Strategic Obfuscation: Influence Through Information Costs

Luis Henrique Linhares Daniel Monte*

August 1, 2026

Abstract

We develop a model of strategic obfuscation in which an interested Principal chooses the cost for a rationally inattentive Decision Maker to process information, making learning costs endogenous. We characterize when the Principal chooses transparency, total obfuscation (no-learning), or partial learning. Endogenous information costs can lead to a counterintuitive implication: increasing the Decision Maker's cost of mistakes can improve her resulting welfare as it leads the Principal to favor more transparency. The framework links persuasion and rational inattention, shifting attention from commitment in disclosure to the design of information costs. The model rationalizes fine print, product complexity, and has implications for disclosure and transparency policy.

Keywords: strategic obfuscation, rational inattention, costly information
JEL Codes: D11, D82, D83

*Linhares: Boston University. E-mail: lhrlinha@bu.edu. Monte: Sao Paulo School of Economics - FGV. E-mail: daniel.monte@fgv.br. We are very thankful for conversations with Henrique de Oliveira, Lucas Maestri, Chiara Margaria, Laurent Mathevet, Jawwad Noor, Juan Ortner, Kareen Rozen, and Adrien Vigier.

1 Introduction

Obfuscation, by which we mean the conscious effort to hinder the collection and processing of information, is ubiquitous. Lawyers required to disclose evidence may bury the relevant material amid a large volume of irrelevant documents. Terms and conditions are long and confusing, so reading them before accepting is the exception, not the rule. Direct product comparison is made harder by the offer of products of different sizes and flavors than competitors (Richards et al., 2020).

We study a setting in which a Principal wishes to induce a Decision Maker (DM) to take a certain action but cannot rely on direct information transmission or on monetary transfers. Instead, the Principal controls how difficult it is for the DM to learn about an exogenous and payoff relevant state of the world that is unknown to both parties. We provide a framework that endogenizes this cost of learning, and characterize the resulting equilibrium degree of transparency.

In a binary actions and states setting, our framework yields a full characterization of when the Principal chooses transparency, total obfuscation, or something in between. Surprisingly enough, even in this simple setting, we show that there are cases where the Principal has incentives to partially obfuscate, inducing the optimal probability of DM's mistakes. Such partial obfuscation is optimal exactly when the DM is skeptical: she rejects the Principal's preferred action under the prior, even though the belief required to persuade her is relatively low. One counterintuitive implication is that policies designed to protect the DM from such costly mistakes can increase the Principal's incentive to obfuscate, so that removing such protections leaves the DM better off.

Formally, the Principal sets the marginal cost of learning in order to influence a rationally inattentive (Sims, 2003) DM's decision. The DM weighs the cost of learning, such as the time, effort, or search costs, against the benefit of avoiding mistakes. Mathematically, the Principal chooses a parameter $\lambda \geqslant 0$ that scales the cost of reducing uncertainty (Shannon, 1948): a lower λ facilitates learning, while a high λ makes information acquisition expensive. By setting $\lambda = 0$, the Principal effectively lets the DM correctly identify the state of the world. By setting a sufficiently large λ, think of $\lambda \to \infty$, the Principal makes any information processing prohibitively costly, and the DM must base her decisions on the common prior alone.

The Principal's efforts to raise costs, or increase the value of λ, can be interpreted in different ways: restricting access to information, decreasing the time available to make a decision, or simply dumping irrelevant data.[1] Unlike other obfuscation papers, we do not

[1]Hiding or providing information may be costly to the Principal. In our model, we abstract from such costs. Online Appendix B briefly discusses an alternative model in which the Principal (in this case, a firm)

focus on a single mechanism. Instead, we generalize how this behavior can appear when facing a rationally inattentive DM.

A key contribution of our paper is to focus on the design of which information can feasibly be learned.[2] As such, we provide an approach to persuasion that does not require a strong commitment assumption.[3] Instead, we assume the cost of information is chosen before the Principal observes the state of the world and thus, the chosen λ is not a signal in itself. We go over a comparison of Strategic Obfuscation with Bayesian Persuasion and Cheap Talk in Section 3.1.

To illustrate our framework, consider the following example in which the Principal is a firm and the DM is a consumer. The firm is attempting to sell a product of uncertain quality. The product may be *Good*, in which case a consumer would enjoy a positive benefit from purchasing it, or *Bad*, in which case consumers would regret the purchase. Both the firm and the consumer share a common prior belief that the product is Good with probability $\mu_0 \in (0, 1)$. In contrast to the consumer whose payoff depends on the realized quality, the firm always benefits from a sale. The payoff structure is summarized as

	not buy	buy
Good	$0, 0$	$1, B$
Bad	$0, 0$	$1, -C$.

Here, the first entry in each cell represents the firm's payoff, while the second entry represents the consumer's payoff. The parameters satisfy $B > C > 0$, so that buying is desirable only when the product is Good. From the consumer's perspective, the decision can be summarized by a threshold belief: if the probability μ_0 that the product is Good exceeds a belief of indifference

$$\mu_{ind} = \frac{C}{B + C},$$

then the expected benefit of buying is positive, and the consumer will purchase if no new information is obtained. In this case, the firm's best strategy is to fully obfuscate, making

incurs a cost when choosing λ. Adding these costs, however, does not alter our main conclusions.

[2]Thus, our model contrasts with the Bayesian Persuasion framework (see Kamenica and Gentzkow (2011) but also Kamenica (2019) and Bergemann and Morris (2019) for comprehensive surveys), in which a Sender commits to an information disclosure policy. In our paper, the Principal (essentially) chooses how the DM will learn, within the rational inattention framework. Indeed, Kamenica and Gentzkow (2011) use commitment in the information acquisition stage instead of commitment in the information transmission stage as a possible motivation for the commitment assumption in Bayesian Persuasion.

[3]We comment further on this assumption when presenting the model and describing practical examples. Martin (2017) and De Clippel and Rozen (2025) present models in which there is a strategic component to the fact that an agent is choosing to obfuscate. In settings where this assumption does not hold, we interpret our results as what the Principal could achieve if it could commit to an obfuscation level, and therefore provide an upper bound to the Principal's payoffs when compared to no commitment.

additional information prohibitively costly, so that the consumer bases her decision solely on the favorable prior. This is the sales adage "don't talk yourself out of a sale".[4]

If the prior μ_0 falls below this threshold, the consumer would not buy without further investigation. The firm faces the following problem. If it provides full transparency, it will enable the consumer to perfectly observe whether the product is Good or Bad. This maximizes efficiency but yields sales only when the product is indeed Good, which may be unattractive to the firm when the prior is low. On the other hand, the firm may release information in a way that is costly to process. By raising the marginal cost of learning, the firm induces the consumer to remain only partially informed. If this cost is appropriately chosen, the consumer buys more frequently than under full transparency.

Consider now a product replacement or recycling policy that cushions losses when the product is Bad, effectively lowering C. In that case, the consumer is more willing to buy under partial information, and the firm benefits from maintaining some degree of obfuscation. Eliminating such a policy increases C. This raises the threshold μ_{ind}, making the consumer more reluctant to buy without clear evidence. The higher stakes from bad outcomes can push the firm to choose full transparency, since obfuscation no longer yields profitable sales. Our results go even further. Our Theorem 2 shows when the consumer's equilibrium welfare is weakly increasing in higher costs of mistakes.

Our results thus carry direct policy implications. Since obfuscation raises the Principal's payoff at the DM's expense, understanding when it arises helps identify policies that improve DM welfare. Policies that seem to reduce consumer payoffs in some states may, in fact, increase overall welfare by inducing firms to simplify disclosure and reveal more information. This is not a general prescription to make mistakes more costly. Rather, it illustrates why interventions must account for the Principal's endogenous response through the information environment. We discuss this example in greater detail in Online Appendix A.

These results highlight the central role of strategic obfuscation. There are cases in which partial information can dominate both full transparency (too few sales) and full secrecy (no sales) for the Principal. Our framework formalizes these trade-offs and provides techniques that allow us to (i) compute exactly how costly the Principal will choose information processing to be; (ii) understand how the DM chooses how much to learn given that cost; and (iii) perform comparative statics on policies aimed at improving welfare.

Finally, we present an example linking obfuscation to consideration sets. When the Principal chooses how much to obfuscate, she indirectly chooses the size of the DM's consideration set. Under full transparency, the DM considers every option; under full obfuscation, she picks the single option that is optimal given the prior alone. Partial

[4] `https://www.inc.com/jim-schleckser/when-you-get-sale-stop-talking.html`

obfuscation can expand the consideration set beyond this singleton, without necessarily including every option. In Section 4.4, we reinterpret an example from Caplin et al. (2019) in our setting, where the DM seeks the only good option among nine bad ones. One result is that expanding the consideration set can benefit the Principal even beyond the set that already contains her preferred option, by adding options she prefers less still. The prospect of reaching those additional options is what makes further experimentation worthwhile.

As the rational inattention literature has moved beyond Shannon Entropy toward more general information-cost structures, we examine which of our results hold under the posterior-separable cost functions studied in Caplin et al. (2022). In particular, we show that partial obfuscation is not idiosyncratic to Shannon Entropy. It is locally robust to the choice of the information cost function.

The rest of the paper is organized as follows. In Section 2, we present a binary model (two actions and two states of the world) in which the Principal has state-independent preferences. Section 3 provides a full characterization of this problem. Theorem 1 shows when the Principal's solution is partial obfuscation, full obfuscation, or full transparency. Theorem 2 presents a result that relates to policy design: whenever the Principal's solution involves partial or full obfuscation, the DM can be made strictly better off by making her own payoff worse in the bad state. The result is counterintuitive because it works through a strategic channel: the Principal responds to the DM's lowered payoff by increasing transparency. In Section 4, we generalize our model to multiple actions and states, allowing the Principal to have state-dependent preferences. Section 5 discusses which of our results can be generalized to other information cost functions. Section 6 concludes. Most proofs are relegated to the Appendix.

Related literature:

Information Design and the commitment assumption - Unlike canonical Bayesian Persuasion models (Kamenica and Gentzkow (2011), Bergemann and Morris (2019), for example), our framework dispenses with commitment in information transmission. In this sense, our paper is related to the works by Lin and Liu (2024), Barros (2026), Lipnowski et al. (2022), Min (2021), Fréchette et al. (2022), Nguyen and Tan (2021), Perez-Richet and Skreta (2022) and Alonso and Camara (2024). Also related are Edmond and Lu (2021), who add confusion to the communication model by making the Sender capable of confusing the DM at a cost, and Bizzotto et al. (2020), who study the case of persuasion when dealing with certifications and add a cost for obtaining information.

Rational Inattention - Particularly useful to our analysis is the branch that developed the foundations for solving the problems of Rationally Inattentive agents (Matějka and McKay, 2015; De Oliveira et al., 2017; Caplin et al., 2019, 2022; Dean and Neligh, 2023). Ellis (2018)

shows an axiomatic model of choice behavior for an agent with limited attention and gives the conditions under which the behavior can be considered the result of a choice of an optimal level of inattention. Following this literature, a growing body of work has applied rational inattention to industrial organization. Examples include Matějka and McKay (2012), Matějka (2016), and Boyacı and Akçay (2018). More recently, Janssen and Kasinger (2024) analyze a duopoly in which firms can obfuscate prices before competing for rationally inattentive consumers, while Cusumano et al. (2024) study competition between two firms selling to a rationally inattentive consumer. Asriyan et al. (2023) show how buyers' imperfect observation of product quality affects the complexity chosen by product designers.

Information Design and Rational Inattention - A few authors have recently combined the study of persuasion with Rational Inattention. Gentzkow and Kamenica (2014) provide a direct extension of the classic Bayesian Persuasion model in which the sender has to buy the signal at a cost given by Shannon Entropy. Matysková and Montes (2023) develop a model in which the Sender sends information at no cost. However, upon seeing a signal realization, the DM may buy further information at a cost. Bloedel and Segal (2021) detail how to optimally persuade a rationally inattentive receiver, as the usual concavification results are not attainable. Other related papers are by Lipnowski et al. (2020), Le Treust and Tomala (2019), Fabbri (2024), and Wei (2020). We differ from these papers mainly in that our information cost is endogenous, which allows for analysis of strategic choice of learning costs.

Strategic Obfuscation - A growing literature on obfuscation—typically in industrial organization contexts—motivates our economic problem. See, for instance, Gabaix and Laibson (2006); Ellison (2005), Ellison and Wolitzky (2012), Kalayci and Potters (2011), Jin et al. (2022), Ben-Shahar and Schneider (2014), and Nocke and Rey (2024). More recently, De Clippel and Rozen (2025) created and empirically tested a model of communication in which obfuscation intensity is endogenous and may appear in settings with mandatory disclosure. They add strategic inference from the agent, i.e., the agent sees obfuscating behavior as a signal and draws inference from that. In contrast, our results characterize outcomes when the cost of information is chosen before the Principal learns the state, so that DMs make no inference from obfuscation per se.

2 The 2x2 Model

In this section, we focus our attention on a simplified class of problems in which there are two states, two actions, and the Principal has state-independent preferences. This allows us to quickly explain how the model works and more directly introduce our main results. We

generalize our model in Section 4.

There are two players, the Principal (*it*) and the Decision Maker (DM, *she*). The DM must choose an action a_i from the set $A = \{a_H, a_L\}$. The set of payoff relevant states of nature is $\Omega = \{\omega_H, \omega_L\}$ and the realization of a state $\omega_i \in \Omega$ is drawn from a distribution $\mu_0 \in int(\triangle(\Omega))$, the prior, known by both players.[5] In this simplified case with two states, we abuse notation and identify a belief $\mu \in \triangle(\Omega)$ by its corresponding probability that the state is ω_H, i.e., with the number $\mu = \mu(\omega_H) \in [0, 1]$.

The Principal and the DM have material utility functions $v : A \to \mathbb{R}$ and $u : A \times \Omega \to \mathbb{R}$, respectively. Without loss, we assume that $v(a_H) > v(a_L)$. We let $a^*(\mu) \equiv \arg\max_{a \in A} \mathbb{E}_\mu[u(a, \omega)]$ represent the set of actions that the DM may take given a belief $\mu \in \triangle(\Omega)$. We break ties in favor of the Principal as this helps ensure that a solution to the Principal's problem exists.[6] Our solution concept is the Principal-preferred subgame perfect equilibrium, which we simply refer as "equilibrium" throughout the paper. We further assume that $u(a_i, \omega_i) > u(a_{-i}, \omega_i)$ for $i \in \{H, L\}$, so no action is optimal for the DM in both states. Finally, to ensure that full obfuscation is possible (as we discuss in Section 4), we assume that the DM is not indifferent between the two actions at the prior, so $a^*(\mu_0)$ is a singleton. In the 2×2 case, the unique belief of indifference is given by

$$\mu_{ind} = \frac{u(a_L, \omega_L) - u(a_H, \omega_L)}{u(a_H, \omega_H) - u(a_H, \omega_L) + u(a_L, \omega_L) - u(a_L, \omega_H)}. \tag{1}$$

The DM may purchase a signal structure (or Blackwell experiment) π that consists of the set A of signals - interpreted as recommended actions, - and a family of distributions $\{\pi(\cdot|\omega)\}_{\omega \in \Omega}$ over A.[7] But instead of working with the experiment π, it will be more convenient to work directly with the corresponding induced distribution of posteriors τ (Kamenica and Gentzkow, 2011), which we will do moving forward.

The DM is rationally inattentive and upon choosing a distribution of posteriors τ, she incurs an information cost given by

$$c_\lambda(\tau) \equiv \lambda(H(\mu_0) - \mathbb{E}_\tau[H(\mu)]), \tag{2}$$

[5]We denote by $\triangle(\Omega)$ the space of all probability distributions on Ω, and by $int(X)$ the interior of set X. For any $M \in \mathbb{N}$, we endow $\mathbb{R}^M$ with the corresponding Lebesgue measure.

[6]While this is a widely used assumption in the Bayesian Persuasion literature (Kamenica and Gentzkow, 2011), its implications in our model are two-fold: when the DM is indifferent among different distributions of posteriors, she chooses the one favored by the Principal; in addition, if the DM is indifferent between any two actions at an induced posterior, she again chooses the one favored by the Principal.

[7]It is without loss to consider the set of signal realizations to be A and treat the signal realization as a recommended action. Suppose more than one signal induces the same action. Then, the DM is paying for unused information (c.f. Matějka and McKay (2015)).

where $\lambda \geqslant 0$ is the marginal cost of information; $\mu \sim \tau$ is the posterior; and $H : \triangle(\Omega) \to \mathbb{R}_+$ is the Shannon Entropy (Shannon, 1948),

$$H(\mu) = -\mu \ln(\mu) - (1-\mu)\ln(1-\mu).$$

We assume the Principal chooses the value of λ; it is the only parameter the Principal has control over, and it is the only way it can persuade the DM to take a certain action. As we mentioned in the introduction, this can be thought of as the Principal hindering or easing the process for the DM to gather new information.

Entropy can be interpreted as a measure of uncertainty, so the cost function charges for how much uncertainty the signal reduces.[8] There are two ways a signal can have zero cost: (i) if the Principal sets $\lambda = 0$ and "reveals the state", or (ii) if the DM chooses an uninformative signal so that the posterior equals the prior and there is no change in entropy. For a given cost λ, the DM chooses τ, and a signal $a \in A$ is realized, leading to a posterior $\mu^a \in \triangle(\Omega)$, calculated by Bayes' rule.[9] She then takes an action a, and utilities $u(a, \omega)$ and $v(a)$ are realized (given that the state is ω).

An increase in λ makes information more expensive and impacts the DM's probability of taking a certain action. The Principal will therefore choose the value of λ that maximizes the odds that a_H is chosen. Note that when the Principal chooses λ, it does not know the true state of the world yet. Thus, this setting is more suited to applications in which there is uncertainty on both sides: think of a new Principal interacting with a possible DM for the first time.[10]

2.1 The DM's Rational Inattention Problem

Since the Principal has to anticipate what the DM will do given a value of λ, it is important to first understand how the DM will optimally choose to learn. Two opposing forces are in play: the DM wants to learn the state of nature to choose the best possible action vs. the DM wants to avoid learning costs. In what follows, let $U(\mu) \equiv \mathbb{E}_{w\sim\mu}[u(a^*(\mu), \omega)]$ be the indirect utility for a belief $\mu \in \triangle(\Omega)$.

[8] For beliefs in the boundary of $\triangle(\Omega)$, we let $0 \cdot \ln(0) \equiv \lim_{x\to 0^+} x \cdot \ln x = 0$ as usual.

[9] If a is chosen with positive probability, and $\pi(a|\omega)$ denotes the conditional probability $\mathbb{P}(a|\omega)$ for a given signal π,

$$\mu^a(\omega) = \frac{\pi(a|\omega)\mu_0(\omega)}{\sum_{j=1}^{2} \pi(a|\omega_j)\mu_0(\omega_j)}. \tag{3}$$

[10] In a similar fashion, Li and Shi (2017) provide a model in which a seller can allow the buyer to obtain more information about the valuation of a good. Neither the seller nor the buyer have previous perfect information about the valuation.

Definition 1 (The DM's Rational Inattention Problem)**.** *For given μ_0 and $\lambda \in \mathbb{R}_+$, the DM's Rational Inattention Problem is*

$$\max_{\tau \in \triangle(\triangle(\Omega))} \mathbb{E}_\tau \left[U(\mu)\right] - \lambda(H(\mu_0) - \mathbb{E}_\tau[H(\mu)]), \tag{4}$$
$$s.t.\ \mathbb{E}_\tau[\mu] = \mu_0.$$

The restriction is the usual requirement of Bayes plausibility. We denote by τ_λ the solution to (4) for a given λ, which identifies the posteriors and the probability that each action is taken.

2.1.1 Solving the DM's Problem

In this section, we go straight to describing a functional form of the solution to the DM's problem based on the characterization from Caplin et al. (2019). We hold on to the details until we go over this solution more explicitly in Section 4, where we also look at the comparative statics for λ and discuss when obfuscation is possible.

The characterization of Caplin et al. (2019)[11] allows us to find the posteriors that are reached following a given signal realization for any value of λ (see Lemma 3). With these posteriors in hand, we then find τ_λ through Bayes' plausibility. In the 2×2 case, we can simplify the characterization of the solution from Caplin et al. (2019) by using the following auxiliary functions:[12]

$$\underline{\mu}(\lambda) \equiv \frac{1 - exp(\frac{u(a_H,\omega_L) - u(a_L,\omega_L)}{\lambda})}{exp(\frac{u(a_H,\omega_H) - u(a_L,\omega_H)}{\lambda}) - exp(\frac{u(a_H,\omega_L) - u(a_L,\omega_L)}{\lambda})}, \tag{5}$$

$$\overline{\mu}(\lambda) \equiv \frac{1 - exp(\frac{u(a_L,\omega_L) - u(a_H,\omega_L)}{\lambda})}{exp(\frac{u(a_L,\omega_H) - u(a_H,\omega_H)}{\lambda}) - exp(\frac{u(a_L,\omega_L) - u(a_H,\omega_L)}{\lambda})}. \tag{6}$$

As illustrated in Figure 1, for a given λ, $\underline{\mu}(\lambda)$ and $\overline{\mu}(\lambda)$ represent the threshold beliefs for which actions a_L and a_H, respectively, are taken with probability 1,[13]. To understand how these thresholds characterize the solution to the DM's problem for a value of λ, we can go case by case.

If (i) $\mu_0 \in [0, \underline{\mu}(\lambda)]$, the cost of obtaining more information outweighs the benefits of learning - she knows enough, - and the DM prefers to purchase a non-informative signal that costs 0. The posterior is equal to the prior. Since $\mu_0 \in [0, \underline{\mu}(\lambda)]$, acting according to the prior means choosing a_L with probability 1 (remember that μ_0 is the probability that the state

[11] To give an intuition, the characterization comes from applying the KKT conditions to Equation (4). Matějka and McKay (2015) and Ravid (2020) also discuss how to solve this problem.

[12] The exponential function arises from the use of the Entropy as a cost function.

[13] For more details see Caplin et al. (2019); Matysková and Montes (2023) and Section 4 in this paper.

is H, so the closer μ_0 is to 0, the more likely it is that action a_L is optimal). Analogously, if (ii) $\mu_0 \in [\overline{\mu}(\lambda), 1]$, the DM acquires a completely non-informative signal, but now she always chooses a_H. If, however, (iii) $\mu_0 \in (\underline{\mu}(\lambda), \overline{\mu}(\lambda))$, then it is profitable to acquire new information. The DM will choose a distribution of posteriors with support $\mu^{a_L} = \underline{\mu}(\lambda)$ and $\mu^{a_H} = \overline{\mu}(\lambda)$. Thus, if the recommended action is a_L the induced posterior will be $\underline{\mu}(\lambda)$ and she chooses a_L; if the realization is a_H the posterior will be $\overline{\mu}(\lambda)$ and she chooses a_H, so both actions are chosen with a positive probability.

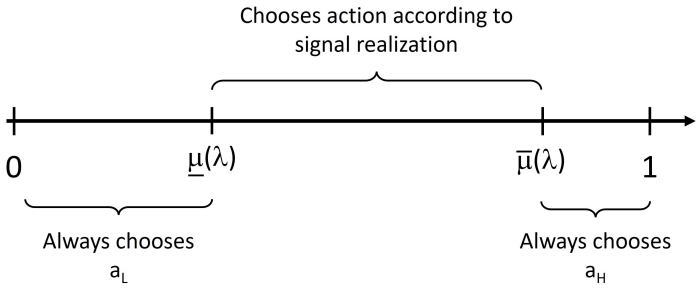

Figure 1: Given λ, the functions $\underline{\mu}$ and $\overline{\mu}$ define the choice by the DM depending on where the prior lays in the $[0, 1]$ interval.

To see how the Principal will choose λ is then sufficient to understand how the thresholds $\underline{\mu}(\lambda)$ and $\overline{\mu}(\lambda)$ change with λ. To ensure continuity, we define

$$\underline{\mu}(0) \equiv \lim_{\lambda \to 0^+} \underline{\mu}(\lambda) = 0, \text{ and } \overline{\mu}(0) \equiv \lim_{\lambda \to 0^+} \overline{\mu}(\lambda) = 1.$$

We get these limits since $u(a_i, \omega_{-i}) - u(a_i, \omega_i) < 0$ and $u(a_i, \omega_i) - u(a_{-i}, \omega_i) > 0$. By calculating the limit when $\lambda \to \infty$, we get

$$\lim_{\lambda \to \infty} \underline{\mu}(\lambda) = \lim_{\lambda \to \infty} \overline{\mu}(\lambda) = \frac{u(a_L, \omega_L) - u(a_H, \omega_L)}{u(a_H, \omega_H) - u(a_H, \omega_L) + u(a_L, \omega_L) - u(a_L, \omega_H)} = \mu_{ind}. \quad (7)$$

When $\lambda = 0$, the thresholds are in the extremities 0 and 1; as λ increases, they move toward μ_{ind} in the interior. As a consequence of these properties, we also have that $\mu_{ind} \in \left[\underline{\mu}(\lambda), \overline{\mu}(\lambda)\right]$ for all $\lambda \in [0, \infty)$. This makes sense: as λ grows and new information becomes more expensive, the regions for which buying new information is beneficial diminish. At the limit, only when the DM is very unsure (μ_0 close to μ_{ind}) will she buy any information - a fact we discuss in more detail in Section 4. Finally, we highlight that $\underline{\mu}$ / $\overline{\mu}$ are strictly increasing/decreasing respectively[14] and continuous.[15] Figure 2 illustrates how the threshold beliefs $\underline{\mu}$ and $\overline{\mu}$ change with λ in an example where $\mu_{ind} = 0.5$.

[14]For more details, see Matysková and Montes (2023). We can also check that the derivative is always

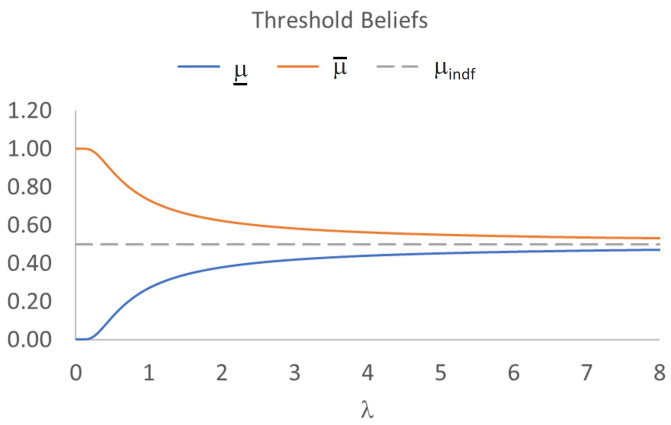

Figure 2: The value of $\overline{\mu}$ and $\underline{\mu}$ as a function of lambda when $\mu_{ind} = 0.5$.

2.2 The Principal's Problem

Now that we know how the DM will react to the cost of information, we are better equipped to understand how the Principal's utility changes with λ, which in turn lets us predict how much the Principal will choose to obfuscate. We define the Principal's indirect expected utility given a belief μ as $V(\mu) \equiv \mathbb{E}_{\mu}[v(a^*(\mu))]$. Anticipating the optimal choice τ_λ by the DM (which is unique in the 2×2 case), we can define the Principal's indirect expected utility from obfuscation as $\tilde{V}_{\mu_0}(\lambda) \equiv E_{\tau_\lambda}[V(\mu)]$. We use the prior as an index to highlight that, because of Bayes' plausibility, it affects the optimal choice of λ.

Intuitively, the Principal will choose a λ that maximizes the probability that its favorite action is played. For instance, if the Principal could agree with the DM in every state, it would choose $\lambda = 0$; if the Principal wants to maximize the chance of mistakes, it will choose full obfuscation. The following Lemma ensures that full obfuscation is possible since $\mu_0 \neq \mu_{ind}$.

Lemma 1. *There exists* $\overline{\lambda} \in \mathbb{R}_{++}$ *such that*

$$\begin{cases} \underline{\mu}(\overline{\lambda}) = \mu_0, \text{ if } \mu_0 < \mu_{ind}, \\ \overline{\mu}(\overline{\lambda}) = \mu_0, \text{ if } \mu_0 > \mu_{ind}. \end{cases} \tag{8}$$

Proof. We highlight three previously mentioned facts: (i) $\underline{\mu}(\lambda)$ and $\overline{\mu}(\lambda)$ are continuous and strictly increasing/decreasing respectively; (ii) $\underline{\mu}(0) = 0$, $\overline{\mu}(0) = 1$, $\lim_{\lambda\to\infty} \underline{\mu}(\lambda) = \lim_{\lambda\to\infty} \overline{\mu}(\lambda) = \mu_{ind}$; (iii) $\mu_0 \neq \mu_{ind}$ by assumption. Now, consider the case where $\mu_0 < \mu_{ind}$. Since $\underline{\mu}$ is continuous, $\underline{\mu}(0) = 0$ and $\lim_{\lambda\to\infty} \underline{\mu}(\lambda) = \mu_{ind}$, by the Intermediate Value Theorem, we deduce

positive/negative.

[15] Continuity is ensured because both $\underline{\mu}(\cdot)$ and $\overline{\mu}(\cdot)$ are composites of continuous functions.

the existence of $\overline{\lambda} \in (0, \infty)$ such that $\underline{\mu}(\overline{\lambda}) = \mu_0$. The case where $\mu_0 > \mu_{ind}$ is analogous. $\square$

We formalize this result for multiple states and actions in Section 4. If $\lambda = 0$, the DM acquires a fully informative signal; if $\lambda \in (0, \overline{\lambda})$, then $\mu_0 \in [\underline{\mu}(\lambda), \overline{\mu}(\lambda)]$ and the DM acquires some information; if $\lambda \geqslant \overline{\lambda}$ information is too expensive and the DM plays according to the prior. It is then clear that $\tilde{V}_{\mu_0}(\lambda) = V(\mu_0)$ for any $\lambda \geqslant \overline{\lambda}$. Therefore, without loss of generality, we consider that the Principal chooses $\lambda \in [0, \overline{\lambda}]$. We are now ready to state the Principal's problem.

Definition 2 (The Principal's Obfuscation Problem)**.** *The Principal's Obfuscation Problem is*

$$\max_{\lambda \in [0, \overline{\lambda}]} \tilde{V}_{\mu_0}(\lambda). \tag{9}$$

In Section 4, we state Proposition 7 that ensures that a solution exists as long as $\mu_0 \neq \mu_{ind}$. For an intuition, when $\mu_0 = \mu_{ind}$, any information, no matter how coarse, is useful for the agent to "break the tie" between actions. So no matter how high the value of λ, the agent will always find it useful to buy a signal coarse enough that she can still afford, making $\overline{\lambda} \to \infty$ and full obfuscation impossible.[16]

2.3 A Geometric Interpretation for the Principal's Problem

In this section, we provide a functional form for $\tilde{V}_{\mu_0}$. This functional form has a similar interpretation as the concavification result in Bayesian Persuasion (BP). Given that the DM is subject to Bayes' plausibility, we can find the value of $\tilde{V}_{\mu_0}(\lambda)$ by evaluating at the prior μ_0 the line in $\mathbb{R}^2$ that connects the points $(\underline{\mu}(\lambda), V(\underline{\mu}))$ and $(\overline{\mu}(\lambda), V(\overline{\mu}))$.

This is illustrated in Figure 3, where we plot the Principal's payoff when $v(a_H) = 1$, $v(a_L) = 0$, $\mu_0 = 0.75$, and $\mu_{ind} = \frac{1}{2}$.[17] The left-hand side of Figure 3 shows the split of posteriors under full-information ($\lambda = 0$). Then, $\tilde{V}_{\mu_0}(\lambda)$ is found as the y-coordinate of the point at the intersection of the vertical line passing through the prior and the steep line connecting $(\underline{\mu}(\lambda), V(\underline{\mu}))$ and $(\overline{\mu}(\lambda), V(\overline{\mu}))$. On the right side of Figure 3 we see what happens when a non-informative signal is acquired ($\lambda = \overline{\lambda}$). The posterior will equal the prior, and we can find $\tilde{V}_{\mu_0}(\overline{\lambda})$ by looking at the y-coordinate of the intersection between the vertical line crossing the prior and the step function $V(\mu)$.

[16] Of course, this is only an issue if the Principal wants to induce full obfuscation.

[17] Note that μ_{ind} is the only information we need from the DM to define the payoffs of the Principal as a function of the beliefs. If $\mu(\omega_H) \geqslant \frac{1}{2}$, the DM chooses a_H and the Principal's payoff is 1. Similarly, if $\mu(\omega_H) < \frac{1}{2}$, the Principal's payoff is 0.

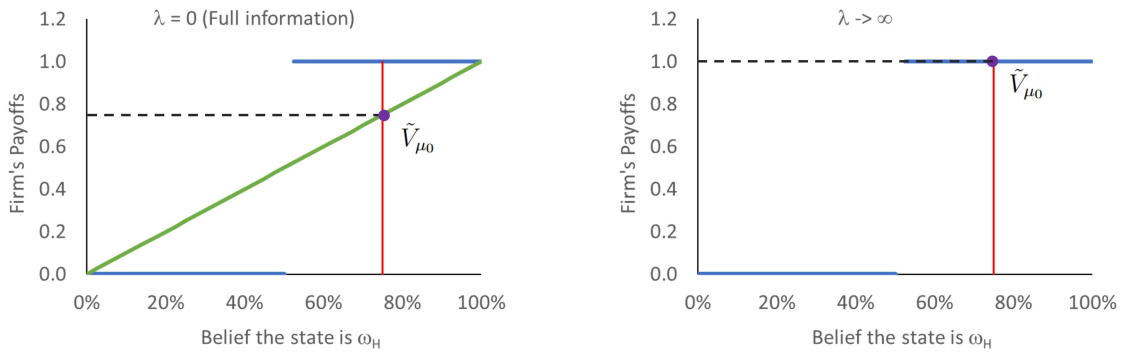

Figure 3: Left: the expected payoff of full information is at the intersection of the steep line with the vertical line of the prior. Right: the posterior is equal to the prior with no new information. In both cases, we set $v(a_H) = 1$, $v(a_L) = 0$, $\mu_0 = 0.75$, and $\mu_{ind} = \frac{1}{2}$.

To illustrate the dynamics of how $\tilde{V}_{\mu_0}$ changes with λ, consider Figure 4. An increase in λ moves the beliefs $\underline{\mu}$ and $\overline{\mu}$ closer to μ_{ind}. The value of $\tilde{V}_{\mu_0}$ then changes as the intersection of the vertical prior line with the line segment connecting V evaluated at the posteriors moves up or down. At the top row of Figure 4, the value of $\tilde{V}_{\mu_0}$ is monotonically increasing as the point of intersection goes up. In the middle row, we have an example where $\tilde{V}_{\mu_0}$ is monotonically decreasing as the point of intersection goes down. At the bottom row, $\tilde{V}_{\mu_0}$ changes non-monotonically: it increases a bit at first, but after a point it decreases with λ.

Define $\mathcal{O} \equiv [0, \overline{\lambda}]$ and, to ease notation, let $m : \mathbb{R}_+ \to \mathbb{R}$ be given by[18]

$$m(\lambda) \coloneqq \left(\frac{V(\underline{\mu}(\lambda)) - V(\overline{\mu}(\lambda))}{\underline{\mu}(\lambda) - \overline{\mu}(\lambda)} \right).$$

The functional form for $\tilde{V}_{\mu_0}$ is thus

$$\tilde{V}_{\mu_0}(\lambda) = \begin{cases} m(\lambda)(\mu_0 - \underline{\mu}(\lambda)) + V(\underline{\mu}(\lambda)) \text{ if } \lambda \in \mathcal{O}, \\ V(\mu_0), \text{ otherwise.} \end{cases} \tag{10}$$

Note that $m(\overline{\lambda})(\mu_0 - \underline{\mu}(\overline{\lambda}) + V(\underline{\mu}(\overline{\lambda})) = V(\mu_0)$ and so $\tilde{V}_{\mu_0}$ is continuous. Indeed, finding $\overline{\lambda}$, and therefore $\mathcal{O}$, is just a matter of discovering the unique λ that satisfies

$$m(\lambda)\left(\mu_0 - \underline{\mu}(\lambda)\right) + V(\underline{\mu}(\lambda)) = V(\mu_0).$$

[18]The function $m(\cdot)$ gives us the tangent of the steep line connecting points $(\underline{\lambda}, \tilde{V}_{\mu_0}(\underline{\mu}(\lambda))$ and $(\overline{\lambda}, \tilde{V}_{\mu_0}(\overline{\mu}(\lambda))$.

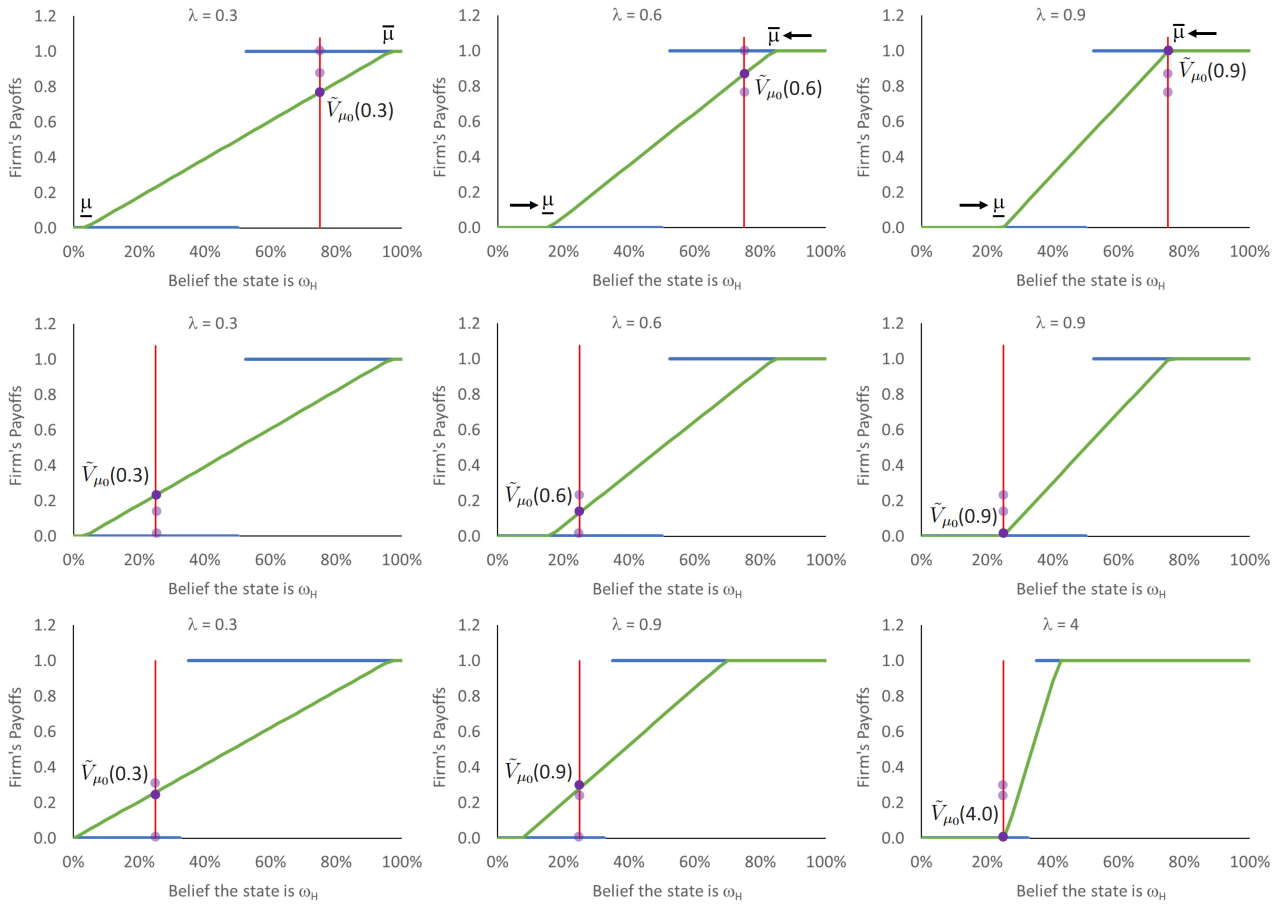

Figure 4: Dynamics of an increase in λ on the Principal's payoff when $\tilde{V}_{\mu_0}$ is increasing (top row); decreasing (middle row); non-monotonic (bottom row). We set $v(a_H) = 1$ and $v(a_L) = 0$ in all cases. In the top row: $\mu_0 = 0.75, \mu_{ind} = 0.5$. In the middle row: $\mu_0 = 0.25, \mu_{ind} = 0.5$. In the bottom row: $\mu_0 = 0.25, \mu_{ind} = 1/3$.

3 Transparency, Obfuscation, and Partial Obfuscation

This section characterizes the equilibrium degree of obfuscation and its welfare implications. Three outcomes are possible: full transparency (in which the Principal chooses $\lambda = 0$, implying complete learning), full obfuscation ($\lambda = \bar{\lambda}$, implying no learning), and partial obfuscation ($\lambda \in (0, \bar{\lambda})$, with some learning). The next definition formalizes these cases and extends naturally to the multi-state model.

Definition 3 (Partial vs. Boundary Obfuscation Solutions)**.** *The Principal's obfuscation problem admits a* partial obfuscation solution *if neither* $\lambda = 0$ *nor* $\lambda = \bar{\lambda}$ *solves the Principal's optimization problem. Otherwise, the equilibrium lies on the boundary, and we refer to it as a* boundary obfuscation solution*, which can take two forms:*

(i) full transparency, when the Principal's optimal choice is $\lambda = 0$*;*

(ii) full obfuscation, when the Principal's optimal choice is $\lambda = \bar{\lambda}$.

A sufficient condition for a problem to have a boundary solution can be determined by examining the shape of the Principal's indirect utility function V. The following result is proven for the 2x2 case we described, but we highlight that it applies when the Principal has state-dependent preferences (for instance, if $v = u$).

Proposition 1 (Concavity and Obfuscation)**.** *Given a Principal's Problem, if* V *is concave (convex), then* $\tilde{V}_{\mu_0}$ *is increasing (decreasing) in* λ.

It extends to the Principal the result that a rationally inattentive agent with convex utility function prefers more information. When the Principal's indirect utility $V(\mu)$ is concave, the expected value of V under any distribution of posteriors is below $V(\mu_0)$. In other words, allowing the DM to learn, thereby spreading her beliefs around the prior, reduces the Principal's expected payoff. The Principal therefore prefers higher information costs and less learning. Conversely, if V is convex, the inequality reverses: the Principal's expected payoff increases when the DM becomes better informed, so it favors transparency.

Does partial obfuscation ever occur? One might expect the simplicity of the binary environment to induce only full transparency or full obfuscation in equilibrium. Yet the model shows that partial obfuscation can emerge even in this simple setting and under fairly general conditions. This case is particularly relevant in real-world situations where frictions or additional costs influence the Principal's choice of λ.[19]

Partial obfuscation appears when $\tilde{V}_{\mu_0}$ is non-monotonic. The key force is that the Principal need not always prefer less learning: full transparency may reveal too much, but full obfuscation may leave the DM acting on an unfavorable prior. In such cases, an intermediate information cost allows the DM to acquire some information while still inducing mistakes in the direction the Principal prefers. This mechanism helps interpret practices such as add-on fees that are formally disclosed but placed in fine print, or product comparisons in which relevant attributes are available but surrounded by many additional features. These examples are best understood as ex-ante choices of complexity, format, or ease of comparison. Although partial obfuscation could also result from full obfuscation being prohibitively costly, our baseline model shows that such costs are not necessary: partial obfuscation can be optimal even when the Principal can choose any λ at no direct cost.

Our main result for this section, Theorem 1, provides a characterization of cases of partial obfuscation in our setting of 2x2 and state-dependent preferences for the Principal. Our main example, presented in Online Appendix A, details the case of a Principal trying to induce the DM to buy a product that may be good or bad. Before going into the theorem itself, we state the following result, which will prove useful.

[19]The version of the model with an explicit cost of obfuscation is detailed in Online Appendix B.

Lemma 2. *If the DM has Symmetric Preferences (i.e., $\mu_{ind} = \frac{1}{2}$), then we have a boundary obfuscation solution.*

In such scenarios, the Principal chooses a boundary solution for any prior. The main idea is that, in this environment, any increase in λ will always result in symmetric movements by $\underline{\mu}$ and $\overline{\mu}$. Since V is a step function when the Principal has state-independent preferences, the resulting $\tilde{V}_{\mu_0}$ will be monotonic. This result provides some insight into how the DM optimally learns based on the belief of indifference, which can be thought of as the ratio of the stakes of each state, as seen in Equation (7). If the stakes are the same in both states, the DM will learn about both states equally. But if one state has costlier mistakes, say ω_H, she will want to be more certain that she is in state ω_H when it happens, even if it means choosing the wrong action more often when ω_L happens instead.[20]

From this result, we get the motivation for looking into the DM from the perspective of the belief that makes her indifferent between the two actions (remember that we consider $v(a_H) > v(a_L)$).

Definition 4 (Skeptical DM)**.** *We say that we have a* Skeptical DM *if the indifference belief is lower than $\frac{1}{2}$ and the initial prior is even lower, i.e., $\mu_0 < \mu_{ind} < 1/2$*

A *Skeptical DM* would not choose the Principal's optimal action absent new information. More precisely, a Skeptical DM is one whose prior belief is so extreme that her optimal action under the prior differs from the Principal's optimal action, even though there is a wide range of beliefs for which the two would agree. Put differently, even a modestly low (high) belief would suffice to induce her to take the Principal's optimal action, but her actual belief falls short of (exceeds) that already-low (already-high) threshold. This definition sets up the main result of this section.

Theorem 1 (Obfuscation Characterization)**.** *Given a Principal's Obfuscation Problem,*

(i) we have a Partial Obfuscation solution if, and only if, we have a Skeptical DM;

(ii) we have a Full Obfuscation solution if, and only if, we have Agreement at the Prior, i.e. $a^*(\mu_0) = \arg\max_{a_i \in A} v(a_i)$;

(iii) otherwise, we have a Full Transparency solution.

For part (i), we prove that Skeptical DM indeed yields Partial Obfuscation, then we proceed to show that all other possible cases result in Boundary Obfuscation. Part (ii) shows

[20] Brown and Jeon (2024) provide empirical evidence of such behavior. They document that when faced with choosing among different types of health insurance, patients with a higher likelihood of needing insurance tend to gather more information about the plans before making a decision.

that the full obfuscation happens if, and only if, absent new information, the DM would choose the Principal's preferred action. Full transparency encompasses all remaining cases.

Leveraging the results from Theorem 1, we can shed light on policy implications aimed at improving DM welfare by inducing full transparency. The next result can help us see how. To simplify exposition, consider the normalization[21]

$$u(a_H, \omega_H) = B > 0, \quad u(a_L, \omega_H) = 0, \quad u(a_H, \omega_L) = -C < 0, \quad u(a_L, \omega_L) = 0,$$

so that $\mu_{ind} = \frac{C}{B+C}$, and let $u_0^* \in \mathbb{R}$ be the DM's ex-ante expected payoff if the Principal were to set $\lambda = 0$. Clearly, u_0^* is the highest payoff the DM can achieve.[22] We can now state our next result.

Theorem 2 (Policy Design for Transparency)**.** *Assume we do not have Full Transparency. Then, there exists $\underline{\varepsilon} > 0$ and an $i \in \{H, L\}$ such that reducing $u(a_i, \omega_{-i})$ by $\varepsilon > \underline{\varepsilon}$, induces a full information equilibrium (the Principal sets $\lambda^* = 0$) that achieves u_0^*. Moreover,*

(i) under Partial Obfuscation, $\underline{\varepsilon} = B - C$, and the result holds with equality, i.e., for all $\varepsilon \geqslant \underline{\varepsilon}$;

(ii) under Full Obfuscation, $\underline{\varepsilon} = \max\{\frac{\mu_0}{1-\mu_0}B - C, B - C\}$.

It may at first be counterintuitive that strictly reducing the DM's payoff from an action in a state can improve the DM's equilibrium welfare. We are effectively increasing the costs of mistakes. However, the intuition is straightforward: increasing the costs of mistakes ensures that the DM will only buy when she has enough information, leaving the Principal no choice but to reduce λ. Moreover, the high cost of a mistake has no direct impact on welfare because under the induced full information, the DM never makes a mistake.

Applying Theorem 2 only requires knowing things that are common knowledge: payoffs and the prior. It does not require experimentation on the part of the policy-maker and still leaves learning in the hands of the DM.

One may question whether $\underline{\varepsilon}$ always has to be so large as to make the result meaningless. The next result shows that it is not necessarily the case. Note that a Skeptical DM is characterized by the condition

$$\mu_0 < \mu_{ind} \iff \frac{\mu_0}{1-\mu_0}B < C < B. \tag{11}$$

[21] Lemma 5 in the Appendix shows that this is without loss of generality.

[22] We have that $u_0^* = \mu_0 u(a_H, \omega_H) + (1 - \mu_0)u(a_L, \omega_L)$.

For such a Skeptical DM, let

$$W(\lambda, C) \coloneqq \max_{\tau} \mathbb{E}_{\mu\sim\tau}[\max\{0, (B+C)\mu - C\}] - \lambda(H(\mu_0) - \mathbb{E}_{\mu\sim\tau}[H(\mu)]),$$
$$\text{s.t. } E_{\mu\sim\tau}[\mu] = \mu_0.$$

denote the DM welfare when parameters are (λ, C). Correspondingly, for a given equilibrium level of obfuscation $\lambda^*(C)$,[23] we can further define $W^*(C) \coloneqq W(\lambda^*(C), C)$.

Proposition 2 (Welfare and Mistake Costs)**.** *Assume we have a Skeptical DM, so* $C \in (\frac{\mu_0}{1-\mu_0}B, B)$*. Then, there exists* $\eta > 0$ *such that if* $C \in \left(\frac{\mu_0}{1-\mu_0}B,\ \frac{\mu_0}{1-\mu_0}B + \eta\right) \cup (B-\eta, B)$*, then*

$$\frac{dW^*(C)}{dC} > 0.$$

Therefore, if the conditions of Proposition 2 hold, then there exists $\delta > 0$ such that the DM's welfare is increasing in ε (the increase in the cost of mistakes) for all $\varepsilon \in (0, \delta)$. While we do not have an analytical proof that applies to all values of C in the relevant interval, since the result applies for boundary values, it indicates that the property should hold more generally. Proposition 2 imbues additional economic meaning to the definition of a Skeptical DM and, therefore, to Theorem 1. Under full obfuscation, a small increase in the cost of mistakes cannot make the DM better-off: if ε is too small, the prior remains in the agreement-at-the-prior region, the Principal continues to choose full obfuscation, and the DM simply faces a worse payoff in the bad state. By contrast, under partial obfuscation and the conditions of Proposition 2, policy makers can rest easy knowing that increasing the cost of mistakes is beneficial for the DM even for very small values of ε. The result may be interpreted as a local comparative static. It highlights a strategic channel through which worsening the DM's payoff off path can improve her equilibrium welfare.

Before moving forward, one might wonder if we could use a First Order Condition to study the optimal value of λ for the Principal. Indeed, that can be done.

Proposition 3 (Uniqueness of Optimal λ)**.** *Given a Principal's Obfuscation Problem, there exists at most one* $\lambda^* \in (0, \overline{\lambda})$ *such that* $\frac{d\tilde{V}_{\mu_0}(\lambda^*)}{d\lambda} = 0$*.*

Proposition 3 implies two things. First, any $\lambda \in (0, \overline{\lambda})$ that satisfies $\frac{d\tilde{V}_{\mu_0}(\lambda)}{d\lambda} = 0$ is not only a local minimum/maximum, but a global minimum/maximum. Therefore, we have Partial Obfuscation only if there exists a $\lambda^* \in (0, \overline{\lambda})$ such that $\frac{d\tilde{V}_{\mu_0}(\lambda^*)}{d\lambda} = 0$. To see if the candidate λ^* is a global maximum instead of a global minimum, it suffices to check if $\tilde{V}_{\mu_0}(\lambda^*) > \tilde{V}_{\mu_0}(0)$.

[23] Proposition 3 shows that the solution to the Principals' problem is unique.

Second, it implies that a Partial Obfuscation solution to the Principal's problem is unique. Since the functional form of $\tilde{V}_{\mu_0}$ is rather convoluted - and its derivative even more so - it may be quite impractical to check the FOC.[24] On the other hand, checking for a Skeptical DM takes very little effort.

3.1 Comparison to Communication Games

Our framework complements existing communication models. In Bayesian Persuasion (Kamenica and Gentzkow, 2011), the sender fully controls the information structure; our sender instead shapes the receiver's cost of acquiring information, yielding weakly lower payoffs than in a full commitment setting.

The comparison with Cheap Talk (Crawford and Sobel, 1982) is more nuanced. Cheap Talk (CT) involves costless, non-binding communication. Relative to CT, obfuscation provides a limited form of commitment: the sender cannot dictate messages but can constrain what the receiver can afford to learn. However, Cheap Talk also offers a type of flexibility that strategic obfuscation lacks. In a CT game, a sender could use a wider range of communication experiments or signals that might not be possible to replicate within the rational inattention framework with information acquisition costs. Indeed, we show in Example 1 below that Cheap Talk can outperform Strategic Obfuscation (SO) when we allow for more than 2 actions.

Example 1. [Cheap Talk outperforming Strategic Obfuscation] There are two players, player 1 (sender or Principal), and player 2 (receiver or DM). Player 2 chooses an action from the set $A = \{a, b, c, d\}$, which determines both players' payoffs. The state $\omega \in \{High, Low\}$ is drawn with prior $\mu_0 = \Pr(\omega = High)$. Table 1 shows the payoffs for each player (the rows represent the state of the world).

Table 1: Payoffs for the players.

		Player 2			
		a	b	c	d
Player 1	High	1, -2	0, -1	1, 0	0, 1
	Low	1, 1	0, 0.9	1, 0	0, -1

Given the optimal action for player 2 at each belief, $a^*(\mu) = \arg\max_{x \in A} E_\mu[u(x, \omega)]$, Figure 5 plots the indirect utility for Player 1, $V(\mu) = \mathbb{E}_\mu[v(a^*(\mu))]$.

[24] One could also check the SOC to separate between the minimum and maximum, but higher order derivatives of $\tilde{V}_{\mu_0}$ are even more computationally extensive.

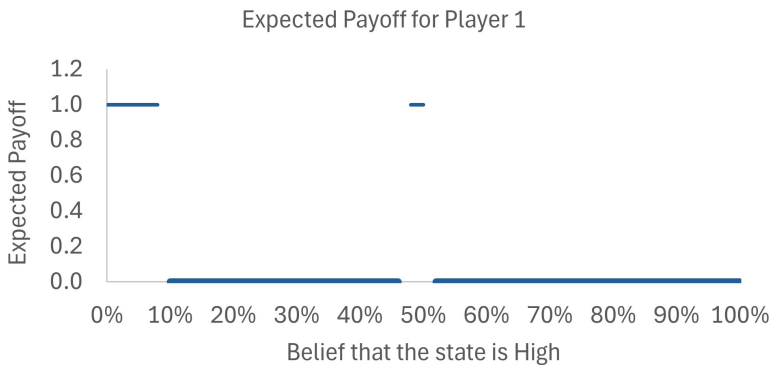

Figure 5: Expected payoffs for player 1.

We fix a prior of 0.4. Under Cheap Talk, we can leverage the results from Lipnowski and Ravid (2020), and there exists an equilibrium in which player 1 induces the distribution of posteriors τ given by $\tau(0.05) = \frac{2}{9}$ and $\tau(0.5) = \frac{7}{9}$ that yields a payoff of 1. The goal is to induce posteriors that fall in the small areas where the payoff for player 1 is 1. Under Strategic Obfuscation, however, a payoff of 1 for player 1 is not possible. For any $\lambda \in \mathbb{R}_+$, the induced distribution of posteriors τ_λ must assign positive probability that action b or d is recommended.

□

Example 2. [Cheap Talk underperforming Strategic Obfuscation]

Consider now that Player 2 chooses an action from the set $A = \{a, b, c\}$, and that the payoffs are given by Table 2.

Table 2: Payoffs for the players.

		Player 2		
		a	b	c
Player 1	High	0.5, -1	0, 0.5	1, 1
	Low	0.5, 1	0, 0.5	1, -1

The corresponding expected utility for Player 2 is plotted in Figure 6. For a prior of 40%, a payoff of 0.5 is the maximum equilibrium payoff that a sender can get under Cheap Talk (Lipnowski and Ravid, 2020). However, under SO, the Principal can choose $\lambda = 0$, which leads to full information and an expected payoff of 0.7.

□

For the remainder of this section, we keep the same environment of binary actions and binary states of the world. Formally, let there be two players: player 1 (the Principal, or

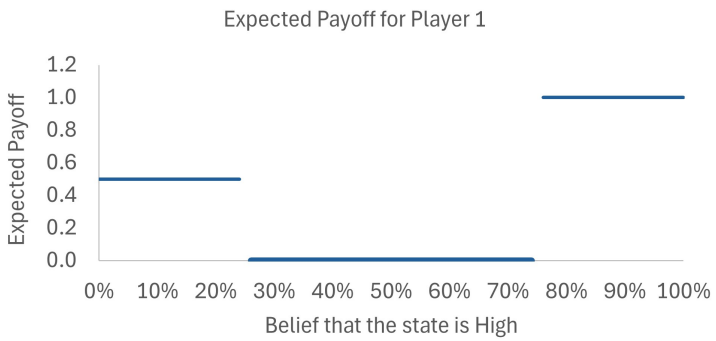

Figure 6: Expected payoffs for player 1.

the sender) wants to induce player 2 (the DM, or the receiver) to take a certain action. Let $\Omega = \{\omega_H, \omega_L\}$ denote the possible states of the world, and $A = \{a_H, a_L\}$ denote the set of actions of player 2. Players 1 and 2 have utility functions $v, u : \Omega \times A \to \mathbb{R}$ respectively, and a common prior $\mu_0 \in \triangle(\Omega)$. Fixing these common objects, we can define the corresponding Bayesian Persuasion, Cheap Talk, and Obfuscation settings.[25]

Definition 5 (Sender Payoffs)**.** *Let s^P be the optimal payoff for the Sender under Bayesian Persuasion; let s^C be the Sender's best equilibrium payoff under Cheap Talk; and let s^O be the optimum payoff for the Principal under Strategic Obfuscation.*

With these in hand, we claim the next result. With state-independent preferences[26], the sender or Principal's payoff only depend on the receiver or DM's action.

Proposition 4 (Payoff Ranking)**.** *The following holds:*

(i) $s^O \leqslant s^P$. Under State-Independent Preferences for the sender/Principal, $s^O = s^P$ if, and only if, we have Agreement at the Prior, i.e., $a^*(\mu_0) \subseteq \arg\max_{a_i \in A} v(a_i)$.

(ii) Under State-Independent Preferences for the sender/Principal, $s^C \leqslant s^O$.

The highest outcome that the Sender can achieve under BP is an upper bound to what can be achieved under SO as the possible induced signals in SO are a subset of the possible induced signals in BP.[27] The optimal experiment under both models will coincide only when the DM chooses the Principal-preferred action absent new information.

In a simplified setting of 2 x 2, the comparison with Cheap Talk is rather straightforward (Cheap Talk would lead to Babbling Equilibria), but we kept the comparison in Proposition 4

[25]We can assume that the set of messages M for Cheap Talk is rich enough, i.e., $|M| \geqslant |A|$.

[26]Also called transparent motives as in Lipnowski and Ravid (2020).

[27]Induced distributions of posteriors in both SO and BP have to be Bayes plausible. While that is the only restriction in BP, in SO you can't set $\overline{\mu}$ and $\underline{\mu}$ independently.

for completeness. In a more complex setting, as the previous examples show, the comparison between Cheap Talk and SO requires a more careful eye.

4 The General Model

Here, we expand the 2×2 model to allow for multiple states and actions and state-dependent preferences for the Principal. The rest of the setting (players, information, and timing) remains the same. Formally, the DM (she) now chooses an action from a finite set $A = \{a_1, \ldots, a_N\}$. We denote by $\Omega = \{\omega_1, \ldots, \omega_M\}$ the finite set of payoff-relevant states of nature, where the realization of a state is drawn from the prior distribution $\mu_0 \in int(\triangle(\Omega))$, known by both individuals.

The DM and the Principal have material utility functions $u : A \times \Omega \to \mathbb{R}$ and $v : A \times \Omega \to \mathbb{R}$ respectively. We let $a^*(\mu) \equiv \arg\max_{a \in A} \mathbb{E}_\mu[u(a, \omega)]$ represent the set of actions that the DM may take given a belief $\mu \in \triangle(\Omega)$, again breaking ties in favor of the Principal. This will help ensure that a solution to the Principal's problem exists. We restrict A to contain only actions that are optimal for the DM for at least one prior belief.[28] We further assume that no two actions give the same payoffs in every state. Finally, to ensure that full obfuscation is possible (as we will see later), we assume that there is only one optimal action at the prior, so that $a^*(\mu_0)$ is a singleton[29] (as in the 2×2 case). This property holds almost-everywhere in the space of beliefs.[30]

As before, the DM buys information $\pi = \{\pi(\cdot|\omega)\}_{\omega \in \Omega}$ with a cost given by the Shannon Entropy, where the marginal cost of information is parametrized by λ.[31] We can again think in terms of the induced distribution of posteriors $\tau \in \triangle(\triangle(\Omega))$. The Principal chooses the value of $\lambda \in \mathbb{R}_+$ and, for a given cost λ, the DM chooses τ. When a signal a_i is realized, she takes action a_i.

4.1 The DM's Rational Inattention Problem

Here, we go into more detail over the solution to the DM's problem. The statement of the DM's problem is the same as given in Equation (4). We use the characterization of the solution for the DM's problem in Caplin et al. (2019), which can be called the posterior approach. However, when we get to the Principal's problem, it will also be worth thinking

[28] The DM would never choose any other actions, and they can be removed without loss of generality.

[29] We could also get rid of this assumption if we assume that there is a cost to obfuscation (no matter how small), as in the extension in Online Appendix B.

[30] To give an intuition why, think of an example with 2 states and 2 actions: the resulting belief of indifference can be thought of as a single point in the $[0, 1]$ interval.

[31] In the general case, the form of the Entropy function is given by $H(\mu) = -\sum_{m=1}^{M} \mu(\omega_m) \ln(\mu(\omega_m))$.

of the solution in terms of the induced action probabilities. Indeed, Matějka and McKay (2015) show that we can alternatively write the DM's problem in terms of the unconditional probabilities of taking a certain action. More formally, it is as if the DM selects a stochastic choice function $P_\lambda : \Omega \to \triangle(A)$ from a set $\mathcal{P} = \{P : \Omega \to \triangle(A)\}$ (Corollary 1 from Matějka and McKay (2015)). For a signal π_λ that induces a posterior distribution τ_λ, the unconditional probability of a is $P_\lambda(a) = \sum_{m=1}^{M} \mu_0(\omega_m)\pi_\lambda(a|\omega_m)$, so $P_\lambda(a) = \tau_\lambda(\mu^a)$ for any action played with positive probability. We let $A(\tau_\lambda) = \{a \in A : P_\lambda(a) > 0\}$ denote the set of actions played with positive probability under a solution τ_λ.

4.1.1 Solving the DM's Problem

The solution to the DM's problem, for any $\lambda > 0$, has been characterized in Caplin et al. (2019), which we rewrite here for completeness.

Lemma 3. *(Proposition 2, Caplin et al. (2019).) A distribution of posteriors τ_λ and posteriors $\{\mu^a : a \in A(\tau_\lambda)\}$ is a solution to the DM's Rational Inattention Problem (with $\lambda > 0$) if and only if, $\sum_{a \in A(\tau_\lambda)} \tau_\lambda(\mu^a)\mu^a(\omega) = \mu_0(\omega)$, and:*

1. *For any $\omega \in \Omega$ and any $a_i, a_j \in A(\tau_\lambda)$,*

$$\frac{\mu^{a_i}(\omega)}{\exp\left(\frac{u(a_i,\omega)}{\lambda}\right)} = \frac{\mu^{a_j}(\omega)}{\exp\left(\frac{u(a_j,\omega)}{\lambda}\right)}. \tag{12}$$

2. *Given $a_i, a_j \in A$ such that $a_i \in A(\tau_\lambda)$, and $a_j \notin A(\tau_\lambda)$,*

$$\sum_{m=1}^{M} \mu^{a_i}(\omega_m) \exp\left(\frac{u(a_j,\omega_m) - u(a_i,\omega_m)}{\lambda}\right) \leqslant 1. \tag{13}$$

We denote by $T_{\mu_0}(\lambda)$ the set of solutions to the DM's problem for a given λ. Again, let $\tau_\lambda \in T_{\mu_0}(\lambda)$ denote an arbitrary solution to the DM's problem.

4.1.2 The Impact of the Cost of Information on the DM's Solution

In this section, we analyze how the solution τ_λ for the DM's Problem changes as we vary λ. To find τ_λ, we must solve the system of equations defined by (12) and (13), which does not provide an easy functional form. Therefore, to think about the dynamics of the solution as λ increases, it is helpful to define, for each $a_i \in A$, the set

$$I(\lambda; a_i) \equiv \left\{ \mu \in \triangle(\Omega) : \sum_{m=1}^{M} \mu(\omega_m) \exp\left(\frac{u(a_j,\omega_m) - u(a_i,\omega_m)}{\lambda}\right) \leqslant 1, \forall j \neq i \right\}. \tag{14}$$

This set characterizes, for a given λ, all prior beliefs in which buying no information and playing a_i with probability 1 is optimal.[32]

We define $I(0, a_i) \equiv \lim_{\lambda \to 0^+} I(\lambda; a_i)$, which has a simple characterization. Note that, as λ approaches $0, I(\lambda; a_i)$ only includes beliefs that put a very low probability on states in which a_i is not the optimal action. Indeed, let $\Omega(a_i) = \{\omega \in \Omega : u(a_i, \omega) \geqslant u(a_j, \omega)\ \forall j \neq i\}$ denote the set of all states in which a_i is an optimal action.

Lemma 4. *We have that* $\lim_{\lambda \to 0^+} I(\lambda; a_i) = \{\mu \in \triangle(\Omega) : \text{supp}(\mu) \subset \Omega(a_i)\}$.

Note that it is possible that $I(0, a_i) = \emptyset$ for some i, but not all. If $\lambda_1 < \lambda_2$, then $I(\lambda_1, a) \subset I(\lambda_2, a)$, which intuitively implies that the regions where not learning is optimal grow (in the set inclusion sense) as information becomes more expensive.[33] Let $\delta_\mu \in \triangle(\triangle(\Omega))$ be the degenerate distribution of posteriors that puts probability 1 on belief μ, and the following Proposition will help us understand the DM's decision when λ is large enough that $\mu_0 \in I(\lambda; a_i)$ for some a_i.

Proposition 5 (Uniqueness of No Learning)**.** *We have that* $\delta_{\mu_0} \in T_{\mu_0}(\lambda)$ *if and only if* $\mu_0 \in I(\lambda; a_i)$ *for some* $a_i \in A$*. Moreover, if* $\mu_0 \in int(I(\lambda; a_i))$*, then* $T_{\mu_0}(\lambda) = \{\delta_{\mu_0}\}$.

Indeed, if $\mu_0 \in I(\lambda; a_i)$, setting $\mu^a = \mu_0$, and $\tau_\lambda(\mu^a) = 1$ automatically satisfies the sufficient and necessary conditions of the solution to the DM's problem. New information is expensive, so the DM prefers to act according to the already available information. In the full proof, we also formalize the intuition that, at any belief $\mu \in I(\lambda; a_i)$, action a_i is optimal. If, furthermore, $\mu \in int(I(\lambda; a_i))$, the DM cannot be indifferent between any two actions at μ (i.e., $a^*(\mu) = \{a_i\}$).

The main contribution of Proposition 5 lies in showing that acquiring no information is the unique solution whenever μ_0 is in the interior of $I(\lambda; a_i)$. This allows us to know the values of the cost of information that induces no learning as a function of the utility of each choice. When $\mu_0 \in int(I(\lambda; a_i))$ for some $\lambda > 0$, the Principal knows the DM will take action a_i with probability 1, and we do not need to consider tie-breaking rules.[34]

We are now equipped to understand how the DM acquires information for any $\lambda \in \mathbb{R}_+$. At $\lambda = 0$, the DM buys a fully informative signal, learns the true state, and takes the

[32]This set is used in Caplin et al. (2019) and Matysková and Montes (2023) to define the (in their case, exogenous) set of prior beliefs in which no-learning is a solution for the DM's problem. Our Proposition 5 shows when it is the unique solution to the DM's problem. We use uniqueness to guarantee that the objective function of the Principal's problem is well behaved. We cannot restrict our attention to the conditions for uniqueness specified in Caplin et al. (2019) Section 4 as those hold only for a fixed λ.

[33]For a formal proof, check Proposition 3 in Matysková and Montes (2023).

[34]Alternatively, there could be other "Principal preferred" solutions where the DM is indifferent between not learning or learning even with a very high λ. Given our tie-breaking rule that favors the Principal, solving the Principal's problem would be much more complicated.

optimal action (this means $\mu^{a_i} \in I(0; a_i)$ for all a_i s.t. $P_\lambda(a_i) > 0$). If instead $\mu_0 \notin \cup_i I(\lambda; a_i)$, the DM will purchase an informative signal and follow the action recommended by the signal realization. Lemma 3 yields the posteriors that follow a signal realization and the unconditional probability that an action is recommended. No matter what happens, for any $\lambda \geqslant 0$, the solution τ_λ satisfies $\text{supp}(\tau_\lambda) \subset \cup_i I(\lambda; a_i)$. Finally, if λ is high enough such that $\mu_0 \in int(I(\lambda; a_i))$ for some action a_i, the DM acquires a completely uninformative signal and chooses action a_i with probability 1.

The last step is to understand what happens as λ grows arbitrarily large. Surprisingly, there are prior beliefs for which, no matter how high the marginal cost of information λ gets, the DM will always optimally choose to acquire some information. Let $C_\mu(A) = \arg\max_{a_i \in A} \mathbb{E}_\mu[u(a_i, \omega)]$ denote the set of optimal actions for the DM given a belief μ.

Proposition 6 (The Limits of Full Obfuscation)**.** *For any $\mu \in int(\triangle(\Omega))$, the following statements are equivalent:*

(i) $a^*(\mu)$ is a singleton;

(ii) there exists $\overline{\lambda} \in \mathbb{R}_{++}$ s.t. $\mu \in int(I(\overline{\lambda}; a_i))$ for some $a_i \in C_\mu(A)$.

Proposition 6 characterizes the prior beliefs for which the DM always learns, no matter the cost. They are exactly the beliefs in which she has more than one optimal action available. We can see why: when she is indifferent between actions, any information, no matter how coarse, will help break the tie. Therefore, even if λ is arbitrarily large, the DM can afford a cheap enough signal (by making it less informative) that is still useful. Consequently, inducing total obfuscation is impossible in these cases, and only in these cases.

Corollary 1. *Full obfuscation is possible, i.e. $\delta_{\mu_0} \in T_{\mu_0}(\lambda)$ for some $\lambda \in \mathbb{R}_+$, if and only if, $a^*(\mu_0)$ is a singleton.*

An important property of the DM's solution is that of Locally Invariant Posteriors (Caplin et al., 2022): the support of the distribution of induced posteriors is the same for any $\mu_0 \notin \cup_i I(\lambda; a_i)$. The prior only affects the solution by altering the distribution of posteriors through Bayes' plausibility. While this is surely still true in our model when looking from the DM's side - i.e., for a fixed λ, - when we consider the Principal, the prior has a much bigger impact on equilibrium outcomes. As we will see in Section 4.2, two different priors, even if both are in $\cup_i I(\lambda; a_i)$, may lead to wildly different obfuscation levels and, consequently, different supports for the distribution of induced posteriors.

4.2 The Principal's Problem

It may be helpful to see that we can rewrite the Principal's indirect utility function $\tilde{V}_{\mu_0}$ as

$$\tilde{V}_{\mu_0}(\lambda) = \sum_{i=1}^{N} P_\lambda(a_i) \sum_{m=1}^{M} \mu^{a_i}(\omega_m) v(a_i, \omega_m). \tag{15}$$

The statement of the Principal's problem then stays the same as in Equation (9). When the Principal has state-independent preferences, we can write $\tilde{V}_{\mu_0}(\lambda) = \sum_{i=1}^{N} P_\lambda(a_i) v(a_i)$. Intuitively, the Principal chooses the λ that maximizes the probability its favorite actions are played, which becomes clearer in Equation (15).

At this point, the assumption that $a^*(\mu_0)$ is a singleton comes into play. If the Principal wants to induce full obfuscation and $a^*(\mu_0)$ is not a singleton, by Proposition 6 that would not be possible. However, since we assume $a^*(\mu_0)$ is a singleton, there exists $\overline{\lambda}$ such that $\mu_0 \in int(I(\lambda; a_i))$ for some $a_i \in A$ and we can then restrict the Principal's choice set to $\mathcal{O} \equiv [0, \overline{\lambda}]$. We are ready to state our next result.

Proposition 7 (Existence of a Solution)**.** *The Principal's Obfuscation Problem has a solution.*

The formal proof is in the Appendix, but it follows the usual arguments. The assumption that $a^*(\mu_0)$ is a singleton ensures that $\mathcal{O}$ is compact and the assumption that ties are broken in the Principal's favor ensures that the maximizing function is upper semicontinuous.

4.3 Transparency and Obfuscation in the General Case

We begin by discussing three classes of games where we can easily determine the equilibrium level of obfuscation. In what follows, let $a_F^*(\mu) = \arg\max_{a \in A} \mathbb{E}_\mu[v(a, \omega)]$ denote the optimal action for the Principal given a belief μ.

Aligned Principal: $\arg\max_{a_i \in A} u(a_i, \omega_m) \subseteq \arg\max_{a_i \in A} v(a_i, \omega_m)$ for all $\omega_m \in \Omega$. With all information available, the DM chooses a Principal-optimal action in every state. Hence, $\lambda = 0$ maximizes the Principal's expected utility and we have *full transparency.* By contraposition, obfuscation can arise only if there exists at least one belief μ in which $a_F^*(\mu) \neq a^*(\mu)$.

Totally misaligned Principal: $u(a_i, \omega_m) + v(a_i, \omega_m) = c$, $\forall a_i, \omega_m$, where c is a constant. We essentially have a zero-sum game: the DM's utility is weakly decreasing in λ, while the Principal's utility is weakly increasing. Therefore, we have full obfuscation.

Agreement at the Prior: the Principal has state-independent preferences ($v(a_i, \omega_m) = v(a_i)$ for all $m = 1, \ldots, M$) and $a^*(\mu_0) \subseteq \arg\max_{a_i \in A} v(a_i)$. Then $\lambda = \overline{\lambda}$ is necessarily optimal, and we have a full obfuscation problem. This case has been discussed earlier in the paper.

As a final remark, consider a paternalistic Principal (for instance, when $v = u$), so that the Principal and the DM agree on the optimal action in every state. Lipnowski et al. (2020) show that full disclosure is optimal if and only if there are two possible states of nature. When the DM exhibits rational inattention, however, the Principal cannot fully disclose the state, because the DM cannot fully process all the information. In contrast, our model allows the Principal to effectively reduce the cost of information acquisition to zero, making full disclosure optimal even when there are more than two states.

4.4 Strategic Obfuscation and Consideration Sets

This section applies our general framework to a well-known example from Caplin et al. (2019, Example 3.1) that illustrates the formation of *consideration sets*. The environment features a DM who must choose the single good option from a menu $A = \{a_1, \ldots, a_N\}$ that also includes $N - 1$ bad options. The state space is $\Omega = \{\omega_1, \ldots, \omega_N\}$, where the realization of ω_m indicates that a_m is the good option. Identifying the state requires costly information acquisition, which is endogenously set by the Principal through its choice of λ. The Principal has a single seller-preferred option,[35] while all other options yield the same payoff for the Principal. The Principal's preferred option may or may not coincide with the DM's preferred option.

In Caplin et al. (2019), the information cost parameter λ is exogenous. Here, we make it a choice variable, allowing us to have a prediction about which options the DM considers. When λ is small, learning is cheap and the DM examines many options; as λ rises, learning becomes harder and her consideration set shrinks. Eventually, when λ is sufficiently high, she focuses only on the option with the highest prior probability of being good and stops learning altogether.

Formally, for a fixed λ, the consideration set consists of the actions chosen with positive probability, $\{a \in A : P_\lambda(a) > 0\}$. The DM's utility is given by $u(a, \omega)$, where a is the chosen action and ω the realized state of the world. Choosing the good (bad) option yields utility u_G (u_B), with $u_G > u_B$. Formally,

$$u(a_i, \omega_m) = \begin{cases} u_G, & \text{if } i = m, \\ u_B, & \text{otherwise.} \end{cases}$$

Given the prior μ_0, we order the states without loss of generality so that $\mu_0(\omega_i) \geqslant \mu_0(\omega_{i+1})$ for all i; that is, option 1 is the most likely to be good, while option N is the least likely.

[35] We can think of an ice cream truck in which the price for all flavors is the same, but the vanilla flavor (the blandest) has a lower marginal cost of production.

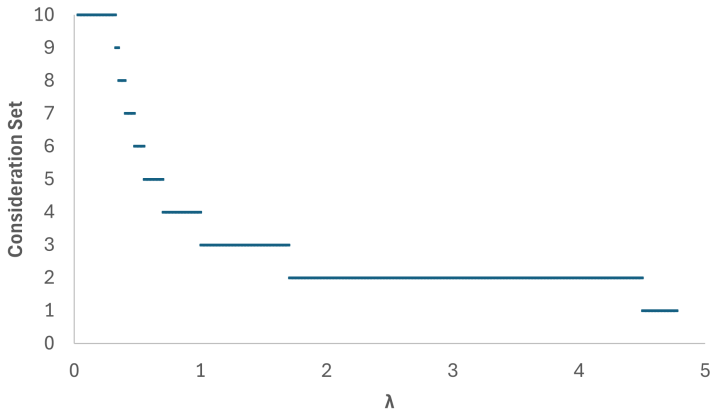

Figure 7: The number of options in the consideration set as a function of λ. For low λ, all 10 options are considered; as λ increases, fewer and fewer options are considered, until around $\lambda \approx 4.5$ when only option 1 is considered and it is bought with probability 1.

Theorem 1 in Caplin et al. (2019) characterizes the optimal choice for the rationally inattentive DM given λ, which in turn defines the corresponding consideration set. To illustrate the impact of λ on the consideration set, Caplin et al. provide the following numerical example: let $u_G = 1, u_B = 0$, and $N = 10$. Moreover, let the prior distribution over the states of the world be $\mu_0(\omega_m) = \alpha 0.8^{m-1}, m = 1, \ldots, 10$ and $\alpha = \frac{1-0.8}{1-0.8^{10}}$. Figure 7 shows how the size of the consideration set varies with the information cost λ.

When the seller-preferred option is 1, the optimal λ induces no-learning: with no new information, the DM goes for the most likely best option according to her prior and buys option 1 with probability 1. When the seller-preferred option is 10, the optimal λ induces full learning. When the Principal wants to sell a product that a DM perceives to be bad, it wants to induce learning to increase the chances of changing her mind. Figure 8 illustrates these results.

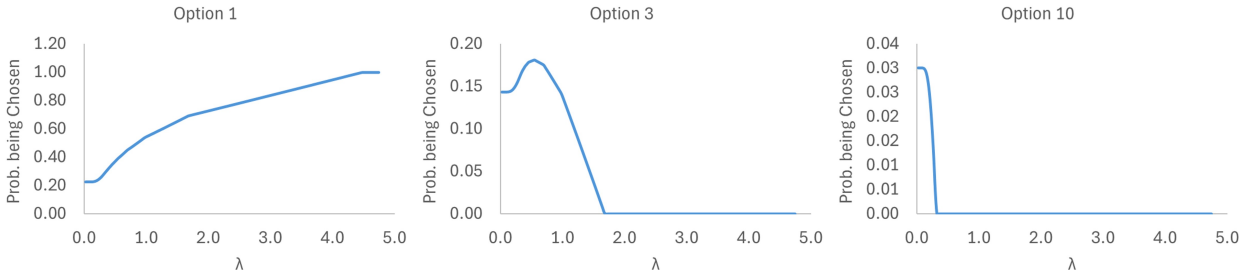

Figure 8: The probability of the DM choosing: (a) option 1, which is increasing in λ; (b) option 3, which is non-monotonic in λ; (c) option 10, which is decreasing in λ.

The cases of option 1 and option 10 illustrate the forces at play. Low λ promotes experimentation, but it enlarges the consideration set and increases the likelihood that

other goods are chosen. This tradeoff generates the non-monotonic relationship between the probability that option 3 is selected and the information cost λ. At first glance, we might expect the optimal λ for selling option 3 to be just low enough to include it in the consideration set while keeping all options $i > 3$ excluded. Interestingly, this is not the case. The value $\lambda \approx 0.55$ that maximizes the probability of selling option 3 induces a consideration set that includes options 4 and 5, as shown in Figure 7. For option 3, allowing for slightly more experimentation — even if this means that additional goods enter the consideration set — remains beneficial.

Two final remarks are in order. First, if the prior were uniform ($\mu_0(\omega_i) = 1/N$ for all ω_i), the consideration set would remain constant and include all options for all values of λ. Proposition 6 tells us that, in this case, the Principal could not induce full obfuscation. Second, from a DM welfare perspective, the best case is when option 10 is seller-preferred as it induces full information and the DM always gets the good option.

This setup reveals how the Principal's preferred product determines the equilibrium level of obfuscation. If the Principal's favored product already enjoys a high prior probability, raising λ—and thus discouraging search—benefits the Principal. In contrast, if the preferred product has a low prior probability, the Principal gains from lowering λ to encourage information acquisition. In some intermediate cases, the relationship between λ and the probability of sale is non-monotonic: a moderate amount of obfuscation can increase sales, but excessive obfuscation drives the DM to rely on her prior and ignore other options. This analysis connects strategic obfuscation to the emergence of limited consideration and choice overload.

5 Generalizing the Information Cost Function

In this section, we comment on how we can generalize our model using the definitions of cost functions introduced in Caplin et al. (2022), and on which results survive this generalization. For simplicity, we focus on the 2x2 case in which the Principal has state-independent preferences. We assume $v(a_H) > v(a_L)$, and again use the normalization $u(a_H, \omega_H) = B,\ u(a_L, \omega_H) = 0,\ u(a_H, \omega_L) = -C,\ u(a_L, \omega_L) = 0$.

Definition 6 (Posterior-Separable Costs, Caplin et al. (2022))**.** *A posterior-separable cost function $c_\lambda : \triangle(\triangle(\Omega)) \to \mathbb{R}$ with parameter $\lambda \in \mathbb{R}_+$ is weakly uniformly posterior-separable if, for each subset $\overline{\Omega} \subset \Omega$, there exists a strictly convex function $S_{\overline{\Omega}} : \triangle(\overline{\Omega}) \to \overline{\mathbb{R}}$ such that, for all $\mu_0 \in \triangle(\Omega)$, and Bayes' plausible $\tau \in \triangle(\triangle(\Omega))$,*

$$c_\lambda(\mu_0, \tau) = \lambda \left(\sum_{\mu \in \operatorname{supp} \tau} \tau(\mu) S_{\operatorname{supp} \mu_0}(\mu) - S_{\operatorname{supp} \mu_0}(\mu_0) \right). \tag{16}$$

The original definition does not have the λ parameter scaling the function, but the change does not affect it in any meaningful way.[36] We made the change so that the Principal's choice is explicit in the cost function. The proposition below, for which we omit the formal proof, shows that full obfuscation and agreement at the prior is a general property of cost functions.

Proposition 8 (Full Obfuscation, Generalized)**.** *In the 2x2 case with the cost function as in Equation* (16)*, we have full obfuscation if, and only if, we have agreement at the prior.*

This is intuitive. If $a^*(\mu_0)$ is the optimal action for the Principal, full obfuscation gives its maximal payoff. If not, then full obfuscation induces the worst action for the Principal with probability 1, so $\lambda = 0$ is strictly better as the prior is interior. Because agreement at the prior still holds, and attains the Bayesian Persuasion optimum, the results from part (i) of Proposition 4 follow immediately in this setting as well.

The following proposition generalizes the result that if $\mu_0 = \mu_{ind}$, then the DM purchases an informative signal for all $\lambda \in \mathbb{R}_+$, so full obfuscation is not feasible.

Proposition 9 (Indifference Blocks Obfuscation)**.** *Assume that $c_\lambda(\mu_0, \tau)$ is given as in Equation* (16) *and that $S_{\operatorname{supp}\mu_0}$ is differentiable at μ_{ind}. If $\mu_0 = \mu_{ind}$, then $\delta_{\mu_0} \notin T_{\mu_0}(\lambda)$ for any $\lambda \in \mathbb{R}_+$.*

One may consider how special partial obfuscation is in this general environment. Are there cost functions other than Entropy that allow for it? The next result gives a positive answer.

Proposition 10 (Robustness to Cost Functions)**.** *Assume that we have a Skeptical DM, i.e., $\mu_0 < \frac{C}{B+C} < \frac{1}{2}$, and let $K \subset (0,1)$ be a compact interval that contains the posteriors induced by partial obfuscation. Then there exists a neighborhood $\mathcal{N}$ of the Shannon entropy cost function S_E in the $C^2(K)$ topology such that for every $S \in \mathcal{N}$, the corresponding cost*

$$c^S_\lambda(\mu_0, \tau) = \lambda\big(\mathbb{E}_\tau[S_{\operatorname{supp}\mu_0}(\mu)] - S_{\operatorname{supp}\mu_0}(\mu_0)\big),$$

also admits a partial obfuscation solution.

This result leads to the conclusion that partial obfuscation is not a "knife's edge" result that depends on the choice of Entropy as the cost function.

6 Conclusion

This paper develops a model of persuasion and influence that operates through the design of learning frictions rather than signal transmission. Our framework shows that a Principal can

[36] Indeed, if we define, for $\lambda > 0$, $T'_{\operatorname{supp}\mu_0}(\mu) = T_{\operatorname{supp}\mu_0}(\mu)/\lambda$, we get the usual representation.

systematically influence a Decision Maker's choices by strategically choosing the marginal cost of learning. Consequently, our approach serves as a bridge between the rational inattention and strategic communication literatures.

Our central contribution demonstrates that rational agents may frequently choose to *partially* shroud information, challenging the binary paradigm of full transparency versus total obfuscation. This insight has direct implications for consumer protection policy: regulatory interventions that mandate disclosure without restricting complexity may be entirely ineffective. Conversely, our results highlight a counterintuitive policy lever. For Skeptical consumers, artificially increasing the cost of consumer mistakes reduces the firm's incentive to obfuscate, ultimately making the consumer better off in equilibrium. We illustrate several mechanisms by which this outcome arises—such as asymmetries in consumer preferences—and discuss economic settings in which such behavior is likely to occur, including firm competition and politics.

We also provide evidence that our results apply to a broader family of information cost specifications. In particular, partial obfuscation is robust to a family of cost functions in a neighborhood of the entropy, and our Online Appendix B shows that allowing for costly obfuscation does not alter the economic intuition we provide.

From a theoretical point of view, our approach sits between Bayesian Persuasion and Cheap Talk. We relax the strong assumption that a Principal can commit to any arbitrary posterior distribution. Instead, we highlight an alternative mechanism of persuasion, focused strictly on the design of information costs, that can still outperform Cheap Talk in different environments.

The framework has a few natural extensions that could be the object of fruitful future research. First, one can study what happens when the choice of the information cost λ is made after the Principal observes the state of the world, i.e., when the Principal has no commitment power. This would turn the choice of λ into a signal in itself. Second, future work could explore the implications of multiple Principals choosing λ simultaneously, or a single Principal setting a uniform λ for heterogeneous decision-makers. Finally, the dynamic implications of the problem, for instance, how a firm can influence consideration sets over time, remain an open question.

References

Alonso, R. and O. Camara (2024). Organizing data analytics. *Management Science 70*(5), 3123–3143.

Asriyan, V., D. Foarta, and V. Vanasco (2023). The good, the bad, and the complex: Product

design with imperfect information. *American Economic Journal: Microeconomics 15*(2), 187–226.

Barros, L. (2026). The limits of information acquisition in cheap talk. *working paper*.

Ben-Shahar, O. and C. E. Schneider (2014). *More Than You Wanted to Know*. Princeton University Press.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Bizzotto, J., J. Rüdiger, and A. Vigier (2020). Testing, disclosure and approval. *Journal of Economic Theory 187*, 105002.

Bloedel, A. W. and I. Segal (2021). Persuading a rationally inattentive agent. *Available at SSRN*.

Boyacı, T. and Y. Akçay (2018). Pricing when customers have limited attention. *Management Science 64*(7), 2995–3014.

Brown, Z. Y. and J. Jeon (2024). Endogenous information and simplifying insurance choice. *Econometrica 92*(3), 881–911.

Caplin, A., M. Dean, and J. Leahy (2019). Rational inattention, optimal consideration sets, and stochastic choice. *Review of Economic Studies 86*(3), 1061–1094.

Caplin, A., M. Dean, and J. Leahy (2022). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy 130*(6), 1676–1715.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431.

Cusumano, C. M., F. Fabbri, and F. Pieroth (2024). Competing to commit: Markets with rational inattention. *American Economic Review 114*(1), 285–306.

De Clippel, G. and K. Rozen (2025). Communication, Perception and Strategic Obfuscation. working paper.

De Oliveira, H., T. Denti, M. Mihm, and K. Ozbek (2017). Rationally inattentive preferences and hidden information costs. *Theoretical Economics 12*(2), 621–654.

Dean, M. and N. Neligh (2023). Experimental tests of rational inattention. *Journal of Political Economy 131*(12), 3415–3461.

Edmond, C. and Y. K. Lu (2021). Creating confusion. *Journal of Economic Theory 191*, 105145.

Ellis, A. (2018). Foundations for optimal inattention. *Journal of Economic Theory 173*, 56–94.

Ellison, G. (2005). A Model of Add-On Pricing. *The Quarterly Journal of Economics 120*(2), 585–637.

Ellison, G. and A. Wolitzky (2012). A search cost model of obfuscation. *RAND Journal of*

Economics 43(3), 417–441.

Fabbri, F. (2024). Attention holdup. *Working Paper*.

Fréchette, G. R., A. Lizzeri, and J. Perego (2022). Rules and commitment in communication: An experimental analysis. *Econometrica 90*(5), 2283–2318.

Gabaix, X. and D. Laibson (2006). Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *The Quarterly Journal of Economics 121*(2), 505–540.

Gentzkow, M. and E. Kamenica (2014). Costly Persuasion. *American Economic Review 104*(5), 457–462.

Janssen, A. and J. Kasinger (2024). Obfuscation and rational inattention. *The Journal of Industrial Economics 72*(1), 390–428.

Jin, G. Z., M. Luca, and D. Martin (2022). Complex disclosure. *Management Science 68*(5), 3236–3261.

Kalayci, K. and J. Potters (2011). Buyer confusion and market prices. *International Journal of Industrial Organization 29*(1), 14–22.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics 11*(1).

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Le Treust, M. and T. Tomala (2019). Persuasion with limited communication capacity. *Journal of Economic Theory 184*, 104940.

Li, H. and X. Shi (2017). Discriminatory information disclosure. *American Economic Review 107*(11), 3363–85.

Lin, X. and C. Liu (2024). Credible persuasion. *Journal of Political Economy 132*(7), 2228–2273.

Lipnowski, E., L. Mathevet, and D. Wei (2020). Attention management. *American Economic Review: Insights 2*(1), 17–32.

Lipnowski, E. and D. Ravid (2020). Cheap talk with transparent motives. *Econometrica 88*(4), 1631–1660.

Lipnowski, E., D. Ravid, and D. Shishkin (2022). Persuasion via weak institutions. *Journal of Political Economy 130*(10), 2705–2730.

Martin, D. (2017). Strategic pricing with rational inattention to quality. *Games and Economic Behavior 104*, 131–145.

Matějka, F. (2016). Rationally inattentive seller: Sales and discrete pricing. *Review of Economic Studies 83*(3), 1125–1155.

Matějka, F. and A. McKay (2012). Simple market equilibria with rationally inattentive consumers. *American Economic Review 102*(1), 24–29.

Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–298.

Matysková, L. and A. Montes (2023). Bayesian persuasion with costly information acquisition. *Journal of Economic Theory 211*, 105678.

Min, D. (2021). Bayesian persuasion under partial commitment. *Economic Theory 72*, 743–764.

Nguyen, A. and T. Y. Tan (2021). Bayesian persuasion with costly messages. *Journal of Economic Theory 193*, 105212.

Nocke, V. and P. Rey (2024). Consumer search, steering, and choice overload. *Journal of Political Economy 132*(5), 1684–1739.

Perez-Richet, E. and V. Skreta (2022). Test design under falsification. *Econometrica 90*(3), 1109–1142.

Ravid, D. (2020). Ultimatum bargaining with rational inattention. *American Economic Review 110*(9), 2948–63.

Richards, T. J., G. J. Klein, C. Bonnet, and Z. Bouamra-Mechemache (2020). Strategic obfuscation and retail pricing. *Review of Industrial Organization 57*, 859–889.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal 27*(4), 623–656.

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

Wei, D. (2020). Persuasion under costly learning. *Journal of Mathematical Economics*, 1–26.

Appendix

Proof of Proposition 1. Assume V is concave (the case for V convex is analogous). We restrict our analysis to $\lambda \in [0, \overline{\lambda}]$ as the function is otherwise constant. In this interval, $\tilde{V}_{\mu_0}(\lambda) = m(\lambda) * \mu_0 + V(\underline{\mu}(\lambda)) - m(\lambda) * \underline{\mu}(\lambda)$. We also know that: (i) in $\mathbb{R}^2$, $\tilde{V}_{\mu_0}(\lambda)$ is equal to the value of the *y-coordinate* of the point at the intersection of the line segment formed by connecting the points $(\underline{\mu}(\lambda), V(\underline{\mu}(\lambda)))$ and $(\overline{\mu}(\lambda), V(\overline{\mu}(\lambda)))$, and a vertical line passing through μ_0; (ii) $\underline{\mu}$ and $\overline{\mu}$ are strictly increasing and decreasing respectively. This means that, for $0 \leqslant \lambda_1 < \lambda_2 \leqslant \overline{\lambda}$, $0 \leqslant \underline{\mu}(\lambda_1) < \underline{\mu}(\lambda_2) \leqslant \mu_0 \leqslant \overline{\mu}(\lambda_2) < \overline{\mu}(\lambda_1) \leqslant 1$, with at most one equality for $\underline{\mu}(\lambda_2) \leqslant \mu_0 \leqslant \overline{\mu}(\lambda_2)$.

Thus, if we prove the general result that, for $a < c \leqslant \mu_0 \leqslant d < b$ (again, with at least one strict inequality), and a concave function $f : \mathbb{R}_+ \to \mathbb{R}$, where $[a, b] \subset \mathbb{R}_+$, the intersection of the line segment connecting $(a, f(a))$ and $(b, f(b))$ has an equal or lower value than the line segment connecting points $(c, f(c))$ and $(d, f(d))$ when both are evaluated at μ_0, we are done. To put it simply, we are considering that f represent $\tilde{V}_{\mu_0}$; $[a, b]$ represents $[\underline{\mu}(\lambda_1), \overline{\mu}(\lambda_1)]$, and $[c, d]$ represents $[\underline{\mu}(\lambda_2), \overline{\mu}(\lambda_2)]$ for arbitrary $\lambda_1 < \lambda_2$.

Let $L1 : [a, b] \to \mathbb{R}$ and $L2 : [c, d] \to \mathbb{R}$ be given by the following formulas.

$$L1(x) : \left(\frac{f(b) - f(a)}{b - a}\right) x + f(a) - \left(\frac{f(b) - f(a)}{b - a}\right) a,$$
$$L2(x) : \left(\frac{f(d) - f(c)}{d - c}\right) x + f(c) - \left(\frac{f(d) - f(c)}{d - c}\right) c.$$

Clearly, $L1$ and $L2$ are linear in x (they represent the line segments mentioned before). Evaluating $L1$ and $L2$ at c, we get

$$L1(c) : (f(b) - f(a)) \frac{(c - a)}{(b - a)} + f(a),$$
$$L2(c) : f(c).$$

But since $c \in (a, b)$, there exists $t_c \in (0, 1)$ such that $c = t_c b + (1 - t_c)a$. We then get that

$$\frac{(c - a)}{(b - a)} = \frac{t_c b + a - t_c a - a}{b - a} = t_c \frac{b - a}{b - a} = t_c,$$

and so:

$$(f(b) - f(a)) \frac{(c - a)}{(b - a)} + f(a) = (f(b) - f(a)) t_c + f(a). \tag{17}$$

Finally, since f is concave,

$$L2(c) = f(c) = f(t_c b + (1-t_c)a) \geqslant t_c f(b) + (1-t_c)f(a) = (f(b) - f(a))\, t_c + f(a) = L1(c).$$

The same argument can be made for the point d. Since $d \in (a,b)$, there exists $t_d \in (0,1)$ such that $d = t_d b + (1-t_d)a$, and we get that:

$$L2(d) = f(d) = f(t_d b + (1-t_d)a) \geqslant t_d f(b) + (1-t_d)f(a) = (f(b) - f(a))\, t_d + f(a) = L1(d)$$

And so, $L1(c) \leqslant L2(c)$ and $L1(d) \leqslant L2(d)$. Since the line segments are by definition linear, they are on the same line or they intersect at most once: $L2$ is never below $L1$ in the interval $[c,d]$. Finally, as $\mu_0 \in [c,d]$, then $L2(\mu_0) \geqslant L1(\mu_0)$. As mentioned before, $L2(\mu_0)$ represent the value of $\tilde{V}_{\mu_0}$ at λ_2 and $L1(\mu_0)$ the value of $\tilde{V}_{\mu_0}$ at λ_1, and we get that $\tilde{V}_{\mu_0}$ is increasing in λ. □

Proof of Lemma 2. First, note that since the DM's preferences are symmetric, $\underline{\mu}(\lambda) = 1 - \overline{\mu}(\lambda)$ for all λ. This can be checked by noting that $\mu_{ind} = 0.5$ implies that $u(a_H, \omega_H) - u(a_L, \omega_H) = u(a_L, \omega_L) - u(a_H, \omega_L)$ and using substitution in Equations 5 and 6. Given that the Principal's preferences are state-independent, V is a step function, so for any belief $\mu < \mu_{ind}$, $V(\mu) = V(0)$ and for any belief $\mu > \mu_{ind}$, $V(\mu) = V(1)$. By assumption, $v(a_H) = V(1) > V(0) = v(a_L)$. Now, fix $\lambda_1, \lambda_2 \in \mathcal{O} = [0, \overline{\lambda}]$ (as outside this interval the function is constant) with $\lambda_1 < \lambda_2$, and let $d \coloneqq \mu_0 - \underline{\mu}(\lambda_1)$, $D \coloneqq \overline{\mu}(\lambda_1) - \mu_0$. By triangle similarity,

$$\frac{d}{V(\underline{\mu}(\lambda_1)) - \tilde{V}_{\mu_0}(\lambda_1)} = \frac{D}{\tilde{V}_{\mu_0}(\lambda_1) - V(\overline{\mu}(\lambda_1))}.$$

Since $\underline{\mu}(\lambda_1) < \mu_{ind}$ and $\overline{\mu}(\lambda_1) > \mu_{ind}$, we have $V(\underline{\mu}(\lambda_1)) = V(0)$ and $V(\overline{\mu}(\lambda_1)) = V(1)$. We can then write

$$\frac{d}{V(0) - \tilde{V}_{\mu_0}(\lambda_1)} = \frac{D}{\tilde{V}_{\mu_0}(\lambda_1) - V(1)} \implies \tilde{V}_{\mu_0}(\lambda_1) = \frac{D}{d+D}V(0) + \frac{d}{d+D}V(1). \tag{18}$$

Now, let $\delta \coloneqq \underline{\mu}(\lambda_2) - \underline{\mu}(\lambda_1)$. Since DM's preferences are Symmetric, $\overline{\mu}(\lambda_1) - \overline{\mu}(\lambda_2) = \delta$, and we can write

$$\tilde{V}_{\mu_0}(\lambda_2) = \frac{D-\delta}{d+D-2\delta}V(0) + \frac{d-\delta}{d+D-2\delta}V(1). \tag{19}$$

But note that, by construction, $\delta < \frac{d+D}{2}$, and so the following holds:

$$\frac{D-\delta}{d+D-2\delta} > \frac{D}{d+D} \iff d < D.$$

We now have two cases. If $D > d$, from Equation (18) to Equation (19), we are increasing the proportion of $V(0)$ and since $V(1) > V(0)$, $\tilde{V}_{\mu_0}(\lambda_1) > \tilde{V}_{\mu_0}(\lambda_2)$ and the problem has a boundary solution as $\tilde{V}_{\mu_0}$ is decreasing. If $d > D$, from Equation (18) to Equation (19) we are decreasing the proportion of $V(0)$ and since $V(1) > V(0)$, this means that $\tilde{V}_{\mu_0}(\lambda_1) < \tilde{V}_{\mu_0}(\lambda_2)$ and $\tilde{V}_{\mu_0}$ is increasing. □

Proof of Theorem 1. For the proof of part (i), we start by showing sufficiency. Our goal is to find a λ that yields a higher payoff for the Principal than either $\lambda = 0$ or $\lambda = \overline{\lambda}$. Again, we assume $v(a_H) = V(1) > V(0) = v(a_L)$, as the other case follows a symmetric argument. Therefore, the goal of the Principal is to maximize the probability that the DM chooses a_H, which is given by

$$p_H(\lambda) = \frac{\mu_0 - \underline{\mu}(\lambda)}{\overline{\mu}(\lambda) - \underline{\mu}(\lambda)},$$

through Bayes' plausibility. If we have a Skeptical DM, $\mu_0 < \mu_{ind} < \frac{1}{2}$, and the DM would choose a_L if they act according to the prior. This means that $\underline{\mu}(\overline{\lambda}) = \mu_0$ and $p_H(\overline{\lambda}) = 0$. We also have that $p_H(0) = \mu_0 > 0$, so we want to show that there exists $\lambda \in (0, \overline{\lambda})$ such that $p_H(\lambda) > \mu_0$. Using the expression for $p_H(\lambda)$, we have

$$\frac{\mu_0 - \underline{\mu}(\lambda)}{\overline{\mu}(\lambda) - \underline{\mu}(\lambda)} > \mu_0 \iff \underline{\mu}(\lambda) < \frac{\mu_0}{1 - \mu_0}\big(1 - \overline{\mu}(\lambda)\big). \tag{20}$$

Define $q \equiv \frac{\mu_0}{1-\mu_0}$. Thus it is enough to find $\lambda \in (0, \bar{\lambda})$ such that $\frac{\underline{\mu}(\lambda)}{1-\overline{\mu}(\lambda)} < q$. Let $D \coloneqq u(a_H, \omega_H) - u(a_L, \omega_H) > 0$ and $d \coloneqq u(a_L, \omega_L) - u(a_H, \omega_L) > 0$. Then $\mu_{\text{ind}} = \frac{d}{D+d}$, and since $\mu_{ind} < \frac{1}{2}$, $d < D$. Using the functional forms for $\underline{\mu}(\lambda)$ and $\overline{\mu}(\lambda)$, we get

$$\underline{\mu}(\lambda) = \frac{1 - e^{-d/\lambda}}{e^{D/\lambda} - e^{-d/\lambda}}, \text{ and } 1 - \overline{\mu}(\lambda) = \frac{1 - e^{-D/\lambda}}{e^{d/\lambda} - e^{-D/\lambda}}.$$

Therefore,

$$\frac{\underline{\mu}(\lambda)}{1 - \overline{\mu}(\lambda)} = \frac{e^{d/\lambda} - 1}{e^{D/\lambda} - 1}.$$

Since $d < D$, we have $\lim_{\lambda \to 0^+} \frac{\underline{\mu}(\lambda)}{1-\overline{\mu}(\lambda)} = 0$. Because $q > 0$, we deduce the existence of $\lambda^* \in (0, \bar{\lambda})$ small enough such that $\frac{\underline{\mu}(\lambda^*)}{1-\overline{\mu}(\lambda^*)} < q$. Hence $p_H(\lambda^*) > \mu_0$, and therefore $\tilde{V}_{\mu_0}(\lambda^*) > \tilde{V}_{\mu_0}(0) > \tilde{V}_{\mu_0}(\bar{\lambda})$. Thus neither boundary solves the Principal's problem, and the solution is partial obfuscation.

Necessity is proved case by case. We show that whenever the DM is not Skeptical, the

solution is a boundary solution. If $\mu_0 > \mu_{ind}$, then the DM chooses a_H at the prior. Since a_H is the Principal's preferred action, this is Agreement at the Prior. Full obfuscation gives payoff $V(1)$, which is the maximal payoff the Principal can obtain. Hence $\overline{\lambda}$ solves the Principal's problem, and we have a boundary solution (full obfuscation). It remains to consider the case when $\mu_0 < \mu_{ind}$, but the DM is not Skeptical, so $\mu_{ind} \geqslant \frac{1}{2}$. Equivalently, $d \geqslant D$, and so

$$\frac{e^{d/\lambda} - 1}{e^{D/\lambda} - 1} \geqslant \frac{d}{D} \text{ for all } \lambda > 0,$$

as $x \mapsto \frac{e^{x/\lambda}-1}{x}$ is increasing for $x > 0$. For any $\lambda \in (0, \overline{\lambda})$, we want to show that $p_H(\lambda) < \mu_0$, which holds if, and only if, $\frac{\underline{\mu}(\lambda)}{1-\overline{\mu}(\lambda)} \geqslant \frac{\mu_0}{1-\mu_0}$, as in Equation (20). But

$$\mu_0 < \mu_{ind} \implies \frac{\mu_0}{1-\mu_0} < \frac{\mu_{ind}}{1-\mu_{ind}} = \frac{d}{D} \leqslant \frac{e^{d/\lambda} - 1}{e^{D/\lambda} - 1} = \frac{\underline{\mu}(\lambda)}{1-\overline{\mu}(\lambda)}.$$

Therefore, $\implies p_H(\lambda) < \mu_0$ for any $\lambda \in (0, \overline{\lambda})$, and this is not a problem of partial obfuscation.

We now prove part (ii). Assume we have agreement at the prior. With a slight abuse of notation, let $\overline{a} = a^*(\mu_0)$ represent the favorite action for the Principal, and we have

$$\tilde{V}_{\mu_0}(\lambda) = P_\lambda(a_H)v(a_H) + P_\lambda(a_L)v(a_L) \leqslant v(\overline{a}) \text{ for every } \lambda.$$

Since we have agreement at the prior, $\tilde{V}_{\mu_0}(\overline{\lambda}) = v(\overline{a})$, and $\overline{\lambda}$ solves the Principal's problem. Now, assume we have full obfuscation, i.e., $\overline{\lambda}$ is a solution to the Principal's problem. As before, assume $v(a_H) > v(a_L)$, or equivalently, that $V(1) > V(0)$. Then, $\{a_H\} = \arg\max_{a \in A} v(a)$. We need to show that $\{a_H\} = a^*(\mu_0) = \arg\max_{a \in A} \mathbb{E}_{\mu_0}[u(a, \omega)]$. Since we have full obfuscation,

$$\begin{aligned} & \tilde{V}_{\mu_0}(\overline{\lambda}) \geqslant \tilde{V}_{\mu_0}(\lambda) \ \forall \lambda \in \mathbb{R}_+ \\ \implies & V(\mu_0) \geqslant \tilde{V}_{\mu_0}(\lambda) \ \forall \lambda \in \mathbb{R}_+ \\ \implies & V(\mu_0) \geqslant \tilde{V}_{\mu_0}(0) = \mu_0 V(1) + (1-\mu_0)V(0) \end{aligned} \tag{21}$$

With state-independent preferences, $V(\mu) \in \{V(0), V(1)\}$ for all $\mu \in \triangle(\Omega)$, so $V(\mu_0) \in \{V(0), V(1)\}$. As μ_0 is interior and $V(1) > V(0)$, Equation (21) then implies that $V(\mu_0) = V(1) = v(a_H)$, which in turn implies that $a^*(\mu_0) = \{a_H\}$ and we have agreement at the prior.

The only other possible case is of full transparency, which gives us (iii). □

Lemma 5 (Normalization)**.** *In the 2×2 case with $v(a_H) > v(a_L)$, the DM's problem and the*

Principal's obfuscation choice are unchanged when we normalize payoffs to

$$u'(a_H, \omega_H) = B, \quad u'(a_L, \omega_H) = 0, \quad u'(a_H, \omega_L) = -C, \quad u'(a_L, \omega_L) = 0,$$

where $B = u(a_H, \omega_H) - u(a_L, \omega_H)$, *and* $C = u(a_L, \omega_L) - u(a_H, \omega_L)$. *Moreover, the DM's ex-ante welfare is unchanged up to a constant that is independent of* λ *and* ε.

Proof of Lemma 5. Subtract $u(a_L, \omega_H)$ from the payoff of both actions in state ω_H, and $u(a_L, \omega_L)$ from the payoff of both actions in state ω_L. This does not change the rank of actions at any belief, so the DM's problem is not affected. The belief of indifference remains the same, and therefore, the Principal's problem also does not change. The DM's ex-ante welfare is offset by the constant $\mu_0 u(a_L, \omega_H) + (1 - \mu_0)u(a_L, \omega_L)$, which is independent of λ and ε. Therefore all welfare comparative statics in relation to λ and ε are unchanged. □

Proof of Theorem 2. Assume that we have a Skeptical DM. Now, define $\underline{\varepsilon} = (u(a_H, \omega_H) + u(a_H, \omega_L)) - (u(a_L, \omega_L) + u(a_L, \omega_H)) = B - C$ and note that since $\mu_{ind} < \frac{1}{2}$, $\underline{\varepsilon} > 0$. If we decrease $u(a_H, \omega_L)$ by any $\varepsilon \geqslant \underline{\varepsilon}$, Theorem 1 tells us that we no longer have a partial obfuscation problem as now $\mu_{ind} \geqslant \frac{1}{2}$. Hence, we only need to check which boundary case we are at. If the Principal sets $\lambda = \overline{\lambda}$, the DM will choose the Principal's least favorite action, a_L, with probability 1 as $\mu_0 < \mu_{ind}$. Setting $\lambda = 0$ yields $\mu_0(\omega_H)v(a_H) + \mu_0(\omega_L)v(a_L)$, which is a strictly better expected payoff for the Principal. It then chooses full transparency, which yields the first-best payoff u_0^* for the DM.

Now, assume we have Agreement at the Prior with $v(a_H) > v(a_L)$. Then, $\mu_{ind} < \mu_0$. Take

$$\varepsilon = \frac{\mu_0}{1 - \mu_0}[u(a_H, \omega_H) - u(a_L, \omega_H)] - [u(a_L, \omega_L) - u(a_H, \omega_L)] = \frac{\mu_0}{1 - \mu_0}B - C \geqslant 0,$$

and since $\mu_0 > \mu_{ind}$, $\varepsilon > 0$. Since the new belief of indifference is higher than μ_0, we do not have agreement at the prior. To rule out the possibility of a Skeptical DM, we must have $\mu_{ind} \geqslant \frac{1}{2}$ which happens when $\varepsilon \geqslant B - C$. Then, if we set

$$\underline{\varepsilon} = \max\{\frac{\mu_0}{1 - \mu_0}B - C, B - C\}$$

we get full transparency for any $\varepsilon > \underline{\varepsilon}$, and the DM's new equilibrium payoff is u_0^*. □

Proof of Proposition 3. For $\lambda \in (0, \overline{\lambda})$, can rewrite Equation (10) as

$$\tilde{V}_{\mu_0}(\lambda) = \frac{\mu_0 - \overline{\mu}(\lambda)}{\underline{\mu}(\lambda) - \overline{\mu}(\lambda)}[v(a_L) - v(a_H)] + v(a_H).$$

We assume, without loss, that $v(a_H) > v(a_L)$. Considering Equations (5) and (6), we get that

$$\frac{d\tilde{V}_{\mu_0}(\lambda)}{d\lambda} = \frac{(v(a_H) - v(a_L))}{4\lambda^2}\Bigg[(-1+\mu_0)(u(a_H,\omega_H) - u(a_L,\omega_H))\operatorname{csch}^2\left(\frac{u(a_H,\omega_H) - u(a_L,\omega_H)}{2\lambda}\right) + \mu_0\left(-u(a_H,\omega_L) + u(a_L,\omega_L)\right)\operatorname{csch}^2\left(\frac{u(a_H,\omega_L) - u(a_L,\omega_L)}{2\lambda}\right)\Bigg].$$

where csch denotes the hyperbolic cosecant. Note that on the LHS the term outside the brackets is always positive and that $\operatorname{csch}(-x) = -\operatorname{csch}(x)$. Rearranging and using our normalization, we have

$$\frac{d\tilde{V}_{\mu_0}(\lambda)}{d\lambda} \gtrless 0 \iff \operatorname{csch}^2\left(\frac{C}{2\lambda}\right) \gtrless \frac{1-\mu_0}{\mu_0}\frac{B}{C}\operatorname{csch}^2\left(\frac{B}{2\lambda}\right).$$

In what follows, let

$$G(\lambda) \equiv \frac{\operatorname{csch}^2\left(\frac{C}{2\lambda}\right)}{\operatorname{csch}^2\left(\frac{B}{2\lambda}\right)} = \left[\frac{\sinh\left(\frac{B}{2\lambda}\right)}{\sinh\left(\frac{C}{2\lambda}\right)}\right]^2,$$

and $K \equiv \frac{1-\mu_0}{\mu_0}\frac{B}{C}$, so that

$$\operatorname{sgn}\left(\frac{d\widetilde{V}_{\mu_0}(\lambda)}{d\lambda}\right) = \operatorname{sgn}\left(G(\lambda) - K\right).$$

Moreover,

$$\frac{d}{d\lambda}\log G(\lambda) = \frac{1}{\lambda^2}\left[C\coth\left(\frac{C}{2\lambda}\right) - B\coth\left(\frac{B}{2\lambda}\right)\right].$$

The function $x \mapsto x\coth x$ is strictly increasing on $(0,\infty)$. Therefore, G is strictly decreasing when $B > C$ and strictly increasing when $B < C$. If $B = C$, then $G(\lambda) = 1$ for every $\lambda > 0$. In this last case, $K = \frac{1-\mu_0}{\mu_0} \neq 1$, because $\mu_0 < \mu_{ind} < 1/2$ by assumption. Consequently, in every case the equation $G(\lambda) = K$ has at most one solution. It follows that there exists at most one $\lambda^* \in (0, \overline{\lambda})$ satisfying

$$\frac{d\tilde{V}_{\mu_0}(\lambda^*)}{d\lambda} = 0.$$

□

Lemma 6. *Fix a Principal's obfuscation problem and assume we have partial obfuscation, so*

$\lambda^*(C) \in (0, \overline{\lambda})$. *Then,* $\lambda^*(C)$ *is continuously differentiable, and*

$$\frac{d\lambda^*(C)}{dC} < 0.$$

Proof of Lemma 6. By Proposition 3, the equilibrium $\lambda^*(C)$ is unique, and since we have a Skeptical agent, $\lambda^*(C) > 0$. Using our normalization and the fact that $\operatorname{csch}^2(x) = \operatorname{csch}^2(-x)$, we can simplify the Principal's FOC to

$$\frac{d\tilde{V}_{\mu_0}(\lambda)}{d\lambda} = \frac{v(a_H) - v(a_L)}{4\lambda^2}\left[-B(1-\mu_0)\operatorname{csch}^2\left(\frac{B}{2\lambda}\right) + C\mu_0 \operatorname{csch}^2\left(\frac{C}{2\lambda}\right)\right].$$

Since, by assumption, $v(a_H) - v(a_L) > 0$, at the optimum $\lambda^*(C)$, the FOC implies

$$\begin{aligned} &-B(1-\mu_0)\operatorname{csch}^2\left(\frac{B}{2\lambda^*(C)}\right) + C\mu_0 \operatorname{csch}^2\left(\frac{C}{2\lambda^*(C)}\right) = 0 \\ \implies &\operatorname{csch}^2\left(\frac{C}{2\lambda^*(C)}\right) = \frac{1-\mu_0}{\mu_0}\frac{B}{C}\operatorname{csch}^2\left(\frac{B}{2\lambda^*(C)}\right). \end{aligned}$$

Using the fact that $\operatorname{csch}(x) = 1/\sinh(x)$ and simplifying,

$$\frac{\sinh\left(\frac{B}{2\lambda^*(C)}\right)}{\sinh\left(\frac{C}{2\lambda^*(C)}\right)} = \sqrt{\frac{1-\mu_0}{\mu_0}\frac{B}{C}}. \tag{22}$$

Now, let

$$h(\lambda, C) \coloneqq 2\log\left(\sinh\left(\frac{B}{2\lambda}\right)\right) - 2\log\left(\sinh\left(\frac{C}{2\lambda}\right)\right) - \log\left(\left(\frac{1-\mu_0}{\mu_0}\frac{B}{C}\right)\right),$$

and the FOC is reduced to $h(\lambda^*(C), C) = 0$. If we differentiate in each case, we get

$$\begin{aligned} h_\lambda(\lambda, C) &= \frac{1}{\lambda^2}\left[C\coth\left(\frac{C}{2\lambda}\right) - B\coth\left(\frac{B}{2\lambda}\right)\right], \text{ and} \\ h_C(\lambda, C) &= \frac{1}{C} - \frac{1}{\lambda}\coth\left(\frac{C}{2\lambda}\right). \end{aligned}$$

Note that $x \mapsto x\coth x$ is strictly increasing on $(0, \infty)$, and since $B > C$, $h_\lambda < 0$. Moreover, letting $x = C/(2\lambda) > 0$, we obtain

$$h_C(\lambda, C) = \frac{1 - 2x\coth x}{C} < 0$$

as $x\coth x \geqslant 1$. Therefore, the Implicit Function Theorem implies that $\lambda^*(C)$ is continuously

differentiable and

$$\frac{d\lambda^*(C)}{dC} = -\frac{h_C(\lambda^*(C), C)}{h_\lambda(\lambda^*(C), C)} < 0.$$

□

Lemma 7. *Under the assumptions of Lemma 6, let $W^*(C)$ denote the DM's ex-ante equilibrium welfare. If we have a Skeptical DM, then*

$$\frac{dW^*(C)}{dC} = -(1-\mu_0)\mathbb{P}_{\lambda,C} + MI^*(C)\left|\frac{d\lambda^*(C)}{dC}\right|, \tag{23}$$

where, given (λ, C), $\mathbb{P}_{\lambda,C}(a_H|\omega_L)$ is the equilibrium probability that the DM chooses a_H given ω_L, and $MI^*(C) = H(\mu_0) - \mathbb{E}_{\mu\sim\tau^*_{\lambda,C}}[H(\mu)]$ is the equilibrium mutual information acquired by the DM.

Proof. Under our new normalization, $a^*(\mu) = a_H$ if $\mu \geqslant \frac{C}{B+C}$, and $a^*(\mu) = a_L$ otherwise. Therefore, for a fixed pair (λ, C), the indirect utility of consumption for the DM is $\max\{0, (B+C)\mu - C\}$, so the DM's welfare is calculated by

$$\begin{aligned} W(\lambda, C) \coloneqq \max_{\tau}\ & \mathbb{E}_{\mu\sim\tau}[\max\{0, (B+C)\mu - C\}] - \lambda(H(\mu_0) - \mathbb{E}_{\mu\sim\tau}[H(\mu)]), \\ \text{s.t. }\ & E_{\mu\sim\tau}[\mu] = \mu_0. \end{aligned}$$

Let $\tau^*_{\lambda,C}$ be an optimal experiment. We again denote the supported elements by $\underline{\mu}$ ($\overline{\mu}$) when the recommended action is a_L (a_H). We then get

$$W(\lambda, C) = \tau^*_{\lambda,C}(\overline{\mu})[(B+C)\overline{\mu} - C] - \lambda(H(\mu_0) - (\tau^*_{\lambda,C}(\overline{\mu})H(\overline{\mu}) + \tau^*_{\lambda,C}(\underline{\mu})H(\underline{\mu})))$$

By the envelope theorem[37], $W_C(\lambda, C) = \tau^*_{\lambda,C}(\overline{\mu})(\overline{\mu} - 1)$. But $\overline{\mu} = \mathbb{P}_{\lambda,C}(\omega_H \mid a_H)$. Therefore, $1 - \overline{\mu} = \mathbb{P}_{\lambda,C}(\omega_L \mid a_H)$, and since $\tau^*_{\lambda,C}(\overline{\mu}) = \mathbb{P}_{\lambda,C}(a_H)$, the unconditional probability that a_H is chosen given (λ, C), we get

$$\tau^*_{\lambda,C}(\overline{\mu})(1-\overline{\mu}) = \mathbb{P}_{\lambda,C}(a_H)\mathbb{P}_{\lambda,C}(\omega_L \mid a_H) = \mathbb{P}_{\lambda,C}(a_H, \omega_L).$$

Therefore,

$$W_C(\lambda, C) = -\mathbb{P}_{\lambda,C}(a_H, \omega_L) = -(1-\mu_0)\mathbb{P}_{\lambda,C}(a_H \mid \omega_L).$$

[37]Note that the maximizing function is smooth at the posteriors when we have partial obfuscation.

Using the envelope theorem again, now for λ, gives

$$W_\lambda(\lambda, C) = -(H(\mu_0) - \mathbb{E}_{\mu\sim\tau^*_{\lambda,C}}[H(\mu)]).$$

Now, let $W(C) \coloneqq W(\lambda^*(C), C)$ denote the DM's change in welfare as a function of C. We then have, from Lemma 6,

$$\begin{aligned}
\frac{dW^*(C)}{dC} &= W_C(\lambda^*(C), C) + W_\lambda(\lambda^*(C), C)\frac{d\lambda^*(C)}{dC} \\
&= -(1-\mu_0)\mathbb{P}_{\lambda,C}(a_H \mid \omega_L) - (H(\mu_0) - \mathbb{E}_{\mu\sim\tau^*_{\lambda,C}}[H(\mu)])\frac{d\lambda^*(C)}{dC} \\
&= -(1-\mu_0)\mathbb{P}_{\lambda,C}(a_H \mid \omega_L) + (H(\mu_0) - \mathbb{E}_{\mu\sim\tau^*_{\lambda,C}}[H(\mu)])\left|\frac{d\lambda^*(C)}{dC}\right| \\
&= -(1-\mu_0)\mathbb{P}_{\lambda,C}(a_H \mid \omega_L) + MI^*(C)\left|\frac{d\lambda^*(C)}{dC}\right|
\end{aligned}$$

□

Proof of Proposition 2. Let $A = \frac{1-\mu_0}{\mu_0}$ and $C_0 = \frac{\mu_0}{1-\mu_0}B$. Note that $C \in (C_0,\ C_0+\eta) \cup (B-\eta, B)$ means that, even though we have a Skeptical DM, we are close to full transparency (if $C \in (C_0, C_0+\eta)$) or to full obfuscation (if $C \in (B-\eta, B)$). We first show that there exist $\eta_1, \eta_2 > 0$ such that $\frac{dW^*(C)}{dC} > 0$ whenever $C \in (C_0, C_0+\eta_1)$ or $C \in (B-\eta_2, B)$. By Lemma 7, it is sufficient to check the values of terms on the right-hand side of Equation (23) near the boundaries.

Consider the case where $C \to B^-$. We know from Equation (22) in the proof of Lemma 6 that as $C \to B^-$, the right-hand side converges to $\sqrt{A} > 1$ (as $\mu_0 < \frac{1}{2}$). This implies $\lambda^*(C) \to 0$. Since $\sinh(x) \simeq \frac{1}{2}e^x$ as $x \to \infty$, we get

$$\lambda^*(C) \simeq \frac{B-C}{\log A} \implies \frac{d\lambda^*(C)}{dC} \to -\frac{1}{\log A}.$$

But as $\lambda^*(C) \to 0$, the DM gets full information and the mistake probability goes away, i.e., $\mathbb{P}(a_H \mid \omega_L) \to 0$. Moreover, $MI^*(C) \to H(\mu_0) > 0$. Therefore,

$$\lim_{C\to B^-} \frac{dW^*(C)}{dC} = \frac{H(\mu_0)}{\log A} > 0.$$

By continuity, there exists $\eta_1 > 0$ such that $\frac{dW^*(C)}{dC} > 0$ for all $C \in (B-\eta_1, B)$.

Now consider the case when $C \to C_0$. Let $r \coloneqq \frac{B}{C} \in (1, A)$ and $z \coloneqq \frac{C}{2\lambda^*(C)}$, so that

$\lambda^*(C) = \frac{B}{2rz}$.[38] The Principal's first-order condition becomes

$$\frac{\sinh(rz)}{\sinh z} = \sqrt{rA}.$$

As $C \to C_0$, we have $r \to A$. We simplify the FOC using a Taylor expansion around $z = 0$, which yields

$$A - r = \frac{A(A^2-1)}{3} z^2 + O(z^4).$$

Moreover, differentiating the first-order condition in (r, z) coordinates gives

$$\frac{d\lambda^*(C)}{dC} = \frac{\frac{1}{2} - z\coth z}{2z^2\big(r\coth(rz) - \coth z\big)}.$$

Further known simplifications $z\coth z = 1 + \frac{z^2}{3} + O(z^4)$ and $r\coth(rz) - \coth z = \frac{r^2-1}{3} z + O(z^3)$ give

$$\left|\frac{d\lambda^*(C)}{dC}\right| = \frac{3}{4(r^2-1)} z^{-3} + O(z^{-1}). \tag{24}$$

We can rewrite $\underline{\mu}, \overline{\mu}$ using r, z as

$$\underline{\mu} = \frac{1 - e^{-2z}}{e^{2rz} - e^{-2z}}, \text{ and } \overline{\mu} = \frac{e^{2rz}(1 - e^{-2z})}{e^{2rz} - e^{-2z}}.$$

A Taylor expansion then gives

$$\overline{\mu} - \underline{\mu} = \frac{2r}{r+1} z + O(z^2) \to \frac{2A}{A+1} z + O(z^2) \text{ as } C \to C_0. \tag{25}$$

Hence, there exists $\iota > 0$ such that $\overline{\mu} - \underline{\mu} \geqslant \iota z$.

To ease notation, let $p \coloneqq \mathbb{P}(a_H)$ denote the unconditional probability that a_H is chosen given a solution to the DM's problem. Bayes plausibility then implies that $\lim_{z\to 0} p \in (0, 1)$, so $p(1-p)$ is bounded away from zero,[39] so there exists $\kappa > 0$ such that $p(1-p) \geqslant \kappa$. Let

[38] Note that as $C \to C_0$, $r \to A$ and $z \to 0$.

[39] Indeed, we can write the Taylor expansions for $\overline{\mu}$ and $\underline{\mu}$ around $z = 0$ and $r = A$ to get

$$\overline{\mu} = \mu_0 + \kappa_1 z + O(z^2) \text{ and } \underline{\mu} = \mu_0 - \kappa_2 z + O(z^2),$$

where $\kappa_1, \kappa_2 > 0$. Substituting into the formula for p then implies $p \to \frac{\kappa_1}{\kappa_1 + \kappa_2} > 0$.

$D(p||q)$ denote the Kullback–Leibler divergence between distributions p, q. Then,

$$MI^*(C) = pD(\overline{\mu}\|\mu_0) + (1-p)D(\underline{\mu}\|\mu_0),$$

Pinsker's inequality gives $D(q\|p) \geqslant 2(q-p)^2$, and, together with Bayes' plausibility, we get

$$\begin{aligned} MI^*(C) &\geqslant 2\left[p(\overline{\mu}-\mu_0)^2 + (1-p)(\underline{\mu}-\mu_0)^2\right] \\ &= 2p(1-p)(\overline{\mu}-\underline{\mu})^2 \geqslant 2\kappa(\iota z)^2 \geqslant cz^2 \end{aligned}$$

by Equation (25) for some $c > 0$ if C is close enough to C_0. Note also that $\lambda^*(C) \to \infty$.[40] Together with Equation (24), all yields

$$MI^*(C)\left|\frac{d\lambda^*(C)}{dC}\right| \geqslant c' z^{-1} \to \infty \text{ as } C \to C_0 \implies z \to 0.$$

Since $0 \leqslant (1-\mu_0)\Pr^*(a_H \mid \omega_L) \leqslant 1$, we get that $\frac{dW^*(C)}{dC} > 0$ for all C sufficiently close to C_0. Thus there exists $\eta_2 > 0$ such that $\frac{dW^*(C)}{dC} > 0$ for all $C \in (C_0, C_0 + \eta_2)$. Finally, setting $\eta \coloneqq \min\{\eta_1, \eta_2\}$ yields the first part of the proof.

Now, note that if the DM is Skeptical, the equilibrium remains interior and the relevant objects vary continuously with the parameters. Hence $\varepsilon \mapsto W^*(C+\varepsilon)$ is differentiable in a neighborhood of 0. Since W^* is differentiable at C and its derivative at C is strictly positive, there exists $\delta > 0$ such that $W^*(C+\varepsilon) > W^*(C)$ for all $\varepsilon \in (0, \delta)$. Therefore, for all such ε, $W^*(C+\varepsilon) > W^*(C)$. □

Proof of Proposition 4. We start by proving (i). Note that in Bayesian Persuasion, the Principal maximizes its utility by choosing a distribution of posteriors in the set

$$BP(\mu_0) \coloneqq \{\tau \in \triangle(\triangle(\Omega)) : \sum_{supp(\tau)} \mu\tau(\mu) = \mu_0\}.$$

So, given commitment, the Sender is only restricted by Bayes plausibility. Under Strategic Obfuscation, the Principal maximizes its utility in the set

$$SO(\mu_0) \coloneqq \{\tau \in \triangle(\triangle(\Omega)) : \exists \lambda \in \mathcal{O} \text{ with } supp(\tau) \subseteq \{\underline{\mu}(\lambda), \overline{\mu}(\lambda)\} \wedge \sum_{supp(\tau)} \mu\tau(\mu) = \mu_0\}.$$

We can clearly see that $SO(\mu_0) \subseteq BP(\mu_0)$, and so the Principal in Strategic Obfuscation cannot do better than the Sender in Bayesian Persuasion.

For the second statement of part (i), suppose we have Agreement at the Prior, with

[40] As $C \to C_0$, we get $\mu_{ind} \to \mu_0$, which implies $\overline{\lambda} \to \infty$, so $\lambda^*(C) \in [0, \overline{\lambda}]$ still holds.

$v(a_H) > v(a_L)$. Then, $s^o = \tilde{V}_{\mu_0}(\overline{\lambda}) = V(\mu_0) = v(a_H)$. Under BP, the optimal signal is fully uninformative, so $s^P = V(\mu_0) = v(a_H)$. Now, assume we do not have Agreement at the Prior. Then, $V(\mu_0) = v(a_L)$. Under BP, the optimal split of posteriors would put positive weight on the belief of indifference, which, as we have seen, is not reached under SO for any $\lambda \in \mathbb{R}_+$. Therefore, the concavization result is not reached.

For the proof of (ii), we go back to the case of state-independent preferences for the Principal. We repeat the claims in Proposition 2, and for any belief $\mu < \mu_{ind}$, $V(\mu) = V(0)$ and for any belief $\mu > \mu_{ind}$, $V(\mu) = V(1)$. Assume that $V(0) > V(1)$ (the other case follows a similar argument). From Lipnowski and Ravid (2020), we know that an outcome (τ, s) is an equilibrium outcome if, and only if, $\tau \in \triangle(\triangle(\Omega))$ satisfies Bayes plausibility and $s \in \cap_{\mu \in supp(\tau)} V(\mu)$. But, given the state-independent preferences, we have that $\forall F \in 2^{\triangle(\Omega)}$, the following holds:

$$\begin{cases} \cap_{\mu \in F} V(\mu) & = \{V(0)\} \text{ if } F \subseteq [0, \mu_{ind}], \\ \cap_{\mu \in F} V(\mu) & = \{V(1)\} \text{ if } F \subseteq (\mu_{ind}, 1], \\ \cap_{\mu \in F} V(\mu) & = \emptyset \text{ otherwise.} \end{cases}$$

And so the only possible payoffs for the Sender in a Cheap Talk equilibrium are $V(0)$ or $V(1)$. When adding Bayes Plausibility, we have one of two cases. If $\mu_0 < \mu_{ind}$, the only possible payoff is $V(0)$, and this can be achieved under Strategic Obfuscation by setting $\lambda = \overline{\lambda}$ (since $\mu_0 < \mu_{ind}$, the decision maker will take the optimum action for the Principal if no new information is acquired); if $\mu_0 > \mu_{ind}$, $V(1)$ is the only achievable payoff in equilibrium for the Sender, i.e., the DM chooses the worst action for the Sender with probability 1. But in Strategic Obfuscation the Principal can always get a better payoff by setting $\lambda = 0$. Since μ_0 is interior, by fully revealing the state the optimal action for the Principal is expected to be chosen with strictly positive probability. □

The next results pertain to the more general case of the model, as described in Section 4. The following Lemmas help understand how the comparative statics of the solution to the DM's problem change as λ varies.

Lemma 4. *We have that* $\lim_{\lambda \to 0^+} I(\lambda; a_i) = \{\mu \in \triangle(\Omega) : \text{supp}(\mu) \subset \Omega(a_i)\}$.

Proof of Lemma 4. Fix $\tilde{\mu} \in \triangle(\Omega)$ such that $\text{supp}(\tilde{\mu}) \subset \Omega(a_i)$. Then,

$$\sum_{m=1}^{M} \tilde{\mu}(\omega_m) \exp\left(\frac{u(a_j, \omega_m) - u(a_i, \omega_m)}{\lambda}\right) = \sum_{\omega \in \Omega(a_i)} \tilde{\mu}(\omega) \exp\left(\frac{u(a_j, \omega) - u(a_i, \omega)}{\lambda}\right).$$

But, for any $\omega \in \Omega(a_i)$,

$$\exp\left(\frac{u(a_j,\omega_m) - u(a_i,\omega_m)}{\lambda}\right) \leqslant 1 \text{ for any } \lambda > 0.$$

Therefore, $\tilde{\mu} \in I(\lambda, a_i)$ for all $\lambda > 0$, which implies $\tilde{\mu} \in \lim_{\lambda\to 0^+} I(\lambda, a_i)$. Now, fix $\tilde{\mu} \in \lim_{\lambda\to 0^+} I(\lambda, a_i)$ and assume, by way of contradiction, that $\tilde{\mu} \notin \{\mu \in \triangle(\Omega) : supp(\tilde{\mu}) \subset \Omega(a_i)\}$. Then, $\exists \omega_m \in \Omega$ and $a_j \in A$ such that $\tilde{\mu}(\omega_m) > 0$, and $u(a_j,\omega_m) > u(a_i,\omega_m)$. But then,

$$\lim_{\lambda\to 0^+} \exp\left(\frac{u(a_j,\omega_m) - u(a_i,\omega_m)}{\lambda}\right) > 1. \text{ Contradiction.}$$

□

Lemma 8. *Fix $a_i \in A$ and $\lambda \geqslant 0$. Then, $\mu \in I(\lambda, a_i) \implies a_i \in a^*(\mu)$.*

Proof of Lemma 8. The result when $\lambda = 0$ is immediate, so let $\lambda > 0$. We prove by taking the contrapositive. Let $\mu \in \triangle(\Omega)$ be such that exists action a_j with $\sum_{j=1}^M \mu(\omega_j)(u(a_j,\omega_j) - u(a_i,\omega_j)) > 0$, so that $a_i \notin a^*(\mu)$. To ease notation, let

$$g_\lambda(a_i, a_j, \omega_j) \equiv \frac{u(a_j,\omega_j) - u(a_i,\omega_j)}{\lambda}, \text{ and}$$
$$\varphi(x) \equiv e^x, \text{ for } x \in \mathbb{R}.$$

We then have that $\mathbb{E}_\mu[g_\lambda(a_i, a_j, \omega_j)] > 0 \implies \varphi(\mathbb{E}_\mu[g_\lambda(a_i, a_j, \omega_j)]) > 1$. By Jensen's Inequality,

$$\sum_{j=1}^M \mu(\omega_j) \exp(g_\lambda(a_i, a_j, \omega_j)) = \mathbb{E}_\mu[\varphi(g_\lambda(a_i, a_j, \omega_j))] \geqslant \varphi(\mathbb{E}_\mu[g_\lambda(a_i, a_j, \omega_j)]) > 1.$$

Therefore, $\mu \notin I(\lambda; a_i)$. □

Lemma 9. *For any prior μ_0 and $\lambda \geqslant 0$, the set of solutions for the DM's problem is convex.*

Proof of Lemma 9. Fix μ_0, $\lambda \geqslant 0$, and the corresponding induced set of solutions to the DM's problem. If the set is a singleton we are done. So let τ^1 and τ^2 be different solutions, and fix $\alpha \in (0,1)$. We define $\tau^3 \in \triangle(\triangle(\Omega))$ as $\tau^3(\mu) = \alpha\tau^1(\mu) + (1-\alpha)\tau^2(\mu)$. Clearly,

$$\mu \in \text{supp}(\tau^3) \iff \tau^3(\mu) > 0 \iff \alpha\tau^1(\mu) + (1-\alpha)\tau^2(\mu) > 0 \iff [\tau^1(\mu) > 0] \vee [\tau^2(\mu) > 0],$$

and thus $\text{supp}(\tau^3) = \text{supp}(\tau^1) \cup \text{supp}(\tau^2)$. We then get

$$\begin{aligned}
\sum_{a \in A(\tau^3)} \tau^3(\mu^a)\mu^a(\omega) &= \sum_{a \in A(\tau^3)} \left(\alpha\tau^1(\mu^a) + (1-\alpha)\tau^2(\mu^q)\right)\mu^a(\omega) \\
&= \alpha \sum_{a \in A(\tau^1)} \tau^1(\mu^a)\mu^a(\omega) + (1-\alpha) \sum_{a \in A(\tau^2)} \tau^2(\mu^a)\mu^a(\omega) \\
&= \alpha\mu_0(\omega) + (1-\alpha)\mu_0(\omega) \text{ (since } \tau^1 \text{ and } \tau^2 \text{ are solutions)} \\
&= \mu_0(\omega) \text{ for all } \omega.
\end{aligned}$$

Therefore, τ^3 satisfies Bayes' plausibility. Furthermore,

$$\begin{aligned}
&\mathbb{E}_{\tau^3}[U(\mu^a)] - \lambda(H(\mu_0) - \mathbb{E}_{\tau^3}[H(\mu^a)]) \\
&\quad= \sum_{a \in A(\tau^3)} \tau^3(\mu^a)U(\mu^a) - \lambda\left(H(\mu_0) - \sum_{a \in A(\tau^3)} \tau^3(\mu^a)H(\mu^a)\right) \\
&\quad= \alpha\left(\sum_{a \in A(\tau^1)} \tau^1(\mu^a)U(\mu^a) - \lambda\left(H(\mu_0) - \sum_{a \in A(\tau^1)} \tau^1(\mu^a)H(\mu^a)\right)\right) \\
&\qquad+ (1-\alpha)\left(\sum_{a \in A(\tau^2)} \tau^2(\mu^a)U(\mu^a) - \lambda\left(H(\mu_0) - \sum_{a \in A(\tau^2)} \tau^2(\mu^a)H(\mu^a)\right)\right) \\
&\quad= \alpha\left(\mathbb{E}_{\tau^1}[U(\mu^a)] - \lambda(H(\mu_0) - \mathbb{E}_{\tau^1}[H(\mu^a)])\right) \\
&\qquad+ (1-\alpha)\left(\mathbb{E}_{\tau^2}[U(\mu^a)] - \lambda(H(\mu_0) - \mathbb{E}_{\tau^2}[H(\mu^a)])\right) \\
&\quad= \mathbb{E}_{\tau^1}[U(\mu^a)] - \lambda(H(\mu_0) - \mathbb{E}_{\tau^1}[H(\mu^a)]).
\end{aligned}$$

The last equality follows from the fact that both τ^1 and τ^2 are solutions. Therefore, τ^3 is also a solution to the DM's problem. □

Corollary 2. *Let τ^1 and τ^2 solve the DM's Problem. Then, Equation* (12) *holds for all $a_i, a_j \in A(\tau^1) \cup A(\tau^2)$.*

Proof of Corollary 2. Fix $\alpha \in (0,1)$. By Lemma 9, τ^3 defined by $\tau^3(\mu) = \alpha\tau^1(\mu) + (1-\alpha)\tau^2(\mu)$ for all μ is also a solution. Necessity implies that Equation (12) holds for any $a_i, a_j \in A(\tau^3) = A(\tau^1) \cup A(\tau^2)$. □

Proof of Proposition 5. That $\delta_{\mu_0} \in T_{\mu_0}(\lambda)$, or equivalently, that $\tau_\lambda(\mu_0) = 1$ is a solution if and only if $\mu_0 \in I(\lambda; a_i)$ is a direct application of Lemma 3: by construction $\sum_{\mu^a \text{supp } \tau_\lambda} \tau_\lambda(\mu_0)\mu^a(\omega) = \mu_0(\omega)$; then, for the necessary part, if $\mu_0 \notin \cup_i I(\lambda; a_i)$, it cannot satisfy (13); for the sufficiency, $\mu_0 \in I(\lambda; a_i)$ implies that Equation (13) is satisfied, and

$\tau_\lambda(\mu_0) = 1$ implies that Equation (12) is vacuously satisfied as $A(\tau_\lambda) = \{a_i\}$. Lemma 8 shows that a_i is the recommended action for the uninformative signal, and a_i is chosen with probability 1.

The rest of the proof shows that the solution is unique whenever $\mu_0 \in int(I(a, \lambda)$, so suppose this statement holds. Since μ_0 is assumed to be interior, $\lambda > 0$. Now, suppose towards a contradiction, that there exists another solution $\tau \in \triangle(\Omega)$ such that $\tau(\mu_0) < 1$. Therefore, there must be another action $a_j \neq a_i$ that is recommended with positive probability ($a_j \in A(\tau)$) and induces posterior μ^{a_j}. By Corollary 2, we must have

$$\frac{\mu_0(\omega)}{\exp\left(\frac{u(a_i,\omega)}{\lambda}\right)} = \frac{\mu^{a_j}(\omega)}{\exp\left(\frac{u(a_j,\omega)}{\lambda}\right)}, \text{ for all } \omega. \tag{26}$$

But $\mu_0 \in int(I(\lambda; a_i))$, and thus

$$\sum_{m=1}^{M} \mu_0(\omega_m) \frac{\exp(u(a_j, \omega_m)/\lambda)}{\exp(u(a_i, \omega_m)/\lambda)} < 1.$$

Substituting according to Equation (26), gives us

$$\sum_{m=1}^{M} \mu^{a_j}(\omega_m) \frac{\exp(u(a_i, \omega_m)/\lambda)}{\exp(u(a_j, \omega_m)/\lambda)} \frac{\exp(u(a_j, \omega_m)/\lambda)}{\exp(u(a_i, \omega_m)/\lambda)} < 1,$$

$$\implies \sum_{m=1}^{M} \mu^{a_j}(\omega_m) < 1. \text{ Contradiction.}$$

□

Proof of Proposition 6. Fix $\mu \in int(\triangle(\Omega))$. We begin by showing that $(i) \implies (ii)$. Assume $a^*(\mu)$ is a singleton, and let $i \in \{1, \ldots, N\}$ be such that $a^*(\mu) = \{a_i\}$. Therefore,

$$\sum_{m=1}^{M} \mu(\omega_m)[u(a_j, \omega_m) - u(a_i, \omega_m)] < 0 \text{ for all } j \neq i. \tag{27}$$

For the rest of the proof, we fix $j \neq i$ and let $z(m) = u(a_j, \omega_m) - u(a_i, \omega_m)$. Now, let $\overline{m} = \arg\min_m z(m)$. We can rewrite Equation (27) as

$$\sum_{m \neq \overline{m}} \mu(\omega_m) z(m) + \mu(\omega_{\overline{m}}) z(\overline{m}) < 0 \iff \sum_{m \neq \overline{m}} \mu(\omega_m) z(m) + \left(1 - \sum_{m \neq \overline{m}} \mu(\omega_m)\right) z(\overline{m}) < 0$$

$$\iff \sum_{m \neq \overline{m}} \mu(\omega_m)[z(m) - z(\overline{m})] < -z(\overline{m}).$$

Take $\tilde{m} \neq \overline{m}$ such that $z(\tilde{m}) > z(\overline{m})$.[41] We then get

$$\sum_{m \neq \tilde{m}, \overline{m}} \mu(\omega_m)[z(m) - z(\overline{m})] + \mu(\omega_{\tilde{m}})[z(\tilde{m}) - z(\overline{m})] < -z(\overline{m})$$

$$\iff \mu(\omega_{\tilde{m}}) < \underbrace{\frac{-z(\overline{m})}{[z(\tilde{m}) - z(\overline{m})]}}_{T_0} - \sum_{m \neq \tilde{m}, \overline{m}} \mu(\omega_m) \underbrace{\frac{[z(m) - z(\overline{m})]}{[z(\tilde{m}) - z(\overline{m})]}}_{T_m}. \tag{28}$$

Now take a belief $\gamma \in \triangle(\Omega)$ and for any $\lambda \in \mathbb{R}_{++}$,

$$\gamma \in I(\lambda; a_i) \iff \sum_{m=1}^{M} \gamma(\omega_m) \exp(z(m)/\lambda) \leqslant 1$$

$$\iff \sum_{m \neq \overline{m}} \gamma(\omega_m) \exp(z(m)/\lambda) + \left(1 - \sum_{m \neq \overline{m}} \gamma(\omega_m)\right) \exp(z(\overline{m})/\lambda) \leqslant 1$$

$$\iff \sum_{m \neq \overline{m}} \gamma(\omega_m)[\exp(z(m)/\lambda) - \exp(z(\overline{m})/\lambda)] + \gamma(\omega_{\tilde{m}}[\exp(z(\tilde{m})/\lambda) - \exp(z(\overline{m})/\lambda)] \leqslant 1 - \exp(z(\overline{m})/\lambda)$$

$$\iff \sum_{m \neq \tilde{m}, \overline{m}} [\exp(z(m)/\lambda) - \exp(z(\overline{m})/\lambda)] + \gamma(\omega_{\tilde{m}})[\exp(z(\tilde{m})/\lambda) - \exp(z(\overline{m})/\lambda)] \leqslant 1 - \exp(z(\overline{m})/\lambda)$$

$$\iff \gamma(\omega_{\tilde{m}}) \leqslant \underbrace{\frac{1 - \exp(z(\overline{m})/\lambda)}{[\exp(z(\tilde{m})/\lambda) - \exp(z(\overline{m})/\lambda)]}}_{\Gamma_0} - \sum_{m \neq \tilde{m}, \overline{m}} \gamma(\omega_m) \underbrace{\frac{[\exp(z(m)/\lambda) - \exp(z(\overline{m})/\lambda)]}{[\exp(z(\tilde{m})/\lambda) - \exp(z(\overline{m})/\lambda)]}}_{\Gamma_m}.$$

$$\iff \gamma(\omega_{\tilde{m}}) \leqslant \Gamma_0 - \sum_{m \neq \tilde{m}, \overline{m}} \gamma(\omega_m) \Gamma_m \tag{29}$$

But note that

$$\Gamma_0 = \frac{1 - \exp(z(\overline{m})/\lambda)}{[\exp(z(\tilde{m})/\lambda) - \exp(z(\overline{m})/\lambda)]} \xrightarrow{\lambda \to \infty} T_0 = \frac{-z(\overline{m})}{[z(\tilde{m}) - z(\overline{m})]},$$

and, for any $m \neq \tilde{m}, \overline{m}$ we get,

$$\Gamma_m = \frac{[\exp(z(m)/\lambda) - \exp(z(\overline{m})/\lambda)]}{[\exp(z(\tilde{m})/\lambda) - \exp(z(\overline{m})/\lambda)]} \xrightarrow{\lambda \to \infty} T_m = \frac{[z(m) - z(\overline{m})]}{[z(\tilde{m}) - z(\overline{m})]}.$$

Thus, for any $\delta > 0$, we then deduce the existence of $\overline{\lambda}$ large enough such that, for $m \neq \tilde{m}, \overline{m}$ we get

$$|T_0 - \Gamma_0| < \frac{\delta}{M-1}, \text{ and } |T_m - \Gamma_m| < \frac{\delta}{M-1}. \tag{30}$$

[41] We know such $\tilde{m}$ exists since $\overline{m}$ minimizes $z(\cdot)$ and we cannot have $z(m) = z(\overline{m})$ for all m, otherwise a_j would be strictly dominated by a_i in all states, which is ruled out by assumption.

Going back to Equation (28), there exists $\delta > 0$ such that $\mu(\omega_{\tilde{m}}) + \delta < T_0 + \sum_{m \neq \tilde{m}, \overline{m}} \mu(\omega_m) T_m$. We then get

$$\begin{aligned}
\mu(\omega_{\tilde{m}}) &< T_0 - \frac{\delta}{M-1} - \sum_{m \neq \tilde{m}, \overline{m}} \left[\mu(\omega_m) T_m + \frac{\delta}{M-1} \right] \\
&= T_0 - \frac{\delta}{M-1} - \sum_{m \neq \tilde{m}, \overline{m}} \mu(\omega_m) \left[T_m + \frac{\delta}{M-1} \right] - \sum_{m \neq \tilde{m}, \overline{m}} (1 - \mu(\omega_m)) \frac{\delta}{M-1} \\
&\leqslant \left(T_0 - \frac{\delta}{M-1} \right) - \sum_{m \neq \tilde{m}, \overline{m}} \mu(\omega_m) \left[T_m + \frac{\delta}{M-1} \right] \\
&\leqslant \Gamma_0 - \sum_{m \neq \tilde{m}, \overline{m}} \mu(\omega_m) \Gamma_m.
\end{aligned}$$

Where the last inequality comes from Equation (30). Therefore, by Equation (29), $\mu \in I(a_i, \overline{\lambda})$. We show that $(ii) \implies (i)$ by taking the contrapositive. Assume $a^*(\mu)$ is not a singleton, so that for two different actions a_i, a_j we get

$$\mathbb{E}_\mu[u(a_j, \omega) - u(a_i, \omega)] = 0 \implies \mathbb{E}_\mu \left[\frac{u(a_j, \omega) - u(a_i, \omega)}{\lambda} \right] = 0 \ \ \forall \lambda > 0.$$

We again denote by $\varphi : \mathbb{R} \to \mathbb{R}_+$ the exponential function: $\varphi(x) = e^x$ for any $x \in \mathbb{R}$. Then, by strict Jensen's inequality

$$\mathbb{E} \left[\varphi \left(\frac{u(a_j, \omega) - u(a_i, \omega)}{\lambda} \right) \right] > \varphi \left(\mathbb{E}_\mu \left[\frac{u(a_j, \omega) - u(a_i, \omega)}{\lambda} \right] \right) = \varphi(0) = 1 \ \ \forall \lambda > 0.$$

Therefore, $\mu \notin I(a, \lambda)$ for any $a \in a^*(\mu)$ and $\lambda > 0$. By Lemma 8, $\mu \notin I(a_k, \lambda)$ for any $a_k \in A \setminus a^*(\mu)$. □

Proof of Proposition 7. Since we assume that $a^*(\mu_0)$ is a singleton, Proposition 6 implies that there exists $\overline{\lambda} > 0$ such that, for every $\lambda \geqslant \overline{\lambda}$, the DM's optimal choice is to acquire no information and choose $a^*(\mu_0)$. Hence, the Principal's payoff is constant for all $\lambda \geqslant \overline{\lambda}$. It is, therefore, without loss to restrict the Principal's choice set to the compact real interval $[0, \overline{\lambda}]$.

For this proof, we will use the stochastic-choice representation of the DM's problem instead of working with the distribution of posteriors.[42] Remember that $\pi(a|\omega)$ represents the probability that a DM chooses action a given that the state is ω. Now, define

$$\Pi \coloneqq \{\pi : A \times \Omega \to [0, 1] : \sum_{a \in A} \pi(a|\omega) = 1\}.$$

[42]For a review, check Kamenica and Gentzkow (2011) and Matějka and McKay (2015).

Since A and Ω are finite, Π is compact. Given $\pi \in \Pi$, we calculate the unconditional probability that action $a \in A$ is chosen as $P_\pi(a) = \sum_\omega \mu_0(\omega)\pi(a|\omega)$. For a fixed π, remember that the probability of a given posterior μ_π^a, that follows a signal realization a, is equal to the unconditional probability that a is chosen. Therefore, the associated entropy-based cost of information is given by $c(\pi) = H(P_\pi) - \sum_{\omega\in\Omega} \mu_0(\omega)H(\pi(\cdot|\omega))$. Since the entropy function is continuous, $c : \Pi \to \mathbb{R}_+$ is continuous.

Now, we rewrite the DM's objective function $f_{\mu_0} : \mathbb{R} \times \Pi \to \mathbb{R}$ as

$$f_{\mu_0}(\lambda, \pi) = \sum_{\omega\in\Omega} \sum_{a\in A} \mu_0(\omega)\pi(a \mid \omega)u(a, \omega) - \lambda c(\pi),$$

which is continuous in both λ and π. Therefore, by the maximum theorem, the solution correspondence

$$T(\lambda) = \arg\max_{\pi\in\Pi} f(\lambda, \pi)$$

is nonempty, compact-valued, and upper hemi-continuous. Correspondingly, the Principal's expected payoff from a signal π is

$$g_{\mu_0}(\pi) = \sum_{\omega\in\Omega} \sum_{a\in A} \mu_0(\omega)\pi(a \mid \omega)v(a, \omega),$$

and so, because ties are broken in favor of the Principal, the Principal's indirect payoff is

$$\tilde{V}_{\mu_0}(\lambda) = \max_{\pi\in T(\lambda)} g_{\mu_0}(\pi).$$

Since $T(\lambda)$ is nonempty and compact-valued and g is continuous, $\tilde{V}_{\mu_0}$ is well defined. We claim that $\tilde{V}_{\mu_0}$ is upper semicontinuous. Indeed, take any $\lambda_n \to \lambda$, and for each n, there exists $\pi_n \in \Pi$ such that $\tilde{V}_{\mu_0}(\lambda_n) = g_{\mu_0}(\pi_n)$. Since Π is compact, there exists a subsequence π_{n_k} that converges to some $\pi^* \in \Pi$. Since T is upper hemi-continuous, $\pi^* \in T(\lambda)$. Then, by continuity of g_{μ_0},

$$\lim_{k\to\infty} \tilde{V}_{\mu_0}(\lambda_{n_k}) = \lim_{k\to\infty} G(\pi_{n_k}) = G(\pi^*) \leqslant \max_{\pi\in T(\lambda)} G(\pi) = \tilde{V}_{\mu_0}(\lambda).$$

and $\tilde{V}_{\mu_0}$ is upper semicontinuous. Finally, since $[0, \overline{\lambda}]$ is compact, $\tilde{V}_{\mu_0}$ attains a maximum, and the Principal's problem has a solution. □

Proof of Proposition 9. Fix $\lambda \in \mathbb{R}_+$. With our normalization, we can write the DM's problem as $\max_\tau \mathbb{E}_\tau[\max\{0, (B + C)\mu - C\}] - \lambda(\mathbb{E}_\tau[S_{\operatorname{supp}\mu_0}(\mu)] - S_{\operatorname{supp}\mu_0}(\mu))$. If the DM acquires an

uninformative signal δ_{μ_0} there is no cost of information, and she gets $\mathbb{E}_{\delta_{\mu_0}}[U(\mu)] = 0$ (given the normalization). To show that δ_{μ_0} is not optimal, we will construct an attention strategy τ_ϵ that yields a strictly positive payoff. Consider a symmetric signal that induces posteriors $\mu_H = \mu_{ind} + \epsilon$ and $\mu_L = \mu_{ind} - \epsilon$, each with probability $\frac{1}{2}$, for some $\epsilon > 0$ small enough such that $\mu_H, \mu_L \in int(\Delta(\Omega))$. This strategy clearly satisfies Bayes' plausibility. The expected utility under τ_ϵ is

$$\begin{aligned}\mathbb{E}_{\tau_\epsilon}[U(\mu)] &= \frac{1}{2}\max\{0, (B+C)(\mu_{ind}+\epsilon) - C\} + \frac{1}{2}\max\{0, (B+C)(\mu_{ind}-\epsilon) - C\} \\ &= \frac{1}{2}\max\{0, (B+C)\epsilon\} + \frac{1}{2}\max\{0, -(B+C)\epsilon\} \\ &= \frac{1}{2}(B+C)\epsilon.\end{aligned}$$

Next, we look at the cost of the signal (where we suppress the index on $S_{\operatorname{supp}\mu_0}$), which gives

$$c_\lambda(\mu_{ind}, \tau_\epsilon) = \lambda\left[\frac{1}{2}S(\mu_{ind}+\epsilon) + \frac{1}{2}S(\mu_{ind}-\epsilon) - S(\mu_{ind})\right].$$

Because S is differentiable at μ_{ind}, we can take the Taylor expansion of S around μ_{ind}:

$$S(\mu_{ind} \pm \epsilon) = S(\mu_{ind}) \pm S'(\mu_{ind})\epsilon + o(\epsilon).$$

Substituting this into the cost function yields $c_\lambda(\mu_{ind}, \tau_\epsilon) = \lambda \cdot o(\epsilon)$. The net expected payoff from strategy τ_ϵ is therefore

$$V(\tau_\epsilon) = \frac{1}{2}(B+C)\epsilon - \lambda \cdot o(\epsilon).$$

Because $\frac{1}{2}(B+C) > 0$, for any $\lambda \in \mathbb{R}_+$, there exists an $\epsilon > 0$ sufficiently small such that $\frac{1}{2}(B+C)\epsilon > \lambda \cdot o(\epsilon)$. Thus, $\delta_{\mu_0} \notin T_{\mu_0}(\lambda)$ for any $\lambda \in \mathbb{R}_+$. □

Proof of Proposition 10. Normalize $v(a_H) = 1$ and $v(a_L) = 0$. In the Skeptical region the Principal's expected payoff under full transparency is μ_0. Under full obfuscation, the DM acquires no information and relies on the prior. Because $\mu_0 < \mu_{ind} = \frac{C}{B+C}$, the DM's optimal action under the prior is a_L, yielding a payoff of 0 for the Principal. Therefore, $\tilde{V}_{\mu_0}(0) > \tilde{V}_{\mu_0}(\overline{\lambda})$.

By hypothesis, the entropy model S_E admits a partial obfuscation equilibrium. This implies there exists $\lambda_0 \in (0, \overline{\lambda})$ such that $\tilde{V}_{\mu_0}(\lambda_0) > \tilde{V}_{\mu_0}(0)$. Hence, the probability the DM chooses a_H under Shannon, denoted $p_S(\lambda_0)$, satisfies $p_S(\lambda_0) > \mu_0$.

For any strictly convex and C^2 posterior-separable cost function S, the DM's optimal

interior posteriors $\underline{\mu}$ and $\overline{\mu}$ for a given λ are characterized by the tangency conditions (equal slopes and intercepts for the net utility functions) in Caplin et al. (2022). This can be rewritten as a system of equations $F(\underline{\mu}, \overline{\mu}; S) = 0$, where

$$F(\underline{\mu}, \overline{\mu}; S) \coloneqq \begin{pmatrix} S'(\overline{\mu}) - S'(\underline{\mu}) - \frac{B+C}{\lambda_0} \\ S(\overline{\mu}) - S(\underline{\mu}) - S'(\underline{\mu})(\overline{\mu} - \underline{\mu}) - \frac{(B+C)\overline{\mu} - C}{\lambda_0} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

At $S = S_E$, let the unique roots be $(\underline{\mu}_S, \overline{\mu}_S)$. The Jacobian matrix of F with respect to $(\underline{\mu}, \overline{\mu})$ evaluated at S_E is given by

$$J = \begin{pmatrix} -S_E''(\underline{\mu}_S) & S_E''(\overline{\mu}_S) \\ -S_E''(\underline{\mu}_S)(\overline{\mu}_S - \underline{\mu}_S) & S_E'(\overline{\mu}_S) - S_E'(\underline{\mu}_S) - \frac{(B+C)}{\lambda_0} \end{pmatrix}$$

But note that at a solution of the first tangency condition, the bottom-right element equals zero. Therefore, $\det J = S''(\underline{\mu})S''(\overline{\mu})(\overline{\mu} - \underline{\mu})$. For the Shannon case,

$$S_E(\mu) = -H(\mu) = \mu \log(\mu) + (1 - \mu) \log(1 - \mu),$$

and at any interior belief,

$$S_E''(\mu) = \frac{1}{\mu(1-\mu)} > 0 \implies \det J = \frac{\overline{\mu} - \underline{\mu}}{\overline{\mu}(1-\overline{\mu})\underline{\mu}(1-\underline{\mu})} > 0$$

Therefore, the Jacobian is non-singular. By the Implicit Function Theorem, there exists a neighborhood $\mathcal{N}$ of S_E in the C^2 topology and continuously differentiable functions $\underline{\mu}(S)$ and $\overline{\mu}(S)$ such that $F(\underline{\mu}(S), \overline{\mu}(S); S) = 0$ for all $S \in \mathcal{N}$, with $\underline{\mu}(S_E) = \underline{\mu}_S$ and $\overline{\mu}(S_E) = \overline{\mu}_S$.

By Bayes' plausibility, the unconditional probability of selling under the cost function S is

$$p_S(\lambda_0) = \frac{\mu_0 - \underline{\mu}(S)}{\overline{\mu}(S) - \underline{\mu}(S)}$$

Because $\underline{\mu}(S)$ and $\overline{\mu}(S)$ are continuous on S, $p_S(\lambda_0)$ is also continuous in S around S_E. Since $p_S(\lambda_0) > \mu_0$, we can choose the neighborhood $\mathcal{N}$ to be sufficiently small such that, for all $S \in \mathcal{N}$, $p_S(\lambda_0) > \mu_0$, and we get partial obfuscation. □

Online Appendix

to

Strategic Obfuscation: Influence Through Information Costs

Luis Henrique Linhares **Daniel Monte**

A Good or Bad Product

Producer, Inc. wants to sell its product to the Customer, who may choose between two actions: *buy* or *not buy*. The quality of the product can be either *Good* or *Bad* when compared to another similar product on the market (made by a different Principal) that the Customer is already familiar with. The common prior that the product is *Bad* is given by $\mu_0 \in (0,1)$. In accordance with our assumption, the prior is interior for both players because the Customer is not acquainted with all the features of the product, and the Producer does not have all the information on the Customer's preferences yet. The prior then represents the fraction of DMs in society that prefer the competitor's products. One can also consider the case of a reseller who was not present in the production process and so is unsure of the product's quality for the final buyer.

The Customer may conduct a private investigation to obtain more information on the quality of the product: cost-benefit analysis, price comparison, search for online reviews, etc. The Producer may facilitate or hinder the quest for information by selling the product in different sizes or batches than competitors, putting the specifications in fine print, or not including the price for add-ons.[43]

We normalize the payoff to zero for both players if the Customer chooses *not buy*. If the Customer chooses *buy*, the Customer's payoff will be based on how much better the Product is compared to the competitors'. If it is *Good*, the benefit is given by $b > 0$, and if it is *Bad*, the payoff is given by $-c$, where $c > 0$. The Producer gets a benefit of $d > 0$ independently of the state of the world. The payoffs are presented in Table A.1, where rows denote the state of the world, not actions.

For a more concrete illustration, we set $\mu_0 = 0.75$, $d = b = 2$ and $c = 1$, which gives $\mu_{ind} \approx 0.667$. We then get the payoffs in Table A.2. Note that we have a certain asymmetry

[43] Kalayci and Potters (2011) describe the ordeal of buying a mobile phone. With more than 30 listed technical attributes (such as weight, memory size, and battery capacity), comparing options becomes an arduous task.

Table A.1: Payoffs table for the Good or Bad Product game.

		Customer	
		not buy	buy
Producer	Good	$0, 0$	d, b
	Bad	$0, 0$	$d, -c$

considering the payoffs for when the Customer chooses *buy*. This can be interpreted as either the Customer not getting very angry when buying a worse product, or as there being in place a recycling policy that returns some benefit for the Customer if they choose to recycle the bad product.

Table A.2: Payoffs table with numerical values.

		Customer	
		not buy	buy
Producer	Good	$0, 0$	2, 2
	Bad	$0, 0$	$2, -1$

Now, to solve the problem, we first plot U, the indirect utility for the Customer as a function of the beliefs, as seen in Figure A.1. The horizontal axis denotes the probability that the state is *Bad*. On the right of Figure A.1, it is shown the split of posteriors when $\lambda = 0$. At this value of λ the Customer will acquire a fully informative signal, and thus, the posterior will be either 1 or 0. If, on the other hand, the cost is too high, the Customer will purchase a non-informative signal and will act according to the prior.

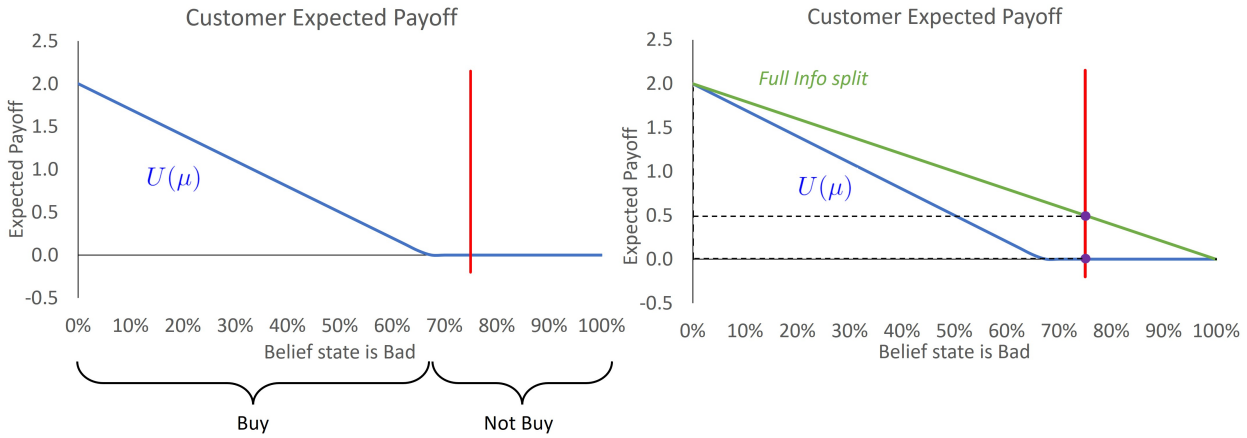

Figure A.1: Payoffs as a function of beliefs for the Customer. The vertical line represents the prior. Left: Curly brackets indicate the optimal action at that belief. Right: the diagonal green line indicates the split of posteriors and resulting expected payoffs when $\lambda = 0$.

Since the beliefs are restricted by Bayes' Plausibility, we can obtain the expected utilities

by looking at where the vertical line of the prior intersects with the posteriors induced by the signal. When $\lambda = 0$, we have full information and the expected payoff is 0.5. When λ is "too high", i.e., $\lambda > \overline{\lambda}$, we have full obfuscation and the expected payoff is 0.

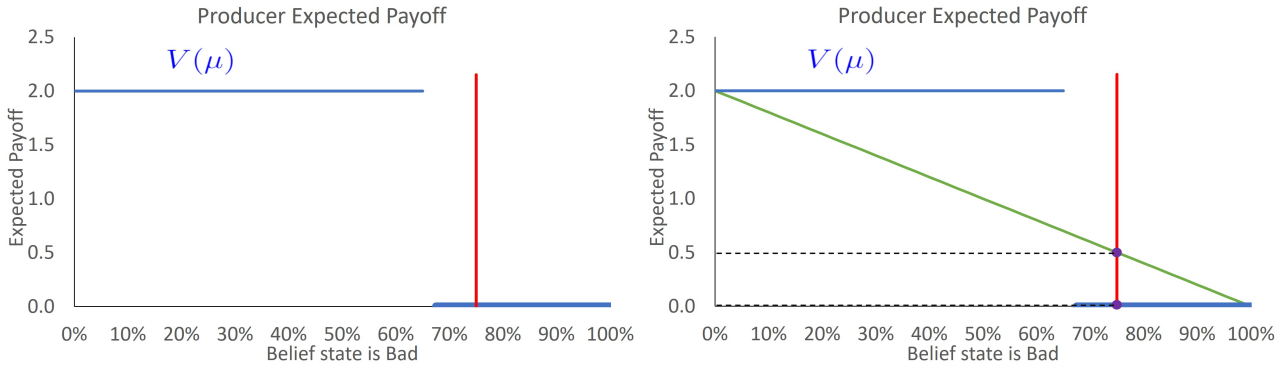

Figure A.2: Payoffs as a function of λ for the Producer. Right: the diagonal green line indicates the split of posteriors for $\lambda = 0$.

Now that we know how the Customer will act at each belief, Figure A.2 shows the expected utility V for the Producer. Again, on the right, we can find the expected payoffs for a given signal at the point where the vertical line of the prior intersects the induced split of posteriors. Under full information, the expected payoff for the Producer is 0.5; under no new information, the expected payoff is 0. Figure A.3 plots the expected utility for the Customer on the left and the cost of information at each belief on the right (remember that $\lambda H(\mu_0)$ is a constant from the point of view of the DM). The sum of these two functions results in the final objective function that the Customer will maximize.

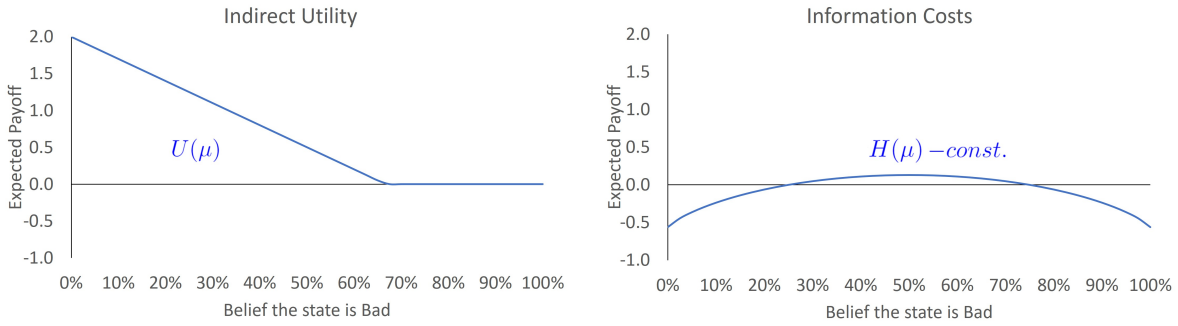

Figure A.3: The DM's objective function is given by the sum of the indirect utility U (plotted on the left) and the entropy learning costs (plotted on the right). We set $\lambda = 1$ and $\mu_0 = 0.75$.

We then can plot the function the Customer will maximize when $\lambda = 1$ as shown in Figure A.4. The Customer chooses a distribution of posteriors with support at $\underline{\mu}(1) = 0.335$ and $\overline{\mu}(1) = 0.91$. Since the prior is between these two values, the Customer will learn by purchasing a signal that, when the realization is *buy*, updates her beliefs to $\underline{\mu}(1)$, and when

the realization is *not buy*, updates her beliefs to $\overline{\mu}(1)$, so her expected payoff is 0.11. Had the prior fallen outside this interval, the Customer would have chosen a fully uninformative signal and acted according to the prior. Figure A.4 also shows that $\underline{\mu}(\lambda)$ and $\overline{\mu}(\lambda)$ indicate where the concavification detaches from the Customer's objective function.

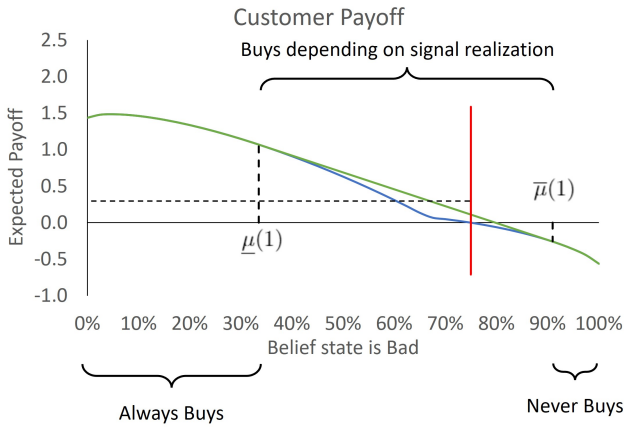

Figure A.4: The maximizing function and its concavification for the Customer for $\lambda = 1$ and $\mu_0 = 0.75$.

As an illustration of how λ affects the maximizing function and the solution, Figure A.5 shows on the left the maximizing function for $\lambda = 0.5$, in which the solution yields a payoff of 0.243 and on the right the maximizing function for $\lambda = 2$, which the solution yields a payoff of 0.02. It is visible that on the right, the concavity of the Entropy has a higher impact on the final function. We can also see that as λ increases, $\underline{\mu}(\lambda)$ and $\overline{\mu}(\lambda)$ approach μ_{ind}.

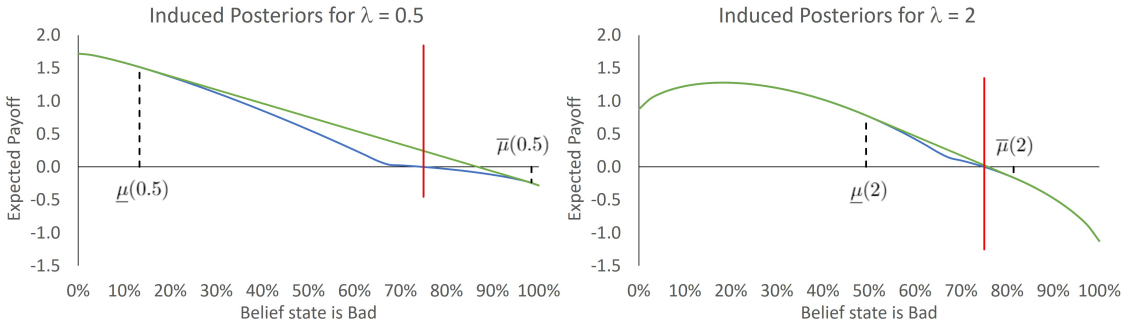

Figure A.5: Customer's maximizing function and solution. Left: Case for $\lambda = 0.5$. Right: Case for $\lambda = 2$.

Now we return to the Producer and try to understand how intermediate values of λ affect its payoffs. Figure A.6 plots the expected payoff for the case in which $\lambda = 1$. The payoff now is 0.556, higher than both the full information and no new information cases.

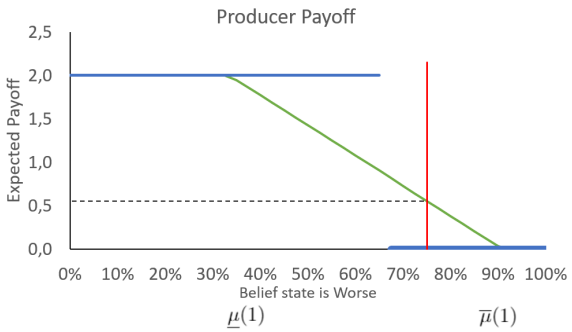

Figure A.6: Producer's indirect utility and the chosen split for $\lambda = 1$.

In fact, since we know the functional form for $\tilde{V}_{\mu_0}$ from Equation (10), we can plot it for the different values of λ. We do this in Figure A.7. Our numerical solution indicates that, at the optimal level of obfuscation ($\lambda^* \approx 0.75$), the Producer should try to engage in partial obfuscation. Therefore, the Producer may divulge all product specifications, but amongst irrelevant information that the Customer has to sift through. In equilibrium, the Customer gets an expected utility of 0.162 while the Producer gets an expected utility of 0.556.

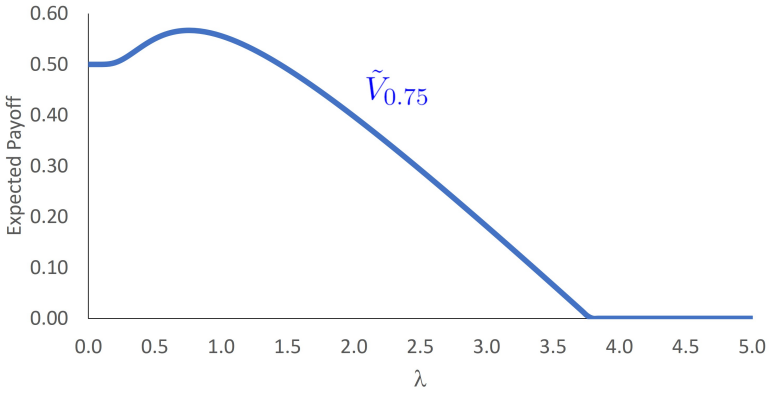

Figure A.7: Producer's indirect utility from λ.

Now, let's see what happens if we tweak the parameters slightly, setting $c = 2$. The new payoff table is shown in Table A.3, and we can see that the Customer has symmetric preferences, i.e., $\mu_{ind} = 0.5$.

Table A.3: Payoffs table with new value of c.

		Customer	
		not buy	buy
Producer	Good	$0, 0$	$2, 2$
	Bad	$0, 0$	$2, -2$

The steps to the solution in this new case are the same. We already know the behavior of the Customer as λ increases, and we have the more general functional form of $\tilde{V}_{\mu_0}$ as given by Equation 10. Hence, we can plot the indirect utility for the Producer in this new context. Figure A.8 shows the result. Now, the indirect utility is actually decreasing, and the optimal choice for the Producer is to engage in no obfuscation ($\lambda = 0$), which clearly makes the Customer better-off.

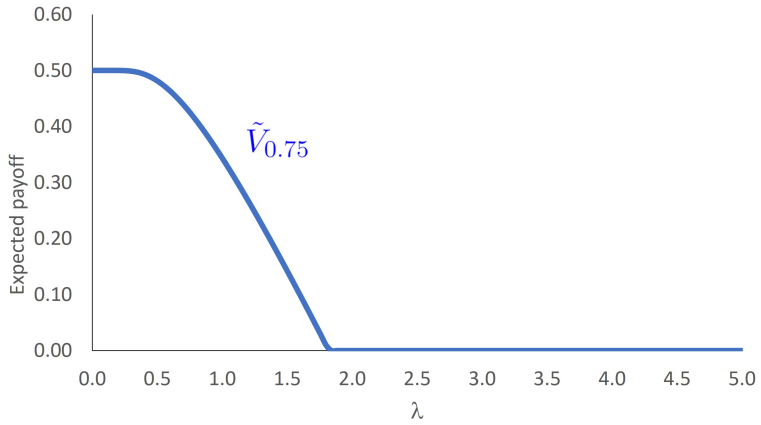

Figure A.8: Producer's indirect utility from λ for the new case in which $c = 2$.

Indeed, let λ^* denote the equilibrium value of λ, and we have the following expected payoffs in equilibrium:

(i) with $u(buy, Bad) = -1$, $\lambda^* \approx 0.75$, the Producer has a payoff of 0.556 and the Customer has a payoff of 0.162;

(ii) with $u(buy, Bad) = -2$, $\lambda^* = 0$, and the Producer has a payoff of 0.50, while the Customer has a payoff of 0.50.

When disregarding the strategic obfuscating behavior, it would seem that revoking the recycling policy would only make the Customer worse. We effectively reduce the agent's utility in a given state with no apparent trade-off. However, this induces full transparency and, consequently, the Customer is actually better off in equilibrium. Another way of interpreting this is that if the DM could commit to "hurting" themselves if they are in the *Bad* state, they could also induce the Principal to choose full disclosure.

In this example, the mechanism behind this phenomenon is the asymmetry in the Customer's payoffs and how it affects the dynamics of the problem. It implies that the Customer's response to higher information costs will not be symmetrical when gathering information. When we set $c = 1$, we make the Customer less averse to buying. As a result, the Customer is more inclined to purchase a signal structure that recommends *buy* more

often, even if it leads to more mistakes. While λ is still low, the Customer has enough leeway to choose this sort of signal. The Producer takes advantage of this fact and induces the Customer to *buy* more. As λ increases, however, this kind of signal becomes impracticable because, to recommend *buy* more and more frequently, the signal must also recommend *not buy* less and less, which requires accuracy, and that's expensive. At a certain point, the costs outweigh the benefits, and the Customer will be inclined to follow the prior, in which *not buy* is the better option. By setting $c = 2$, we are effectively making the Customer more "afraid of mistakes", and she will be less inclined to choose a signal biased towards recommending *buy*: there is no space for exploiting the Customer.[44] The best course of action for the Producer is then full transparency to at least induce *buy* when the state is actually *Good*.

Finally, we also point out that depending on the parameters of the model, we can have a boundary obfuscation problem. For instance, if the prior were 0.25 (instead of 0.75), Theorem 1 says that the best result is total obfuscation, or $\lambda = \overline{\lambda}$. What changed? The idea is that at the new prior of 0.25, the default action (i.e. if the DM plays according to the prior) is *buy*. Thus, the Principal (the Producer) does not want the DM to learn anything new and chooses $\lambda = \overline{\lambda}$.

B Costly Obfuscation

In this section, we extend the model to allow for costly obfuscation and briefly discuss our conclusions. We consider that it is costly to hide or provide information relative to a status quo, which depends, for instance, on the technology available. Think of a couple planning where to go out for dinner at the cheapest restaurant in the neighborhood. In the past, finding such a restaurant required making a list of local places, then calling each in turn; now, all it takes is a quick search on Google Maps.

To allow for this "status-quo", we add a parameter $\lambda_0 \in \mathbb{R}_+$, which we call the *natural* marginal cost of information. Obfuscation (or revealing) becomes costlier the further away the chosen λ is from λ_0. In the restaurant example, before Google Maps was invented, we had a high λ_0. After Google Maps, we have a low λ_0.

Definition 7. *The Principal's Costly Obfuscation Problem is*

$$\max_{\lambda \in [0,\overline{\lambda}]} \tilde{V}_{\mu_0}(\lambda) - \kappa(\lambda - \lambda_0)^2. \tag{B.1}$$

Here, $\kappa \geqslant 0$ represents the marginal cost of pushing obfuscation away from the natural level

[44] This is an analogous result to the one in Martin (2017), albeit not in a pricing context, in which an increase in search costs simply induces the buyer to take the outside option, which is bad for the seller.

and $\overline{\lambda}$ is defined as before. It is easy to see that the existence of a solution for the costly version is guaranteed by the existence of a solution for the original problem. We can further remove the assumption that $a^*(\mu_0)$ is a singleton whenever $\kappa > 0$, and the costly version of the problem will still always have a solution.[45] Indeed, as long as $\lambda_0 \in \mathbb{R}_+$, setting $\lambda \to \infty$ will never be optimal for the Principal.

We can look at the impact of λ_0 on the equilibrium level of obfuscation. Let λ^* be the solution to the Principal's problem when $\kappa = 0$. The closer λ_0 is to λ^*, the less of an effect κ has on the equilibrium level of obfuscation. Indeed, if $\lambda^* = \lambda_0$, the equilibrium level of obfuscation does not change with adding a cost of obfuscation κ.

Let us look at a simple example of costly obfuscation building upon the application in Section A. We will start with the payoff structure in Table A.3 and the same prior of $\mu_0 = 0.75$. We know from our Theorem 1 that, with $\kappa = 0$, we have a full transparency problem. Now, let us set $\kappa = 0.2$ and $\lambda_0 = 1$, so both full transparency and full obfuscation are costly for the Principal. Figure B.1 shows the payoffs for the Principal as a function of λ. We can see that we do not get full transparency anymore (compare Figure B.1 with Figure A.8).

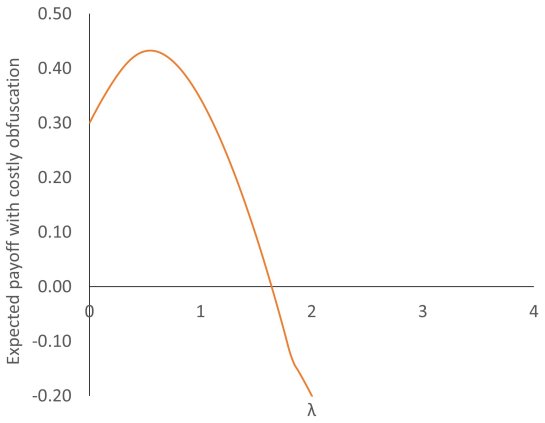

Figure B.1: The Principal's payoffs as a function of λ under costly obfuscation. We set $\lambda_0 = 1$ and $\kappa = 0.2$. There is a kink around $\lambda = 1.75$ because that is the point where $\tilde{V}_{\mu_0}$ becomes constant.

In this example, where the Principal would choose full transparency if it could, policies that reduce κ and λ_0 would result in more transparency, but that is not always the case. It is easy to see that, with costly obfuscation, inducing full transparency needs to take into account the values of both κ and λ_0 simultaneously.

[45]The corresponding Principal's Problem is then $\max_{\lambda \in \mathbb{R}_+} \tilde{V}_{\mu_0}(\lambda)$.