Procuring unverifiable information*

Salil Sharma†
Stockholm School of Economics

Elias Tsakas‡
Maastricht University

Mark Voorneveld§
Stockholm School of Economics

December, 2023

Abstract

We study settings where information in the form of Bayesian signals is acquired by an expert on behalf of a principal. Information acquisition is costly for the expert, and crucially not verifiable by the principal. The expert is compensated by the principal with a menu of state-contingent payments. We provide a full characterization of the set of all menus that implement (resp., strictly implement) each signal. Moreover, we provide a closed-form characterization for the expected cost for the cheapest such menu, which we call *proxy cost* of the signal. Surprisingly, in general, the proxy cost is neither increasing in the Blackwell order, nor posterior-separable, even when the expert's cost function is posterior-separable itself. Subsequently, we study the full agency problem (by introducing a downstream decision), thus endogenizing the signal. We show that there is always an optimal signal that can be strictly implemented, meaning that it is without loss of generality to exogenously restrict attention to strict implementation. As a result, similarly to Bayesian persuasion, the complexity of the principal's optimal signal is bounded by the cardinality of the state space. Finally, we present some applications of interest.

Keywords: Costly information acquisition, verifiability, implementation, posterior-separability, proxy cost, Blackwell monotonicity, complexity.
JEL codes: D46, D47, D81, D82, D83, D84.

*We are greatly indebted to an Associate Editor and two anonymous referees for their valuable input. We also thank Mike Burkart, Michael Greinecker, Christoph Kuzmics, Andy Mackenzie, Marcus Opp, Ronald Peeters, Chris Stapenhurst, Stefan Terstiege and the audience at the 2020 Mechanism and Institution Design Conference for helpful comments. Financial support by the Wallander-Hedelius Foundation under grants P2016-0072:1 and P20-0035 (Voorneveld, Sharma) is gratefully acknowledged.

†Department of Economics, Stockholm School of Economics, Box 6501, 113 83 Stockholm, Sweden; E-mail: salil.sharma@phdstudent.hhs.se

‡Department of Microeconomics and Public Economics, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; Homepage: www.elias-tsakas.com; E-mail: e.tsakas@maastrichtuniversity.nl

§Department of Economics, Stockholm School of Economics, Box 6501, 113 83 Stockholm, Sweden; E-mail: mark.voorneveld@hhs.se

1. Introduction

Relying on information provision by experts or analysts is a central characteristic of several economic environments (e.g., Bergemann and Bonatti, 2019, and references therein). Depending on the application at hand, there are multiple reasons why a (female) principal may be interested in acquiring information from a (male) expert. For instance, consider cases where the principal wants to acquire information before making a downstream decision (e.g., an investor receives advise from a financial analyst about the success of a project before deciding whether to invest in it), or the principal needs arguments in order to persuade an agent to take a certain action in turn (e.g., a pharmaceutical company presents evidence acquired from a respected doctor on the safety of a drug in order to persuade the regulators to give approval), or the principal is required by established institutions to rely on the professional assessment of certified experts (e.g., a real-estate developer must consult a civil engineer regarding the stability of the ground before building house). In all these settings, information acquisition is typically costly for the expert, as it involves cognitively taxing processes. As a result, without a contract that establishes the terms of information acquisition and the subsequent reporting, it is often reasonable to expect the expert to free-ride and not obtain the desired information.

However, even if the principal decides to write a contract, there is a fundamental problem that naturally arises. Namely, it is difficult to verify what information the expert has obtained (Kashyap and Kovijnykh, 2016; Rappoport and Somma, 2017; Yoder, 2022). This is because neither the information acquisition technology nor the realized outcome are observable, due to the cognitive nature of the process. Hence, they are not contractible either. Thus, contracts which aim to reward experts for information acquisition in an incentive compatible way are inherently difficult to write, as there is very little one can actually condition on.

In this paper, we address this problem from a mechanism-design point of view (e.g., Laffont and Tirole, 1986), viz., the expert's (state-contingent) payment depends only on his report, similarly in some ways to Kashyap and Kovijnykh (2016), Rappoport and Somma (2017) and Yoder (2022). Within such a setting, we first want to characterize the contracts that guarantee that the expert has indeed acquired the principal's desired information (Section 2). Second, we study the properties of the cost that the principal incurs in the form of payments to expert (Section 3). Finally, we endogenize downstream decisions, and we seek to identify what information the principal may optimally seek to acquire via the expert (Section 4). Let us elaborate on each of these three general questions.

Formally, we consider a general model with finitely many states of nature and a commonly known prior, where the expert is required to acquire information in the form of a Bayesian signal and report its realization to the principal. Following the surging literature on rationally inattention, we assume that the expert incurs posterior-separable costs to acquire information (e.g., Caplin et al., 2022), while his subsequent reporting remains costless. Both the choice of the signal and its realization are private information of the expert. To incentivize the expert to choose the desired signal, the principal offers a menu of state-contingent acts.[1]

Given an arbitrary menu, the expert picks a signal, thereby incurring the respective cost. After observing the realization of the signal, he updates his beliefs about the state of the world and then chooses an act so as to maximize his expected utility. Thus, for a given menu, the expert solves an optimization exercise of choosing a signal so as to maximize net expected utility. We say that a

[1] In the language of mechanism design, each act is a message, and the menu is the mechanism. One can label each act as a posterior belief, thus implicitly resorting to a direct mechanism.

menu implements (resp., strictly implements) a signal if it is optimal (resp., uniquely optimal) for the expert to choose this signal in response to this menu. Thus, the principal's problem boils down to designing a menu that implements (resp., strictly implements) the desired signal.[2] Actually, we can characterize the set of signals that can be implemented (resp., strictly implemented) by some menu (Theorem 1). Moreover, for each such signal, we can characterize the set of menus that implement (resp., strictly implement) this particular signal (Theorem 2). Each such menu can be decomposed into a fixed and a variable component, where the variable component guarantees incentive compatibility, while the fixed component determines how the total surplus is divided between the principal and the expert.

Minimizing the expert's expected payment across the menus that implement a particular signal yields a new type of cost for information (Theorem 3).[3] We refer to this minimal payment as the cost for *proxy attention*. The reasoning behind this term is that the principal hires a proxy (viz., a self-interested expert) in order to incur the attention cost which are associated with the information acquisition process on her behalf. Then, we proceed to study the basic properties of the cost for *proxy attention*. Quite surprisingly, we show that it violates the most basic condition which is shared among all existing cost functions, i.e., Blackwell monotonicity (Remark 2). Furthermore, it is not posterior-separable, even though the expert's cost function is (Corollary 3).

Then, using our new cost function, we study a full agency problem. That is, we explicitly introduce a downstream decision (to be taken either by the principal herself or by a third agent), thus making the choice of the desired signal endogenous. Within this framework one can study a wide range of problem, including delegated information acquisition and delegated persuasion. Despite the generality of our setting, we show that among the principal's optimal menus (at the first stage of the principal-expert interaction), there is always one that induces a unique optimal signal for the expert. Hence, it is without loss of generality to exogenously restrict focus to strict implementation (Theorem 5). But even more interestingly, the latter implies that the complexity of the principal's optimal signal is bounded by the cardinality of the state space. Note that our theorem bears a striking similarity with the well-known result of Kamenica and Gentzkow (2011) from the literature on Bayesian Persuasion. Remarkably, in our case, our result is not driven by posterior-separability, which —as we have already discussed— is often violated by the cost for proxy attention.

In the final part of the paper, we apply our results in the context of two applications of interest. First, we consider the case of decentralized information acquisition, where the principal decides how to allocate information acquisition across a number of experts. Proposition 1 provides conditions under which delegation to one expert may be strictly preferred, that is, the proxy cost are superadditive. This is in contrast to the usual cost functions for information, which are either linear (Pomatto et al., 2023) or subadditive (Sims, 2003). In our second application, we show that when the stakes are high enough for the principal, she prefers to acquire the information herself (Proposition 2). Thus, in this case the proxy cost lies in the same family of costs as the expert's.

The paper is structured as follows. The following subsection surveys related literature. In Section 2, we introduce our model and present the results for the existence and characterization of incentive-compatible menus. Section 3 provides a definition of the cost of proxy attention, an examination of the properties its violates, and surplus analysis. Section 4 provides the endogenized version of the problem. In Section 5 we present the remaining two applications. All proofs are

[2] We consider a wide range of feasibility constraints corresponding to different forms of liability.

[3] Here we no longer need to make the distinction between weak and strict implementation as the respective cheapest menus would cost the same in expectation.

relegated to the Appendices.

1.1. Related Literature

Three related papers to ours are Kashyap and Kovijnykh (2016), Yoder (2022) and Rappoport and Somma (2017). All of them essentially take a mechanism design approach, similarly to us, in order to overcome the non-verifiability problem that we face too. Let us elaborate on the relationship of our work to each of them separately.

Kashyap and Kovijnykh (2016) analyze a model of a credit rating agency which exerts costly effort in order to acquire a signal about the quality of a project. As is the case here, the choice of the signal and its realization are private information of the agency. Moreover, the contracts are contingent both on the rating and the project's performance. Thus they solve a very similar problem to ours, albeit only with a binary state space. However, their focus of analysis differs from ours. In particular, they focus on how the optimal contracts differ depending on who pays, i.e., the investors or a planner. They also study the differences between new and established securities. On the other hand, we also focus on the properties of the proxy cost and the complexity of optimal signal in the agency problem.

Yoder (2022) considers the problem of contracting for experiments in binary state space. The expert exerts costly effort to carry out the experiment. This cost, however, depends on the type of the expert which is private information. This constitutes a problem of moral hazard and adverse selection. Moreover, the expert produces hard/verifiable evidence. Allowing for two contracting environments, where the experiment or the outcome of the experiment is contractible, the paper shows that the additional incentive compatibility constraints are essentially redundant. An important difference in our analysis is that we show when contracting is contingent on the outcome of the experiment and the realized state, incentive compatibility can be achieved even in the absence of hard evidence.

Rappoport and Somma (2017) discuss the alternate problem of incentivizing an agent to perform a costly experiment when the target of the information is a third-party rather than the principal herself. Similarly to Yoder (2022), in their model too the principal can contract on the expert's posterior beliefs. Using this they characterize the conditions under which the principal can achieve her first-best outcome. In particular, they focus on the role of risk preferences and symmetry of the experiments for efficiency, under limited-liability. Once again, our focus is very different as we study the properties of the cost function and the complexity of optimal signal.

This paper can be seen as part of the (ex ante) mechanism design approach to selling/buying information. This literature is reviewed by Bergemann and Bonatti (2019, Ch. 3), and more recently by Bergemann and Ottaviani (2021). As opposed for instance to markets for data, the main premise within this literature is that the compensation scheme is set before signals have been realized. Contrary to our work, most papers within this literature fix at least some market characteristics. For instance, Bergemann et al. (2018) study a model of a monopolist expert offering a menu of verifiable signals together with a price for each of them to a principal with private information over his willingness to pay. Babaioff et al. (2012) study a similar model with two-sided private information. Both these papers also differ from ours in that information is verifiable. Esö and Szentes (2007) consider a consultant (expert) who receives a single signal and decides whether to disclose it to the client (principal). Within the same stream of literature belongs the work of Hörner and Skrzypacz (2016) that considers a principal who is interested in hiring an expert without knowing his competence, and the expert tries to gradually persuade the principal that he is of a good type. Related to our work is also the paper of Ali et al. (2020)

who study markets for information with multiple principals who can resell information among themselves at a later stage. In their benchmark case, they show that the expert can extract strictly positive surplus from at most one principal. Furthermore, the problem compensating the sale of information has also been recently studied by Haghpanah et al. (2022). Unlike our work, in their setting information does not have any instrumental value. Moreover, instead of making compensation state-contingent, they divide it in two stages, one after the signal has been realized, and one after it has been decided by the principal whether it will be publicized or not.

Our work also relates to the stream of literature providing microfoundations to attention costs based on the distribution of posteriors. Papers in this literature include Morris and Strack (2019), Pomatto et al. (2023), Zhong (2022) and, Bloedel and Zhong (2021), among others. The central theme in this literature is to derive static costs of information under varying assumptions on the dynamics of the learning processes and costs. The crucial difference between this literature and ours is that the cost to the principal is derived from delegation. In effect, the principal incurs indirect attention costs in terms of compensation to the agent who, as we assume, is rationally inattentive. Relatedly, our work also contributes to the stream within the marketing literature that focuses on pricing of information. For instance, Arora and Fosfuri (2005) study pricing of diagnostic information (i.e., information that allows to predict if a project will be successful or not), and in particular they are interested in the role of prior beliefs on the value of diagnostic information. In a similar framework, Chang and Lee (1994) and Iyer and Soberman (2000) study optimal pricing of information provided by a marketing consultant to a set of principals. The difference between our work and these papers is once again that we remain agnostic on the underlying market characteristics, unlike all these contributions which specify certain characteristics (something understandable of course, as they are interested in information acquisition in a specific marketing context).

Our treatment of the expert's optimization problem is broadly related to a larger literature on the tradeoff between material payoffs and cognitive costs. The foundations of this tradeoff have been extensively studied by Alaoui and Penta (2016, 2022) and Alaoui et al. (2020).

There is also a relation between our paper and a stream within the information design literature that focuses on applications where information acquisition is delegated by the principal to some third agent, other than the receiver (Bizzotto et al., 2020). While this setting differs from ours in many central aspects, the two share an important common feature, namely that the principal wants the expert to acquire a specific signal.

Our paper is also related – albeit more tangentially – to parts of the contract theory literature on optimal incentive schemes for acquiring information. Some recent contributions include Zermeño (2011, 2012) which consider a linearly ordered subset of signals that are represented by the expert's effort, Carroll (2019) who restricts the set of signals and considers a principal who is maxmin expected utility maximizer, and Lindbeck and Weibull (2020) who restrict attention to binary menus and entropic costs. But our mechanism design approach to selling/buying information is very different from the contract theory literature, for a variety of reasons. First, the latter literature studies the problem of an optimal incentive scheme from the point of view of the principal, thus implicitly postulating specific market characteristics (i.e., the principal always makes a take-it-or-leave-it offer as it is often the case in contract theory). Second, in our work a specific signal is traded, whereas the contract theory literature focuses on the principal's optimization problem which determines the equilibrium traded signal and compensation scheme endogenously. Finally, the menu of contracts which correspond to the menus of acts used in our analysis have been extensively applied in the literature on procurement and regulation beginning with Laffont and Tirole (1986).

2. Signal implementation

2.1. Costly information acquisition

Consider a finite state space Ω, and let $\mu \in \Delta(\Omega)$ denote an arbitrary subjective belief. Throughout the paper, for notation simplicity, we often index states by $\Omega = \{\omega_1, \ldots, \omega_K\}$, and we denote by $\mu^k := \mu(\omega_k)$ the probability that belief μ attaches to the k-th state. The belief that attaches probability 1 to state ω will be denoted by $[\omega]$.

Information is acquired by means of Bayesian signals chosen by a risk-neutral expert, who has a full-support prior belief $\mu_P \in \mathrm{int}(\Delta(\Omega))$. Each signal is a stochastic mapping from the set of states to a compact set of signal realizations. So we can represent the set of signals as the set of mean-preserving distributions of posterior beliefs:

$$\Pi(\mu_P) := \left\{ \pi \in \Delta(\Delta(\Omega)) : \int_{\Delta(\Omega)} \mu \, d\pi = \mu_P \right\}. \tag{1}$$

Call signal $\pi \in \Pi(\mu_P)$ completely uninformative if $\mathrm{supp}(\pi) = \{\mu_P\}$. For a set of posteriors $M \subseteq \Delta(\Omega)$ and a prior $\mu_P \in \mathrm{relint}(\mathrm{conv}(M))$, the signal $\pi \in \Pi(\mu_P)$ with $\mathrm{supp}(\pi) = M$ is unique if and only if the beliefs in M are linearly independent in $\mathbb{R}^\Omega$. As usual, the set of Bayesian signals in $\Pi(\mu_P)$ is endowed with the partial Blackwell order, $\succeq$ (Blackwell, 1953).

The expert's cost for acquiring a signal $\pi \in \Pi(\mu)$ is equal to

$$C(\pi) := \mathbb{E}_\pi(c) - c(\mu), \tag{2}$$

where $c : \Delta(\Omega) \to \mathbb{R}$ is a continuous strictly convex function. Throughout the paper, we assume that c is continuously differentiable in the interior of $\Delta(\Omega)$, with $\nabla c(\mu)$ denoting the gradient at μ. At a boundary point μ, we define the subdifferential of c by

$$\partial c(\mu) := \{t \in \mathbb{R}^\Omega : c(\mu') \geq c(\mu) + \langle t, \mu' - \mu \rangle \text{ for all } \mu' \in \Delta(\Omega)\}.$$

Each vector in this set is called a subderivative or subtangent at μ. At every interior μ, by convexity of c it follows that $\partial c(\mu) \neq \emptyset$, and by differentiability it follows that $\partial c(\mu)$ is a singleton containing only $\nabla c(\mu)$. At boundary points on the other hand, $\partial c(\mu)$ will either be empty or it will contain infinitely many subtangents. In the latter case, we denote by $\nabla c(\mu) \in \partial c(\mu)$ the limit of $\nabla c(\mu_k)$, where $(\mu_k)_{k=1}^\infty$ is a sequence of full-support beliefs converging to μ. Geometrically, $\nabla c(\mu)$ will be the flattest of all subtangent at μ. Thus, it is straightforward to verify that $\mathbb{E}_{\tilde{\mu}}(\nabla c(\mu)) \geq \mathbb{E}_{\tilde{\mu}}(t)$ for every $t \in \partial c(\mu)$ and every $\tilde{\mu} \in \Delta(\Omega)$.

A cost function that takes the form of (2) is called posterior-separable, and it is widely-used within the rational inattention literature.[4] For a graphical illustration, see Figure 1. The most common special case is the entropic cost specification, which is obtained by setting $c(\mu) := \kappa H(\mu)$ for some $\kappa > 0$, where

$$H(\mu) := \sum_{\omega \in \Omega} \mu(\omega) \log \mu(\omega)$$

is the Shannon entropy function, with the usual convention $0 \log 0 = 0$. This specific function is used in many applications throughout the literature, and we will also adopt it in our applications section.

[4] The rational inattention literature was initiated by Sims (2003) in the context of macroeconomics, before attracting attention among microeconomists (Caplin and Dean, 2015; De Oliveira et al., 2017; Ellis, 2018) and being widely used in applications (Hu et al., 2023; Li and Hu, 2023; Lipnowski et al., 2020; Tsakas, 2020). Partly, its appeal stems from the fact that it has strong theoretical foundations (Caplin et al., 2022; Zhong, 2022; Denti, 2022; Tsakas, 2020) and is supported by recent experimental evidence (Dean and Neligh, 2023).

2.2. Problem definition

Suppose that two risk-neutral agents, a (female) principal and a (male) expert, share the prior μ_P. The principal wants to acquire information in the form of a signal about the state space. Information acquisition is delegated to the expert. Following a mechanism design approach, the principal incentivizes the expert to acquire specific signal(s). Incentives are based on what the expert reports to the principal and the eventual state realization.

Formally, an act is a state-contingent payment $a \in \mathbb{R}^\Omega$ with $a(\omega)$ being the state-payment at ω. A menu is a compact set of acts $A \subseteq \mathbb{R}^\Omega$, while $\mathcal{A}$ denotes the set of all compact menus. Each menu $A \in \mathcal{A}$ is interpreted as a mechanism, and each act $a \in A$ is interpreted as a message that yields the corresponding state-contingent payment. Whenever the expert faces A, he first chooses a signal $\pi \in \Pi(\mu_P)$, then he updates to some posterior $\mu \in \operatorname{supp}(\pi)$, and he finally picks an act $a \in A$ that maximizes the expected payoff

$$\mathbb{E}_\mu(a) := \sum_{\omega \in \Omega} \mu(\omega) a(\omega).$$

Which is the optimal menu from the principal's point of view? Of course, the answer to this question depends on the set of feasible menus. Feasibility constraints typically take the form of some liability specification. That is, for some lower bound $b \in \mathbb{R}^\Omega$ we define the set of feasible menus

$$\mathcal{A}^b := \left\{ A \in \mathcal{A} : a \geq b \text{ for all } a \in A \right\}. \tag{3}$$

Examples of such feasibility constraints include no liability (viz., $b(\omega) = 0$ at all $\omega \in \Omega$), limited liability (viz., $b(\omega) < 0$ for some $\omega \in \Omega$), and guaranteed flat fee (viz., $b(\omega) > 0$ for all $\omega \in \Omega$). Notice that in all of the above cases, the liability constraint need not be a constant act. A second type of feasibility constraints sometimes arises in the literature when a payment scheme $A \in \mathcal{A}$ has been already set by the expert, and the principal simply decides whether to accept or reject it, as in Kashyap and Kovijnykh (2016), i.e., the principal can effectively choose a menu from $\{A, \{\mathbf{0}\}\}$, where $\mathbf{0}$ is the constant act that pays 0 at every state. In this last case the feasibility constraint is endogenized, as there is third stage added in the beginning of the game where the expert sets A. In this paper, we will focus on the earlier exogenous liability constraints.

Before thinking about the principal's optimization problem within the constrained set of menus, we first need to understand how the expert best responds to each menu. In particular, we want to characterize the set of signals that the expert chooses in response to each menu $A \in \mathcal{A}$.

2.3. The expert's maximization problem

For an arbitrary menu $A \in \mathcal{A}$, the expert's indirect payoff as a function of each posterior belief $\mu \in \Delta(\Omega)$ is given by the convex function

$$\phi_A(\mu) := \max_{a \in A} \mathbb{E}_\mu(a). \tag{4}$$

This means that his expected payoff from choosing a signal π is equal to $\mathbb{E}_\pi(\phi_A)$. As we will see, from the expert's point of view, the function ϕ_A is the only relevant part of the menu, i.e., the expert's optimization problem will remain invariant across different menus $A, A' \in \mathcal{A}$ for which $\phi_A = \phi_{A'}$. Nevertheless, later in the paper we will see that the choice between A and A' will matter for the principal.

Subtracting the cost $C(\pi) = \mathbb{E}_\pi(c) - c(\mu_P)$, we find the expert's net expected payoff

$$V_A(\pi) := \mathbb{E}_\pi(\phi_A) - C(\pi) = \mathbb{E}_\pi(\phi_A - c) + c(\mu_P).$$

Thus, whenever the expert faces a menu A, his problem boils down to maximizing V_A over the set $\Pi(\mu_P)$ of Bayesian signals corresponding with prior μ_P. For notation simplicity, we define the function

$$\psi_A(\mu) := \phi_A(\mu) - c(\mu). \tag{5}$$

Since $c(\mu_P)$ is just an additive constant, the expert's maximization problem reduces to maximizing $\mathbb{E}_\pi(\psi_A)$ over $\Pi(\mu_P)$. Throughout the paper, for notation simplicity, we denote the set of optimal signals for some menu $A \in \mathcal{A}$ by

$$\Pi_A(\mu_P) := \arg \max_{\pi \in \Pi(\mu_P)} \mathbb{E}_\pi(\psi_A).$$

Solving this maximization problem can be done using the concavification technique which was first introduced in the repeated games literature by Aumann and Maschler (1995) and was later extensively used in the Bayesian persuasion literature following the seminal contribution of Kamenica and Gentzkow (2011): we take the concave closure of ψ_A,

$$\overline{\psi}_A(\mu) := \max_{\pi \in \Pi(\mu)} \mathbb{E}_\pi(\psi_A), \tag{6}$$

and we find the largest set $M_A \subseteq \Delta(\Omega)$ containing the prior μ_P where $\overline{\psi}_A$ is linear (Tsakas, 2020). It is well-known that the optimal signals are exactly those distributions $\pi \in \Pi(\mu_P)$ that put positive probability only to points $\mu \in M_A$ such that $\overline{\psi}_A(\mu) = \psi_A(\mu)$.

In Figure 1, menu $A = \{a_P, a_L, a_H\}$ is feasible menu given the constraint $b \in \mathbb{R}^\Omega$, i.e., no act intersects the interior of the grey-shaded area. The longest interval where $\bar{\psi}_A$ is linear is $M_A = [\mu_L^1, \mu_H^1]$. The only points in this interval where $\bar{\psi}_A$ coincides with ψ_A itself are $\{\mu_P, \mu_L, \mu_H\}$. Hence, the set of optimal signals $\Pi_A(\mu_P)$ contains the completely uninformative signal, the unique signal with support $\{\mu_L, \mu_H\}$, as well as infinitely many signals with support $\{\mu_P, \mu_L, \mu_H\}$.

Recall that one of our stated objectives is that the chosen act —or equivalently the reported message— reveals the actual signal realization to the principal. As it turns out, this is not really a concern, i.e., as also pointed out by Rappoport and Somma (2017), non-observability of the posteriors does not really pose a constraint. By simply observing the expert's chosen act, the principal is able to unambiguously pin down the realized posterior belief. In this sense, throughout the paper, choosing an act is equivalent to (implicitly) reporting his true posterior. In the context of our previous example, a_L will be chosen if and only if the posterior is μ_L; a_H will be chosen if and only if the posterior is μ_H; and finally a_P will be chosen if and only if the prior μ_P is realized.

Let us now do a bit of reverse engineering. That is, we fix a signal $\pi \in \Pi(\mu_P)$ and try to identify the menus under which this signal will possibly/definititely be chosen. As less of a mouthful, we will use the following terminology to describe such menus:

Definition 1. For a signal $\pi \in \Pi(\mu_P)$, the following is said:

(i) Menu $A \in \mathcal{A}$ *implements* signal π if $\pi \in \Pi_A(\mu_P)$.

(ii) Menu $A \in \mathcal{A}$ *strictly implements* signal π if $\{\pi\} = \Pi_A(\mu_P)$.

Moving forward, our first set of questions (*which signals can be implemented/strictly implemented?*) is answered in Section 2.4, while the answer to our second set of questions (*which menus implement/strictly implement each signal?*) is found in Section 2.5.

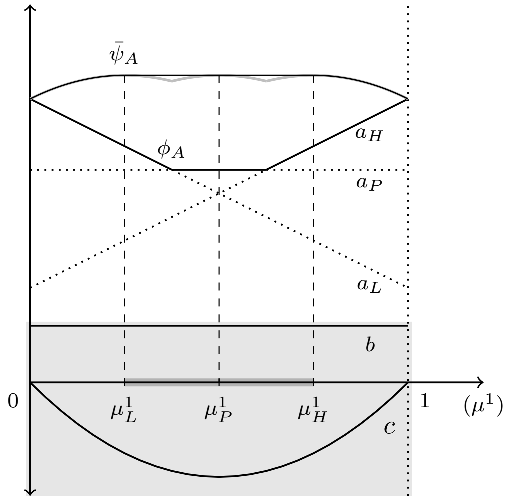

Figure 1: Concavification: The concave closure $\bar{\psi}_A$ is linear in the interval $[\mu_L^1, \mu_H^1]$, and within this interval it coincides with ψ_A only ay μ_L, μ_H and μ_P. Thus, every optimal signal for the expert will necessarily yield of these posteriors.

2.4. Which signals can be implemented?

We first establish necessary and sufficient conditions for the existence of menus that implement, and respectively strictly implement a signal. Of course, if the signal is the completely uninformative one, this can be done by picking an arbitrary singleton menu. So we look beyond this trivial case.

Let us start by introducing the following two properties of a signal:

(P_1) Subdifferentiability: At every $\mu \in \text{supp}(\pi)$ the function c is subdifferentiable.

(P_2) Independence: The beliefs in $\text{supp}(\pi)$ are linearly independent.

Then, we can characterize the set of signals that can be implemented (in the weak and strict sense) in terms of these two properties.

Theorem 1. *Let π be a signal that is not completely uninformative.*

(a) Signal π is implemented by some menu in $\mathcal{A}$ if and only if it satisfies (P_1).

(b) Signal π is strictly implemented by some menu in $\mathcal{A}$ if and only if it satisfies (P_1) − (P_2).

Regarding the first part of the previous theorem, by c being convex, it follows it is subdifferentiable in the interior of $\Delta(\Omega)$. So, only signals that put positive probability to the boundary of the simplex could in principle violate (P_1). This is for instance the case if the expert has entropic information cost, while at the same time the principal wants to implement the perfectly informative signal which reveals the true state with probability 1.

Regarding the second part, it trivially implies that all binary signals can be strictly implemented when the state space is binary. However, things get a bit more complicated in higher dimensions, e.g., signals with three possible realizations can only be strictly implemented if the realizations are not collinear. Thus, $|\text{supp}(\pi)| \leq |\Omega|$ is only a necessary condition for π to be strictly implemented, but it is not sufficient when $|\Omega| \geq 3$.[5]

[5] Part (b) of our result can also follow from Winkler (1988).

2.5. Which menus implement a signal?

Now we turn to our second question: which are the exact menus that implement (resp., strictly implement) a given signal?

To state our result we need to introduce some additional machinery. A convex function $g : \Delta(\Omega) \to \mathbb{R}$ will be said to support c at $M \subseteq \Delta(\Omega)$ whenever $g(\mu) \leq c(\mu)$ at all $\mu \in \Delta(\Omega)$ with equality holding if and only if $\mu \in M$. An example of such a function g is the pointwise maximum of a collection of subtangents at M, i.e., $g(\mu) := \max\{\mathbb{E}_\mu(t) | t \in \partial c(\mu),\ \mu \in M\}$. Then, we are ready to characterize the set of menus that implement (resp., strictly implement) signal π.

Theorem 2. *For a signal $\pi \in \Pi(\mu_P)$ the following hold:*

(a) Suppose that $\pi \in \Pi^1(\mu_P)$. Then, menu A implements signal π if and only if there exists some linear functional L_A such that $\phi_A - L_A$ supports c at some $M \supseteq \operatorname{supp}(\pi)$.

(b) Suppose that $\pi \in \Pi^2(\mu_P)$. Then, menu A strictly implements signal π if and only if there exists some linear functional L_A such that $\phi_A - L_A$ supports c at $\operatorname{supp}(\pi)$.

In Figure 2, our theorem implies that menu $A = \{a_P, a_L, a'_L, a_H\}$ implements signal π with support $\operatorname{supp}(\pi) = \{\mu_L, \mu_H\}$. This is because $\phi_A - L_A$ supports c at $M = \{\mu_P, \mu_L, \mu_H\}$, which is a superset of $\operatorname{supp}(\pi)$.

This last example provides a clear roadmap on how to construct menus that implement a signal $\pi \in \Pi^1(\mu_P)$:

- Step 1: We take (at least) one subtangent of c at each posterior in $\operatorname{supp}(\pi)$. In our previous example, these are the hyperplanes in $\mathbb{E}_\mu(t_L)$, $\mathbb{E}_\mu(t'_L)$ and $\mathbb{E}_\mu(t_H)$.

- Step 2: We may take a second collection of hyperplanes, all of which lie below c. In our previous example, these are the hyperplane $\mathbb{E}_\mu(t_P)$.

- Step 3: We add to each hyperplane from the previous two steps the same linear functional L_A. In our previous example, this yields the hyperplanes $\mathbb{E}_\mu(a_L)$, $\mathbb{E}_\mu(a'_L)$, $\mathbb{E}_\mu(a_H)$ and $\mathbb{E}_\mu(a_P)$.

The menu that implements π is not unique. This is due to a variety of reasons. First of all, in Step 1, for each posterior in $\operatorname{supp}(\pi)$ we can add as many tangent hyperplanes as we like. This type of multiplicity disappears if we focus on signals with full-support posteriors. Second, in Step 2 we may arbitrarily add hyperplanes that are dominated by c. Such additional acts may be tangents at beliefs outside $\operatorname{supp}(\pi)$ or they may lie strictly below c. In the first case they will induce more optimal signals (with support larger than π), whereas in the second case the additional hyperplanes will be completely inconsequential, as they will eventually correspond to irrelevant acts that will never be used anyway. This second type of multiplicity disappears if we focus on strict implementation. Finally, in Step 3, we can rescale the menu by adding the same act to each of the acts that we obtained from the previous two steps. Of course, there are infinitely many such acts which guarantee that the liability condition is satisfied. This third type of multiplicity disappears if we require that the menu is the cheapest one from the set of feasible menus. We further elaborate in the next section.

Remark 1. The previous discussion implies that every menu that strictly implements a signal can be decomposed into a flat and a variable payment. In the previous example, the menu $\{a_L, a_H\}$

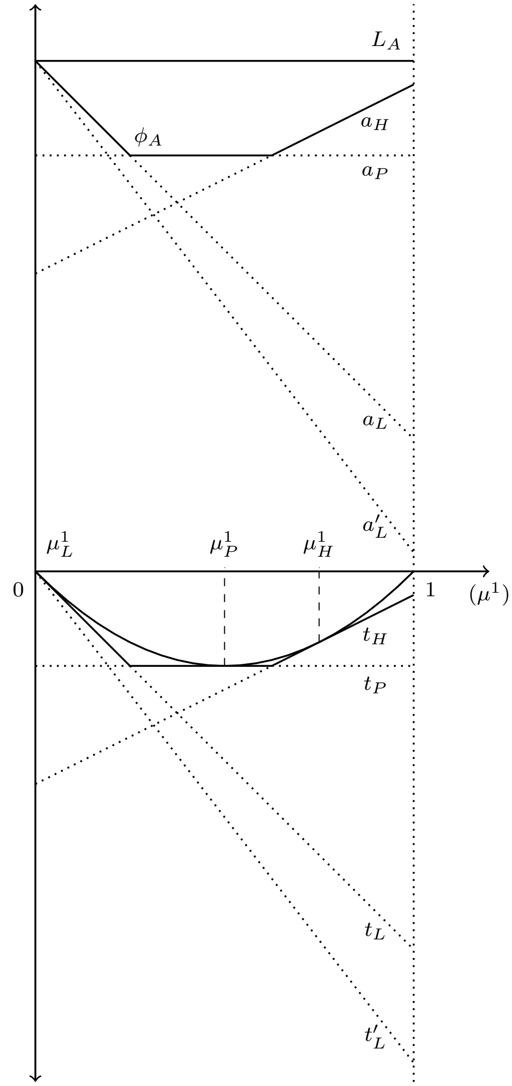

Figure 2: Implementation: The menus that implement signal π with $\mathrm{supp}(\pi) = \{\mu_L, \mu_H\}$ are obtained by first taking tangents of c at $\{\mu_L, \mu_P, \mu_H\} \supseteq \mathrm{supp}(\pi)$ and then adding the same linear functional L_A.

that strictly implements signal π is decomposed into the flat payment $L_A(\mu)$ which is paid to the expert irrespective of which action he takes, and the variable payment $\max\{\mathbb{E}_\mu(t_L), \mathbb{E}_\mu(t_H)\}$ which depends on the expert's action. Importantly, if the signal has full support, all menus that strictly implement the signal agree on the variable payment and differ only on the flat payment. Thus, the following interpretation is given: *the variable payment compensates for the lack of verifiability by providing the exact incentives that lead to the specific signal being chosen, whereas the flat payment determines how the total surplus is split between the two agents.* We elaborate on the division of the surplus later in the paper. $\triangleleft$

3. Supply of information

3.1. Cheapest menu

As we have already discussed, there are multiple menus that implement (resp., strictly implement) signal π. *Which is the cheapest one for the principal?* Answering this question will turn out to be crucial in applications where we specify the principal's objective function, and we treat the principal as the designer who chooses the menu.

Unfortunately, the definition of the "cheapest menu" is not straightforward, particularly when the signal cannot be strictly implemented. Indeed, whenever $\pi \in \Pi^1(\mu_P) \setminus \Pi^2(\mu_P)$, there are additional optimal signals besides π. Thus, from the principal's point of view, it is ambiguous which of these optimal signals will be eventually chosen by the expert. As a result it is ambiguous how much the menu will end costing in expectation.

Despite this difficulty, it is still possible to find a cheapest menu among those that weakly implement π. The idea is that this menu will be preferred by the principal compared to any alternative menu that implements π, not only if π itself is eventually chosen, but also if any other signal is picked among those that are optimal for the expert. In this sense our notion of "a cheapest menu" is robust with respect to the expert's optimal choices. And of course, this definition is directly extended to signals that can be strictly implemented.

Here is how such a menu is obtained. First, we define the gradient at each posterior in the support of π:

$$A_\pi := \{\nabla c(\mu) | \mu \in \operatorname{supp}(\pi)\}. \tag{7}$$

Of course, if μ is an interior point, by differentiability of c, the definition of $\nabla c(\mu)$ is standard. On the other hand, for boundary points, we use the the one which is given in Section 2.1. That is, for each μ in the boundary of $\Delta(\Omega)$ with $\partial c(\mu) \neq \emptyset$, we take a sequence of full-support beliefs $(\mu_k)_{k=1}^\infty$, and we define $\nabla c(\mu) := \lim_{k\to\infty} \nabla c(\mu_k)$. Then, we define the act $f_\pi^b \in \mathbb{R}^\Omega$ by

$$f_\pi^b(\omega) := b(\omega) - \min_{t \in A_\pi} t(\omega) \tag{8}$$

Notice that by continuity of ∇c together with the fact that $\operatorname{supp}(\pi)$ is compact, it follows that A_π is compact too. Hence, the minimum in the previous definition always exists, and a fortiori f_π^b is well-defined. Finally, we are ready to introduce the subset of acts

$$A_\pi^b := A_\pi + f_\pi^b. \tag{9}$$

Obviously, A_π^b is compact, and therefore it is a menu in $\mathcal{A}$.

Theorem 3. *For any signal $\pi \in \Pi^1(\mu_P)$ and liability constraint $b \in \mathbb{R}^\Omega$, we have*

$$\mathbb{E}_{\pi'}(\phi_{A_\pi^b}) \leq \mathbb{E}_{\pi'}(\phi_A), \tag{10}$$

for all $A \in \mathcal{A}$ that implement π, and all optimal signals $\pi' \in \Pi_A(\mu_P)$.

Corollary 1. *For any signal $\pi \in \Pi^2(\mu_P)$ and liability constraint $b \in \mathbb{R}^\Omega$, we have*

$$\mathbb{E}_{\pi}(\phi_{A_\pi^b}) \leq \mathbb{E}_{\pi}(\phi_A), \tag{11}$$

for all $A \in \mathcal{A}$ that strictly implement π.

The proof of the previous corollary is omitted as it follows trivially from the preceding theorem, by simply noticing that π is the only optimal signal for any menu that strictly implements π itself. Hence, the last quantifier in Theorem 3 is trivially satisfied.

Let us illustrate the previous result in the context of the example in Figure 2. Suppose that we want to strictly implement signal π with $\text{supp}(\pi) = \{\mu_L, \mu_H\}$ in the cheapest possible way subject to the liability constraint imposed by $b = \mathbf{0}$. As we have already discussed in the previous section, the two ingredients of a menu that strictly implements π is a collection of tangents of c and a hyperplane L_A that is added to each of these tangents in order to satisfy the liability constraint. Starting with the first ingredient, notice that there is a unique subtangent at μ_H, so the only question is which of the infinitely many subtangents at μ_L we will pick. According to our last result, we will pick the flattest one, i.e., t_L instead of t'_L. In order to understand why, we need to look at the second ingredient of the cheapest menu, according to which we need to push them both upwards (by adding the same linear function) until they become non-negative. Clearly, if we have already picked t_L we are already closer to the horizontal axis than if we had picked t'_L, and as a result the entire menu becomes cheaper.

3.2. Cost of proxy attention

Henceforth, for each $\pi \in \Pi^1(\mu_P)$ and each liability constraint $b \in \mathbb{R}^\Omega$, we define the *cost for proxy attention (abbrev., proxy cost)* of signal π by

$$K(\pi) := \mathbb{E}_\pi(\phi_{A^b_\pi}), \tag{12}$$

The reason we refer to $K(\pi)$ as the cost of proxy attention because this is the (minimum) cost that the principal must pay in order to contract an agent who acts as a proxy in acquiring information and reporting the posterior.

The previous formula holds also for signals that can be strictly implemented. It is not difficult to show that the proxy cost of a signal has the following elegant characterization.

Theorem 4. *For a liability constraint $b \in \mathbb{R}^\Omega$, the proxy cost of signal $\pi \in \Pi^1(\mu_P)$ is given by*

$$K(\pi) = \mathbb{E}_\pi(c) + \mathbb{E}_{\mu_P}(f^b_\pi). \tag{13}$$

This last result provides a microeconomic foundation for the proxy cost of information, in the sense that $K(\pi)$ is the expert's willingness to accept for π under the constraints imposed by the primitives by the cost function c and the liability constraint b. Importantly, the supply side (viz., the expert's willingness to accept) is disentangled from the demand side (viz., the principal's willingness to pay for information). The later will be formally introduced down the stretch. For now, let us state some basic results that follow directly from the previous characterization result.

Corollary 2. *For a binary state space, K is strictly increasing in the Blackwell order.*

Remark 2. (Non-increasing proxy cost of information with respect to the Blackwell order). Surprisingly, the previous result does not carry over to state spaces with larger cardinality. The reason is that the posterior beliefs are not aligned with the extreme points of the simplex. As a result, the second term in (13) could be significantly higher for the less informative signal, thus making the function K non-increasing with respect to the Blackwell order. The following example illustrates such a case. $\triangleleft$

Example 1. Take the state space $\Omega = \{\omega_1, \omega_2, \omega_3\}$ together with the prior $\mu_P(\omega_1) = \mu_P(\omega_2) = \varepsilon/2$ and $\mu_P(\omega_3) = 1-\varepsilon$ for some $\varepsilon \in (0,1)$. Consider the perfectly informative signal π^*, as well as the signal π that puts probability $1-\varepsilon$ to $[\omega_3]$ and probability ε to $\mu_0 = (\frac{1}{2} \times \omega_1, \frac{1}{2} \times \omega_2)$. So, let us show that there exists a strictly convex function c such that $K(\pi) > K(\pi^*)$, although obviously π^* is strictly more informative than π. The way we proceed is by defining a hyperplane for each of the four relevant posteriors, viz., the three extreme points $[\omega_1], [\omega_2], [\omega_3]$ as well as μ_0. In particular, we have:

$$\begin{aligned}
t_{[\omega_1]}(\mu) &= 4\mu^1 + \mu^2 + 9\mu^3, \\
t_{[\omega_2]}(\mu) &= \mu^1 + 4\mu^2 + 9\mu^3, \\
t_{[\omega_3]}(\mu) &= \mu^1 + \mu^2 + 10\mu^3, \\
t_{\mu_0}(\mu) &= 3.
\end{aligned}$$

Note that for each of the four posteriors, the corresponding hyperplanes lies strictly higher compared to the other three. As a result, there exists a strictly convex function $c : \mathbb{R}^3 \to \mathbb{R}$, such that the respective hyperplane supports c at the corresponding posterior. An example, take the set of relevant beliefs $M = \{\mu_0, [\omega_1], [\omega_2], [\omega_3]\}$, and let the strictly convex cost function be given by

$$c(\mu) := \max_{\nu \in M} \left\{ \, t_\nu(\mu) + \|\mu - \nu\|^2 \, \right\} .$$

Then, using formula (13) with the no-liability condition (i.e., $b = \mathbf{0}$), we obtain

$$\begin{aligned}
K(\pi^*) &= \underbrace{4\varepsilon + 10(1-\varepsilon)}_{\mathbb{E}_{\pi^*}(c)} \underbrace{-\varepsilon - 9(1-\varepsilon)}_{\mathbb{E}_{\mu_P}(f^b_{\pi^*})} = 3\varepsilon + (1-\varepsilon), \\
K(\pi) &= \underbrace{3\varepsilon + 10(1-\varepsilon)}_{\mathbb{E}_{\pi}(c)} \underbrace{-\varepsilon - 3(1-\varepsilon)}_{\mathbb{E}_{\mu_P}(f^b_{\pi})} = 2\varepsilon + 7(1-\varepsilon).
\end{aligned}$$

Finally, it is not difficult to verify that for sufficiently small ε, we get $K(\pi^*) < K(\pi)$. ◁

Corollary 3. *K is not posterior-separable.*

Intuitively, this is because $\mathbb{E}_{\mu_P}(f^b_\pi)$ is equal to $\int_{\Delta(\Omega)} \mathbb{E}_\mu(f^b_\pi)\, d\pi$, and crucially $\mathbb{E}_\mu(f^b_\pi)$ does not depend only on μ, but also on the other posteriors in the support of π, i.e., the posteriors cannot be separated.[6]

The last two points (Remark 2 and Corollary 3) are quite interesting as they show that, from the principal's point of view, the cost of information is qualitatively very different when information acquisition is carried out by an expert (like in our model) compared to situations where it is carried out by the principal herself (like in standard models of information acquisition).

3.3. Welfare analysis

With the characterization of the proxy cost of a signal at hand, we can identify the expert's minimum surplus from procuring signal π. This is obviously equal to

$$S(\pi) := K(\pi) - C(\pi). \tag{14}$$

Then, it follows directly from Theorem 4 that $S(\pi)$ has the following elegant characterization.

[6]We are grateful to an anonymous referee for suggesting this point.

Corollary 4. *For signal $\pi \in \Pi^1(\mu_P)$ and liability constraint $b \in \mathbb{R}^\Omega$, the expert's guaranteed surplus is given by*

$$S(\pi) = \mathbb{E}_{\mu_P}(f_\pi^b) + c(\mu_P). \tag{15}$$

In principle, it is not surprising that the expert's expected surplus is always strictly positive: a similar conclusion is drawn in problems with limited liability (e.g., in principal-agent models with moral hazard and adverse selection).

Whenever the liability constraint is non-negative (i.e., $b \geq 0$), the expert's surplus is always positive in our case is because *the principal can verify neither the signal she is buying nor the realized posterior (i.e., information is unverifiable).* Thus, the principal compensates the lack of verifiability with incentive-compatibility. However, the incentives that are needed in order to guarantee that the expert will indeed choose the signal π are exogenously given and the principal cannot do anything to affect them. Thus, she always ends up paying a guaranteed premium to the expert for the traded signal, in relation to what she would have paid him if she could monitor his information-acquisition process and verify the posterior that he would have formed himself. Such conclusion also follows from the (similar) analysis of Rappoport and Somma (2017).

Corollary 5. *For a binary state space, S is strictly increasing in the Blackwell order.*

This last result follows again from the fact that $\mathbb{E}_{\mu_P}(f_\pi^b)$ is strictly increasing in the Blackwell order (see proof of Corollary 2). However, once again, for larger cardinality the result no longer holds. For instance, in Example 1, the expert's guaranteed surplus for signal π is larger than the guaranteed surplus for the perfectly informative signal.

3.4. Comparative statics

How sensitive is our analysis to changes in the main fundamental parameter of our model, the expert's cost function? In order to make such comparison, we focus on the family of cost functions

$$\mathcal{C} := \{\kappa c \mid \kappa > 0\}, \tag{16}$$

for some strictly convex c. For example let $\mathcal{C}$ be the class of entropic cost functions. One interpretation is that lower κ corresponds to cheaper information acquisition technology, and a fortiori to higher level of expertise.

Can we then conclude that higher level of expertise leads to a higher compensation for π? Or to higher guaranteed surplus for the expert? In other words, does the expert internalize the benefits of his higher expertise level? This turns out not be the case: for any given signal, increased competence of the expert leads to a lower price for the principal, as well as to a lower guaranteed surplus for the expert himself.

Corollary 6. *Within the class $\mathcal{C}$, for any signal $\pi \in \Pi^1(\mu_P)$, the following hold:*

(a) The proxy cost $K(\pi)$ is strictly increasing with respect to κ.

(b) The expert's guaranteed surplus $S(\pi)$ is strictly increasing with respect to κ.

To better understand the result, denote by $K_\kappa(\pi)$ and $S_\kappa(\pi)$ the proxy cost and the expert's guaranteed surplus for signal π given the parameter $\kappa > 0$. Then, by (8) and (13), we have

$$K_\kappa(\pi) = \mathbb{E}_{\mu_P}(b) + \kappa\Big(\mathbb{E}_\pi(c) - \mathbb{E}_{\mu_P}\big(\min_{t \in A_\pi} t\big)\Big), \tag{17}$$

where the second component that is multiplied by κ is strictly positive (by strict convexity of c), and therefore $K_\kappa(\pi)$ is linearly increasing in κ. Likewise, by (14), we have

$$S_\kappa(\pi) = \mathbb{E}_{\mu_P}(b) + \kappa\Big(c(\mu_P) - \mathbb{E}_{\mu_P}\big(\min_{t \in A_\pi} t\big)\Big),$$

where once again the second component that is multiplied by κ is strictly positive (by strict convexity of c), and therefore $S_\kappa(\pi)$ is linearly increasing in κ. Intuitively, the incentives induced by the variable payment become stronger, as c becomes "more convex".

4. Endogenized signals

So far, we have considered the signal π that the principal wants to acquire as an exogenously given primitive. Let us now relax this assumption by endogeneizing π. To this end, assume that there is a downstream choice to be made from a compact set $X \subseteq \mathbb{R}^\Omega$, after signal π is realized and the expert having picked some action a from the menu that was offered to him by the principal. This setting naturally incorporates two classes of models:

- Delegated information acquisition: the principal chooses from X herself (e.g., Carroll, 2019; Deimen and Szalay, 2019).
- Delegated persuasion: the choice is made by a third agent, viz., the receiver. This setting is similar to costly Bayesian persuasion (Gentzkow and Kamenica, 2014; Matysková and Montes, 2023), with the difference being that information acquisition is not carried out by the sender herself, but rather by an expert.

In either case, the crucial assumption is that the expert does not have any stakes in the eventual choice from X, and only cares about his own choice from A.

The principal's indirect payoff function is denoted by $\phi_X(\mu)$ for each $\mu \in \Delta(\Omega)$. In our first class of models, where the principal chooses himself, his indirect utility function is given by

$$\phi_X(\mu) := \max_{x \in X} \mathbb{E}_\mu(x). \tag{18}$$

In the second class of models the receiver will choose some $x_\mu \in \arg\max_{x \in X} \mathbb{E}_\mu(v(x))$, where $v : X \to \mathbb{R}^\Omega$ is the receiver's (continuous) payoff function. Aligned with the Bayesian persuasion literature, whenever the receiver has multiple optimal choices, we assume that she breaks the tie by picking the principal's preferred choice. As a result, the principal's indirect utility function becomes

$$\phi_X(\mu) := \mathbb{E}_\mu(x_\mu), \tag{19}$$

which, by our tie-breaking assumption, is upper semi-continuous.

Thus, for each signal $\pi \in \Pi^1(\mu_P)$, we can define the principal's value,

$$V_X(\pi) := \mathbb{E}_\pi(\phi_X) - K(\pi). \tag{20}$$

The principal's optimization problem boils down to maximizing V_X over $\Pi^1(\mu_P)$. Note that V_X is not defined for signals in $\Pi(\mu_P) \setminus \Pi^1(\mu_P)$, as such signals cannot be implemented, and therefore from the principal's point of view are not feasible. As a result the domain of V_X is not necessarily compact. Nevertheless, we can still establish existence of an optimal signal for the principal (Proposition A1 in the Appendix). But more interestingly, we can also prove that there is always an optimal signal for the principal which can be strictly implemented.

Theorem 5. *There exists some $\pi \in \Pi^2(\mu_P)$ such that $V_X(\pi) \geq V_X(\pi')$ for all $\pi' \in \Pi^1(\mu_P)$.*

A direct consequence of the previous result is that the complexity of the principal's optimal signal will be bounded by the cardinality of Ω, similarly to the well-known result from the Bayesian persuasion literature (Kamenica and Gentzkow, 2011). However, despite the striking conceptual similarity, our proof is significantly different. The reason is that Kamenica and Gentzkow (2011) rely on the fact that the principal's (viz., the sender's) indirect utility function is posterior separable, something which no longer holds in our case (Corollary 3).

This last result allows us to significantly simplify the computation of the principal's optimal signal, and in many cases to obtain tractable closed form solutions that allow us to do comparative statics.

5. Applications

In this section we will focus on some applications where our earlier characterization results yield specific testable hypotheses and explain certain stylized facts. Throughout the section, we consider a parsimonious model with a binary state space $\Omega = \{\omega_1, \omega_2\}$, a no-liability constraint $b = 0$, and an entropic cost function $c = H$ with multiplier parameter $\kappa = 1$ for the expert.

5.1. Decentralization of information acquisition

A principal is tasked to obtain N independent data points from the same experiment. There is a set of identical experts I, to whom data collection can be delegated. The question is how to optimally allocate the collection of data across these experts, i.e., should the principal centralize the information acquisition by asking few experts to obtain many data points each, or to decentralize it by asking many experts to obtain few data points each? This question is quite common in clinical trials where a pharmaceutical company must decide whether to delegate the data collection to few selected hospitals or to many distinct hospitals (e.g., Masri et al., 2012, and references therein).

Let us address this question in a parsimonious setting, with a uniform prior $\mu_P(\omega_1) = \mu_P(\omega_2) = 1/2$, and the data generating process being described by a binary symmetric channel with error probability $\varepsilon < 1/2$. The idea is that each data point will take the value ω_1 or ω_2, and it will coincide with the true state with probability $1 - \varepsilon$. A signal that yields $n \leq N$ independent observations from this binary symmetric signal is denoted by π_n.

Since the principal wants to obtain N independent data points in total, her problem comes down to deciding how many observations from each $i \in I$ to request from each expert. That is formally, the principal must request $n_i \in \{0, 1, \ldots, N\}$ data points from each $i \in I$, so that $\sum_{i \in I} n_i = N$. The question the becomes: which is the optimal allocation $(n_i)_{i \in I}$ of data points across experts, so that she minimizes her cost?

Proposition 1. *For sufficiently large noise parameter ε, it is the case that $K(\pi_n) > nK(\pi_1)$ for every $n > 1$. Hence, the principal strictly prefers to request a single data point from each one of N distinct experts.*

A very interesting implication of our result is that, in our context, superadditive cost for information can naturally emerge. This is the first model to the best of our knowledge with this feature, which is in contrast to other cost functions that are either subadditive (Shannon, 1948) or additive (Pomatto et al., 2023). Intuitively the driving force behind this difference is the fact that c becomes very steep as we approach the boundary. As a result, the cost of multiple signals coming from the same source increases non-linearly.

5.2. High-stakes decisions

Suppose that the principal wants to make a downstream decision from a set $X = \{(0,0),(x,-x)\}$ for some $x > 0$, as in Section 4. The strength of the principal's stakes are described by the magnitude of x. Suppose for instance that x is the size of an investment that the principal is considering to make, while ω_1 is the good state where the investment pays out and ω_2 is the bad state where it does not. Clearly, regardless how large the stakes are, the principal wants to invest if and only if the probability of the good state is larger or equal than 50%. We assume that the prior belief μ_P is uniformly distributed.

Suppose that it is uniformly more costly for the principal to produce a signal than for the expert in the sense that we defined in Section 3.4, i.e., the principal's cost parameter is $\tilde{\kappa} > \kappa = 1$. Let π be the principal's optimal signal when information acquisition is delegated to the expert (as in Section 4), and $\tilde{\pi}$ be the principal's optimal signal when she acquires information herself (as in a standard costly information acquisition setting). We say that the principal prefers to acquire information herself if and only if

$$V_X(\pi) \leq \mathbb{E}_{\tilde{\pi}}(\phi_X) - \mathbb{E}_{\tilde{\pi}}(\tilde{c}) + \tilde{c}(\mu_P), \tag{21}$$

where the left hand side is the principal's net expected payoff if she delegates information acquisition to an expert, whereas the right hand side is the principal's net expected payoff if she acquires her optimal signal herself.

Proposition 2. *For every $\tilde{\kappa} > \kappa$, there is some $x_0 > 0$, such that for every $x > x_0$ the principal prefers to acquire information herself.*

The previous result suggests that no matter how much more skilled the expert is in acquiring information, when the stake becomes large, the principal will always prefer to acquire information herself. Consider for instance a high-end R&D process: companies often outsource some of their activities, but when the stakes become large, they would rather keep it in house. Common wisdom suggests that this is because of intellectual property concerns. However, here we propose an alternative explanation: high stakes decisions often require very informative signals, the cost of which increases much faster when delegated to experts compared to when acquired by the principal herself. As a result, signals that are acquired via the expert are too expensive in such cases.

A. Proofs

Proof of Theorem 1. Part (a). (a) Necessity: Take an arbitrary menu $A \in \mathcal{A}$, and an optimal signal $\pi \in \Pi_A(\mu_P)$. Consider a linear functional $L_A : \Delta(\Omega) \to \mathbb{R}$ that supports the concave function $\overline{\psi}_A$ at μ_P. Then, by concavity of $\overline{\psi}_A$, it is the case that

$$L_A(\mu) \geq \overline{\psi}_A(\mu) \geq \psi_A(\mu) \tag{A.1}$$

for all $\mu \in \Delta(\Omega)$. By optimality of π, equalities hold at all $\tilde{\mu} \in \text{supp}(\pi)$. Hence, we have

$$c(\mu) \geq \phi_A(\mu) + L_A(\mu). \tag{A.2}$$

Finally, for any $\tilde{\mu} \in \text{supp}(\pi)$, pick an act $\tilde{a} \in \arg\max_{a\in A} \mathbb{E}_{\tilde{\mu}}(a)$. Thus, we get

$$c(\mu) \geq \mathbb{E}_{\mu}(\tilde{a}) + L_A(\mu), \tag{A.3}$$

at all $\mu \in \Delta(\Omega)$, with equality holding at $\tilde{\mu}$. Hence, c is subdifferentiable at $\tilde{\mu}$.

Sufficiency: Let c be subdifferentiable at every $\tilde{\mu} \in \text{supp}(\pi)$, and define the menu

$$A_\pi := \{\nabla c(\tilde{\mu}) | \tilde{\mu} \in \text{supp}(\pi)\}. \tag{A.4}$$

By ∇c being continuous and $\text{supp}(\pi)$ being closed, we have that A_π is compact, and therefore it is a well-defined menu. Moreover, by construction, we have $\phi_{A_\pi}(\mu) \leq c(\mu)$ with equality holding if and only if $\mu \in \text{supp}(\pi)$. As a result,

$$\mathbb{E}_{\pi'}(\psi_{A_\pi}) \leq 0, \tag{A.5}$$

with equality holding if and only if $\mu \in \text{supp}(\pi')$. Hence, we obtain

$$\mathbb{E}_{\pi'}(\psi_{A_\pi}) \leq \mathbb{E}_{\pi}(\psi_{A_\pi}) \tag{A.6}$$

meaning that π is implemented by A_π.

Part (b). Sufficiency: Take the menu A_π that we defined in (A.4), which implements π. Since the points in $\text{supp}(\pi)$ are linearly independent vectors in $\mathbb{R}^\Omega$, there is a unique convex combination of points in $\text{supp}(\pi)$ that yield the prior μ_P. Thus, the only signal in $\Pi(\mu_P)$ that puts probability 1 to $\text{supp}(\pi)$ is π itself, meaning that A_π strictly implements π.

Necessity: Let π be strictly implemented by some A. By Part (a) above, (P_1) holds. So, let us prove (P_2): we proceed by contradiction, starting with the assumption that the vectors in $\text{supp}(\pi)$ are not linearly independent. We distinguish two cases:

- All points in $\text{supp}(\pi)$ are extreme points in $\text{conv}(\text{supp}(\pi))$. This means that $|\text{supp}(\pi)| > |\Omega|$. But then, by Carathéodory's theorem, there exists a strict subset $M \subsetneq \text{supp}(\pi)$ such that $\mu_P \in \text{conv}(M)$.
- There exists some $\mu \in \text{supp}(\pi)$ which is not an extreme point in $\text{conv}(\text{supp}(\pi))$. Thus, if we take $M := \text{supp}(\pi) \setminus \{\mu\}$, it will be the case that $\mu_P \in \text{conv}(M)$.

So in either case, we can find a signal $\pi' \in \Pi(\mu_P)$ with $\text{supp}(\pi') = M \subsetneq \text{supp}(\pi)$. This signal will also be optimal, thus reaching a contradiction. □

Proof of Theorem 2. Part (a). Fix an arbitrary $\pi \in \Pi^1(\mu_P)$.

Sufficiency: Take a menu $A \in \mathcal{A}$ and a linear functional $L_A : \Delta(\Omega) \to \mathbb{R}$, such that the (convex) function $\phi_A - L_A$ supports c at $M \supseteq \text{supp}(\pi)$. Thus, we obtain $\phi_A(\mu) - L_A(\mu) - c(\mu) \leq 0$ for all $\mu \in \Delta(\Omega)$, with equality holding if and only $\mu \in M$. So, it is the case that

$$\mathbb{E}_{\pi'}(\psi_A) = \mathbb{E}_{\pi'}(\phi_A - c) \leq \mathbb{E}_{\pi'}(L_A) = L_A(\mu_P) \tag{A.7}$$

with equality holding if and only if $\text{supp}(\pi') \subseteq M$. Therefore, we have $\pi \in \Pi_A(\mu_P)$, i.e., π is implemented by A.

Necessity: Let π be implemented by A. Take a tangent L_A of $\overline{\psi}_A$ at μ_P. Recall from the proof of Theorem 1.(a) that $\phi_A(\mu) + L_A(\mu) \leq c(\mu)$, with equality holding at every $\mu \in \text{supp}(\pi)$. The latter implies that $\phi_A + L_A$ supports c.

Part (b). Fix an arbitrary $\pi \in \Pi^2(\mu_P)$.

SUFFICIENCY: Repeat the steps of Part (a) for $M = \text{supp}(\pi)$. Since there is no signal $\pi' \in \Pi^2(\mu_P) \setminus \{\pi\}$ with $\text{supp}(\pi') \subseteq M$, we conclude that menu A strictly implements π.

NECESSITY: Repeat the steps of Part (a), noticing that the only posteriors such that $\overline{\psi}_A(\mu) = \psi_A(\mu)$ are those in $\text{supp}(\pi)$, and therefore $\phi_A - L_A$ support c at $\text{supp}(\pi)$. $\square$

Proof of Theorem 3. Fix some $\pi \in \Pi^1(\mu_P)$. By the proof of Theorem 1, menu $A_\pi = \{\nabla c(\mu) | \mu \in \text{supp}(\pi)\}$ implements π. By Theorem 2, menu $A_\pi^b = A_\pi + f_\pi^b$ implements π.

Take another $A \in \mathcal{A}$ that satisfies the liability constraint b and implements π. Then, again by Theorem 2, there exists some $M \supseteq \text{supp}(\pi)$ and some $T_\mu \subseteq \partial c(\mu)$ for each $\mu \in M$, and some $f \in \mathbb{R}^\Omega$, such that menu A can be rewritten as

$$A = \{t + f | t \in T_\mu \text{ for some } \mu \in M\}. \tag{A.8}$$

We will now proceed in three steps. As a first step, take the alternative menu

$$A_1 = \{\nabla c(\mu) + f | \mu \in M\}. \tag{A.9}$$

that we obtain by replacing T_μ with the singleton $\{\nabla c(\mu)\}$ for each $\mu \in M$. By Theorem 2, the last two menus yield the same set of optimal signals, i.e., $\Pi_A(\mu_P) = \Pi_{A_1}(\mu_P)$. Moreover, observe that, for every $\mu \in M$, it is the case that $\phi_A(\mu) = \phi_{A_1}(\mu)$. As a result, for every $\pi' \in \Pi_A(\mu_P)$ it will be the case that

$$\mathbb{E}_{\pi'}(\phi_A) = \mathbb{E}_{\pi'}(\phi_{A_1}). \tag{A.10}$$

As a second step, remove from A_1 the tangents that correspond to posteriors that are not in $\text{supp}(\pi)$, thus obtaining the new menu

$$A_2 = \{\nabla c(\mu) + f | \mu \in \text{supp}(\pi)\}. \tag{A.11}$$

Since $A_2 \subseteq A_1$, it will be the case that $\phi_{A_2} \leq \phi_{A_1}$. Hence, for each $\pi' \in \Pi_A(\mu_P)$ we get

$$\mathbb{E}_{\pi'}(\phi_{A_2}) \leq \mathbb{E}_{\pi'}(\phi_A). \tag{A.12}$$

Finally, as a third step, we replace f with f_π^b in A_2, thus obtaining A_π^b. Observe that for any $\mu \in \text{supp}(\pi)$ at the boundary of $\Delta(\Omega)$, our definition of $\nabla c(\mu)$ implies $\mathbb{E}_{\tilde{\mu}}(t) \leq \mathbb{E}_{\tilde{\mu}}(\nabla c(\mu))$ for every $t \in \partial c(\mu)$, with equality holding if $\tilde{\mu} = \mu$. This is because c is strictly convex and $\nabla c(\mu)$ is the flattest among all tangents at μ. Thus, by setting $\tilde{\mu} := [\omega]$, we obtain

$$\begin{aligned} f_\pi^b(\omega) &= b(\omega) - \min_{t \in A_\pi} t(\omega) \\ &\leq b(\omega) - \min_{t \in A_1 - f} t(\omega) \\ &= b(\omega) - \min_{t \in A_1} t(\omega) + f(\omega) \\ &\leq f(\omega). \end{aligned}$$

The first equality is by definition of f_π^b, the first inequality is due to $A_\pi = A_2 - f \subseteq A_1 - f$, and the second inequality is due to A_1 satisfying the liability condition. Therefore, we have

$$\begin{aligned} \mathbb{E}_{\pi'}(\phi_{A_\pi^b}) &= \mathbb{E}_{\pi'}(\phi_{A_\pi}) + \mathbb{E}_{\mu_P}(f_\pi^b) \\ &= \mathbb{E}_{\pi'}(\phi_{A_2}) - \mathbb{E}_{\mu_P}(f) + \mathbb{E}_{\mu_P}(f_\pi^b) \\ &\leq \mathbb{E}_{\pi'}(\phi_{A_2}), \end{aligned}$$

which completes the proof. $\square$

Proof of Theorem 4. By result follows directly from the following chain of equalities

$$\begin{aligned}\mathbb{E}_\pi(\phi_{A_\pi^b}) &= \mathbb{E}_\pi(\phi_{A_\pi}) + \mathbb{E}_\pi(\phi_{\{f_\pi^b\}}) \\ &= \mathbb{E}_\pi(\phi_{A_\pi}) + \mathbb{E}_{\mu_P}(f_\pi^b) \\ &= \mathbb{E}_\pi(c) + \mathbb{E}_{\mu_P}(f_\pi^b),\end{aligned}$$

where the first equality follows from $\phi_{A_\pi^b} = \phi_{A_\pi} + \phi_{\{f_\pi^b\}}$, the second equality holds because $\phi_{\{f_\pi^b\}}$ is linear and π is a mean-preserving distribution, and the third inequality holds because ϕ_{A_π} supports c at $\text{supp}(\pi)$. □

Proof of Corollary 2. Take any two signals such that $\pi \succ \pi'$, and let $\{\mu_L^1, \mu_H^1\}$ be the probabilities assigned to ω_1 by the two extreme posteriors in the support of π. Since the state space is binary and $\pi \succ \pi'$, it follows that for every $\mu \in \text{supp}(\pi')$ it will be the case that $\mu_L^1 \leq \mu^1 \leq \mu_H^1$. Hence, by Aliprantis and Border (1994, Thm 7.22), we obtain $\mathbb{E}_{\mu_P}(f_\pi^b) \geq \mathbb{E}_{\mu_P}(f_{\pi'}^b)$. Moreover, by Blackwell's theorem, we have $\mathbb{E}_\pi(c) > \mathbb{E}_{\pi'}(c)$. Plugging the two inequalities into (13) directly yields $K(\pi) > K(\pi')$. □

Proof of Corollary 3. The proof is based on the observation that K does not satisfy dynamic consistency (Tsakas, 2020). In particular, in a binary state space, consider the following compound signal: first take a signal $\tilde{\pi}$ with $\text{supp}(\tilde{\pi}) = \{\tilde{\mu}_L, \tilde{\mu}_H\}$, and then given each posterior $\tilde{\mu}_k \in \text{supp}(\tilde{\pi})$ take a signal $\pi_k \in \Pi(\tilde{\mu}_k)$ with $\text{supp}(\pi_k) = \{\mu_L, \mu_L\}$, where $\mu_L^1 < \tilde{\mu}_L^1 < \mu_P^1 < \tilde{\mu}_H^1 < \mu_H^1$. This compound signal is equivalent to $\pi \in \Pi(\mu_P)$ in terms of the total probability that it attaches to each posterior, i.e.,

$$\pi(\cdot) = \tilde{\pi}(\tilde{\mu}_L)\pi_L(\cdot) + \tilde{\pi}(\tilde{\mu}_H)\pi_H(\cdot).$$

Nevertheless, since π, π_L and π_H have the same support, they will share the same cheapest menu, and therefore we obtain $f_\pi^b = f_{\pi_L}^b = f_{\pi_H}^b$. As a result, we will have

$$\begin{aligned}K(\pi) &= \mathbb{E}_\pi(c) + \mathbb{E}_{\mu_P}(f_\pi^b) \\ &= \tilde{\pi}(\tilde{\mu}_L)\big(\mathbb{E}_{\pi_L}(c) + \mathbb{E}_{\tilde{\mu}_L}(f_{\pi_L}^b)\big) + \tilde{\pi}(\tilde{\mu}_H)\big(\mathbb{E}_{\pi_H}(c) + \mathbb{E}_{\tilde{\mu}_H}(f_{\pi_H}^b)\big) \\ &= \tilde{\pi}(\tilde{\mu}_L)K(\pi_L) + \tilde{\pi}(\tilde{\mu}_H)K(\pi_H) \\ &< \tilde{\pi}(\tilde{\mu}_L)K(\pi_L) + \tilde{\pi}(\tilde{\mu}_H)K(\pi_H) + K(\tilde{\pi}).\end{aligned}$$

This implies that although the compound signal is equivalent with the reduced one-stage signal, their respective costs differ. □

Proposition A1. *The function V_X achieves a maximum in $\Pi^1(\mu_P)$.*

Proof. First note that V_X is continuous in $\Pi^1(\mu_P)$. Then, take the continuous extension $V_X : \Pi(\mu_P) \to [-\infty, \infty)$, which is defined as follows: for each $\pi \in \Pi(\mu_P) \setminus \Pi^1(\mu_P)$, take a sequence of signals $(\pi_k)_{k=1}^\infty$ in $\Pi^1(\mu_P)$ such that $\pi_k \to \pi$, and subsequently set

$$V_X(\pi) = \lim_{k\to\infty} V_X(\pi_k) = \mathbb{E}_\pi(\phi_X - c) - \lim_{k\to\infty} \mathbb{E}_{\mu_P}(f_{\pi_k}^b),$$

which (by Theorem 4) implies $V_X(\pi) = -\infty$. Then, by a standard compactness-plus-continuity argument, we conclude that the extended function V_X achieves a maximum in $\Pi(\mu_P)$. However, since $V_X(\pi) > -\infty$ if and only if $\pi \in \Pi^1(\mu_P)$, the maximum is achieved in $\Pi^1(\mu_P)$. □

Proof of Theorem 5. Suppose $\pi' \in \Pi^1(\mu_P)$ is a maximizer of V_X, which by Proposition A1 always exists. Using Theorem 4, it is the case that

$$V_X(\pi') = \mathbb{E}_{\pi'}(\phi_X - c) - \mathbb{E}_{\mu_P}(f^b_{\pi'}). \tag{A.13}$$

Then, there exists some $\pi \in \Pi^2(\mu_P)$ with $\text{supp}(\pi) \subseteq \text{supp}(\pi')$, such that

$$\mathbb{E}_{\pi}(\phi_X - c) \geq \mathbb{E}_{\pi'}(\phi_X - c). \tag{A.14}$$

The latter follows from the fact that there exists some $M \subseteq \text{supp}(\pi')$ satisfying (P_2), and some signal π with $\text{supp}(\pi) = M$ such that (A.14) is satisfied. At the same time, since $\text{supp}(\pi) \subseteq \text{supp}(\pi')$, it will also be the case that $A_\pi \subseteq A_{\pi'}$, and therefore by Equation (8) we obtain $f^b_\pi \leq f^b_{\pi'}$, and a fortiori

$$\mathbb{E}_{\mu_P}(f^b_\pi) \leq \mathbb{E}_{\mu_P}(f^b_{\pi'}). \tag{A.15}$$

Finally, putting (A.14) and (A.15) together yields $V_X(\pi) \geq V_X(\pi')$. $\square$

Proof of Proposition 1. Given the signal π_n, let

$$\mu^n_k := \frac{(1-\varepsilon)^k \varepsilon^{n-k}}{\varepsilon^k (1-\varepsilon)^{n-k} + (1-\varepsilon)^k \varepsilon^{n-k}}, \tag{A.16}$$

be the posterior probability attached to ω_1 if the realized dataset contains k times ω_1 and $n-k$ times ω_2. Using Equation (13), we obtain

$$\begin{aligned} K(\pi_n) &= \mathbb{E}_\pi(c) + \mathbb{E}_{\mu_P}(f^b_\pi) \\ &> c(\mu_P) - \log(\mu^n_0) \\ &= \log\frac{1}{2} - \log\frac{\varepsilon^n}{\varepsilon^n + (1-\varepsilon)^n}. \end{aligned}$$

Applying again Equation (13) for $n = 1$, we obtain

$$K(\pi_1) = (1-\varepsilon)\log\frac{1-\varepsilon}{\varepsilon}. \tag{A.17}$$

Now, let us define the difference

$$\Delta(\varepsilon) := K(\pi_n) - nK(\pi_1), \tag{A.18}$$

and observe that as ε becomes large, we will have

$$\lim_{\varepsilon \to 1/2} \Delta(\varepsilon) > 0. \tag{A.19}$$

Hence, there is a neighborhood of ε close to 1/2, where $K(\pi_n) > nK(\pi_1)$. $\square$

Proof of Proposition 2. By Theorem 5, signal π puts positive probability to two symmetric posteriors, that put probability μ^1_x and $1 - \mu^1_x$ to ω_1 respectively, for some $\mu^1_x > 1/2$. Hence, the principal's optimal signal (when she delegates information acquisition to the expert), is the maximizer of the function:

$$V_X(\pi) := \underbrace{\mu_x^1 x}_{\mathbb{E}_\pi(\phi_X)} \underbrace{- \mu_x^1 \log \frac{\mu_x^1}{1-\mu_x^1}}_{K(\pi)}. \tag{A.20}$$

Taking first order condition yields

$$x = 1 + \frac{\mu_x^1}{1-\mu_x^1} + \log \frac{\mu_x^1}{1-\mu_x^1}. \tag{A.21}$$

Hence, $\lim_{x\to\infty} \mu_x^1 = 1$. So, there exists some $x_0 > 0$ such that for all $x > x_0$,

$$K(\pi) > \tilde{\kappa} \log 2 \geq \mathbb{E}_\pi(\tilde{c}) - \tilde{c}(\mu_P). \tag{A.22}$$

Finally, it is obviously the case that

$$\begin{aligned} V_X(\pi) &= \mathbb{E}_\pi(\phi_X) - K(\pi) \\ &< \mathbb{E}_\pi(\phi_X) - \mathbb{E}_\pi(\tilde{c}) + \tilde{c}(\mu_P) \\ &\leq \mathbb{E}_{\tilde{\pi}}(\phi_X) - \mathbb{E}_{\tilde{\pi}}(\tilde{c}) + \tilde{c}(\mu_P) \end{aligned}$$

which completes the proof. □

References

Alaoui, L. & Penta, A. (2016). Endogenous depth of reasoning. *Review of Economic Studies* 83, 1297–1333.

——— (2022). Cost-benefit analysis in reasoning. *Journal of Political Economy* 130(4), 881–925

Alaoui, L., Janezic, K. & Penta, A. (2020). Reasoning about others' reasoning. *Journal of Economic Theory* 189, 105091.

Ali, N., Chen-Zion, A. & Lillethun, E. (2022). Reselling information. *Working Paper.*

Aliprantis, C. & Border, K. (1994). *Infinite dimensional analysis.* Springer Verlag, Berlin.

Argenziano, R., Severinov, S. & Squintani, F. (2016). Strategic information acquisition and transmission. *American Economic Journal: Microeconomics* 8, 119–155.

Arora, A. & Fosfuri, A. (2005). Pricing diagnostic information. *Management Science* 51, 1092–1100.

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. *The Rate and Direction of Inventive Activity: Economic and Social Factors, National Bureau of Economic Research* 609–625.

——— (1963). Uncertainty and the welfare economics of medical care. *American Economic Review* 53, 941–973.

Aumann, R. & Maschler, M. (1995). *Repeated games with incomplete information.* MIT Press.

Babaioff, M., Kleinberg, R. & Paes Leme, R. (2012). Optimal mechanisms for selling information. *Proceedings of the 13th ACM Conference on Electronic Commerce*, 92–109.

Bergemann, D. & Bonatti, A. (2019). Markets for information: an introduction. *Annual Review of Economics* 11, 1–23.

Bergemann, D., Bonatti, A. & Smolin, A. (2018). The design and price of information. *American Economic Review* 108, 1–48.

Bergemann, D. & Ottaviani, M. (2021). Information markets and nonmarkets. *Handbook of industrial organization* Vol. 4, No. 1, 593-672

Bizzotto, J., Perez-Richet, E. & Vigier, A. (2020). Information design with agency. *Working Paper.*

Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics* 24, 265–272.

Bloedel, A.W. & Zhong,W. (2021). The cost of optimally-acquired information. *Working Paper*

Bolton, P. & Dewatripont, M. (2005). *Contract theory.* MIT Press.

Caplin, A. (2016). Measuring and modeling attention. *Annual Review of Economics* 8, 379–403.

Caplin, A. & Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review* 105, 2183–2203.

Caplin, A., Dean, M. & Leahy, J. (2022). Rationally inattentive behavior: characterizing and generalizing Shannon entropy. *Journal of Political Economy.*

Carroll, G. (2019). Robust incentives for information acquisition. *Journal of Economic Theory* 181, 382–420.

Chang, C. & Lee, C. (1994). Optimal pricing strategy in marketing research consulting. *International Economic Review* 35, 463–478.

Coase, R. (1974). The market for goods and the market for ideas. *American Economic Review: Papers and Proceedings* 64, 384–391.

De Oliveira, H., Denti, T., Mihm, M. & Ozbek, K. (2017). Rationally inattentive preferences and hidden information costs. *Theoretical Economics* 12, 621–654.

Dean, M. & Neligh, N. (2023). Experimental tests of rational inattention. *Journal of Political Economy* 131.

Deimen, I. & Szalay, D. (2019). Delegated expertise, authority, and communication. *American Economic Review* 109, 1349–1374.

Denti, T. (2022). Posterior-separable cost of information. *American Economic Review* 112, 3215–3259

Denti, T., Marinacci, M. & Rustichini, A. (2022). Experimental cost of information. *American Economic Review* 112, 3106-3123.

Dulleck, U. & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: the economics of credence goods. *Journal of Economic Literature* 44, 5–42.

Ellis, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory* 73, 56–94.

Esö, P. & Szentes, B. (2007). The price of advice. *RAND Journal of Economics* 38, 863–880.

Financial Times (2010). Pay the rating agencies according to results. June 3, 2010.

Gentzkow, M. & Kamenica, E. (2014). Costly persuasion. *American Economic Review* 104, 457–462.

Haghpanah, N., Ali, N., Lin, X. & Siegel, R. (2022). How to sell hard information. *The Quarterly Journal of Economics* 137, 619-678.

Hu, L., Li, A. & Segal I. (2023). The politics of personalized news aggregation. *Journal of Political Economy Microeconomics* 1, 463-505.

Hörner, J. & Skrzypacz, A. (2016). Selling information. *Journal of Political Economy* 124, 1515–1562.

Iyer, G. & Soberman, D. (2000). Markets for product modification information. *Marketing Science* 19, 203–225.

Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review* 101, 2590–2615.

Kashyap, A. & Kovrijnykh, N. (2016). Who should pay for credit rating and how? *Review of Financial Studies* 29, 420–456.

Laffont, J.J. & Martimort, D. (2002). *The theory of incentives: the principal-agent model.* Princeton University Press.

Laffont, J.J. & Tirole,J. (1986). Using cost observation to regulate firms. *Journal of Political Economy* 94, 614–641.

Li, A.& Hu, L. (2023). Electoral accountability with personalized news aggregation. *Games and Economic Behavior* 140, 296-315.

Lindbeck, A. & Weibull, J. (2020). Delegation of investment decisions, and optimal renumeration of agents. *European Economic Review* 129, 103559.

Lipnowski, E., Mathevet, L. & Wei, D. (2020). Attention management. *American Economic Review: Insights* 2, 17–32.

Maćkowiak, B., Matějka, F. & Wiederholt, M. (2023). Rational inattention: a review. *Journal of Economic Literature* 61, 226-273.

Masri, M.D., Ramirez, B., Popescu, C. & Reggie, E.M. (2012). Contract research organizations: an industry analysis. *International Journal of Pharmaceutical and Healthcare Marketing* 6, 336–350.

Matysková, L. & Montes, A. (2023). Bayesian persuasion with costly information acquisition. *Journal of Economic Theory*, 105678.

Morris, S. & Strack, P. (2019). The wald problem and the relation of sequential sampling and ex-ante information costs. *Working Paper*

Pomatto, L., Strack, P. & Tamuz, O. (2023). The cost of information: The case of constant marginal costs. *American Economic Review* 112, 1360-1393.

Rappoport, D., Somma,V. (2017). Incentivizing information design. *Working Paper.*

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423, 623– 656.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50, 665–690.

——— (2006). Rational inattention: beyond the linear-quadratic case. *American Economic Review* 96, 158–163.

Tsakas, E. (2020). Robust scoring rules. *Theoretical Economics* 15, 955–987.

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

Winkler, G. (1988). Extreme points of moment sets. *Mathematics of Operations Research* 13, 581–587.

Yoder, N. (2022). Designing incentives for heterogeneous researchers. *Journal of Political Economy* 130, 2018–2054.

Zermeño, L. (2011). A principal-expert model and the value of menus. *Working Paper.*

——— (2012). Essays on the principal-expert problem. *PhD Dissertation, MIT.*

Zhong, W. (2022). Optimal dynamic information acquisition. *Econometrica* 90, 1537–1582 .