action $a \in \mathscr{A}$ and the true state is (θ_1, θ_2):

	(0,0)	(0,1)	(1,0)	(1,1)
a_1	$u_{00}(a_1)$	$u_{01}(a_1)$	$u_{10}(a_1)$	$u_{11}(a_1)$
a_2	$u_{00}(a_2)$	$u_{01}(a_2)$	$u_{10}(a_2)$	$u_{11}(a_2)$
...	...	...	...	...

[7] Non-linear effect of attention on information precision is studied in Moscarini and Smith (2001) for the optimal stopping problem with Brownian motion. Abandoning linearity is not a technically trivial exercise and therefore it is outside of the scope of this paper. See Section 6.4 in Che and Mierendorff (2019) for that extension in the special case $\theta_1 + \theta_2 = 1$.

I will refer to this matrix as the **payoff matrix**. Since the goal of learning is to take an action with the highest payoff at the true state, the payoff matrix is the set of parameters that characterize the *goal of learning*.

Information collection is costly. For simplicity, assume the flow cost of each information source is 1 (see Appendix C for asymmetric costs). Then the total payoff the agent gets is

$$u_{\theta_1\theta_2}(a)-\tau,$$

where $\tau \geq 0$ is the time when the agent takes action $a \in \mathscr{A}$. Note that I assume no discounting.[8]

A strategy of the agent is a triple (x, τ, α), where α is a function that tells what action the agent takes given the information he has by the stopping time τ.

Denote by $p_{\theta_1\theta_2}$ the agent's belief that the state is (θ_1, θ_2). An optimal strategy maximizes the expected payoff

$$\sup_{(x,\tau,\alpha)} \mathbb{E}_p\left[u_{\theta_1\theta_2}(\alpha)-\tau\right] \tag{1}$$

[8]Solving the model with (exponential) discounting is technically challenging (Nikandrova and Pancs (2018) had to resort to numerical analysis in the special case of my model, see Section 5 of their paper). Having said that, I am not claiming that my results are robust to introducing discounting. On the contrary, I conjecture that the main result, — a source index depends only on the payoff in one state, — will not hold for a discounting case. To see that, let us suppose for simplicity that $\theta_1 = 1$ and $\theta_2 = 1$ are mutually exclusive, so that the probability of state (1,1) is zero. Then the index policy prescribes to pay attention to the most promising direction, that is, to pay attention to source 1 if $\theta_1 = 1$ is more likely than $\theta_2 = 1$, and vice versa. Suppose the payoff matrix is such that knowing θ_1 is more important for the agent than knowing θ_2. Conditional on discovering the state, an infinitely patient agent would search for the truth in the most efficient way: he chooses the most promising direction. However, when the agent is impatient, he would want to learn a more important state sooner rather than later and therefore choose source 1 even if the probability of (0,1) is slightly higher than the probability of (1,0). A formal argument appeals to Che and Mierendorff (2019) analysis of the special case with two states, (1,0) and (0,1), where they characterize the solution for the discounting case as well (see formula (A.13) in Che and Mierendorff (2019)).

3 An Optimal Strategy

The optimization problem (1) has the Markov property, with the belief vector $p = (p_{00}, p_{01}, p_{10}, p_{11})$ playing the role of a state variable with dimensionality 3. When full attention is paid to source 1, the belief $p_{10}+p_{11}$ about the state $\theta_1 = 1$ decreases, while the conditional beliefs

$$\mathbb{P}[\theta_2 = 1 \mid \theta_1 = 1] = \frac{p_{11}}{p_{10}+p_{11}}, \quad \mathbb{P}[\theta_2 = 1 \mid \theta_1 = 0] = \frac{p_{01}}{p_{00}+p_{01}}$$

stay the same. This process is illustrated in Figure 1.

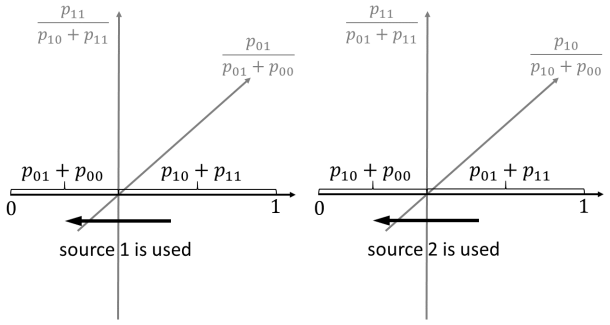

Figure 1: Belief movement in the absence of jumps when full attention is paid to one source.

The class of optimal strategies is best described through "regimes." A regime itself is a Markovian strategy which is a function of the current state p.

Regime i, $i \in \{1,2\}$ The agent uses only source i and ignores the other source. More precisely, this regime is the solution to the optimal stopping problem when only source i is available.

Regime $(0, a_1, a_2)$, $a_1, a_2 \in \mathscr{A}$. The agent is indifferent between both sources

(any attention rule can be used) whenever

$$p_{10}+p_{11}u_{11}(a_1) < p_{11}\left(\max_{a\in\mathscr{A}} u_{11}(a)-1\right), \quad p_{01}+p_{11}u_{11}(a_2) < p_{11}\left(\max_{a\in\mathscr{A}} u_{11}(a)-1\right).$$

Otherwise, the agent must use source 1 if

$$p_{10} + p_{11}u_{11}(a_1) > p_{01} + p_{11}u_{11}(a_2),$$

source 2 if

$$p_{10} + p_{11}u_{11}(a_1) < p_{01} + p_{11}u_{11}(a_2),$$

and split his attention according to $x_1 = \frac{p_{10}}{p_{10}+p_{01}}$, $x_2 = \frac{p_{01}}{p_{10}+p_{01}}$ on the line

$$p_{10} + p_{11}u_{11}(a_1) = p_{01} + p_{11}u_{11}(a_2).$$

Figures 2 and 3 illustrate the regime through belief movements $p(t)$ in the absence of jumps. Suppose the agent starts with $p_{10}(0)+p_{11}(0)u_{11}(a_1) > p_{01}(0)+p_{11}(0)u_{11}(a_2)$. Then he uses source 1 until $p_{10}(t) + p_{11}(t)u_{11}(a_1) = p_{01}(t) + p_{11}(t)u_{11}(a_2)$ (or he leaves the regime earlier). Then he splits his attention according to $x_1(t) = \frac{p_{10}(t)}{p_{10}(t)+p_{01}(t)}$, $x_2(t) = \frac{p_{01}(t)}{p_{10}(t)+p_{01}(t)}$, so that to stay on the line $p_{10}(t) - p_{01}(t) - \text{const} \times p_{11}(t) = 0$.[9,10] This pattern continues until the agent leaves the regime.

Theorem 1 formalizes the main result of the paper. It describes a class of strategies to which any optimal strategy belongs.

[9] By Bayes' rule, $dp_{11} = -p_{11}(p_{00}+p_{01}x_1+p_{10}x_2)dt$, $dp_{10} = p_{10}((p_{11}+p_{01})x_2-(p_{00}+p_{01})x_1)dt$, $dp_{01} = p_{01}((p_{11}+p_{10})x_1-(p_{00}+p_{10})x_2)dt$. Then $d(p_{10}-p_{01}-\text{const}\times p_{11}) = (p_{10}x_2-p_{01}x_1)dt$ taking into account $x_1+x_2=1$, $p_{00}+p_{01}+p_{10}+p_{11}=1$ and $p_{10}-p_{01}-\text{const}\times p_{11}=0$.

[10] The line $p_{10} + p_{11}u_{11}(a_1) = p_{01} + p_{11}u_{11}(a_2)$ along which the agent splits attention between the sources is called a *turnpike* in Presman and Sonin (1990): in a neighborhood of a turnpike the system always comes to it and subsequently moves along it (p. 29).

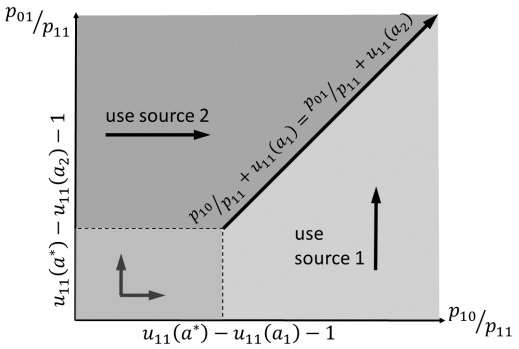

Figure 2: Regime $(0, a_1, a_2)$ for $p_{11} > 0$. Here $u_{11}(a^*) = \max_{a \in \mathcal{A}} u_{11}(a)$.

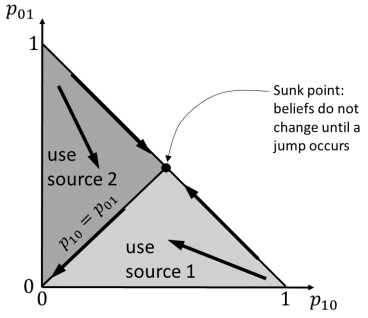

Figure 3: Regime $(0, a_1, a_2)$ for $p_{11} = 0$.

Theorem 1. *Any optimal strategy consists of two phases: it follows the phase 1 rule up until a certain moment in time t^* or until a jump occurs (whichever happens first), then switches to phase 2.*

Phase 1: For any current belief p, regime $(0, a_1(p), a_2(p))$ is used. Action $a_i(p)$

is defined as the action that is optimal to take if source i produces a jump at the moment when the agent's current belief is p.[11]

Phase 2: If the agent enters the phase after source 1 (2) produced a jump, regime 2 (1) is used. Otherwise, regime i is used.

Theorem 1 does not give the full characterization of the set of optimal strategies, leaving some variables undefined. First of all, one such variable is the switching time $t^* \geq 0$ between the phases. Another variable is $i \in \{1,2\}$, which characterizes the regime used during phase 2 after the switch at moment t^*. Moreover, regimes 1 and 2 themselves are defined in an indirect way as the solution to the optimal stopping problem. The precise algorithm how to find the set of optimal strategies in a given decision problem is given in Appendix A.

Theorem 1 deals with the general payoff matrix. In applications, the number of options and their relationship with the state are often subject to very restrictive application-dependent assumptions. When these assumptions are given, Theorem 1 could be used as a tool to find the full characterization of the set of optimal strategies. Below I consider two examples taken from Che and Mierendorff (2019) and Nikandrova and Pancs (2018).

3.1 Example 1: Che and Mierendorff (2019)

There are only two states, (1,0) and (0,1), meaning that $p_{00} = p_{11} = 0$, $p_{10} + p_{01} = 1$. The agent must choose between two actions, $\mathscr{A} = \{a_1, a_2\}$.

[11] More precisely, actions a_1 and a_2 are defined as follows. If source 1 produces a jump, it becomes useless and the agent faces the optimal stopping problem with source 2. The solution to the optimal stopping problem is characterized by the belief threshold $\underline{p}_{11}$ and action a_1 such that the agent uses source 2 until p_{11} becomes as low as $\underline{p}_{11}$, at which moment he takes action a_1, or a jump occurs, at which moment he takes action a^* that is optimal in state (1,1). Action a_2 is defined in a symmetric way.

Taking action a_1 [a_2] is optimal when the state is (1,0) [(0,1)]:

$$u_{10}(a_1) > u_{10}(a_2), \quad u_{01}(a_2) > u_{01}(a_1).$$

The phase 1 rule corresponds to the regime $(0, a_1, a_2)$. Indeed, if source 1 produces a jump at any current belief, state (1,0) is revealed and therefore it is optimal to take action a_1.

Figure 3 illustrates the regime $(0, a_1, a_2)$ when $p_{11} = 0$. The assumption $p_{00} = 0$ simplifies the picture even further — see Figure 4.

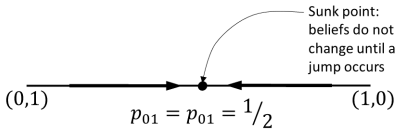

Figure 4: Regime $(0, a_1, a_2)$ for $p_{11} = p_{00} = 0$.

Two observations simplify the application of Theorem 1. First, it is optimal to stop whenever a jump is observed. Thus, if the agent enters phase 2 after observing a jump, he immediately stops and takes an action. Second, the Markov property guarantees that either phase 1 has zero length ($t^* = 0$), or it has an infinite length ($t^* = +\infty$), or the agent does not change how he splits his attention between the sources as he moves from phase 1 ($t < t^*$) to phase 2 ($t > t^*$). Indeed, suppose $p_{10}(t^*) < p_{01}(t^*)$. Then the phase 1 rule implies using only source 2. If the agent switches to source 1 at moment t^*, his belief starts moving in the opposite direction. By the Markov property, an optimal strategy is a function of beliefs, which implies two different attention plans are optimal for the same belief. Contradiction. Note that when the agent does not change his

allocation of attention when entering phase 2, there is an ambiguity to how the moment t^* should be defined. Without loss of generality, let t^* be the earliest moment. Then in this example phase 1 has either zero length ($t^* = 0$), or an infinite length ($t^* = +\infty$).

According to Theorem 1, any optimal strategy takes one of three forms (their names are taken from Che and Mierendorff (2019)):

a) *no learning*: the agent does not use any source and takes an action right away (both phases have zero length);

b) *own-biased learning*: the agent pays full attention to only one source and ignores the other one (phase 1 has zero length, phase 2 has a positive length);

c) *opposite-biased learning*: the agent pays full attention to source 1 when $p_{10} > p_{01}$, he pays full attention to source 2 when $p_{10} < p_{01}$, and he splits his attention in half between the sources when $p_{10} = p_{01}$ until the state is revealed (phase 1 has an infinite length).

3.2 Example 2: Nikandrova and Pancs (2018)

The state components θ_1 and θ_2 are independent, meaning that $p_{11} = (p_{10} + p_{11})(p_{01} + p_{11})$. The agent must choose between two actions, $\mathscr{A} = \{a_1, a_2\}$, with payoffs $u_{\theta_1\theta_2}(a_i) = \theta_i$.

The phase 1 rule corresponds to the regime $(0, a_1, a_2)$. Indeed, if source i produces a jump at any current belief, state $\theta_i = 1$ is revealed and therefore action a_i leads to the highest possible payoff 1. Since $u_{11}(a_1) = u_{11}(a_2)$, the phase 1 rule is to use the source that corresponds to the most likely state.

According to Theorem 1, any optimal strategy takes one of three forms:

a) the agent does not use any source and takes an action right away;

b) the agent pays full attention to only one source and ignores the other one;

c) there exists a moment $t^* > 0$ such that before that moment and before the state is revealed,

   (a) if $p_{10} > p_{01}$, then the agent pays full attention to source 1,

   (b) if $p_{10} < p_{01}$, then the agent pays full attention to source 2,

   (c) if $p_{10} = p_{01}$, then the agent splits his attention in half between the sources,

   and the agent pays full attention to only one source after the moment t^* until it is optimal to stop learning and make the decision.

Note that the same conclusion remains true even if the state components are not independent (though the threshold t^* and the optimal stopping time that Nikandrova and Pancs (2018) have found will be qualitatively different).

3.3 Discussion

The phase 1 rule features an *index* policy: use source i with the highest $\mathbb{P}(\theta_i = 1) + \mathbb{P}(\theta_1 = \theta_2 = 1)\, u_{11}(a_i)$.[12,13] This rule is intuitive: choose source i

[12] When both $p_{10} + p_{11}u_{11}(a_1)$ and $p_{01} + p_{11}u_{11}(a_2)$ are less than $p_{11}\left(\max_{a \in \mathscr{A}} u_{11}(a) - 1\right)$, the agent is indifferent between the sources. For simplicity of exposition, I succumb to a slight abuse by sometimes referring to the optimal policy as an index policy within phase 1.

[13] I apologize for using the term "index" (due to lack of my imagination to come up with a better alternative) to those readers who are familiar with *Gittins indices*, or *dynamic allocation indices* (Gittins et al. (2011)). Gittins proved that an index policy, — a policy where at any moment the arm

that reveals the state θ_i with higher probability ($\mathbb{P}(\theta_i = 1)$), with some adjustment. Digging deeper, the rule combines four possibilities. First, when the true state is $\theta_1 = \theta_2 = 0$, both sources are useless, so that possibility leaves the agent indifferent between them:

$$p_{00}: \quad 1 \quad \text{vs} \quad 1.$$

Second, when the true state is $\theta_1 = 1$, $\theta_2 = 0$, only source 1 can produce a jump, so the agent prefers source 1:

$$p_{10}: \quad 1 \quad \text{vs} \quad 0.$$

Third, when the true state is $\theta_1 = 0$, $\theta_2 = 1$, only source 2 can produce a jump, so the agent prefers source 2:

$$p_{01}: \quad 0 \quad \text{vs} \quad 1.$$

Finally, when the true state is $\theta_1 = \theta_2 = 1$, the agent prefers the source that gives him the higher payoff in case of a jump:

$$p_{11}: \quad u_{11}(a_1) \quad \text{vs} \quad u_{11}(a_2).$$

Weighting these possibilities with their probability, I get

$$\underbrace{p_{10} + p_{11}u_{11}(a_1)}_{\text{source 1 index}} \quad \text{vs} \quad \underbrace{p_{01} + p_{11}u_{11}(a_2)}_{\text{source 2 index}}.$$

with the highest Gittins index should be chosen, — is optimal for the multi-armed bandit problem. While the expression for the index is different in my model, an index policy is still optimal, though only conditional on being in phase 1. Moreover, another key property of the Gittins index that my index does not inherit is its independence of the characteristics of the other arms (in particular, their running time through p_{11}). What my index does inherit from the Gittins index is that they both are strictly increasing in the probability the arm generates success (see Banks and Sundaram (1992) for a generalization of this property for the Gittens index).

Another way to interpret the optimality of two-phase strategy is to think of two types of incentives it should balance. By focusing on the goal to reveal the state, the agent follows *belief*-based incentives: the best source is the one that reveals the state with higher probability according the agent's *beliefs*. For example, if $p_{10} > p_{01}$, then source 1 is better according to belief-based incentives. By focusing on the goal to learn about the most payoff-relevant state, the agent follows *payoff*-based incentives. For example, if $u_{1\theta_2}(a) = u_{0\theta_2}(a)$ for all $a \in \mathscr{A}$ and $\theta_2 \in \{0,1\}$, then source 2 is better according to payoff-based incentives. Loosely speaking, the phase 1 rule puts more weight on belief-based incentives, saying that the agent should focus only on what happens when a jump occurs. In contrast, the phase 2 rule brings all considerations on the table, so that the last source should be chosen with all of them in mind. Such priority ordering — first focus more on finding the truth by following belief-based incentives, then more on the goal of learning by following payoff-based incentives — follows intuitively from the dynamics itself: when the agent is about to abandon the search without finding any conclusive evidence, the goal of learning matters the most.[14,15]

An interesting thing to note here is that when the agent splits attention between the sources, he pays **more** attention to the source with **smaller** payoff

[14]Belief vs payoff-based incentives trade-off loosely corresponds to accuracy vs speed trade-off in Che and Mierendorff (2019): when high accuracy is needed, belief-based incentives prevail; when delay is intolerable, payoff-based incentives play an important role. The connection between accuracy and phase 1 is formalized in Theorem 2.

[15]There is an interesting intuitive connection with Forand (2015) who studied a three-armed bandit model with maintenance costs. He showed that a project that is less likely to succeed could be optimally chosen for development. That happens when this project is about to be irreversibly abandoned. Similarly, a source with a lower index could be chosen during phase 2 which starts when the learning process is about to be irreversibly abandoned.

in case of a jump: $x_1 - x_2 = \frac{p_{10}-p_{01}}{p_{10}+p_{01}} = \frac{p_{11}(u_{11}(a_2)-u_{11}(a_1))}{p_{10}+p_{01}}$. If I put it that way, that sounds counterintuitive. The "missing" part in this logic is the fact that the indifference happens when the indices are equal. That means the source with **smaller** payoff in case of a jump is the source with **higher** probability of producing a jump. Thus, the agent faces a trade-off: pay more attention to the source with higher payoff in case of a jump (payoff-based incentives) or to the source with higher probability of producing a jump (belief-based incentives). Since the agent should choose the latter, this reinforces my message: the agent puts more weight on belief-based incentives during phase 1.

I conclude the discussion by singling out one **special case** of my model when there are only three states: (0,0), (1,0), and (0,1). This case is similar to the settings studied in Klein and Rady (2011) (Section 5) and in Klein (2013) within the bandit framework and generalizes Che and Mierendorff (2019). In that case the phase 1 rule is simplified to using the source with the highest probability of generating a signal:

$$\underbrace{p_{10}}_{\text{source 1 index}} \quad \text{vs} \quad \underbrace{p_{01}}_{\text{source 2 index}} .$$

This special case is interesting because the phase 1 rule does not depend on the payoff matrix at all here. As long as the agent is in phase 1 — the condition for *being* in phase 1 *does* depend on the payoff matrix, incidentally, because t^* depends on the payoff matrix — his optimal allocation of attention does not depend on the payoff matrix, he simply wants to find the truth. Note that this special case is about having two competing hypotheses (like two causes of an accident), plus "everything else" that cannot be discovered by the information

sources — a situation that is not so uncommon in real life (see Klein and Rady (2011) for examples).[16]

4 Proof of Theorem 1

4.1 Change of variables

The belief vector p plays the role of a state variable with dimentionality 3. It turns out that it is possible to decrease the dimensionality of the state space to 2 by introducing a change of variables.

Assume that the initial belief p belongs to[17]

$$\mathscr{P} = \{(p_{00}, p_{01}, p_{10}, p_{11}) \mid p_{00} > 0,\ p_{01} > 0,\ p_{10} > 0,\ p_{11} \geq 0,\ p_{00}+p_{01}+p_{10}+p_{11} = 1\}.$$

For any $p \in \mathscr{P}$, denote $\rho(p) = (q_1(p), q_2(p), \xi(p))$, where

$$q_1 = \frac{p_{00}}{p_{10}}, \quad q_2 = \frac{p_{00}}{p_{01}}, \quad \xi = \frac{p_{11} - (p_{10} + p_{11})(p_{01} + p_{11})}{p_{10} p_{01}}. \tag{2}$$

Then function $\rho : \mathscr{P} \mapsto (0, +\infty)^2 \times [-1, +\infty)$ is one-to-one:

$$p_{00} = \frac{q_1 q_2}{1 + \xi + q_1 + q_2 + q_1 q_2},\ p_{01} = \frac{q_1}{1 + \xi + q_1 + q_2 + q_1 q_2},$$
$$p_{10} = \frac{q_2}{1 + \xi + q_1 + q_2 + q_1 q_2},\ p_{11} = \frac{1 + \xi}{1 + \xi + q_1 + q_2 + q_1 q_2} \tag{3}$$

[16] With three states, state (0,0) can be interpreted as "none of the above", in the spirit of being aware of one's unawareness (as in Karni and Vierø (2017)), or in the spirit of "unknowable" state (as in Dumav and Stinchcombe (2013)).

[17] I omit the proof for other cases. If $p_{11} = p_{10} = 0$ or $p_{11} = p_{01} = 0$, then one component of the state is known to be 0 and it is an optimal stopping problem. If $p_{00} = p_{10} = 0$ or $p_{00} = p_{01} = 0$, then one component of the state is known to be 1 and it is again an optimal stopping problem. Case $p_{11} = p_{00} = 0$ is solved in Che and Mierendorff (2019). Case $p_{01} = p_{10} = 0$ effectively features only one source and therefore leads to an optimal stopping problem. When $p_{00} \geq 0$ and $p_{10} p_{01} p_{11} > 0$, one can set $q_1 = \frac{p_{01}}{p_{11}}$, $q_2 = \frac{p_{10}}{p_{11}}$, $\xi = \frac{p_{11} - (p_{10}+p_{11})(p_{01}+p_{11})}{p_{10} p_{01}}$. When $p_{10} = 0$ and $p_{01} p_{00} p_{11} > 0$, one can work with $q_1 = \frac{p_{01}}{p_{11}}$ and $q_2 = \frac{p_{00}}{p_{01}}$. Similarly for $p_{01} = 0$. Note that in all three cases the state variable $q = (q_1, q_2)$ moves as $dq_1 = q_1 x_1 dt$, $dq_2 = q_2 x_2 dt$ in the absence of jumps.

Lemma 1 shows how $\rho(p(t))$ changes over time.

Lemma 1. *In the absence of jumps, the belief vector p stays in $\mathscr{P}$ and $d\xi = 0$, $dq_1 = q_1 x_1 dt$, $dq_2 = q_2 x_2 dt$.*

Remarkably, the variable ξ stays constant. That allows one to decrease the dimensionality of the state space by taking $q = (q_1, q_2)$ as a state variable. Moreover, if the agent pays attention only to source i, only variable q_i changes.

Once $\theta_1 = 1$ is revealed, the dimentionality of the belief space shrinks to 1. A jump in $N_1(t)$ changes $p \in \mathscr{P}$ to $\tilde{p}$ with $\tilde{p}_{00} = \tilde{p}_{01} = 0$, $\tilde{p}_{10} = \frac{p_{10}}{p_{10}+p_{11}}$, $\tilde{p}_{11} = \frac{p_{11}}{p_{10}+p_{11}}$. Based on (3), $\tilde{p}_{10} = \frac{p_{10}}{p_{10}+p_{11}} = \frac{q_2}{1+\xi+q_2}$ and $\tilde{p}_{11} = \frac{p_{11}}{p_{10}+p_{11}} = \frac{1+\xi}{1+\xi+q_2}$. Note that substitution of $q_1 = 0$ to the expressions for p_{00}, p_{01}, p_{10} and p_{11} in (3) gives the same expressions as for $\tilde{p}_{00}$, $\tilde{p}_{01}$, $\tilde{p}_{10}$ and $\tilde{p}_{11}$. This observation triggers the assumption that the jump in $N_1(t)$ corresponds to the change from (q_1, q_2) to $(0, q_2)$. Similarly, a jump in $N_2(t)$ corresponds to the change from (q_1, q_2) to $(q_1, 0)$. This assumption allows me to keep using q as a state variable even after a jump, with the following transition rules: (1) ξ never changes and is defined by prior beliefs $p(0) \in \mathscr{P}$ according to (2); (2) $q_1(0)$ and $q_2(0)$ are defined by prior beliefs $p(0) \in \mathscr{P}$ according to (2); (3) in the absence of a jump from $N_i(t)$, $q_i(t)$ is changing according to $dq_i = q_i x_i dt$; (4) a jump from $N_i(t)$ sets $q_i(t) = 0$; (5) at any moment, the belief vector can be recovered using (3).[18]

[18] Note that $1 + \xi + q_1 + q_2 + q_1 q_2 = 0$ happens only if $\xi = -1$, $q_1 = q_2 = 0$. But if $\xi = -1$, then $p_{11} = 0$, which excludes the event that both sources produce jumps. Thus, $q_1 + q_2 > 0$ whenever $\xi = -1$ and therefore $1 + \xi + q_1 + q_2 + q_1 q_2 > 0$ at any moment.

4.2 Optimal strategy on the boundaries

Once a jump is observed, the state q jumps to one of the boundaries, — $(0, q_2)$ for a jump in source 1 and $(q_1, 0)$ for a jump in source 2, — and the agent faces the decision problem with only one information source, source 2 for boundary $(0, q_2)$ and source 1 for boundary $(q_1, 0)$. This is a classical optimal stopping problem with the following solution.

Let us start with $(q_1(0), 0)$. If $q_1(0) = 0$, then it is optimal to take action a^* right away, where a^* that maximizes $u_{11}(a)$ over $a \in \mathscr{A}$:

$$a^* \in \mathscr{A}: \quad u_{11}(a^*) = \max_{a \in \mathscr{A}} u_{11}(a).$$

Suppose $q_1(0) > 0$. Recall that in the absence of a jump from source 1, $q_1(t)$ is increasing. Fix any stopping threshold $\bar{q}_1 \geq q_1(0)$. The agent uses source 1 as long as $0 < q_1(t) < \bar{q}_1$. He stops when he observes a jump in $N_1(t)$ or when $q_1(t)$ reaches $\bar{q}_1$, whichever happens first. Once a jump is observed, the state moves to $q_1(t) = 0$, the agent stops and takes action a^*. Once $q_1(t) = \bar{q}$, the agent stops learning and takes some action $a \in \mathscr{A}$.

Lemma 2. *The expected payoff from the strategy described above is*

$$U(q_1(0), 0, a) + \frac{(\bar{q}_1 - q_1(0))R(a)}{\bar{q}_1(1 + \xi + q_1(0))} - \frac{q_1(0)}{1 + \xi + q_1(0)} \log\left(\frac{\bar{q}_1}{q_1(0)}\right), \tag{4}$$

where

$$U(q_1, q_2, a) = \frac{q_1 q_2 u_{00}(a) + q_1 u_{01}(a) + q_2 u_{10}(a) + (1 + \xi) u_{11}(a)}{1 + \xi + q_1 + q_2 + q_1 q_2}$$

is the expected payoff from action $a \in \mathscr{A}$ given the state (beliefs) (q_1, q_2), and

$$R(a) = (1 + \xi)(u_{11}(a^*) - u_{11}(a) - 1).$$

Maximization of (4) with respect to $\bar{q}_1 \geq q_1(0)$ leads to the unique optimal threshold $\bar{q}_1 = \max\{R(a), q_1(0)\}$. At this threshold, the expected payoff (4) becomes

$$\frac{q_1(0)}{1+\xi+q_1(0)}\left(f_2(a,q_1(0)) + \log(q_1(0)) + \frac{(1+\xi)(u_{11}(a^*)-1)}{q_1(0)}\right), \qquad (5)$$

where

$$f_2(a,q) = \begin{cases} u_{01}(a) - \log R(a) - 1, & R(a) \geq q, \\ u_{01}(a) - \frac{R(a)}{q} - \log(q), & R(a) < q. \end{cases}$$

Maximization of (5) with respect to $a \in \mathscr{A}$ leads to $a_2(q_1(0))$, which is defined as $a \in \mathscr{A}$ that maximizes $f_2(a, q_1(0))$. Note that in consensus with the dynamic programming principle, as long as $0 < q_1(t) \leq R(a_2(q_1(t)))$, the optimal threshold $\bar{q}_1 = R(a_2(q_1(t)))$ does not change with time t.

Similarly, for the boundary $(0, q_2)$, it is optimal to use source 2 as long as $0 < q_2 < R(a_1(q_2))$, stop and take action a^* at $q_2 = 0$, and stop and take action $a_1(q_2)$ at any $q_2 \geq R(a_1(q_2))$. Action $a_1(q)$ is defined as an action that maximizes[19]

$$f_1(a,q) = \begin{cases} u_{10}(a) - \log R(a) - 1, & R(a) \geq q, \\ u_{10}(a) - \frac{R(a)}{q} - \log(q), & R(a) < q. \end{cases}$$

4.3 The Hamilton-Jacobi-Bellman equation

Let $V(q_1, q_2)$ be the expected payoff from an optimal strategy given $q_1(0) = q_1$, $q_2(0) = q_2$. Following the literature on dynamic programming, let us call this function the value function.

Let us fix an optimal strategy. The agent should not be ex ante strictly better off by paying full attention to source 1 during an infinitely small interval $t \in$

[19] Note that $\lim_{q\to 0} a_1(q) = \lim_{q\to 0} a_2(q) = a^*$. Define $a_1(0) = a_2(0) = a^*$.

$[0, dt]$ and then using the optimal strategy, than using the optimal strategy from the moment $t = 0$. If the agent pays full attention to source 1 during interval $t \in [0, dt]$, he spends $-dt$ and observes a jump with probability $(p_{10} + p_{11})dt = \frac{(1+\xi+q_2)dt}{1+\xi+q_1+q_2+q_1q_2}$:

$$V(q_1, q_2) \geq -dt + \frac{(1+\xi+q_2)V(0,q_2)}{1+\xi+q_1+q_2+q_1q_2}dt + \left(1 - \frac{1+\xi+q_2}{1+\xi+q_1+q_2+q_1q_2}dt\right)V(q_1+q_1dt, q_2).$$

The Taylor expansion $V(q_1 + q_1dt, q_2) = V(q_1, q_2) + \frac{\partial V(q_1,q_2)}{\partial q_1}q_1dt$ gives

$$V(q_1, q_2) \geq V(q_1, q_2) + \mathscr{L}_1(q_1, q_2; V)dt,$$

where

$$\mathscr{L}_1(q_1, q_2; V) = q_1\frac{\partial V(q_1,q_2)}{\partial q_1} + \frac{(1+\xi+q_2)(V(0,q_2) - V(q_1,q_2))}{1+\xi+q_1+q_2+q_1q_2} - 1.$$

Similarly, the agent should not be ex ante strictly better off by using only source 2:

$$V(q_1, q_2) \geq V(q_1, q_2) + \mathscr{L}_2(q_1, q_2; V)dt,$$

where

$$\mathscr{L}_2(q_1, q_2; V) = q_2\frac{\partial V(q_1,q_2)}{\partial q_2} + \frac{(1+\xi+q_1)(V(q_1,0) - V(q_1,q_2))}{1+\xi+q_1+q_2+q_1q_2} - 1,$$

or stop and take the optimal action:

$$V(q_1, q_2) \geq U(q_1, q_2),$$

where

$$U(q_1, q_2) = \max_{a \in \mathscr{A}} U(q_1, q_2, a).$$

That leads us to the sufficient conditions for optimality:

Lemma 3. *Function* $V : [0, +\infty)^2 \mapsto (-\infty, +\infty)$ *is the value function if*

1. *it is continuous,*
2. *it is continuously differentiable everywhere on* $[0, +\infty)^2$ *except for a set of Lebesgue measure 0,*
3. *it is the expected payoff from some strategy with the initial beliefs that correspond to the argument of this function,*
4. *and in all points of differentiability it satisfies the Hamilton-Jacobi-Bellman equation:*

$$\max\{\mathscr{L}_1(q_1, q_2; V),\ \mathscr{L}_2(q_1, q_2; V),\ U(q_1, q_2) - V(q_1, q_2)\} = 0. \tag{6}$$

Note that the above derivations also show $\mathscr{L}_i(q_1, q_2; V) = 0$ whenever V is the expected payoff from a strategy that prescribes using only source i at the beginning.

4.4 Description of an optimal strategy

In this section I reformulate Theorem 1 using the change of variables introduced in Section 4.1.

Regime $(0, a_1, a_2)$, $a_1, a_2 \in \mathscr{A}$. The agent is indifferent between both sources (any attention rule can be used) whenever

$$q_1 < R(a_2), \quad q_2 < R(a_1).$$

Otherwise, the agent must use source 1 if

$$q_1 + R(a_1) < q_2 + R(q_2),$$

source 2 if

$$q_1 + R(a_1) > q_2 + R(q_2),$$

and split his attention according to $x_1 = \frac{q_2}{q_1+q_2}$, $x_2 = \frac{q_1}{q_1+q_2}$ on the line

$$q_1 + R(a_1) = q_2 + R(q_2).$$

By Lemma 1, the attention rule $x_1 = \frac{q_2}{q_1+q_2}$, $x_2 = \frac{q_1}{q_1+q_2}$ leaves the agent on the line $q_1 - q_2 = \text{const}$ (this fact has already been proven for the original belief space in footnote 9). Figure 5 illustrates the movement of $q(t)$ in regime $(0, a_1, a_2)$ (this is an analog of Figure 2).

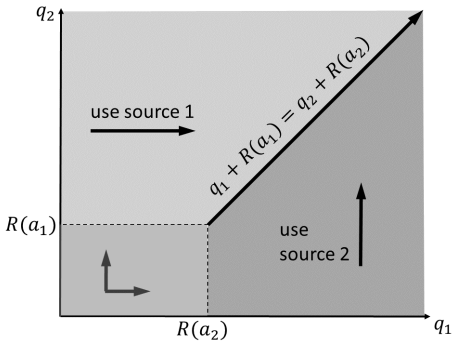

Figure 5: Regime $(0, a_1, a_2)$ in (q_1, q_2) space.

Theorem 1. *Any optimal strategy consists of two phases: it follows the phase 1 rule up until a certain moment in time t^* or until a jump occurs (whichever happens first), then switches to phase 2.*

Phase 1: For any current state (q_1, q_2), regime $(0, a_1(q_2), a_2(q_1))$ is used.

Phase 2: If the agent enters the phase after source 1 (2) produced a jump, regime 2 (1) is used. Otherwise, regime i is used.

4.5 Verification

Let (q_1, q_2) be the initial state. Denote by $W(q_1, q_2, t^*, i)$ the expected payoff from a strategy described in Theorem 1. Maximization over $t^* \geq 0$ and $i \in \{1, 2\}$ gives a candidate for an optimal strategy. Let $V(q_1, q_2)$ be its expected payoff:

$$V(q_1, q_2) = W(q_1, q_2, t^*, i) = \max_{(\tilde{t}^*, \tilde{i})} W(q_1, q_2, \tilde{t}^*, \tilde{i}).$$

A maximizer (t^*, i) is not necessarily unique. This nonuniqueness is twofold: different maximizers might correspond to different strategies or to the same strategy. When it is the latter, assume $t^* \geq 0$ is the largest moment in time when switching from phase 1 to phase 2 is optimal.

This strategy is optimal if V is the value function. The only two conditions in Lemma 3 that require some work to prove is $\mathscr{L}_1(q_1, q_2; V) \leq 0$ and $\mathscr{L}_2(q_1, q_2; V) \leq 0$.

Suppose the alleged optimal strategy prescribes the agent to stop and take action a right away, so that $V(q_1, q_2) \equiv U(q_1, q_2, a)$ in some neighborhood of the initial state.[20] By definition, this strategy is weakly better than the strategy that prescribes the agent to pay full attention to a single source for an infinitely small interval of time and then stop and take action a if no jump occurs, and to

[20] Considering the neighborhood is without loss of generality. Indeed, the indifference curves where the alleged optimal strategy prescribes indifference between two or more courses of actions, have measure zero in the state space (q_1, q_2) and therefore could be ignored during the proof.

follow the optimal strategy in case of the jump:

$$\underbrace{U(q_1,q_2,a)}_{=V(q_1,q_2)} \geq -dt + \frac{(1+\xi+q_2)V(0,q_2)}{1+\xi+q_1+q_2+q_1q_2}dt + \left(1-\frac{1+\xi+q_2}{1+\xi+q_1+q_2+q_1q_2}dt\right)\underbrace{U(q_1+q_1dt,q_2,a)}_{=V(q_1+q_1dt,q_2)}.$$

The Taylor expansion $V(q_1+q_1dt,q_2) = V(q_1,q_2)+\frac{\partial V(q_1,q_2)}{\partial q_1}q_1dt$ gives $\mathscr{L}_1(q_1,q_2;V) \leq 0$. The symmetric argument proves $\mathscr{L}_2(q_1,q_2;V) \leq 0$.

Suppose the alleged optimal strategy prescribes to learn but has only phase 2, that is, $t^* = 0$ (again in some neighborhood of the initial state). Without loss of generality, assume it is optimal to use source 1, that is, $i = 1$. Then $\mathscr{L}_1(q_1,q_2;V) = 0$ (see a note after Lemma 3). Since t^* is the largest moment in time when switching from phase 1 to phase 2 is optimal, the phase 1 rule implies paying positive attention $x_2 > 0$ to source 2 at point (q_1,q_2). Deviating to the phase 1 rule is not locally optimal:

$$V(q_1,q_2) \geq -dt + \frac{(1+\xi+q_2)V(0,q_2)}{1+\xi+q_1+q_2+q_1q_2}x_1dt + \frac{(1+\xi+q_1)V(q_1,0)}{1+\xi+q_1+q_2+q_1q_2}x_2dt$$
$$+\left(1-\frac{(1+\xi+q_2)x_1}{1+\xi+q_1+q_2+q_1q_2}dt - \frac{(1+\xi+q_1)x_2}{1+\xi+q_1+q_2+q_1q_2}dt\right)V(q_1+q_1x_1dt,q_2+q_2x_2dt).$$

The Taylor expansion $V(q_1+q_1x_1dt,q_2+q_2x_2dt) = V(q_1,q_2)+\frac{\partial V(q_1,q_2)}{\partial q_1}q_1x_1dt+\frac{\partial V(q_1,q_2)}{\partial q_2}q_2x_2dt$ gives

$$V(q_1,q_2) \geq V(q_1,q_2) + x_1\underbrace{\mathscr{L}_1(q_1,q_2;V)}_{=0}dt + \underbrace{x_2}_{>0}\mathscr{L}_2(q_1,q_2;V)dt.$$

That implies $\mathscr{L}_2(q_1,q_2;V) \leq 0$.

Finally, suppose $t^* > 0$. What makes things tricky here is that there might be multiple regime changes during phase 1: $a_1(q_2(t))$ and $a_2(q_1(t))$ are not

necessarily constant. Moreover, switching between different attention rules is possible within the same regime. Two observations save me from considering all possible cases separately. First, by Bellman's principle of optimality, if (t^*, i) maximizes $W(q_1(0), q_2(0), \tilde{t^*}, \tilde{i})$ and no jump is observed by $t \leq t^*$, then $(t^* - t, i)$ maximizes $W(q_1(t), q_2(t), \tilde{t^*}, \tilde{i})$ (here, I use the notation $(q_1(t), q_2(t))$ for the *deterministic* belief path under the condition that no jump has been observed by the moment t). Second, at point $t = t^*$, the agent is indifferent between the phase 1 and the phase 2 rules, which means $\mathscr{L}_1(q_1(t^*), q_2(t^*); V) = \mathscr{L}_2(q_1(t^*), q_2(t^*); V) = 0$. Based on these two observations, I argue that it is sufficient to prove $\mathscr{L}_1(q_1(0), q_2(0); V) \leq 0$ and $\mathscr{L}_2(q_1(0), q_2(0); V) \leq 0$ under the assumption that $\mathscr{L}_1(q_1(t), q_2(t); V) \leq 0$ and $\mathscr{L}_2(q_1(t), q_2(t); V) \leq 0$ where $0 < t \leq t^*$ is the moment of either the regime change or the attention rule change.

Lemma 4 says that as long as it is optimal to use source 1 on the boundary $(q_1, 0)$, source 2 cannot be better than source 1.

Lemma 4. *If $q_1 < \bar{q}_1 \leq R(a_2(q_1))$ and source 1 is used until $\bar{q}_1$ for all $\tilde{q}_2 \in (q_2 - \Delta, q_2 + \Delta)$ for some $\Delta > 0$, then $\mathscr{L}_2(q_1, q_2; V) \leq 0$ as long as $\mathscr{L}_2(\bar{q}_1, q_2; V) \leq 0$.*

By symmetry, the same is true for source 2, so that Lemma 4 covers three cases:

1) $q_1 < R(a_2(q_1))$, $q_2 < R(a_1(q_2))$ where any attention rule could be used,

2) $q_1 < R(a_2(q_1))$, $q_2 \geq R(a_1(q_2))$ where source 1 is used,

3) $q_1 \geq R(a_2(q_1))$, $q_2 < R(a_1(q_2))$ where source 2 is used.

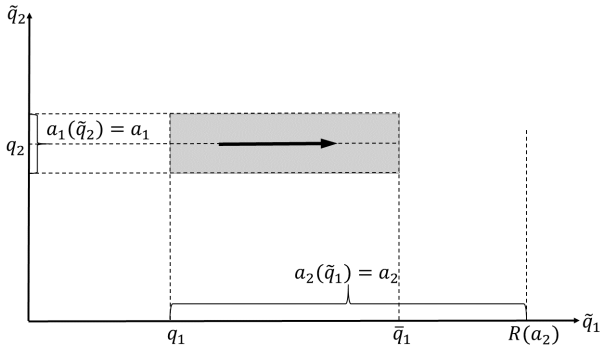

Figure 6: Illustration for Lemma 4.

The last case, when $q_1 \geq R(a_2(q_1))$ and $q_2 \geq R(a_1(q_2))$, is covered by Lemmas 5 and 6.

Lemma 5. *If $R(a_2(q_1)) \leq q_1 < \bar{q}_1$, $R(a_1(q_2)) \leq q_2$, $a_2(\tilde{q}_1) = a_2(q_1)$ for all $q_1 \leq \tilde{q}_1 \leq \bar{q}_1$, and source 1 is used until $\bar{q}_1$ for all $\tilde{q}_2 \in (q_2 - \Delta, q_2 + \Delta)$ for some $\Delta > 0$, then $\mathscr{L}_2(q_1, q_2; V) \leq 0$ as long as $\mathscr{L}_2(\bar{q}_1, q_2; V) \leq 0$ and $\bar{q}_1 + R(a_1(q_2)) \leq q_2 + R(a_2(q_1))$.*

Lemma 6. *If $R(a_2(q_1)) \leq q_1$, $q_1 + R(a_1(q_2)) < q_2 + R(a_2(q_1))$, and source 1 is used until $\breve{q}_1 = q_2 + R(a_2(q_1)) - R(a_1(q_2))$, then the attention is split according to $x_1 = \frac{\tilde{q}_2}{\tilde{q}_1 + \tilde{q}_2}$, $x_2 = \frac{\tilde{q}_1}{\tilde{q}_1 + \tilde{q}_2}$ to stay on the line $\tilde{q}_1 + R(a_1(q_2)) = \tilde{q}_2 + R(a_2(q_1))$ until $(\bar{q}_1, \bar{q}_2)$, $a_1(\tilde{q}_2) = a_1(q_2)$ for all $q_2 \leq \tilde{q}_2 \leq \bar{q}_2$, $a_2(\tilde{q}_1) = a_2(q_1)$ for all $q_1 \leq \tilde{q}_1 \leq \bar{q}_1$, then $\mathscr{L}_2(q_1, q_2; V) \leq 0$.*

Finally, suppose there is a strategy that is optimal but does not belong to the class described in Theorem 1. Naturally, the expected payoff from that strategy V is the same as from the two-phase optimal strategy. Let $t^* > 0$ be the small-

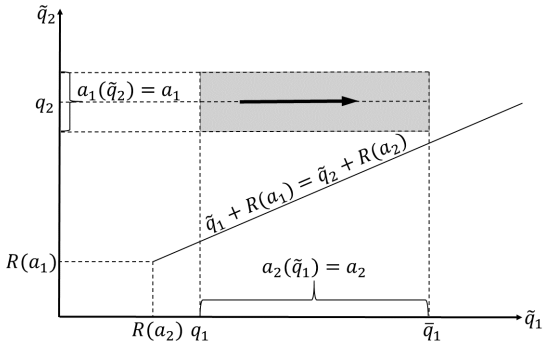

Figure 7: Illustration for Lemma 5.

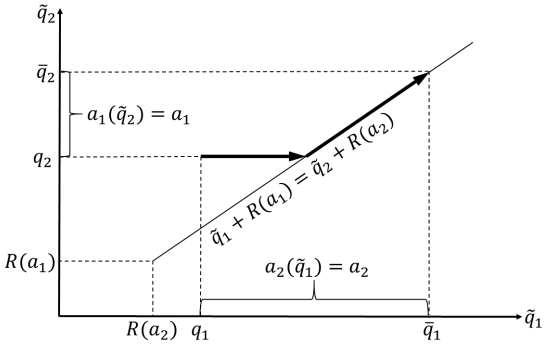

Figure 8: Illustration for Lemma 6.

est moment in time such that one of the sources are permanently abandoned after that moment (it is positive since that strategy does not belong to the described class). Following the proofs for Lemmas 4, 5, and 6, one can verify that $\mathscr{L}_2(q_1, q_2; V) < 0$ ($\mathscr{L}_1(q_1, q_2; V) < 0$) whenever the phase 1 rule unambiguously prescribes exclusive use of source 1 (2). That means using the other source would be locally suboptimal, given the expected payoff V. Contradiction.

5 Applications

5.1 Disagreement

Continuing with the example from the introduction, assume the city council has four options for the area: announce the area as a public beach (a_1), create a reservation (a_2), develop the area for housing (a_3), or build a national park (a_4). There are two groups in the council, one representing ecologists and the other one representing construction companies. If the environment is not unique ($\theta_1 = 1$) and the construction is safe ($\theta_2 = 0$), then the ecologists have no interest in the matter and the median voter strongly opposes giving up the area to expensive housing. Thus, both groups have aligned interests with the general public who wants the public beach:

$$\text{group 1}: \quad u_{10}^{(1)}(a_1) = \alpha, \quad u_{10}^{(1)}(a_2) = 0, \quad u_{10}^{(1)}(a_3) = 0, \quad u_{10}^{(1)}(a_4) = 0;$$

$$\text{group 2}: \quad u_{10}^{(2)}(a_1) = \alpha, \quad u_{10}^{(2)}(a_2) = 0, \quad u_{10}^{(2)}(a_3) = 0, \quad u_{10}^{(2)}(a_4) = 0.$$

In the reverse scenario, — when the construction is dangerous ($\theta_2 = 1$) and the environment is unique ($\theta_1 = 0$), — only the reservation is a good option:

$$\text{group 1}: \quad u_{01}^{(1)}(a_1) = 0, \quad u_{01}^{(1)}(a_2) = \alpha, \quad u_{01}^{(1)}(a_3) = 0, \quad u_{01}^{(1)}(a_4) = 0;$$

$$\text{group 2}: \quad u_{01}^{(2)}(a_1) = 0, \quad u_{01}^{(2)}(a_2) = \alpha, \quad u_{01}^{(2)}(a_3) = 0, \quad u_{01}^{(2)}(a_4) = 0.$$

If the environment is not unique ($\theta_1 = 1$) and the construction is dangerous ($\theta_2 = 1$), then all options are equally bad:

$$\text{group 1}: \quad u_{11}^{(1)}(a_1) = 0, \quad u_{11}^{(1)}(a_2) = 0, \quad u_{11}^{(1)}(a_3) = 0, \quad u_{11}^{(1)}(a_4) = 0;$$

$$\text{group 2}: \quad u_{11}^{(2)}(a_1) = 0, \quad u_{11}^{(2)}(a_2) = 0, \quad u_{11}^{(2)}(a_3) = 0, \quad u_{11}^{(2)}(a_4) = 0.$$

Finally, when the area is both critical for the ecology and safe for the construction ($\theta_1 = \theta_2 = 0$), the groups have misaligned interests. Both the ecologists and the construction companies lobby their causes and they are ready to reward their group in the council in case of a favorable outcome. That means the group that represents the ecologists prefers either the reservation or the national park (group 2), while the other group prefers to give the area for the public beach or housing (group 1):

$$\text{group 1}: \quad u_{00}^{(1)}(a_1) = \gamma, \quad u_{00}^{(1)}(a_2) = 0, \quad u_{00}^{(1)}(a_3) = \alpha, \quad u_{00}^{(1)}(a_4) = 0;$$

$$\text{group 2}: \quad u_{00}^{(2)}(a_1) = 0, \quad u_{00}^{(2)}(a_2) = \gamma, \quad u_{00}^{(1)}(a_3) = 0, \quad u_{00}^{(1)}(a_4) = \alpha.$$

I assume $\alpha > 0$, $\gamma \geq 0$.

When $\gamma = 0$, the following statement is trivially true: it is easier to agree on the type of information to invest in than on the optimal action to take. Mathematically, the optimization problems for the groups are identical, up to renaming actions a_3 and a_4. Thus, the solutions must be identical as well: both groups must agree on whether to learn or not and if yes, which source to use.

While being obvious for the case $\gamma = 0$, this statement becomes questionable (and sometimes false) for $\gamma > 0$. Intuitively, both groups' solutions should not be too different when γ is close to 0. Theorem 1 implies that even if $\gamma > 0$, the groups would agree on the source as long as they agree that both sources could potentially be used. Formally, that happens when both solutions include phase 1 ($t^* > 0$).

For the sake of example, assume $\alpha = 5$ and $\gamma = 1$. Suppose a priori all states are possible, that is, $p_{ij} > 0$, $i, j = 0, 1$. By Lemma 1, $\xi = \frac{p_{11} - (p_{10}+p_{11})(p_{01}+p_{11})}{p_{10}p_{01}}$ remains constant throughout learning, until a jump occurs. Suppose $\xi = 0$ for

the prior. Figure 9 shows the optimal action (a_1, a_2, a_3, or a_4) for each group and the disagreement area, the area where the optimal action is different for the groups.

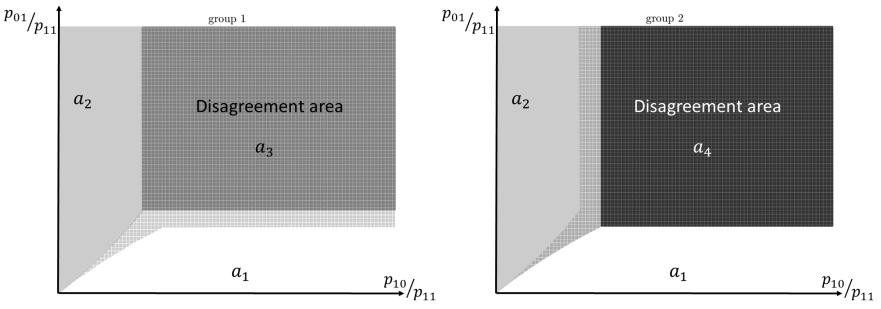

Figure 9: The best action given the beliefs. Group 1 is on the left, group 2 is on the right. Parameters: $\alpha = 5$, $\gamma = 1$, $\xi = 0$.

Now let us see what happens with the disagreement area when the information sources are available. Figure 10 shows the optimal strategy for each group. Comparing all four pictures in Figures 9 and 10 we see that in the upper right corner of $\frac{p_{01}}{p_{11}} \times \frac{p_{10}}{p_{11}}$ graph both groups choose the same information source despite disagreeing on the optimal action in the absence of the information sources (see Figure 11 on the left). Intuitively, when there is enough uncertainty about both θ_1 and θ_2, both information sources are important for each group since different actions are optimal in (1,0) (action a_1), (0,1) (action a_2), and (0,0) (action a_3 or a_4 depending on the group). Once the groups agree that both sources are important, they both start with phase 1 and therefore agree on the source to use.

Theorem 2 provides the sufficient conditions for $t^* > 0$ for an arbitrary pay-

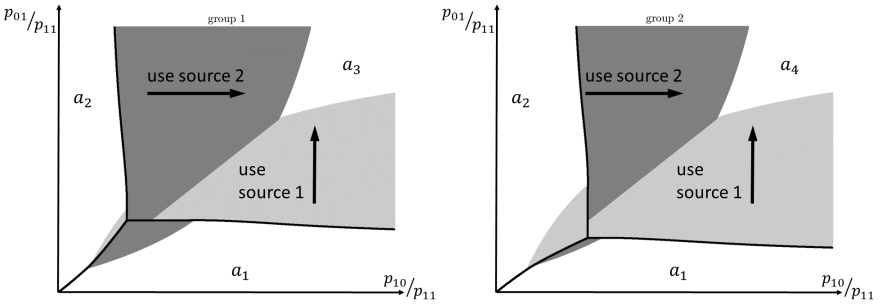

Figure 10: The optimal strategy. Group 1 is on the left, group 2 is on the right. Parameters: $\alpha = 5$, $\gamma = 1$, $\xi = 0$.

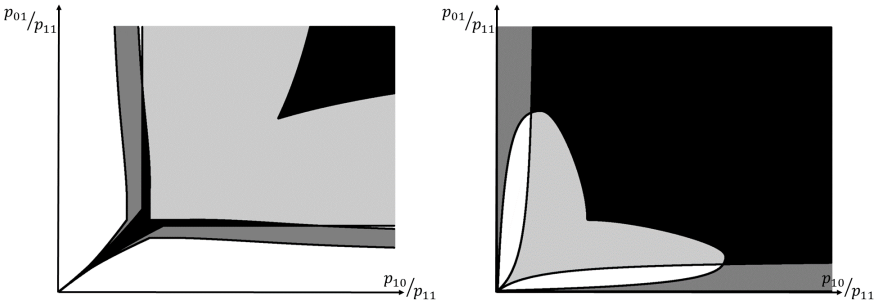

Figure 11: Black and light gray areas: both groups disagree on the optimal action in the absence of the information sources. Black and dark gray areas: both groups disagree on what to do when the information sources are available. Parameters: $\alpha = 5$, $\gamma = 1$, $\xi = 0$ (on the left); $\alpha = 50$, $\gamma = 100$, $\xi = 0$ (on the right).

off matrix and a set of actions.

Theorem 2. *Suppose $p_{01}p_{10} > 0$. Consider a sequence of payoff matrices $\{u^{(n)}_{\theta_1\theta_2}(a)\}_{a\in\mathscr{A}}$, $n = 1, 2, \ldots$. Denote by $a^{(n)}_1$ the best action in state $\theta_1 = 1$, $\theta_2 = 0$, and by $a^{(n)}_2$ the*

best action in state $\theta_1 = 0$, $\theta_2 = 1$. *Suppose* $a_1^{(n)} \neq a_2^{(n)}$. *As long as*

$$\lim_{n\to+\infty} \left\{ u_{10}^{(n)}\left(a_1^{(n)}\right) - u_{10}^{(n)}(a) \right\} = +\infty \quad \forall\, a \in \mathscr{A} \setminus \left\{ a_1^{(n)} \right\}, \tag{7}$$

$$\lim_{n\to+\infty} \left\{ u_{01}^{(n)}\left(a_2^{(n)}\right) - u_{01}^{(n)}(a) \right\} = +\infty \quad \forall\, a \in \mathscr{A} \setminus \left\{ a_2^{(n)} \right\}, \tag{8}$$

$$\limsup_{n\to+\infty} \left\{ \max_{a^*\in\mathscr{A}} u_{11}^{(n)}(a^*) - u_{11}^{(n)}\left(a_i^{(n)}\right) \right\} < +\infty, \quad i = 1,2, \tag{9}$$

starting from some $N \leq n$ *phase 1 has positive length* ($t^* > 0$).

Conditions (7) and (8) say that the knowledge of the optimal action in states $(1,0)$ and $(0,1)$ has an infinite value as $n \to +\infty$. In the example they become $\alpha \to +\infty$. Condition (9) guarantees that once a positive signal from one source is observed, the benefit from the other source is finite. In the example this condition holds automatically for all α and γ.

According to Theorem 2, the groups would agree on the source even if γ is much larger than α (and therefore actions a_3 and a_4 are irrelevant), as long as α is sufficiently high (see Figure 11 on the right).

5.2 Delegation

A policy maker often has to delegate information collection to an external expert (or to a group of experts). Since a seminal paper by Crawford and Sobel (1982), the literature on optimal delegation does not restrict the set of messages for the expert. This flexibility effectively means that the expert simply tells the policy maker what action to take. I restrict the set of messages for the expert to be the set of states (or more precisely, to the set of search results from each information source). This restriction allows the policy maker not to disclose the set of actions to the expert — a desirable feature of a contract for military or

security-related actions. Restricting the expert to report the state also prevents him from revealing the details of the search (in particular, how much time has passed before source i was abandoned with no evidence in favor of $\theta_i = 1$ found) — a realistic situation when the policy maker does not have time or sufficient expertise to read all details in the expert's report.

Creating incentives to conduct the right type of search is trivial when the state is uni-dimensional: with appropriate monetary or non-monetary incentives, the expert devotes exactly the desired amount of attention to the search in a given direction. However, the multidimentional state with different information sources corresponding to different state components, might create inefficiency in the direction of search. For example, a secret service agent must report to a command center whether a threat is found from group 1 or / and group 2. The command center sets the priorities for these groups, so the agent could adjust his effort in response to these priorities. Yet, this adjustment might not be optimal in a dynamic model when the priorities do not change over time and therefore do not change with the current beliefs. Applying Theorems 1 and 2, I show that just setting the priorities right might be enough to incentive the agent to allocate his effort in a desired way. Intuitively, the priorities play the role of indices for the sources, and as long as the index strategy is optimal, priorities form sufficient statistics for the optimal strategy.

Let $\theta_i = 1$ correspond to the presence of a threat from group i, $i = 1, 2$. Let $\{u_{\theta_1\theta_2}(a)\}_{a\in\mathscr{A}}$ be the payoff matrix of the command center with an arbitrary set of actions. The secret service agent reports $(z_1, z_2) \in \{0, 1\}^2$ to the center, where $z_i = 0$ means "there is no threat found from group i" and $z_i = 1$ means "group i is proved to pose a threat". So, the agent's set of actions is $\{\zeta_0 = (0, 0),\ \zeta_1 =$

$(1,0)$, $\zeta_2 = (0,1)$, $\zeta_3 = (1,1)\}$. Assume $p_{ij} > 0$, $i,j = 0,1$. The center sets the priorities $(\alpha_1, \alpha_2, \alpha_3)$ to the agent, so that the agent's payoff matrix is

	(0,0)	(1,0)	(0,1)	(1,1)
ζ_0	0	0	0	0
ζ_1	$-\gamma$	α_1	$-\gamma$	α_1
ζ_2	$-\gamma$	$-\gamma$	α_2	α_2
ζ_3	$-\gamma$	$-\gamma$	$-\gamma$	α_3

where $\alpha_i \geq 0$ is the reward from truthfully reporting a threat from group 1, 2 or both, and $\gamma > 0$ is the reputation loss from falsifying the report. I assume γ is higher enough compare to α_i so that the agent would never choose to falsify the report.

What can the center achieve with various priorities $(\alpha_1, \alpha_2, \alpha_3)$?

Consider the first best when the center collects information itself. By Theorem 1, her optimal strategy has two-phase form. If $t^* = 0$, then only one type of threat should be investigated (at most one source i is used). Setting α_i appropriately, the center achieves the first best.[21]

Suppose $t^* > 0$. The first best behavior up until some moment is to use an index policy with

$$p_{10} + p_{11}u_{11}(a_1)$$

as the source 1 index and

$$p_{01} + p_{11}u_{11}(a_2)$$

as the source 2 index. By setting

$$\alpha_1 = u_{11}(a_1) + \alpha, \quad \alpha_2 = u_{11}(a_2) + \alpha, \quad \alpha_3 = 0,$$

[21] Suppose the first best is such that $t^* = 0$, $i = 1$ and the optimal action at the stopping time conditional on no jumps observed is a. The center could set $\alpha_2 = 0$, $\alpha_1 = \alpha_3 = \max\left\{\frac{u_{11}(a^*) - u_{11}(a) + \frac{p_{10}}{p_{11}}\left(F_1\left(a_1, \frac{p_{10}}{p_{11}}\right) - u_{10}(a)\right)}{1 + \frac{p_{10}}{p_{11}}} - 1, 0\right\}$ (see Lemma 9 in Appendix A).

with α high enough, the center achieves this first best behavior: action ζ_i is optimal when source i produces a jump, which implies an index policy with

$$p_{10} + p_{11}\alpha_1$$

as the source 1 index and

$$p_{01} + p_{11}\alpha_2$$

as the source 2 index. By Theorem 2, there exists α large enough so that the agent's optimal strategy starts with the above index policy.

That was good news. Bad news is that in general the center cannot incentivize the optimal behavior all the time. Once the phase 1 ends or (a_1, a_2) change, the agent might deviate from the first best. This is not surprising since the center is restricted to set only 3 parameters in the payoff matrix of the agent. Yet, I want to emphasize that the optimal attention splitting between two sources could always be achieved during some positive interval of time by manipulating only $(\alpha_1, \alpha_2, \alpha_3)$.

A special case is when $\theta_1 = 1$ and $\theta_2 = 1$ are mutually exclusive, that is, $p_{11} = 0$. In that case the phase 1 rule is always the same for any payoff matrix: use source 1 when $p_{10} > p_{01}$ and use source 2 when $p_{10} < p_{01}$. In that case the first best behavior during phase 1 is always achieved with $\alpha_1 = \alpha_2 = \alpha$ large enough.

5.3 Multi-armed bandits

Separating the goal of learning gives a different perspective on some applications that got a lot of attention in the multi-armed bandit literature (see Berry

and Fristedt (1985) and Gittins et al. (2011) for review).[22] First introduced by Robbins (1952), the multi-armed bandit problems capture the trade-off between getting new knowledge from the environment (exploration) and using obtained knowledge (exploitation) when the means of exploration and exploitation are the same. In contrast, I separate the means of exploration (associated with a set of states of the world and information sources) and the means of exploitation (associated with a set of actions and the payoff matrix). That approach leads to a model with a different set of primitives to begin with, which allows to capture some aspects that cannot be easily incorporated into the multi-armed bandit setup.

For example, one of the practical problems motivating research on bandits is the design of clinical trials where the best way to treat a disease is studied by trials and errors.[23] In practice, the stage of clinical trials is preceded by the laboratory research that aims to select a set of treatments for the trials. The bandit problem takes treatments as given. Ideally, we should study both stages, the research stage and the clinical trails stage, within the same model. Indeed, the more you learn about the disease and the better you select the treatments during the research stage, the more successful clinical trials will be. This paper takes one step in this direction by focusing on the research stage exclusively. I show that as long as the stakes are high (in the sense that is formalized in

[22] The connection with the multi-armed bandit literature goes beyond mere similarity of applications. Mathematically, my model can be formulated as a special case of a general multi-armed bandit problem where information sources play the role of arms. However, this case falls outside of the set of the multi-armed bandit problems that could be solved using so called the Gittins index technique primary because the arms are not independent (see the discussion in *Preface to the English Edition* in Presman and Sonin (1990) book that deals with dependent arms).

[23] Gittins and Jones (1979) state that the multi-armed bandit problem's "chief practical motivation comes from clinical trials."

Theorem 2), — which is usually the case for medical research, — the right choice of the research direction for the pretrial stage depends only on the optimal set of treatments and the payoffs from them in case of a *successful* finding.[24]

To economics, bandit problems came with the market pricing application (Rothschild (1974)), experimental consumption, and the research and development (R&D) problem (Roberts and Weitzman (1981), Choi (1991), Keller et al. (2005)).[25] Most papers with the R&D application use the Poisson arms because it allows one to stay within the bandit framework and at the same time capture the pure learning nature of R&D (see, for example, Klein and Rady (2011) and Klein (2013)). Each arm (research direction) can lead to a discovery at a random time but the intensity with which discoveries occur is unknown (it is called the state of an arm). Following the mainstream in the bandit literature, they have to assume that the payoff from one arm in a fixed state does not depend on the states of the other arms (at the same time, the arms' states can be correlated). Working outside of this literature, I do not have to make that assumption and show that it dramatically affects the solution. For example, suppose a firm has to choose between two products to invest in. Before making the choice, it can research each product, which can be either in high demand ($\theta_i = 1$) or in low demand ($\theta_i = 0$). In the absence of any competing firm, the payoff from product i is 0 if it is in low demand and $v_i > 0$ otherwise:

[24] With application to clinical trials, Henry and Ottaviani (2019) model the research stage (they call it the clinical trials stage but from the bandit problem perspective it is the research stage because the state-dependent payoffs come only at the end) with the optimal stopping problem. Damiano et al. (2019) incorporate both research and clinical trials in a model with one risky arm and one safe arm (these two arms correspond to two treatments in the clinical trials), one "positive" information source (that can deliver a breakthrough only if the risky arm is good), and one "negative" information source (that can deliver a breakthrough only if the risky arm is bad).

[25] See Bergemann and Valimaki (2008) for review.

	(0,0)	(1,0)	(0,1)	(1,1)
invest in product 1	0	v_1	0	v_1
invest in product 2	0	0	v_2	v_2

This situation **can** be studied within two-armed exponential bandit setting. In contrast, the following situation **cannot** be studied within the traditional bandit framework. Suppose now the market is competitive so that if a product is in high demand ($\theta_i = 1$) there is always a firm that invests in it. Assume that at most one firm can invest in a product (it becomes protected by patent rights). Then the payoff matrix could look like this:

	(0,0)	(1,0)	(0,1)	(1,1)
invest in product 1	0	v_1	0	$v_1 - d_1$
invest in product 2	0	0	v_2	$v_2 - d_2$

where $d_i > 0$ accounts for competition between the products (for example, if a firm invests in product 1 which is in high demand then it gets $v_1 - d_1$ if there is another competing product that is in high demand). Now the payoff from arm i depends not only on this arm state (θ_i) but also on the other arm state (θ_{3-i}). This difference implies different index policies for phase 1. For the monopoly market source 1 index is $p_{10} + p_{11}v_1$ and source 2 index is $p_{01} + p_{11}v_2$. For the competitive market source 1 index is $p_{10} + p_{11}(v_1 - d_1)$ and source 2 index is $p_{01} + p_{11}(v_2 - d_2)$.

6 Conclusion

I presented the general form of an optimal strategy for a signal-agent decision problem with an arbitrary payoff matrix and two information sources. Each source corresponds to a binary state meaning that it can provide conclusive evidence only if the state is 1. The main feature of an optimal strategy is that as

long as both sources could potentially be used (that is, phase 1 has a positive length), the optimal allocation of attention at a given moment does not depend on the payoff in state (0,0). Moreover, once the optimal actions in case of both types of discoveries are fixed (a_1 and a_2), the optimal allocation of attention at a given moment does not depend on the payoff in states (1,0) and (0,1) as well. Such independence means that agents with quite different goals (preferences, interests) might easily agree on the type of information to invest in or on the direction of a discussion (providing these agents have common prior).

Enriching and changing the information environment of the decision problem leads to many promising directions for future research. One natural extension is to include more information sources (results from Austen-Smith and Martinelli (2018) allows speculation about what would happen in that case). Other types of learning include gradual learning — when a source is modeled as Brownian motion with state-dependent drift (Ke and Villas-Boas (2019) studied this case with the payoff matrix restrictions similar to Nikandrova and Pancs (2018)) — mixed-type learning — when a source is modeled as the Poisson process generating non-conclusive evidence (see Sections 5 and 6 in Che and Mierendorff (2017a)) — and breakthrough learning with state-independent intensity — when a source is modeled as the Poisson process generating conclusive evidence for both 0 and 1 (see Section E in Che and Mierendorff (2017b)).

In the introduction I mentioned that formulating the options and figuring out the payoff matrix might not be an easy task by itself. Introducing additional information sources that allow one to search for additional options and / or to learn the payoffs from the existing ones (the latter type of information sources is studied in Fudenberg et al. (2018) within the optimal stopping problem frame-

work and in Ke et al. (2016) within the bandit-like framework) might be a better way to capture the difference between learning about fundamentals (the state) and decision options (the payoff matrix).

The nature of strategic interactions between players in multi-agent extension of my model is another question I am leaving for future research. Actively studied within the bandit framework, this question has largely remained unexplored for environments with many information sources.

References

Austen-Smith, David and César Martinelli, "Optimal exploration," 2018. GMU Working Paper in Economics No 18-25. 9, 45

Banks, Jeffrey S and Rangarajan K Sundaram, "A class of bandit problems yielding myopic optimal strategies," *Journal of Applied Probability*, 1992, *29* (3), 625–632. 19

Bergemann, Dirk and Juuso Valimaki, "Bandit problems," in Steven N Durlauf and Lawrence Blume, eds., *The New Palgrave Dictionary of Economics*, 2nd ed., Basingstoke and New York: Palgrave Macmillan Ltd., 2008. 43

Berry, Donald A and Bert Fristedt, *Bandit problems: Sequential allocation of experiments*, Vol. 5, Springer, 1985. 41

Chaloner, Kathryn and Isabella Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, August 1995, *10* (3), 273–304. 7

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer, "Loss attitudes in the US population: Evidence from dynamically optimized

sequential experimentation (DOSE)," 2018. NBER Working Paper No 25072. 7

Chatterjee, Kalyan and Robert Evans, "Rivals' search for buried treasure: competition and duplication in R&D," *RAND Journal of Economics*, 2004, pp. 160–183. 9

Che, Yeon-Koo and Konrad Mierendorff, "Optimal sequential decision with limited attention," July 2017. Working Paper. 45

_ and _ , "Supplemental material for 'Optimal sequential decision with limited attention'," July 2017. Working Paper. 45

_ and _ , "Optimal sequential decision with limited attention," *American Economic Review*, 2019. Forthcoming (December 2018). 9, 10, 11, 15, 17, 20, 21, 22

Chernoff, Herman, "Sequential design of experiments," *Annals of Mathematical Statistics*, 1959, *30* (3), 755–770. 7

Choi, Jay P, "Dynamic R&D competition under 'hazard rate' uncertainty," *RAND Journal of Economics*, 1991, pp. 596–610. 43

Crawford, Vincent P and Joel Sobel, "Strategic information transmission," *Econometrica*, 1982, pp. 1431–1451. 38

Damiano, Ettore, Hao Li, and Wing Suen, "Learning while experimenting," *Economic Journal*, 2019. Forthcoming (December 2018). 9, 43

Dumav, Martin and Maxwell Stinchcombe, "The von Neumann/Morgenstern approach to ambiguity," 2013. Institute of Mathematical Economics Working Paper No 480. 22

Fleming, Wendell H and Raymond W Rishel, *Deterministic and stochastic optimal control*, Vol. 1, Springer Science & Business Media, 2012. 55

Forand, Jean Guillaume, "Keeping your options open," *Journal of Economic Dynamics and Control*, 2015, *53*, 47–68. 20

Francetich, Alejandro et al., "Efficient multi-agent experimentation and multi-choice bandits," *Economics Bulletin*, 2018, *38* (4), 1757–1761. 9

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Speed, accuracy, and the optimal timing of choices," *American Economic Review*, 2018, *108* (12), 3651–84. 8, 45

Gittins, John C and David M Jones, "A dynamic allocation index for the discounted multiarmed bandit problem," *Biometrika*, 1979, *66* (3), 561–565. 42

Gittins, John, Kevin Glazebrook, and Richard Weber, *Multi-armed bandit allocation indices*, John Wiley & Sons, 2011. 18, 42

Henry, Emeric and Marco Ottaviani, "Research and the approval process: The organization of persuasion," *American Economic Review*, 2019, *109* (3), 911–55. 43

Karni, Edi and Marie-Louise Vierø, "Awareness of unawareness: a theory of decision making in the face of ignorance," *Journal of Economic Theory*, 2017. 22

Ke, T Tony and J Miguel Villas-Boas, "Optimal learning before choice," *Journal of Economic Theory*, 2019, *180*, 383–437. 8, 45

_ , Zuo-Jun Max Shen, and J Miguel Villas-Boas, "Search for information on multiple products," *Management Science*, 2016, *62* (12), 3576–3603. 8, 46

Keller, Godfrey, Sven Rady, and Martin Cripps, "Strategic experimentation with exponential bandits," *Econometrica*, 2005, *73* (1), 39–68. 43

Klein, Nicolas, "Strategic learning in teams," *Games and Economic Behavior*, 2013, *82*, 636–657. 21, 43

_ and Sven Rady, "Negatively correlated bandits," *Review of Economic Studies*, 2011, *78* (2), 693–732. 9, 21, 22, 43

Liang, Annie and Xiaosheng Mu, "Complementary information and learning traps," 2017. PIER Working Paper No 18-008. 8

_ , _ , and Vasilis Syrgkanis, "Optimal and myopic information acquisition," 2017. Working Paper. 8

_ , _ , and _ , "Dynamically aggregating diverse information," 2019. PIER Working Paper No 19-005. 8

Moscarini, Giuseppe and Lones Smith, "The optimal level of experimentation," *Econometrica*, 2001, *69* (6), 1629–1644. 10

Naghshvar, Mohammad and Tara Javidi, "Active sequential hypothesis testing," *Annals of Statistics*, 2013, *41* (6), 2703–2738. 7

Nikandrova, Arina and Romans Pancs, "Dynamic project selection," *Theoretical Economics*, 2018, *13* (1), 115–143. 9, 11, 15, 17, 18, 45

Peskir, Goran and Albert Shiryaev, *Optimal stopping and free-boundary problems*, Springer, 2006. 7

Presman, Ernst L. and Isaak M. Sonin, *Sequential control with incomplete information: The Bayesian approach to multi-armed bandit problems*, Academic Press, 1990. 9, 13, 42

Robbins, Herbert, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, 1952, *58* (5), 527–535. 42

Roberts, Kevin and Martin L Weitzman, "Funding criteria for research, development, and exploration projects," *Econometrica*, 1981, pp. 1261–1288. 43

Rothschild, Michael, "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 1974, *9* (2), 185–202. 43

Sims, Christopher A, "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690. 7

Steiner, Jakub, Colin Stewart, and Filip Matějka, "Rational inattention dynamics: Inertia and delay in decision-making," *Econometrica*, 2017, *85* (2), 521–553. 7

Wald, Abraham, "Foundations of a general theory of sequential decision functions," *Econometrica*, 1947, *15* (4), 279–313. 7

Zhong, Weijie, "Optimal dynamic information acquisition," 2017. Working Paper. 7

A Cookbook

Here is the **cookbook** to characterize the class of optimal strategies in a given decision problem. Let $p = (p_{00}, p_{01}, p_{10}, p_{11})$ be the belief vector.

1. Let $a^* \in \mathscr{A}$ be one of the best actions in state $(1,1)$. Define $a_1\left(\frac{p_{10}}{p_{11}}\right)$ and $a_2\left(\frac{p_{01}}{p_{11}}\right)$ as follows. When the state is fully known, take the best action: $a_1(0) = a_2(0) = a^*$, $a_1(+\infty) \in \arg\max_{a\in\mathscr{A}} u_{10}(a)$, $a_2(+\infty) \in \arg\max_{a\in\mathscr{A}} u_{01}(a)$. For $0 < x < +\infty$, define

$$a_1(x) \in \arg\max_{a\in\mathscr{A}} F_1(a,x) \equiv u_{10}(a) - f\left(\frac{u_{11}(a^*) - u_{11}(a) - 1}{x}\right),$$

$$a_2(x) \in \arg\max_{a\in\mathscr{A}} F_2(a,x) \equiv u_{01}(a) - f\left(\frac{u_{11}(a^*) - u_{11}(a) - 1}{x}\right),$$

where $f(y) = \begin{cases} \log(y), & y > 1 \text{ (learning)} \\ y - 1, & y \le 1 \text{ (stop)} \end{cases}$.

2. Draw the phase 1 rule on $\frac{p_{10}}{p_{11}} \times \frac{p_{01}}{p_{11}}$ plane (see Figure 2). On the boundaries, mark $a_1\left(\frac{p_{10}}{p_{11}}\right)$ and $a_2\left(\frac{p_{01}}{p_{11}}\right)$. For each square with $a_1\left(\frac{p_{10}}{p_{11}}\right) = \text{const}$ and $a_2\left(\frac{p_{01}}{p_{11}}\right) = \text{const}$, separate four regions (skipping the regions that fall outside the square):

   (a) $u_{11}(a^*) - u_{11}(a_1) - 1 > \frac{p_{10}}{p_{11}}$, $u_{11}(a^*) - u_{11}(a_2) - 1 > \frac{p_{01}}{p_{11}}$
   $\Rightarrow$ the agent is indifferent between both sources

   (b) $u_{11}(a^*) - u_{11}(a_1) - 1 < \frac{p_{10}}{p_{11}}$, $u_{11}(a^*) - u_{11}(a_2) - 1 > \frac{p_{01}}{p_{11}}$
   $\Rightarrow$ source 1 should be used

   (c) $u_{11}(a^*) - u_{11}(a_1) - 1 > \frac{p_{10}}{p_{11}}$, $u_{11}(a^*) - u_{11}(a_2) - 1 < \frac{p_{01}}{p_{11}}$
   $\Rightarrow$ source 2 should be used

   (d) $u_{11}(a^*) - u_{11}(a_1) - 1 < \frac{p_{10}}{p_{11}}$, $u_{11}(a^*) - u_{11}(a_2) - 1 < \frac{p_{01}}{p_{11}}$
   $\Rightarrow$ source 1 should be used for $\frac{p_{10}}{p_{11}} + u_{11}(a_1) > \frac{p_{01}}{p_{11}} + u_{11}(a_2)$ and source 2 for $\frac{p_{10}}{p_{11}} + u_{11}(a_1) < \frac{p_{01}}{p_{11}} + u_{11}(a_2)$

   Complete phase 1 description by drawing the phase 1 rule on

   - $p_{01} \times p_{10}$ plane for the case $p_{11} = 0$: source 1 should be used for $p_{10} > p_{01}$ and source 2 source be used for $p_{10} < p_{01}$ (see Figure 3);
   - $\frac{p_{00}}{p_{10}} \times \frac{p_{10}}{p_{11}}$ plane for case $p_{01} = 0$: on the boundary $\frac{p_{00}}{p_{10}} = 0$ mark $a_1\left(\frac{p_{10}}{p_{11}}\right)$, and for each region $a_1\left(\frac{p_{10}}{p_{11}}\right) = \text{const}$ source 1 should be used for $\frac{p_{10}}{p_{11}} > u_{11}(a^*) - u_{11}(a_1)$ and source 2 should be used otherwise (see Figure 12);
   - $\frac{p_{01}}{p_{11}} \times \frac{p_{00}}{p_{01}}$ plane for case $p_{10} = 0$: on the boundary $\frac{p_{00}}{p_{01}} = 0$ mark $a_2\left(\frac{p_{01}}{p_{11}}\right)$, and for each region $a_2\left(\frac{p_{01}}{p_{11}}\right) = \text{const}$ source 1 should be used for $\frac{p_{01}}{p_{11}} < u_{11}(a^*) - u_{11}(a_2)$ and source 2 should be used otherwise (see Figure 12).

3. Calculate the expected payoff for regime $(0, a_1, a_2)$ using Lemmas 7 and 8:

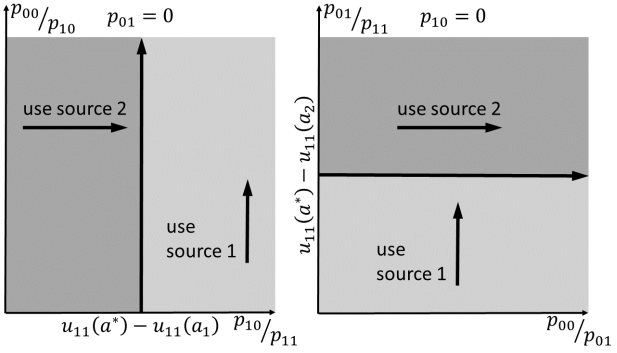

Figure 12: Regime $(0, a_1, a_2)$ for $p_{01} = 0$ and for $p_{10} = 0$.

Lemma 7. *The expected payoff from using source 1 until $\bar{p}$ is*

$$V(p) = \frac{p_{00}+p_{01}}{\bar{p}_{00}+\bar{p}_{01}} V(\bar{p}) - (p_{00}+p_{01}) \log\left(\frac{(\bar{p}_{00}+\bar{p}_{01})(p_{10}+p_{11})}{(p_{00}+p_{01})(\bar{p}_{10}+\bar{p}_{11})}\right) + \left(1 - \frac{(p_{00}+p_{01})(\bar{p}_{10}+\bar{p}_{11})}{(\bar{p}_{00}+\bar{p}_{01})(p_{10}+p_{11})}\right)\left(p_{11}(u_{11}(a^*)-2) + p_{10}\left(F_1\left(a_1, \frac{p_{10}}{p_{11}}\right)-2\right)\right).$$

The expected payoff from using source 2 until $\bar{p}$ is

$$V(p) = \frac{p_{00}+p_{10}}{\bar{p}_{00}+\bar{p}_{10}} V(\bar{p}) - (p_{00}+p_{10}) \log\left(\frac{(\bar{p}_{00}+\bar{p}_{10})(p_{01}+p_{11})}{(p_{00}+p_{10})(\bar{p}_{01}+\bar{p}_{11})}\right) + \left(1 - \frac{(p_{00}+p_{10})(\bar{p}_{01}+\bar{p}_{11})}{(\bar{p}_{00}+\bar{p}_{10})(p_{01}+p_{11})}\right)\left(p_{11}(u_{11}(a^*)-2) + p_{01}\left(F_2\left(a_2, \frac{p_{01}}{p_{11}}\right)-2\right)\right).$$

Lemma 8. *Suppose $p_{01} > 0$, $p_{10} > 0$. The expected payoff from splitting attention $x_1 = \frac{p_{10}}{p_{10}+p_{01}}$, $x_2 = \frac{p_{01}}{p_{10}+p_{01}}$ along the line $p_{10} + p_{11}u_{11}(a_1) = p_{01} + p_{11}u_{11}(a_2)$ until $\bar{p}$ is*

$$\begin{aligned} V(p) = {} & \frac{p_{01}p_{10}\bar{p}_{11}}{\bar{p}_{01}\bar{p}_{10}p_{11}} V(\bar{p}) - p_{00} \log\left(\frac{\bar{p}_{01}\bar{p}_{10}p_{11}^2}{p_{01}p_{10}\bar{p}_{11}^2}\right) - \frac{p_{10}p_{01}}{p_{10}-p_{01}} \log\left(\frac{p_{10}\bar{p}_{01}}{\bar{p}_{10}p_{01}}\right) \\ & - \left(\frac{p_{10}+p_{01}}{p_{11}} - \frac{\bar{p}_{10}+\bar{p}_{01}}{\bar{p}_{11}}\right)\frac{\bar{p}_{11}}{2}\left\{\frac{p_{10}+p_{11}}{\bar{p}_{01}}\left(\frac{p_{10}u_{10}(a_1)}{p_{10}+p_{11}} + \frac{p_{11}u_{11}(a_1)}{p_{10}+p_{11}} - \frac{1}{2}\right)\right. \\ & \left. + \frac{p_{01}+p_{11}}{\bar{p}_{10}}\left(\frac{p_{01}u_{01}(a_2)}{p_{01}+p_{11}} + \frac{p_{11}u_{11}(a_2)}{p_{01}+p_{11}} - \frac{1}{2}\right) - \frac{(1-\bar{p}_{00})(p_{10}+p_{01})}{2\bar{p}_{10}\bar{p}_{01}}\right\} \\ & - \left(\frac{p_{10}+p_{01}}{p_{11}} - \frac{\bar{p}_{10}+\bar{p}_{01}}{\bar{p}_{11}}\right)^2 \frac{\bar{p}_{11}^2 p_{11}}{8\bar{p}_{10}\bar{p}_{01}}(u_{11}(a_1)+u_{11}(a_2)) \end{aligned} \quad (10)$$

if $p_{11}u_{11}(a_1) \neq p_{11}u_{11}(a_2)$. *If* $p_{11}u_{11}(a_1) = p_{11}u_{11}(a_2)$ *and* $p_{00} + p_{11} > 0$, *then*

$$V(p) = \frac{1-p_{00}-p_{11}}{1-\bar{p}_{00}-\bar{p}_{11}} \upsilon V(\bar{p}) + 2p_{00}\log(\upsilon) + \frac{1-p_{00}-p_{11}}{2}(1-\upsilon)(u_{10}(a_1)+u_{01}(a_2)-4) + p_{11}\left(1-\upsilon^2\right)\left(\frac{u_{11}(a_1)+u_{11}(a_2)}{2}-1\right), \tag{11}$$

where[26]

$$\upsilon = \begin{cases} \frac{(1-p_{00}-p_{11})\bar{p}_{11}}{(1-\bar{p}_{00}-\bar{p}_{11})p_{11}}, & p_{11} > 0, \\ \frac{(1-\bar{p}_{00}-\bar{p}_{11})p_{00}}{(1-p_{00}-p_{11})\bar{p}_{00}}, & p_{00} > 0. \end{cases}$$

If $p_{00} = p_{11} = 0$, *then*

$$V(p) = \frac{u_{10}(a_1)+u_{01}(a_2)}{2} - 1.$$

4. For each initial belief p, four variables complete the description of the optimal strategy: the time $t^* \geq 0$ of switching to phase 2, the source i used during phase 2, the action a taken at the stopping time conditional on no jumps observed, and the stopping threshold $\underline{p}$ at which phase 2 ends. The last step is to find the optimal values for these variables (for each belief p!) with the help of Bellman's principle of optimality (see discussion on page 31) and Lemmas 9 and 10.

   Lemma 9. *If the agent could use only source* i *and had to take action* a *at the stopping time conditional on no jumps observed, then he would use source* i *until his belief about* $\theta_i = 1$ *becomes as low as* $\underline{\pi}$ *and his expected payoff is*

$$V(p) = p_{00}u_{00}(a) + p_{01}u_{01}(a) + p_{10}u_{10}(a) + p_{11}u_{11}(a) + \frac{\pi-\underline{\pi}}{\underline{\pi}} - (1-\pi)\log\left(\frac{\pi(1-\underline{\pi})}{\underline{\pi}(1-\pi)}\right),$$

where $\pi = \begin{cases} p_{11}+p_{10}, & i=1 \\ p_{11}+p_{01}, & i=2 \end{cases}$ *and the optimal threshold is* $\underline{\pi} = \begin{cases} \frac{1}{r^*}, & r^* > \frac{1}{\pi}, \\ \pi, & \text{otherwise} \end{cases}$, *where*

$$r^* = \begin{cases} \frac{p_{11}(u_{11}(a^*)-u_{11}(a))}{p_{10}+p_{11}} + \frac{p_{10}}{p_{10}+p_{11}}\left(F_1\left(a_1, \frac{p_{10}}{p_{11}}\right) - u_{10}(a)\right) - 1, & i=1 \\ \frac{p_{11}(u_{11}(a^*)-u_{11}(a))}{p_{01}+p_{11}} + \frac{p_{01}}{p_{01}+p_{11}}\left(F_2\left(a_2, \frac{p_{01}}{p_{11}}\right) - u_{01}(a)\right) - 1, & i=2 \end{cases}$$

   Lemma 10. *If it is optimal for the agent to switch to phase 2 at time* $t^* > 0$, *then*

$$1 = p_{11}\left(\frac{(1-\pi)\underline{\pi}}{\pi(1-\underline{\pi})}(u_{11}(a^*)-u_{11}(a)-1) + \frac{\pi-\underline{\pi}}{\pi(1-\underline{\pi})}\min\left[u_{11}(a^*)-u_{11}(a_i), \frac{\pi}{p_{11}}\right]\right) + \begin{cases} p_{01}\left(F_2\left(a_2, \frac{p_{01}}{p_{11}}\right) - u_{01}(a) - 1 + \log\left(\frac{\pi(1-\underline{\pi})}{\underline{\pi}(1-\pi)}\right)\right) & i=1, \\ p_{10}\left(F_1\left(a_1, \frac{p_{10}}{p_{11}}\right) - u_{10}(a) - 1 + \log\left(\frac{\pi(1-\underline{\pi})}{\underline{\pi}(1-\pi)}\right)\right), & i=2, \end{cases} \tag{12}$$

   where I write p *for* $p(t^*)$, $i \in \{1,2\}$ *for the source the agent uses during phase 2,* a *for the action he takes at the stopping time conditional on no jumps observed. Notations* π, $\underline{\pi}$ *and* r^* *have the same meaning as in Lemma 9.*

[26] If $p_{00}p_{11} > 0$, then $\frac{(1-p_{00}-p_{11})\bar{p}_{11}}{(1-\bar{p}_{00}-\bar{p}_{11})p_{11}} = \frac{(1-\bar{p}_{00}-\bar{p}_{11})p_{00}}{(1-p_{00}-p_{11})\bar{p}_{00}}$.

B Proofs

B.1 Lemma 1

Bayes' rule gives $p_{\theta_1\theta_2}(t) = \frac{p_{\theta_1\theta_2}(0)e^{-\int_0^t(\theta_1 x_1(t)+\theta_2 x_2(t))dt}}{p_{11}(0)e^{-t}+p_{10}(0)e^{-\int_0^t x_1(t)dt}+p_{01}(0)e^{-\int_0^t x_2(t)dt}+p_{00}(0)}$, which implies $q_1(t) = \frac{p_{00}(0)}{p_{10}(0)e^{-\int_0^t x_1(t)dt}}$ and then $\frac{dq_1(t)}{dt} = q_1(t)x_1(t)$. Differential equations for q_2 and ξ are derived in the same way.

B.2 Lemma 2

Let $\Upsilon(q_1(0), \bar{q}_1, \theta_1)$ be the expected payoff given the initial belief state $q_1(0)$, stopping threshold $\bar{q}_1 \geq q_1(0)$ and state of the world θ_1.

If $q_1(0) = \bar{q}_1$, then $\Upsilon(q_1(0), \bar{q}_1, \theta_1) = u_{\theta_1 1}(a)$.

Fix any $\Delta > 0$ and suppose $q_1(0) \leq \bar{q}_1 e^{-\Delta}$. If no jump occurs during $t \in [0, \Delta]$, then $q_1(\Delta) = q_1(0)e^{\Delta}$. If $\theta_1 = 0$, then no jump is possible:

$$\Upsilon(q_1(0), \bar{q}_1, 0) = \Upsilon\left(q_1(0)e^{\Delta}, \bar{q}_1, 0\right) - \Delta. \tag{13}$$

If $\theta_1 = 1$, then the probability that no jump occurs during $[0, t]$ is equal to e^{-t}:

$$\Upsilon(q_1(0), \bar{q}_1, 1) = (1 - e^{-\Delta})u_{11}(a^*) + e^{-\Delta}\Upsilon\left(q_1(0)e^{\Delta}, \bar{q}_1, 1\right) - \int_0^{\Delta} te^{-t}dt - e^{-\Delta}\Delta. \tag{14}$$

Expressions (13) and (14) allow me to express the derivative of Υ with respect to $q_1(0) < \bar{q}_1$:

$$\begin{aligned}\frac{\partial \Upsilon(q_1(0), \bar{q}_1, \theta_1)}{\partial q_1(0)} &= \lim_{\Delta \to 0} \frac{\Upsilon\left(q_1(0)e^{\Delta}, \bar{q}_1, \theta_1\right) - \Upsilon(q_1(0), \bar{q}_1, \theta_1)}{q_1(0)e^{\Delta} - q_1(0)} \\ &= \lim_{\Delta \to 0} \frac{\Upsilon\left(q_1(0)e^{\Delta}, \bar{q}_1, \theta_1\right) - \Upsilon(q_1(0), \bar{q}_1, \theta_1)}{q_1(0)\Delta} = \begin{cases} \frac{1}{q_1(0)}, & \theta_1 = 0 \\ \frac{\Upsilon(q_1(0), \bar{q}_1, 1) - u_{11}(a^*) + 1}{q_1(0)}, & \theta_1 = 1 \end{cases}\end{aligned}$$

Solving this differential equation with the boundary condition $\Upsilon(\bar{q}_1, \bar{q}_1, \theta_1) = u_{\theta_1 1}(a)$, I get

$$\Upsilon(q_1(0), \bar{q}_1, 0) = u_{01}(a) - \log\left(\frac{\bar{q}_1}{q_1(0)}\right), \quad \Upsilon(q_1(0), \bar{q}_1, 1) = u_{11}(a) + \frac{\bar{q}_1 - q_1(0)}{\bar{q}_1}(u_{11}(a^*) - u_{11}(a) - 1).$$

Given $(q_1(0), 0)$, the belief vector recovered from (3) is $p_{00} = 0$, $p_{01} = \frac{q_1(0)}{1+\xi+q_1(0)}$, $p_{10} = 0$, $p_{11} = \frac{1+\xi}{1+\xi+q_1(0)}$. Thus, the expected payoff is

$$\begin{aligned}&\frac{q_1(0)}{1+\xi+q_1(0)}\Upsilon(q_1(0), \bar{q}_1, 0) + \frac{1+\xi}{1+\xi+q_1(0)}\Upsilon(q_1(0), \bar{q}_1, 1) \\ &\qquad = U(q_1(0), 0, a) + \frac{(\bar{q}_1 - q_1(0))R(a)}{\bar{q}_1(1+\xi+q_1(0))} - \frac{q_1(0)}{1+\xi+q_1(0)}\log\left(\frac{\bar{q}_1}{q_1(0)}\right).\end{aligned}$$

B.3 Lemma 3

Take any initial state $q(0) \in [0,+\infty)^2$ and any strategy (x,τ,α). The state variable $q(t)$ complies

$$dq_i(t) = q_i(t)x_i(t)dt - q_i(t)dN_i(t).$$

Suppose function V is continuously differentiable along the whole trajectory $q(t)$, $0 \le t \le \tau$, except maybe a countable set of points. Then Itô's formula gives

$$V(q(\tau)) = V(q(0)) + \int_0^\tau \left(\frac{\partial V(q_1(t),q_2(t))}{\partial q_1} q_1(t)x_1(t) + \frac{\partial V(q_1(t),q_2(t))}{\partial q_2} q_2(t)x_2(t) \right) dt$$
$$+ \int_0^\tau (V(0,q_2(t)) - V(q_1(t),q_2(t)))\, dN_1(t) + \int_0^\tau (V(q_1(t),0) - V(q_1(t),q_2(t)))\, dN_2(t) \quad (15)$$

Taking conditional expectations, I get

$$\mathbf{E}_{q(0)}[V(q(\tau))] = V(q(0)) + \mathbf{E}_{q(0)}\left[\int_0^\tau (\mathscr{L}_1(q_1(t),q_2(t);V)+1)\, x_1(t)dt \right]$$
$$+ \mathbf{E}_{q(0)}\left[\int_0^\tau (\mathscr{L}_2(q_1(t),q_2(t);V)+1)\, x_2(t)dt \right]$$

using $\mathbf{E}_{q(t)}[dN_i(t)] = \frac{1+\xi+q_{3-i}(t)}{1+\xi+q_1(t)+q_2(t)+q_1(t)q_2(t)} x_i(t)dt$. By (6) and $x_1(t)+x_2(t)=1$, the last equality implies

$$V(q(0)) \ge \mathbf{E}_{q(0)}[U(q(\tau)) - \tau]. \quad (16)$$

However, function V might not be continuously differentiable for a whole interval $[t_1,t_2] \subseteq [0,\tau]$, which makes (15) invalid. Denote the set where V is not continuously differentiable as $\mathscr{R} \subset [0,+\infty)^2$. Following the ideas presented in Theorem 7.1, Chapter IV in Fleming and Rishel (2012), I perturb the original strategy (x,τ,α) as follows. Each time $q(t)$ is going to follow *along* a continuous part of the set $\mathscr{R}$, the agent uses some attention rule for some period of time to get away from all continuous parts of $\mathscr{R}$. Since $\mathscr{R}$ is of measure 0, I can always find a decreasing sequence $\{\Delta_m\}_{m=1}^{+\infty}$, $\Delta_m > 0$, with $\lim_{m\to+\infty} \Delta_m = 0$, and a sequence of perturbations, such that the total amount of time the agent spends deviating from the original strategy is Δ_m and (15) holds for each perturb trajectory $q^m(t)$, $0 \le t \le \tau + \Delta_m$. Then (16) holds as well:

$$V(q(0)) \ge \mathbf{E}_{q(0)}[U(q^m(\tau+\Delta_m)) - \tau - \Delta_m]. \quad (17)$$

Taking $m \to +\infty$, I get (16) for the original strategy (the limit can be moved inside the integral by Lebesgue's Dominated Convergence Theorem, where $\max_{\substack{a\in\mathscr{A},\\ i,j\in\{0,1\}}} |u_{ij}(a)|$ plays the role of a dominating function for $U(q^m(\tau+\Delta_m))$).

So, (16) holds for any strategy (x,τ,α). Combining it with the definition of $U(q)$ as the expected payoff from the best action, I get:

$$V(q(0)) \ge \sup_{(x,\tau,\alpha)} \mathbf{E}_{q(0)}\left[u_{\theta_1\theta_2}(\alpha) - \tau\right]. \quad (18)$$

The left hand side of (18) is equal to the value function at point $q(0)$. By assumption, there exists a strategy that gives the expected payoff $V(q(0))$. Thus, (18) holds as equality for all $q(0)$. That establishes the equivalence of V and the value function.

B.4 Lemma 4

When source 1 is used, $\mathscr{L}_1(q_1,q_2;V)=0$ gives

$$V(q_1,q_2)=\frac{q_1(1+\xi+\bar{q}_1+q_2+\bar{q}_1q_2)}{\bar{q}_1(1+\xi+q_1+q_2+q_1q_2)}V(\bar{q}_1,q_2)$$
$$+\frac{(\bar{q}_1-q_1)(1+\xi+q_2)}{\bar{q}_1(1+\xi+q_1+q_2+q_1q_2)}(V(0,q_2)-1)-\frac{q_1(1+q_2)}{1+\xi+q_1+q_2+q_1q_2}\log\frac{\bar{q}_1}{q_1}. \quad (19)$$

Since $\bar{q}_1\le R(a_2(q_1))$, (5) becomes

$$V(\tilde{q}_1,0)=\frac{\tilde{q}_1\,(u_{01}(a_2)-\log R(a_2)-1+\log\tilde{q}_1)+(1+\xi)(u_{11}(a^*)-1)}{1+\xi+\tilde{q}_1},\quad \tilde{q}_1\in[q_1,\bar{q}_1],$$

where $a_2=a_2(q_1)$.

Put it all together to get

$$\mathscr{L}_2(q_1,q_2;V)=\frac{q_1(1+\xi+\bar{q}_1+q_2+\bar{q}_1q_2)}{\bar{q}_1(1+\xi+q_1+q_2+q_1q_2)}\mathscr{L}_2(\bar{q}_1,q_2;V)+\frac{(\bar{q}_1-q_1)(1+\xi+q_2)}{\bar{q}_1(1+\xi+q_1+q_2+q_1q_2)}\mathscr{L}_2(0,q_2;V).$$

$\mathscr{L}_2(\bar{q}_1,q_2;V)\le 0$ by assumption, and $\mathscr{L}_2(0,q_2;V)\le 0$ because $V(0,q_2)$ satisfies (6).

B.5 Lemma 5

When source 1 is used, (19) holds.

Since $R(a_2(q_1))\le q_1<\bar{q}_1$, $a_2(\tilde{q}_1)=a_2(q_1)\equiv a_2$ for all $q_1\le\tilde{q}_1\le\bar{q}_1$, (5) becomes

$$V(\tilde{q}_1,0)=\frac{\tilde{q}_1u_{01}(a_2)+(1+\xi)u_{11}(a_2)}{1+\xi+\tilde{q}_1},\quad \tilde{q}_1\in[q_1,\bar{q}_1]. \quad (20)$$

Symmetrically, since $R(a_1(q_2))\le q_2$,

$$V(0,q_2)=\frac{q_2u_{10}(a_1)+(1+\xi)u_{11}(a_1)}{1+\xi+q_2},\quad \text{where } a_1=a_1(q_2). \quad (21)$$

Put it all together to get

$$\mathscr{L}_2(q_1,q_2;V)=\frac{q_1(1+\xi+\bar{q}_1+q_2+\bar{q}_1q_2)}{\bar{q}_1(1+\xi+q_1+q_2+q_1q_2)}\mathscr{L}_2(\bar{q}_1,q_2;V)$$
$$+\frac{(\bar{q}_1-q_1)(\bar{q}_1+R(a_1)-q_2-R(a_2))}{\bar{q}_1(1+\xi+q_1+q_2+q_1q_2)}+\frac{\bar{q}_1\left(\frac{q_1}{\bar{q}_1}\left(\frac{q_1}{\bar{q}_1}-\log\frac{q_1}{\bar{q}_1}-1\right)-\left(1-\frac{q_1}{\bar{q}_1}\right)^2\right)}{1+\xi+q_1+q_2+q_1q_2}.$$

$\mathscr{L}_2(\bar{q}_1,q_2;V)\le 0$ and $\bar{q}_1+R(a_1)-q_2-R(a_2)\le 0$ by assumption. The last term is also non-positive because function $f(x)=x(x-\log x-1)-(1-x)^2$ is increasing on $0<x\le 1$ to $f(1)=0$.

B.6 Lemma 6

When the attention is split according to $x_1=\frac{q_2}{q_1+q_2}$, $x_2=\frac{q_1}{q_1+q_2}$, V solves $\frac{q_2}{q_1+q_2}\mathscr{L}_1(q_1,q_2;V)+\frac{q_1}{q_1+q_2}\mathscr{L}_2(q_1,q_2;V)=0$, or equivalently

$$\frac{2q_1q_2}{q_1+q_2}\frac{\partial\Upsilon(q_1+q_2,q_1-q_2)}{\partial\zeta}+\frac{(1+\xi+q_1)q_1\,(V(q_1,0)-\Upsilon(q_1+q_2,q_1-q_2))}{(q_1+q_2)(1+\xi+q_1+q_2+q_1q_2)}$$
$$+\frac{(1+\xi+q_2)q_2\,(V(0,q_2)-\Upsilon(q_1+q_2,q_1-q_2))}{(q_1+q_2)(1+\xi+q_1+q_2+q_1q_2)}=1, \quad (22)$$

where $\Upsilon(\zeta,\eta)=V\left(\frac{\zeta+\eta}{2},\frac{\zeta-\eta}{2}\right)$, $\zeta=q_1+q_2$, $\eta=q_1-q_2$. Note that the state vector $q(t)$ is moving along the line $q_1-q_2=\text{const}$. When this constant is equal to η and this rule is used for $\zeta\le\bar{\zeta}$, differential equation (22) gives

$$\Upsilon(\zeta,\eta)=\frac{(\zeta^2-\eta^2)(4\bar{\zeta}+\bar{\zeta}^2-\eta^2+4(1+\xi))}{(\bar{\zeta}^2-\eta^2)(4\zeta+\zeta^2-\eta^2+4(1+\xi))}\Upsilon(\bar{\zeta},\eta)+\frac{2(\zeta^2-\eta^2)}{4\zeta+\zeta^2-\eta^2+4(1+\xi)}\times$$
$$\times\int_{\zeta}^{\bar{\zeta}}\frac{(\tilde{\zeta}-\eta)(2(1+\xi)+\tilde{\zeta}-\eta)V\left(0,\frac{\tilde{\zeta}-\eta}{2}\right)+(\tilde{\zeta}+\eta)(2(1+\xi)+\tilde{\zeta}+\eta)V\left(\frac{\tilde{\zeta}+\eta}{2},0\right)-\tilde{\zeta}\left(4\tilde{\zeta}+\tilde{\zeta}^2-\eta^2+4(1+\xi)\right)}{(\tilde{\zeta}^2-\eta^2)^2}d\tilde{\zeta} \tag{23}$$

Expressions for $V\left(0,\frac{\tilde{\zeta}-\eta}{2}\right)$ and $V\left(\frac{\tilde{\zeta}+\eta}{2},0\right)$ are defined by (20) and (21).

$V(q_1,q_2)$ is defined by (19), where $\check{q}_1$ is used instead of $\bar{q}_1$, and $\Upsilon(\check{q}_1+q_2,R(a_2)-R(a_1))$ instead of $V(\bar{q}_1,q_2)$.

Put it all together to get

$$\mathscr{L}_2(q_1,q_2;V)=\frac{\check{q}_1\left(\frac{q_1}{\check{q}_1}\left(\frac{q_1}{\check{q}_1}-\log\frac{q_1}{\check{q}_1}-1\right)-\left(1-\frac{q_1}{\check{q}_1}\right)^2\right)}{1+\xi+q_1+q_2+q_1q_2}\le 0.$$

B.7 Lemma 7

Denote by $\Upsilon(\pi,\theta_i)$ the expected payoff from using source i until $\bar{\pi}$, given the initial belief $\pi=\mathbb{P}(\theta_i=0)$ and true state θ_i. Bayes' rule gives $\pi(t)=\frac{\pi(0)}{\pi(0)+(1-\pi(0))e^{-t}}$ in the absence of jumps. Then the stopping threshold $\bar{\pi}$ corresponds to the stopping time $\bar{t}=\log\left(\frac{\bar{\pi}(1-\pi)}{(1-\bar{\pi})\pi}\right)$. If $\theta_i=0$, then no jump is possible:

$$\Upsilon(\pi,0)=\Upsilon(\bar{\pi},0)-\log\left(\frac{\bar{\pi}(1-\pi)}{(1-\bar{\pi})\pi}\right).$$

If $\theta_i=1$, then the probability that no jump occurs during $[0,\bar{t}]$ is equal to $e^{-\bar{t}}=\frac{(1-\bar{\pi})\pi}{\bar{\pi}(1-\pi)}$:

$$\Upsilon(\pi,1)=\left(1-\frac{(1-\bar{\pi})\pi}{\bar{\pi}(1-\pi)}\right)\Upsilon(0,1)+\frac{(1-\bar{\pi})\pi}{\bar{\pi}(1-\pi)}\Upsilon(\bar{\pi},1)-\int_0^{\bar{t}}te^{-t}dt-\bar{t}e^{-\bar{t}}$$
$$=\left(1-\frac{(1-\bar{\pi})\pi}{\bar{\pi}(1-\pi)}\right)(\Upsilon(0,1)-1)+\frac{(1-\bar{\pi})\pi}{\bar{\pi}(1-\pi)}\Upsilon(\bar{\pi},1).$$

Then the expected payoff is

$$\Upsilon(\pi)=\pi\Upsilon(\pi,0)+(1-\pi)\Upsilon(\pi,1)=\frac{\pi}{\bar{\pi}}\Upsilon(\bar{\pi})-\pi\log\left(\frac{\bar{\pi}(1-\pi)}{(1-\bar{\pi})\pi}\right)+\left(1-\pi-\frac{\pi}{\bar{\pi}}(1-\bar{\pi})\right)(\Upsilon(0,1)-1). \tag{24}$$

If $i=1$, then $\pi=p_{00}+p_{01}$ and $\Upsilon(0,1)=\frac{p_{11}(u_{11}(a^*)-1)}{p_{10}+p_{11}}+\frac{p_{10}}{p_{10}+p_{11}}\left(u_{10}(a_1)-1-f\left(\frac{u_{11}(a^*)-u_{11}(a_1)-1}{p_{10}/p_{11}}\right)\right)$.

If $i=2$, then $\pi=p_{00}+p_{10}$ and $\Upsilon(0,1)=\frac{p_{11}(u_{11}(a^*)-1)}{p_{01}+p_{11}}+\frac{p_{01}}{p_{01}+p_{11}}\left(u_{01}(a_2)-1-f\left(\frac{u_{11}(a^*)-u_{11}(a_2)-1}{p_{01}/p_{11}}\right)\right)$.

B.8 Lemma 8

First, suppose $p_{11}=0$ (see Figure 3). Then the rule is to split attention equally to stay on the line $p_{10}=p_{01}$ (and take action a_i if a positive signal from source i is received). Note that in this case p_{00} can be taken as a state variable: $p_{10}=p_{01}=\frac{1-p_{00}}{2}$

Denote $\Upsilon(p_{00},\theta_1,\theta_2)$ the expected payoff from following this strategy until $\bar{p}_{00}$, given the initial belief p_{00} and true state (θ_1,θ_2). Bayes' rule gives $p_{00}(t)=\frac{p_{00}(0)}{(1-p_{00}(0))e^{-t/2}+p_{00}(0)}$ in the absence of jumps. Then the stopping time is $\bar{t}=2\log\left(\frac{\bar{p}_{00}(1-p_{00})}{(1-\bar{p}_{00})p_{00}}\right)$.

$$\Upsilon(p_{00},0,0)=\Upsilon(\bar{p}_{00},0,0)-\bar{t},$$

$$\Upsilon(p_{00},1,0)=\int_0^{\bar{t}}(u_{10}(a_1)-t)\frac{e^{-t/2}}{2}dt+(\Upsilon(\bar{p}_{00},1,0)-\bar{t})e^{-\bar{t}/2}=\Upsilon(\bar{p}_{00},1,0)e^{-\bar{t}/2}+\left(1-e^{-\bar{t}/2}\right)(u_{10}(a_1)-2),$$

$$\Upsilon(p_{00},0,1)=\Upsilon(\bar{p}_{00},0,1)e^{-\bar{t}/2}+\left(1-e^{-\bar{t}/2}\right)(u_{01}(a_2)-2).$$

Then the expected payoff is

$$\begin{aligned}\Upsilon(p_{00})&=p_{00}\Upsilon(p_{00},0,0)+\frac{1-p_{00}}{2}\left(\Upsilon(p_{00},1,0)+\Upsilon(p_{00},0,1)\right)\\&=\frac{p_{00}}{\bar{p}_{00}}\Upsilon(\bar{p}_{00})+2p_{00}\log\left(\frac{(1-\bar{p}_{00})p_{00}}{\bar{p}_{00}(1-p_{00})}\right)+\frac{1-p_{00}}{2}\left(1-\frac{(1-\bar{p}_{00})p_{00}}{\bar{p}_{00}(1-p_{00})}\right)(u_{10}(a_1)+u_{01}(a_2)-4)\end{aligned}$$

When $p_{00}=0$, point p is steady state and $\bar{t}=+\infty$. Thus,

$$\Upsilon(0)=\frac{u_{10}(a_1)+u_{01}(a_2)}{2}-1.$$

Now suppose $p_{11}>0$ (see Figure 2). Then $\pi=\frac{p_{10}+p_{01}}{p_{11}}$ can be taken as a state variable:

$$\begin{cases}p_{10}+p_{11}u_{11}(a_1)=p_{01}+p_{11}u_{11}(a_2)\\p_{11}-(p_{10}+p_{11})(p_{01}+p_{11})=\xi p_{10}p_{01}\\p_{10}+p_{01}=\pi p_{11}\end{cases}\quad\Leftrightarrow\quad\begin{cases}p_{10}=\frac{2(\pi-u_{11}(a_1)+u_{11}(a_2))}{(2+\pi)^2+\pi^2\xi-(1+\xi)(u_{11}(a_1)-u_{11}(a_2))^2}\\p_{01}=\frac{2(\pi+u_{11}(a_1)-u_{11}(a_2))}{(2+\pi)^2+\pi^2\xi-(1+\xi)(u_{11}(a_1)-u_{11}(a_2))^2}\\p_{11}=\frac{4}{(2+\pi)^2+\pi^2\xi-(1+\xi)(u_{11}(a_1)-u_{11}(a_2))^2}\end{cases}\tag{25}$$

Source 1 gets attention $x_1=\frac{p_{10}}{p_{10}+p_{01}}=\frac{1}{2}-\frac{u_{11}(a_1)-u_{11}(a_2)}{2\pi}$, source 2 gets attention $x_2=\frac{p_{01}}{p_{10}+p_{01}}=\frac{1}{2}+\frac{u_{11}(a_1)-u_{11}(a_2)}{2\pi}$.

As usual, introduce the notation $\Upsilon(\pi,\theta_1,\theta_2)$ for the expected payoff. Bayes' rule gives

$$T_1(t)\equiv\int_0^t x_1(\tau)d\tau=\log\left(\frac{p_{01}(t)p_{11}(0)}{p_{01}(0)p_{11}(t)}\right)\overset{(25)}{=}\log\left(\frac{\pi(t)+u_{11}(a_1)-u_{11}(a_2)}{\pi(0)+u_{11}(a_1)-u_{11}(a_2)}\right),$$

$$T_2(t)\equiv\int_0^t x_2(\tau)d\tau=\log\left(\frac{p_{10}(t)p_{11}(0)}{p_{10}(0)p_{11}(t)}\right)\overset{(25)}{=}\log\left(\frac{\pi(t)-u_{11}(a_1)+u_{11}(a_2)}{\pi(0)-u_{11}(a_1)+u_{11}(a_2)}\right)$$

$$\begin{aligned}\Rightarrow\quad t=T_1(t)+T_2(t)&=\log\left(\frac{\pi^2(t)-(u_{11}(a_1)-u_{11}(a_2))^2}{\pi^2(0)-(u_{11}(a_1)-u_{11}(a_2))^2}\right)\\&=2T_1(t)+\log\left(\frac{\pi(0)+\left(1-2e^{-T_1(t)}\right)(u_{11}(a_1)-u_{11}(a_2))}{\pi(0)-u_{11}(a_1)+u_{11}(a_2)}\right),\end{aligned}$$

where $T_i(t)$ is the total amount of attention paid to source i up until moment t.

$$\Upsilon(\pi,0,0)=\Upsilon(\bar{\pi},0,0)-\bar{t},$$

$$\Upsilon(\pi,1,0)=\int_0^{\bar{t}}(u_{10}(a_1)-t)\,x_1(t)e^{-T_1(t)}dt+(\Upsilon(\bar{\pi},1,0)-\bar{t})\,e^{-T_1(\bar{t})}$$

$$=\int_0^{T_1(\bar{t})}\left(u_{10}(a_1)-2t_1-\log\left(\frac{\pi+\left(1-2e^{-t_1}\right)(u_{11}(a_1)-u_{11}(a_2))}{\pi-u_{11}(a_1)+u_{11}(a_2)}\right)\right)e^{-t_1}dt_1+(\Upsilon(\bar{\pi},1,0)-\bar{t})\,e^{-T_1(\bar{t})}$$

$$=\begin{cases}\Upsilon(\bar{\pi},1,0)e^{-T_1(\bar{t})}+\frac{\pi+u_{11}(a_1)-u_{11}(a_2)}{2(u_{11}(a_1)-u_{11}(a_2))}\,(T_1(\bar{t})-T_2(\bar{t}))+\frac{(\bar{\pi}-\pi)(u_{10}(a_1)-1)}{\bar{\pi}+u_{11}(a_1)-u_{11}(a_2)}, & u_{11}(a_1)\neq u_{11}(a_2)\\ \Upsilon(\bar{\pi},1,0)\frac{\pi}{\bar{\pi}}+\left(1-\frac{\pi}{\bar{\pi}}\right)(u_{10}(a_1)-2), & u_{11}(a_1)=u_{11}(a_2)\end{cases}$$

$$\Upsilon(\pi,0,1)=\begin{cases}\Upsilon(\bar{\pi},0,1)e^{-T_2(\bar{t})}+\frac{\pi-u_{11}(a_1)+u_{11}(a_2)}{2(u_{11}(a_1)-u_{11}(a_2))}\,(T_1(\bar{t})-T_2(\bar{t}))+\frac{(\bar{\pi}-\pi)(u_{01}(a_2)-1)}{\bar{\pi}-u_{11}(a_1)+u_{11}(a_2)}, & u_{11}(a_1)\neq u_{11}(a_2)\\ \Upsilon(\bar{\pi},0,1)\frac{\pi}{\bar{\pi}}+\left(1-\frac{\pi}{\bar{\pi}}\right)(u_{01}(a_2)-2), & u_{11}(a_1)=u_{11}(a_2)\end{cases}$$

$$\Upsilon(\pi,1,1)=\int_0^{\bar{t}}(u_{11}(a_1)x_1(t)+u_{11}(a_2)x_2(t)-t)\,e^{-t}dt+(\Upsilon(\bar{\pi},1,1)-\bar{t})\,e^{-\bar{t}}$$

$$=\frac{1}{2}\int_0^{\bar{t}}\left(u_{11}(a_1)+u_{11}(a_2)-\frac{e^{-\frac{t}{2}}\,(u_{11}(a_1)-u_{11}(a_2))^2}{\sqrt{\pi^2-(1-e^{-t})(u_{11}(a_1)-u_{11}(a_2))^2}}-2t\right)e^{-t}dt+(\Upsilon(\bar{\pi},1,1)-\bar{t})\,e^{-\bar{t}}$$

$$=\begin{cases}\Upsilon(\bar{\pi},1,1)e^{-\bar{t}}-\frac{\pi^2-(u_{11}(a_1)-u_{11}(a_2))^2}{4(u_{11}(a_1)-u_{11}(a_2))}\,(T_1(\bar{t})-T_2(\bar{t}))-\frac{(\bar{\pi}-\pi)\left(\bar{\pi}\pi+(\bar{\pi}+\pi)(2-u_{11}(a_1)-u_{11}(a_2))+(u_{11}(a_1)-u_{11}(a_2))^2\right)}{2\left(\bar{\pi}^2-(u_{11}(a_1)-u_{11}(a_2))^2\right)}, & u_{11}(a_1)\neq u_{11}(a_2)\\ \Upsilon(\bar{\pi},1,1)\left(\frac{\pi}{\bar{\pi}}\right)^2+\left(1-\left(\frac{\pi}{\bar{\pi}}\right)^2\right)\left(\frac{u_{11}(a_1)+u_{11}(a_2)}{2}-1\right), & u_{11}(a_1)=u_{11}(a_2)\end{cases}$$

Then the expected payoff is

$$\Upsilon(\pi)=\frac{(1+\xi)\left(\pi^2-(u_{11}(a_1)-u_{11}(a_2))^2\right)\Upsilon(\pi,0,0)+2(\pi-u_{11}(a_1)+u_{11}(a_2))\,\Upsilon(\pi,1,0)}{(2+\pi)^2+\pi^2\xi-(1+\xi)(u_{11}(a_1)-u_{11}(a_2))^2}$$
$$+\frac{2(\pi+u_{11}(a_1)-u_{11}(a_2))\,\Upsilon(\pi,0,1)+4\Upsilon(\pi,1,1)}{(2+\pi)^2+\pi^2\xi-(1+\xi)(u_{11}(a_1)-u_{11}(a_2))^2},$$

which leads to expressions (10) and (11) after simplifications.[27]

B.9 Lemma 9

Substitute $\Upsilon(\bar{\pi})=\bar{\pi}\Upsilon_0+(1-\bar{\pi})\Upsilon_1$ into (24) and maximize over $\bar{\pi}\geq\pi$:

$$\bar{\pi}=\begin{cases}1-\frac{1}{\Upsilon(0,1)-\Upsilon_1}, & \Upsilon(0,1)-\Upsilon_1>\frac{1}{1-\pi}\\ \pi, & \text{otherwise}\end{cases}$$

[27] A keen reader notices that the same expression can be obtained by using 23. Here I offer an alternative proof that is longer but more intuitive.

For the optimal threshold $\bar{\pi}$, the expected payoff

$$\Upsilon(\pi) = \pi\Upsilon_0 + (1-\pi)\Upsilon_1 + \frac{\bar{\pi}-\pi}{1-\bar{\pi}} - \pi\log\left(\frac{\bar{\pi}(1-\pi)}{(1-\bar{\pi})\pi}\right)$$

If $i=1$, then $\Upsilon_0 = \frac{p_{00}}{p_{01}+p_{00}}u_{00}(a) + \frac{p_{01}}{p_{01}+p_{00}}u_{01}(a)$, $\Upsilon_1 = \frac{p_{11}}{p_{10}+p_{11}}u_{11}(a) + \frac{p_{10}}{p_{10}+p_{11}}u_{10}(a)$. If $i=2$, then $\Upsilon_0 = \frac{p_{00}}{p_{10}+p_{00}}u_{00}(a) + \frac{p_{10}}{p_{10}+p_{00}}u_{10}(a)$, $\Upsilon_1 = \frac{p_{11}}{p_{01}+p_{11}}u_{11}(a) + \frac{p_{01}}{p_{01}+p_{11}}u_{01}(a)$.

B.10 Lemma 10

By Bayes' rule,

$$\begin{array}{rcl} dp_{00} &=& p_{00}(p_{11}+p_{10}x_1+p_{01}x_2)dt, \\ dp_{01} &=& p_{01}((p_{11}+p_{10})x_1-(p_{00}+p_{10})x_2)dt, \\ dp_{10} &=& p_{10}((p_{11}+p_{01})x_2-(p_{00}+p_{01})x_1)dt, \\ dp_{11} &=& -p_{11}(p_{00}+p_{01}x_1+p_{10}x_2)dt. \end{array} \tag{26}$$

Fix current belief $\tilde{p}$. Let $V_{12}(p)$ be the expected payoff from using source 1 until p, then permanently switching to source 2 until it is optimal to stop (since the goal is to explore the optimality of p, the dependence on $\tilde{p}$ is omitted). By Lemmas 7 and 9, and Bayes' rule (26) with $x_1=1$ and $x_2=0$,

$$\begin{aligned} dV_{12}(p) = \frac{\tilde{p}_{00}}{p_{00}}\Bigg\{ & p_{11}\left(\frac{(1-\pi)\underline{\pi}}{\pi(1-\underline{\pi})}(u_{11}(a^*)-u_{11}(a)-1) + \frac{\pi-\underline{\pi}}{\pi(1-\underline{\pi})}\min\left[u_{11}(a^*)-u_{11}(a_2), \frac{\pi}{p_{11}}\right]\right) \\ & + p_{10}\left(F_1\left(a_1, \frac{p_{10}}{p_{11}}\right) - u_{10}(a) - 1 + \log\left(\frac{\pi(1-\underline{\pi})}{\underline{\pi}(1-\pi)}\right)\right) - 1\Bigg\}dt. \end{aligned} \tag{27}$$

Note the redundancy of calculating dV for the case when it is optimal to use both sources before switching to phase 2. Due to continuity of the optimal strategy in belief space, the switching point lies at the intersection of line $p_{10}+p_{11}u_{11}(a_1) = p_{01}+p_{11}u_{11}(a_2)$ and curve (12).

B.11 Theorem 2

For the proof below, I am writing $h(n) = O(1)$ meaning that $|h(n)| \le M$ for some M as $n \to +\infty$.

Let us fix to a the action taken at the stopping time conditional on no jumps observed (if $t^* > 0$ as $n \to +\infty$ for any fixed a, $t^* > 0$ as $n \to +\infty$ for the optimal a). By contradiction, suppose $t^* = 0$. WLOG, suppose the source used in phase 2 is source 2. Then the payoff increase from using source 1 for dt before permanently switching to source 2 is

$$dV^{(n)}(p) \ge p_{10}\left(\max_{\tilde{a}\in\mathscr{A}} F_1^{(n)}\left(\tilde{a}, \frac{p_{10}}{p_{11}}\right) - u_{10}^{(n)}(a) + \log\max\left\{\max_{\tilde{a}\in\mathscr{A}} F_2^{(n)}\left(\tilde{a}, \frac{p_{01}}{p_{11}}\right) - u_{01}^{(n)}(a) + O(1), \frac{1}{p_{01}}\right\}\right) + O(1),$$

(see (27)). By (9),

$$\max_{\tilde{a}\in\mathscr{A}} F_1^{(n)}\left(\tilde{a}, \frac{p_{10}}{p_{11}}\right) - u_{10}^{(n)}(a) \ge u_{10}^{(n)}\left(a_1^{(n)}\right) - u_{10}^{(n)}(a) - f\left(\frac{\max_{a^*\in\mathscr{A}} u_{11}^{(n)}(a^*) - u_{11}^{(n)}\left(a_1^{(n)}\right) - 1}{\frac{p_{10}}{p_{11}}}\right) = u_{10}^{(n)}\left(a_1^{(n)}\right) - u_{10}^{(n)}(a) + O(1),$$

$$\max_{\tilde{a}\in\mathscr{A}} F_2^{(n)}\left(\tilde{a}, \frac{p_{01}}{p_{11}}\right) - u_{01}^{(n)}(a) \ge u_{01}^{(n)}\left(a_2^{(n)}\right) - u_{01}^{(n)}(a) + O(1).$$

Since $a_1^{(n)} \ne a_2^{(n)}$, conditions (7) and (8) guarantee that $dV^{(n)}(p) \to +\infty$, meaning that it is optimal to start with source 1, that is, $t^* > 0$.

C Asymmetric Costs

In this section I comment on the general case where

- the intensity of the process N_i is $\theta_i \lambda_i x_i(t)$, with $\lambda_i > 0$
- the flow cost of source i is $c_i > 0$

What changes is the definition of regime $(0, a_1, a_2)$. The agent is indifferent between both sources whenever $q_1 < R_1(a_2)$, $q_2 < R_2(a_1)$, where

$$R_i(a) = (1+\xi)\left(\frac{\lambda_i (u_{11}(a^*) - u_{11}(a))}{c_i} - 1\right).$$

Otherwise, the agent must use source 1 if $\frac{c_1(q_1 - R_1(a_2))}{\lambda_1} < \frac{c_2(q_2 - R_2(a_1))}{\lambda_2}$, source 2 if $\frac{c_1(q_1 - R_1(a_2))}{\lambda_1} > \frac{c_2(q_2 - R_2(a_1))}{\lambda_2}$, and split attention according to $x_1 = \frac{c_2 q_2}{c_1 q_1 + c_2 q_2}$, $x_2 = \frac{c_1 q_1}{c_1 q_1 + c_2 q_2}$. See Figure 13.

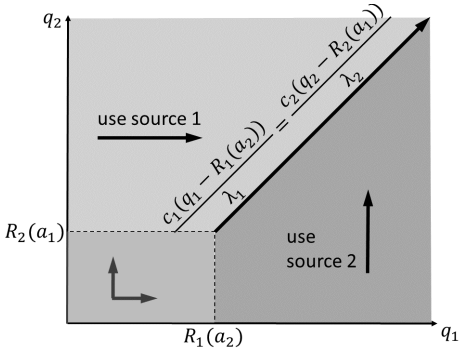

Figure 13: Illustration for Regime $(0, a_1, a_2)$.

Moreover, action $a_i(q)$ is now defined as an action that maximizes $f_i(a,q)$ where

$$f_1(a,q) = \begin{cases} u_{10}(a) - \frac{c_2}{\lambda_2}(\log R_2(a) + 1), & R_2(a) \geq q, \\ u_{10}(a) - \frac{c_2}{\lambda_2}\left(\frac{R_2(a)}{q} + \log(q)\right), & R_2(a) < q; \end{cases} \quad f_2(a,q) = \begin{cases} u_{01}(a) - \frac{c_1}{\lambda_1}(\log R_1(a) + 1), & R_1(a) \geq q, \\ u_{01}(a) - \frac{c_1}{\lambda_1}\left(\frac{R_1(a)}{q} + \log(q)\right), & R_1(a) < q. \end{cases}$$

All this means that now the expression for the *index* is different:[28]

$$\underbrace{\frac{\lambda_1}{c_1}\left(p_{10} + p_{11} + p_{11}\frac{\lambda_2 u_{11}(a_1)}{c_2}\right)}_{\text{source 1 index}} \quad \text{vs} \quad \underbrace{\frac{\lambda_2}{c_2}\left(p_{01} + p_{11} + p_{11}\frac{\lambda_1 u_{11}(a_2)}{c_1}\right)}_{\text{source 2 index}}.$$

First, if the agent gets a positive signal when the true state is (1,1), he might continue with the other source. That explains the correction to the payoff $u_{11}(a_i)$. Second, both indices are adjusted by the efficiency ratio λ_i / c_i, so that a more cost-efficient source has a higher index.

[28] Obviously, there are many ways to define an index here. Ideally, the definition should be consistent with the one derived for many sources case. Since I do not know yet the solution for more than 2 sources, I chose the definition based on whatever is easier to come up with intuition for.

Майская, Татьяна.
Динамический выбор информационных ресурсов [Электронный ресурс] : препринт WP9/2019/05 / Т. Майская ; Нац. исслед. ун-т «Высшая школа экономики». – Электрон. текст. дан. (500 Кб). – М. : Изд. дом Высшей школы экономики, 2019. – (Серия WP9 «Исследования по экономике и финансам»). – 63 с. (На англ. яз.)

Состояние мира состоит из двух (возможно, скоррелированных) бинарных компонент, (θ_1, θ_2), $\theta_i \in \{0, 1\}$. Перед тем как выбрать действие, агент имеет возможность потратить какое-то время на поиск доказательств того, что $\theta_i = 1$. Вне зависимости от того, из каких действий агент должен в итоге выбрать, и вне зависимости от того, какой выигрыш он получает от этих действий в разных состояниях мира, любая оптимальная стратегия агента состоит из двух фаз. В частном случае, когда $\theta_1 + \theta_2 \leq 1$, агент ведёт поиск в самом перспективном направлении во время первой фазы (с возможным изменением направления поиска с течением времени) и полностью игнорирует одну из компонент состояния мира во время второй фазы. Как следствие, когда ставки высоки, группы с противоположными интересами всё же придут к согласию по поводу оптимального направления поиска.

Татьяна Майская
МИЭФ, Национальный исследовательский университет «Высшая школа экономики», Российская Федерация.
Адрес: Российская Федерация, 119049, Москва, ул. Шаболовка, д. 26, каб. 3221.
E-mail: skivinen@hse.ru

Препринты Национального исследовательского университета «Высшая школа экономики» размещаются по адресу: http://www.hse.ru/org/hse/wp

Препринт WP9/2019/05
Серия WP9
Исследования по экономике и финансам

Майская Татьяна

Динамический выбор информационных ресурсов

(*на английском языке*)

Изд. № 2314