Attention Management*

Elliot Lipnowski† Laurent Mathevet‡ Dong Wei§

University of Chicago New York University UC Berkeley

February 8, 2019

Abstract. Attention costs can cause some information to be ignored and decisions to be imperfect. Can we improve the material welfare of a rationally inattentive agent by restricting his information in the first place? In our model, a well-intentioned principal provides information to an agent for whom information is costly to process, but the principal does not internalize this cost. We show that full information is universally optimal if and only if the environment comprises one issue. With multiple issues, attention management becomes optimal: the principal restricts some information to induce the agent to pay attention to other aspects.

Keywords: information disclosure, rational inattention, costly information processing, paternalistic information design.

JEL Codes: D82, D83, D91.

*We would like to thank David Ahn, Nemanja Antic, Ben Brooks, Andrew Caplin, Sylvain Chassang, Piotr Dworczak, Haluk Ergin, Brett Green, Yingni Guo, Filip Matějka, David Pearce, Doron Ravid, Chris Shannon, and Philipp Strack for their feedback, as well as audiences at the NASMES (Davis), the ES Summer School (Singapore), Stony Brook Summer Festival, UC Berkeley, and Northwestern University.

†lipnowski@uchicago.edu

‡lmath@nyu.edu

§dong.wei@berkeley.edu

> "This assumption, then, must be made, and also the following: that it is easier to discern each object of sense when in its simple form than when an ingredient of a mixture; easier, for example, to discern wine when neat than when blended, . . . or to discern the *nêtê* by itself alone".
>
> Aristotle.

1. INTRODUCTION

Information is a gift that may not always be accepted and, hence, useful. Speaking to a toddler about grammar may not improve his linguistic abilities, just as an adult may learn less from a book, an email, or a contract that contains too much detail. Simon (1971) offers an explanation for why less detailed communication may convey more: "What information consumes is rather obvious: it consumes the attention of its recipients." Failures to recognize this fact can have counterproductive effects: consumers are frequently confused by nutritional labels; patients can be overwhelmed in parsing side effects of medications; and so on (see Ben-Shahar and Schneider, 2014). As Simon (1996, p.144) puts it: "The real design problem is not to provide more information to people . . . but [to design] intelligent information-filtering systems."

This paper examines the merits of information filtering for attention purposes. When is it useful? After all, people could themselves filter information that they do not need. In the standard paradigm of rational inattention (Sims, 1998, 2003; Caplin and Dean, 2015; etc.), adding information can never hurt the agent's total welfare, because discarded information is ignored at *no cost*. The agent only pays for the informative content of what he chooses to heed. In such a world, where discarded information is costless, one may suspect that more information can never hinder decision making. But this intuition omits part of the equation. In this paradigm, the risk is not that some provided information will be superfluous, but rather that it will be "inferior" and crowd out other, more useful, information.

In the absence of attention concerns, if a principal (she) and an agent (he) disagree on the best action at a given state, filtering can be used for instrumental purposes: by restricting the agent's information (that is, providing partial information), the principal can *persuade* him to choose an action she prefers (see Kamenica and Gentzkow (2011) and the Bayesian persuasion literature). In the presence of attention concerns, to isolate optimal filtering for attention purposes, we separate it from its persuasive purpose. Thus, we assume that the principal and the agent have the same material motives, so that they agree on the best action in all states.

In our model, a principal provides information to a rationally inattentive agent about some state of the world. The agent cares about his material benefit, but also about the cost of processing information. Justifying this cost is almost a matter of introspection: people and organizations have limited information processing capacity (Sims, 1998, 2003), so that any attention one allocates has an opportunity cost. The principal, however, is motivated only by the agent's material benefit, as a teacher is motivated by her student's test score or a doctor by the fitness of her patient's medical decision. In this paternalistic, benevolent design problem, the principal does not internalize the agent's attention cost. From a subjective cost-benefit analysis, the agent chooses how much of the principal's provided information to process. Put differently, the agent decides how informed he wants to be, taking whatever information the principal makes available as an upper bound. Choosing that upper bound in an optimal way is our design problem.

As hinted above, attention management is about replacing "inferior" information with "better" information, while ensuring the latter's use. To make this more precise, let q^F be the information the agent would acquire given full disclosure. Note that the principal never has a reason to provide more (in the Blackwell sense) or less information than q^F, as more information would be ignored, by a standard revealed preference argument, and less would harm the principal. Therefore, any scope for improvement upon full disclosure must come from providing incomparable information, neither more nor less than q^F, but targeting the most pertinent dimensions for the task at hand. This manipulation prevents the agent from choosing q^F, while preserving some other policies that result in better decision-making.

Our main result shows that attention management is fundamentally about trading off issues. Formally, we prove that full disclosure is universally optimal (i.e., for all action sets, material objectives, and convex attention costs) if and only if the state is binary. The intuition for the positive result is that, when there are only two states, information can never be misused, since its only possible use is to separate one state from the other. No information is "inferior" in this case. Therefore, full information is always optimal for the principal, even though the agent will typically discard some of it (for free). By contrast, with three states or more (call them -1, 0 and 1), information can be used in multiple ways, such as separating 0 and -1, 1 and -1, 0 and $\{-1,1\}$, etc. Call each of these an "issue." When left to his own devices, the agent may choose to pay attention to the wrong issues.

In the other direction, we show that information filtering dominates full disclosure in a canonical example with three states, Shannon attention cost, and a motive for matching the state.[1] When all information is available, the agent pays a lot of attention to the moderate state, at the expense of the extreme states. He learns to discern the former relatively well from the extreme states, but he does not learn to discern the latter from each other, an issue he does not deem worth the attention cost. De facto, the information about the moderate state crowds out the information about the extreme states, even though the latter could save the agent from harmful mistakes. This phenomenon is supported by strong evidence, despite its stylized incarnation in our example. Ben-Shahar and Schneider (2014) present a plethora of decision scenarios, including medical choices, retirement planning, and loan contracting, in which mandated disclosures are counterproductive. A key channel is a crowding-out phenomenon: "Because disclosers can proffer, and dis-

1. This example is readily generalized to more than three states.

closees can receive, only so much information, mandated disclosures effectively keep disclosees from acquiring other information." Ben-Shahar and Schneider (2014, p.737) take brokerage-fee disclosures as a specific example. Given the above, by excluding cheaper information from consideration, a filtering strategy can encourage the agent to pay more attention (bearing a greater cost), specifically to issues he might otherwise ignore.

The remainder of the paper is organized as follows. Below, we discuss related literature. Section 2 sets up a general model of attention management and proves existence of solution. Section 3 presents our main result, establishing the *equivalence* between the optimality of full disclosure and one-issue environments. Section 4 concludes.

Related Literature. Our paper lies at the interface of two literatures: persuasion of decision makers through flexible information (Kamenica and Gentzkow, 2011; Aumann and Maschler, 1995) and rational inattention (Sims, 1998, 2003).

Among other generalizations, the Bayesian persuasion framework has been extended to include costly information provision by the principal (Gentzkow and Kamenica, 2014) and parallel costly information acquisition by the agent (Matyskova, 2018). Other works study persuasion games with departures from "classical" preferences, such as psychological preferences (Lipnowski and Mathevet, 2018), ambiguity aversion (Beauchene, Li, and Li, 2017), and heterogenous beliefs (Alonso and Câmara, 2016; Galperti, 2017). Outside of the flexible information framework, works such as Crémer, Garicano, and Prat (2007) and Glazer and Rubinstein (2012) study communication with a boundedly rational audience. A key feature of all aforementioned papers is that the receiver is a passive learner of provided information: he automatically processes whatever information is revealed by the sender. In our paper, the receiver will actively filter his own information to limit information processing costs. While the constraints this imposes on the principal are nontrivial, our analysis illustrates that the belief-based approach commonly adopted to study persuasion games is still applicable here.

The rational inattention literature (Sims, 1998, 2003; Caplin and Dean, 2015; Caplin and Martin, 2015; Matějka and McKay, 2015; etc.) studies optimal decision making by agents who incur an attention cost (or face an attention constraint), and so decide which of the available information to process before acting. These models are the building blocks of our agent's problem, given the principal's disclosure choice.

A key lesson from our paper is that attention management is fundamentally about choosing which aspects of the state to reveal. The literature on multidimensional cheap talk (Battaglini, 2002; Levy and Razin, 2007; Chakraborty and Harbaugh, 2007) also focuses on revealing lower dimensional aspects of a state. While that literature focuses on trading off dimensions as a way to relax a sender's incentive constraints, some dimensions of information are restricted in our paper to restrict the receiver's latitude to pay partial attention.

Our paper also contributes, through its main strategic tension, to the literature on costly information acquisition under moral hazard. Previous work (e.g., Dewatripont and Tirole (1999) and Li (2001)) has identified various ways to provide better incentives for information acquisition. More directly pertinent, in a setting of delegated decision-making, Szalay (2005) illustrates that eliminating "safe" actions from the agent's choice set can help align incentives to seek useful information, and that this may be valuable even if the principal never benefits from restricting the agent's behavior ex-post. In our model, limiting the information available to the agent *endogenously* eliminates such safe behavior (see Section 3.2).

The most related works, featuring information transmission under some form of inattention, are those of Bloedel and Segal (2018), Lester, Persico, and Visschers (2012), and Wei (2018).[2] In contemporaneous

2. Less related are Persson (2017) and Hirshleifer, Lim, and Teoh (2004). The former studies a model in which competing firms exploit consumers' limited attention through deliberate information overload. The latter shows, in a verifiable disclosure setting, how a simple form of inattention (specifically, some fraction of receivers being exogenously perfectly inattentive) can break standard equilibrium unraveling results.

work, Bloedel and Segal (2018) also study a setting in which a principal chooses which information to give an agent, who flexibly decides how to allocate attention in advance of a decision. In addition to important modeling differences,[3] their work also has a different purpose. We ask when information filtering can aid decision making, insisting on the role of multiple issues. In contrast, Bloedel and Segal (2018) apply the toolbox of Dworczak and Martini (2018) to explicitly solve for an optimally persuasive principal's policy in a canonical entropic-cost, binary-action model. Lester, Persico, and Visschers (2012) analyze a model of evidence exclusion in courts of law. In their model, a judge chooses which of finitely many pieces of evidence should be considered by the jury, who then choose a subset of those to examine at a cost. The authors provide examples in which evidence exclusion leads to fewer sentencing errors. Our paper studies this same basic tradeoff in a flexible information-choice framework, showing that optimal attention management is fundamentally about disclosure with multiple issues. Finally, Wei (2018) extends our framework to misaligned preferences, studying the optimal information choice of a seller trying to persuade a rationally inattentive buyer.

2. The Attention Management Problem

2.1 Our Model

Let Θ and A be a compact metrizable spaces of states and actions, respectively, with at least two elements. An agent must make a decision $a \in A$ in a world with uncertain state $\theta \in \Theta$ distributed according to some prior $\mu \in \Delta\Theta$. When the agent chooses a in state θ, his material payoff is given by $u(a,\theta)$, where $u : A \times \Theta \to \mathbb{R}$ is continuous. The principal's payoff

3. Beyond their focus on binary actions and entropic attention costs, and the possibility of misaligned material motives, Bloedel and Segal (2018) use a qualitatively different cost specification from ours. In our model, the agent's cost of attention concerns the degree to which he reduces his uncertainty about the state. Their agent, instead, bears a cost to reduce uncertainty about the realized message from the principal's chosen experiment.

is equal to the agent's material utility, u.[4]

In addition to his material utility, the agent also incurs an attention cost. As in the rational inattention literature, this cost is interpreted as the utility loss from processing information. To define it, first let

$$\mathscr{R}\left(\mu\right) := \left\{p \in \Delta\Delta\Theta : \int_{\Delta\Theta} \nu \, \mathrm{d}p(\nu) = \mu\right\}$$

be the set of **(information) policies**, which are the distributions over the agent's beliefs such that the mean equals the prior. It is well-known, for example from the work of Kamenica and Gentzkow (2011), that signal structures and information policies are equivalent formalisms. For the purpose of this paper, an attention cost function is a mapping $C : \Delta\Delta\Theta \to \mathbb{R}_+$ such that for every policy p,

$$C(p) = \int_{\Delta\Theta} c \, \mathrm{d}p \tag{1}$$

for some convex continuous $c : \Delta\Theta \to \mathbb{R}_+$. Jensen's inequality tells us that an agent who increases his attention, in the sense of obtaining a policy p that is more (Blackwell) informative than q, denoted $p \succeq^B_\mu q$,[5] incurs a higher cost for p than for q.

The timing of the game is as follows:

– The principal first commits to an information policy $p \in \mathscr{R}(\mu)$.

– The agent then decides to what extent he should pay attention to p: he chooses a policy $q \in \mathscr{R}(\mu)$ such that $q \preceq^B_\mu p$. Such a policy q is called an **(attention) outcome**.

– Finally, the agent's belief is drawn from q, at which point he takes an action $a \in A$. The agent's belief is his updated belief following receipt of a message sent from the principal's signal structure.

We study principal-preferred subgame perfect equilibrium.

4. For reasons outside the model, we take it as given that the agent must make the decision and cannot cede responsibility to the principal.

5. For any $p, q \in \mathscr{R}(\mu)$, $p \succeq^B_\mu q$ (or simply $p \succeq^B q$) if p is a mean-preserving spread of q.

It is convenient to work with the principal's indirect utility at $\nu \in \Delta\Theta$

$$U_P(\nu) = U(\nu) := \max_{a \in A} \int_{\Theta} u(a, \cdot)\, \mathrm{d}\nu,$$

and the agent's indirect utility

$$U_A(\nu) = U(\nu) - c(\nu).$$

Note that the attention cost does not affect the agent's optimal choice of a conditional on a given belief. The principal's problem can therefore be formalized as follows:

$$\begin{aligned} &\sup_{p,q} \int_{\Delta\Theta} U_P \, \mathrm{d}q \\ &\quad \text{s.t. } \; p \in \mathscr{R}\left(\mu\right) \; \text{ and } \; q \in G^*(p) \end{aligned} \tag{2}$$

where

$$G^*(p) := \operatorname*{argmax}_{q \in \mathscr{R}(\mu):\; q \preceq^B p} \left\{ \int_{\Delta\Theta} U \, \mathrm{d}q - C(q) \right\} = \operatorname*{argmax}_{q \in \mathscr{R}(\mu):\; q \preceq^B p} \int_{\Delta\Theta} U_A \, \mathrm{d}q$$

is the agent's optimal garbling correspondence. An information policy $p^* \in \mathscr{R}(\mu)$ is **(principal-) optimal** if (p^*, q^*) solves (2) for some outcome $q^* \in \Delta\Delta\Theta$. The corresponding q^* is an **optimal (attention) outcome**.

Note that, in this problem, the policy p chosen by the principal only appears in the constraint and does not *directly* affect any party's payoff. In choosing which information to make available, the principal proposes a menu of information policies and lets the agent pick his favorite one among them.

2.2 Existence

Given the sequential nature of our game and the infinite number of alternatives in the agent's menu of attention policies, it is not immediate that a solution to the principal's problem exists.[6] Our first result

6. Harris (1985) establishes existence of subgame perfect equilibrium in a class of games which, given results from our appendix, applies to ours. Rather than appealing to that paper's result, we prove directly

shows that it does indeed exist, and additionally shows that some optimum takes a convenient form. Say that an information policy $p \in \mathscr{R}(\mu)$ is **incentive compatible (IC)** if the agent finds it optimal to pay full attention to it, i.e., if $p \in G^*(p)$.

LEMMA 1. There exists a solution q^* to

$$\begin{aligned} &\sup_{q \in \mathscr{R}(\mu)} \int_{\Delta\Theta} U_P \, \mathrm{d}q \\ &\text{s.t.} \quad q \text{ is IC.} \end{aligned} \tag{3}$$

Moreover, q^* solves (3) if and only if (p^*, q^*) is a solution to (2) for some p^*. Finally, if Θ is finite, then there is a solution to (3) such that $|\mathrm{supp}(q^*)| \leq |\Theta|$.

Existence follows from a continuity argument, relying on the observation that the garbling correspondence is continuous. That IC policies are without loss, analogous to the revelation principle, relies on simple revealed preference reasoning: if q is an optimal attention outcome, then it must be an optimal garbling of itself, $q \in G^*(q)$. Finally, we show in the appendix how an appeal to Choquet's theorem enables the restriction to equilibria with few messages.

3. FILTERING AND THE NUMBER OF ISSUES

Given the limited conflict between our two players, one might conjecture that there is no cause for the principal to filter the agent's information. Indeed, the only divergence in preferences is that the principal does not internalize the agent's cost of attention. Thus, a sensible intuition is that the agent will either (i) use information the principal provides, in which case the principal benefits from providing it, or (ii) ignore the principal, in which case the principal bears no cost from providing it. In one-issue environments, where there are only two possible states, this intuition turns out to be exactly correct.

that a principal-optimal SPE exists.

With three or more states, however, the principal can withhold some aspects of the state and, in doing so, give the agent a higher marginal value for the information that the principal does make available. As we will show later in this section, withholding some information can indeed induce the agent to pay "better attention" and improve his decision-making.

Let the full disclosure policy, $p^F \in \mathscr{R}(\mu)$, be that with $p^F(\{\delta_\theta\}_{\theta\in\Theta}) = 1$, where $\delta_\theta \in \Delta\Theta$ is the belief that puts probability 1 on state θ. Let q^F be an attention outcome from full information.

THEOREM 1. Given Θ, the following are equivalent:

1. Full disclosure is optimal for every $\langle A, u, c, \mu\rangle$.
2. The state is binary.

The rest of this section is devoted to illustrating the two directions of the above theorem. In the first subsection, we present an intuitive argument for why full disclosure is optimal with two states. In the second, we demonstrate that full disclosure can be suboptimal with three states or more by counterexample. We provide formal counterparts of these two arguments in the Appendix.

3.1 Full Disclosure in Binary-State Environments

The principal never has a reason to provide more or less information in the Blackwell sense than what the agent would acquire given full information.[7] More information would be ignored, and less would harm the principal. The only way to have the agent bear a greater cost of attention and make a better decision, is to provide a policy that is *incomparable* to q^F (the attention outcome from full information).

When the state is binary, Lemma 1 indicates that we can without loss focus on information policies with binary support. The Blackwell ranking

7. In knife-edge cases, the agent might have many best responses to full information; in that case, we take q^F to be the most informative one (which exists in the binary-state setting).

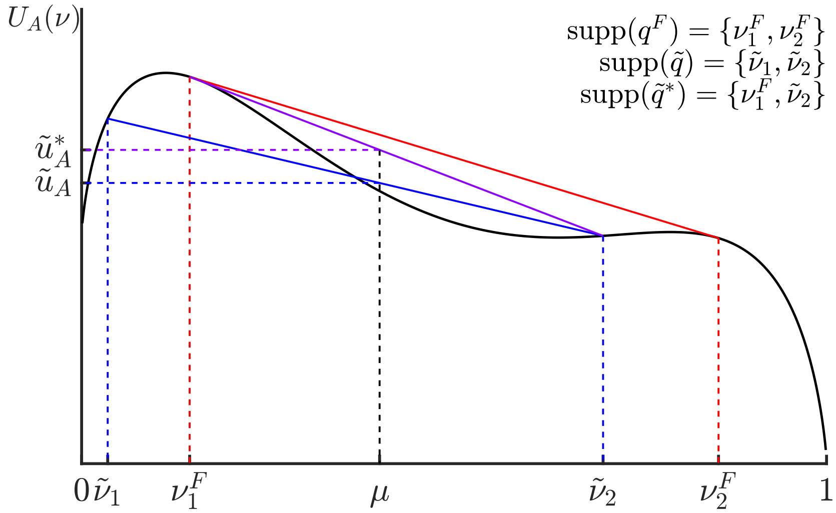

FIGURE I: OPTIMALITY OF FULL INFORMATION WHEN $|\Theta| = 2$

of such policies enjoys a simple characterization (see Lemma 2): for any $p, q \in \mathcal{R}(\mu)$ with $\text{supp}(p) = \{\nu_1^p, \nu_2^p\}$ and $\text{supp}(q) = \{\nu_1^q, \nu_2^q\}$,

$$p \succeq^B q \iff \text{supp}(q) \subseteq \text{co}[\text{supp}(p)] \iff \nu_1^p \leq \nu_1^q \leq \mu \leq \nu_2^q \leq \nu_2^p. \quad (4)$$

It turns out that any such policy that is incomparable to q^F is *not* IC. To see this, let q^F be represented by the red line in Figure I. When full information is disclosed, q^F can be found by the standard concavification technique. Now, suppose that the principal offered some other policy, say $\tilde{q}$ (in blue) in Figure I, which is incomparable to q^F by condition (4). Then, the agent would not pay full attention to $\tilde{q}$, because $\tilde{q}^*$ (in purple) is a garbling of $\tilde{q}$ by (4), and it clearly gives the agent a strictly higher payoff than $\tilde{q}$ (since $\tilde{u}_A^* > \tilde{u}_A$).

The above argument suggests that the principal can only induce the agent to pay attention to q^F or to less informative policies than q^F. Given their preference alignment, the principal finds it optimal to induce attention outcome q^F by providing full information.

3.2 Information Filtering in Multi-State Environments

Consider now a canonical ternary-state specialization of our model. As we shall see, information filtering can strictly benefit the agent's decision making. Whenever the state is not binary, this example is readily extended to an example in which full disclosure is suboptimal.

There are three states and a symmetric prior; each state has one associated action tailored to match it; and attention costs are proportional to the reduction in Shannon entropy. Formally, let the state and action spaces be given by $\Theta = A = \{-1, 0, 1\}$; the prior belief be given by $\mu = \left(\frac{1-\mu_0}{2}, \mu_0, \frac{1-\mu_0}{2}\right)$ for some $\mu_0 \in (0,1)$; the material utility be given by $u(\theta, a) = -(a-\theta)^2$; and the attention cost be given by

$$c(\nu) = \kappa\left[\mathbf{H}(\mu) - \mathbf{H}(\nu)\right] \quad \text{where} \quad \mathbf{H}(\nu) = -\sum_{\theta} \nu(\theta) \log[\nu(\theta)]$$

and $\kappa > 0$.

Rather than solving for the optimal policy in this model (which is not needed to establish the theorem), we show that a particular form of information filtering generates a strictly higher payoff to the principal than does full disclosure, for a range of parameters.

To this end, consider temporarily a simpler delegation problem: what would happen if the principal could, instead of restricting information (assumed to be full), restrict the agent to a nonempty set of *actions*, $B \subseteq A$, and the agent would optimally allocate his attention to choose from B? We prove a useful lemma for this auxiliary problem, which says that, for some parameter values (μ_0, κ):

(1) The principal would benefit from restricting the agent to actions $\{-1, 1\}$ (rather than allowing unrestricted choice).

(2) The agent would rather be restricted to actions $\{-1, 1\}$ than be restricted to $\{0\}$.

While the direct calculations are more delicate under entropic cost, the intuition for (1) is Szalay's (2005) familiar insight. As he shows, there is a benefit to forcing an agent to choose from extreme options when information acquisition is subject to moral hazard. By removing safe actions from an agent's choice set (here, action 0), the principal makes the marginal value of information higher to the agent, because mistakes become more harmful, for example choosing -1 in state 1. This strengthening of incentives for information acquisition can outweigh the ex-post payoff losses from not being able to perfectly adapt to the state.

What (2) enables is a bridge between Szalay's (2005) intuition and the attention management framework. Specifically, we exhibit an information policy p^O such that, if (2) holds in the simpler problem, then the agent's optimal behavior (i.e., his state-contingent choice distribution) in the original problem, given p^O but unrestricted choice, is identical to what he would do given choice restriction $\{-1,1\}$ and unrestricted information. So, by restricting information, the principal can *endogenously* restrict the agent's choice set by (2), which serves the principal's objective by (1).

What information filtering strategy might the principal want to adopt in the original problem? Quite simply, suppose that she reveals only the sign of the state, with uniform mixing if $\theta = 0$. This strategy restricts the agent's attention to a particular issue: he can learn nothing about whether the state is 0, but he can otherwise expend attention flexibly to rule out either extreme state. This strategy corresponds to policy $p^O \in \mathscr{R}(\mu)$ such that $\text{supp}(p^O) = \{(1-\mu_0, \mu_0, 0), (0, \mu_0, 1-\mu_0)\}$. See Figure II for a representation in the belief simplex.

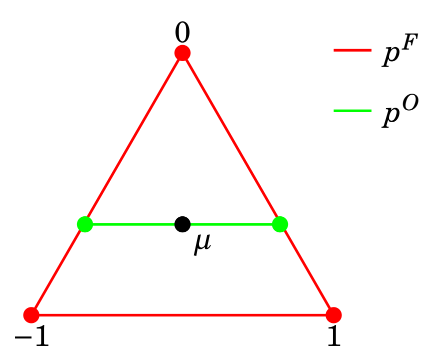

FIGURE II: p^F AND p^O

Let us now observe how, given (2), the agent's behavior under restricted information p^O coincides exactly with his behavior under restricted actions $\{-1,1\}$. As p^O has binary support S, Lemma 2 in the Appendix tells us that the agent's optimal value given p^O is simply the concave envelope of:

$$\begin{aligned} \text{co}(S) &\to \mathbb{R} \\ \nu &\mapsto \max\left\{\int_\Theta u(-1,\cdot)\,\mathrm{d}\nu, \int_\Theta u(0,\cdot)\,\mathrm{d}\nu, \int_\Theta u(1,\cdot)\,\mathrm{d}\nu\right\}, \end{aligned} \tag{5}$$

the restriction of U_A to co(S) (the green line segment in Figure II). But (5) is an even (i.e., symmetric) function, which is the maximum of three strictly concave functions—one for each action. Thus, its concave envelope about μ (a symmetric prior) is either (i) the value of the peak of the middle function, which would be exactly the agent's value if only action $\{0\}$ were available; or (ii) the value of the peaks of the other two functions, which would be exactly the agent's value if only actions $\{-1,1\}$ were available.[8] Therefore, if the agent strictly prefers restriction $\{-1,1\}$ to restriction $\{0\}$, then his behavior under p^O perfectly coincides with that under restriction $\{-1,1\}$.

Beyond conceptually tying attention management to the world of delegated choice, the above reduction is mathematically convenient. Indeed, the agent's problem given full information (with any restricted set of actions) is a standard discrete choice problem with rational inattention, for which his optimal behavior can be explicitly derived from the method in Caplin and Dean (2013). We can therefore verify all the required payoff comparisons by direct computation.

When the marginal attention cost κ is not too high,[9] the situation is described by Figure III. As foreshadowed, q^O and q^F capture different "issues," i.e., dimensions of uncertainty. In technical terms, they are

8. To see the latter, notice that the indirect utility U_A is symmetric about the line $\{\nu \in \Delta\Theta : \nu(-1) = \nu(1)\}$, and so the (unique) agent best response to restricted action set $\{-1,1\}$ is symmetric with binary support, and is therefore supported on co(S).

9. If κ is very large, q^O will be no information (different from what Figure III draws), and the agent always takes action 0.

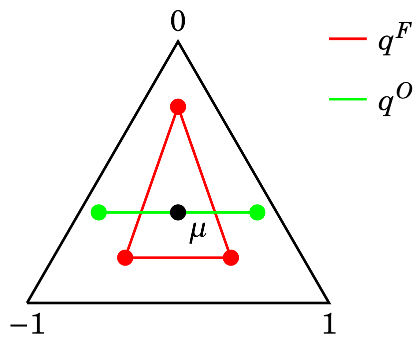

FIGURE III: q^F AND q^O

Blackwell incomparable by condition (4).[10] Under q^F, the agent pays much more attention to state 0 than under q^O, but he learns less about the extreme states. At a high level, the agent lets the "minor issues" (represented by state 0) steal his attention when unsupervised, whereas q^O redirects his attention toward the "big issues" — those with greater marginal material reward (represented by states -1 and 1).

4. CONCLUSION

We study the design problem of a well-intentioned principal who paternalistically seeks to help an inattentive agent make informed decisions. Even though the principal unequivocally wants the agent to be better informed, we find that withholding information can be optimal, helping guide the agent to make better decisions. A key takeaway from our analysis is that attention management is fundamentally about choosing the right "issues" on which the agent should focus. We convey this point by showing that single-issue information should never be withheld and by demonstrating the possibility of fruitful information withholding in a canonical multi-issue example.

10. Crucially, in contrast to the binary-state world, this richer environment exhibits information policies (for example, q^O) which are both IC *and* incomparable to the attention outcome q^F arising from from full disclosure.

References

Aliprantis, Charalambos D. and Kim C. Border. 2006. *Infinite Dimensional Analysis*. Berlin: Springer, second ed.

Alonso, Ricardo and Odilon Câmara. 2016. “Bayesian persuasion with heterogeneous priors.” *Journal of Economic Theory* 165 (C):672–706.

Aumann, Robert J. and Michael B. Maschler. 1995. *Repeated Games with Incomplete Information*. Cambridge, MA: MIT Press.

Battaglini, Marco. 2002. “Multiple Referrals and Multidimensional Cheap Talk.” *Econometrica* 70 (4):1379–1401.

Beauchene, Dorian, Jian Li, and Ming Li. 2017. “Ambiguous Persuasion.” Working paper.

Ben-Shahar, O. and C. Schneider. 2014. “The Failure of Mandated Disclosure.” John M. Ohlin Law Econ. Work. Pap. 526, Law Sch., Univ. of Chicago.

Bloedel, Alexander and Ilya Segal. 2018. “Persuasion with Rational Inattention.” Working Paper.

Caplin, Andrew and Mark Dean. 2013. “Behavioral Implications of Rational Inattention with Shannon Entropy.” Working Paper 19318, National Bureau of Economic Research. URL http://www.nber.org/papers/w19318.

———. 2015. “Revealed Preference, Rational Inattention and Costly Information Acquisition.” *American Economic Review* 105 (7):2183–2203.

Caplin, Andrew and Daniel Martin. 2015. “A Testable Theory of Imperfect Perception.” *The Economic Journal* 125:184–202.

Chakraborty, Archishman and Rick Harbaugh. 2007. “Comparative cheap talk.” *Journal of Economic Theory* 132 (1):70–94.

Chatterji, Srishti Dhar. 1960. “Martingales of Banach-valued random variables.” *Bulletin of the American Mathematical Society* 66 (5):395–398.

Crémer, Jacques, Luis Garicano, and Andrea Prat. 2007. “Language and the Theory of the Firm.” *The Quarterly Journal of Economics* 122 (1):373–407.

Csiszár, I. 1974. “On an extremum problem of information theory.” *Studia Scientiarum Mathematicarum Hungarica* 9 (1):57–71.

Dewatripont, Mathias and Jean Tirole. 1999. “Advocates.” *Journal of Political Economy* 107 (1):1–39.

Dworczak, Piotr and Giorgio Martini. 2018. “The simple economics of optimal persuasion.” .

Galperti, Simone. 2017. “Persuasion: The Art of Changing Worldviews.” Working paper.

Gentzkow, Matthew and Emir Kamenica. 2014. “Costly Persuasion.” *American Economic Review Papers and Proceedings* 104 (5):457–62.

Glazer, Jacob and Ariel Rubinstein. 2012. “A Model of Persuasion with Boundedly Rational Agents.” *Journal of Political Economy* 120 (6):1057 – 1082.

Harris, Christopher. 1985. “Existence and characterization of perfect equilibrium in games of perfect information.” *Econometrica* :613–628.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh. 2004. “Disclosure to an Audience with Limited Attention.” Game Theory and Information 0412002, EconWPA.

Kallenberg, Olav. 2006. *Foundations of modern probability*. Springer Science & Business Media.

Kamenica, Emir and Matthew Gentzkow. 2011. “Bayesian Persuasion.” *American Economic Review* 101 (6):2590–2615.

Lester, Benjamin, Nicola Persico, and Ludo Visschers. 2012. “Information Acquisition and the Exclusion of Evidence in Trials.” *Journal of Law, Economics, and Organization* 28 (1):163–182.

Levy, Gilat and Ronny Razin. 2007. “On the Limits of Communication in Multidimensional Cheap Talk: A Comment.” *Econometrica* 75 (3):885–893.

Li, Hao. 2001. “A Theory of Conservatism.” *Journal of Political Economy* 109 (3):617–636.

Lipnowski, Elliot and Laurent Mathevet. 2018. “Disclosure to a Psychological audience.” *AEJ: Microeconomics.* Forthcoming.

Matějka, Filip and Alisdair McKay. 2015. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model.” *American Economic Review* 105 (1):272–298.

Matyskova, Ludmila. 2018. “Bayesian Persuasion With Costly Information Acquisition.” Working paper.

O’Brien, Richard C. 1976. “On the Openness of the Barycentre Map.” *Mathematische Annalen* 223 (3):207–212.

Persson, Petra. 2017. “Attention Manipulation and Information Overload.” NBER Working Papers 23823, National Bureau of Economic Research, Inc.

Phelps, Robert R. 2001. *Lectures on Choquet’s Theorem*. Berlin: Springer, second ed.

Simon, Herbert A. 1971. “Designing Organizations for an Information Rich World.” In *Computers, Communications, and the Public Interest*, edited by Martin Greenberger. Baltimore, 37–72.

———. 1996. *The Sciences of the Artificial*. Cambridge, MA, USA: MIT Press.

Sims, Christopher. 1998. “Stickiness.” *Carnegie-Rochester Conference Series on Public Policy* 49 (1):317–356.

———. 2003. “Implications of Rational Inattention.” *Journal of Monetary Economics* 50 (3):665–690.

Szalay, Dezsö. 2005. “The Economics of Clear Advice and Extreme Options.” *Review of Economic Studies* 72 (4):1173–1198.

Wei, Dong. 2018. “Persuasion Under Costly Learning.” Working paper.

Wu, Wenhao. 2018. “Sequential Bayesian Persuasion.” Working paper.

A. For Online Publication: Proofs

A.1 Toward the proof of Lemma 1

We first introduce some additional notation. Given compact metrizable spaces X and Y, a map $f : X \to \Delta Y, x \in X$, and Borel $B \subseteq Y$, let $f(B|x) := (f(x))(B)$. Define the barycentre map $\beta_X : \Delta\Delta X \to \Delta X$ by $\beta_X(\hat{X}|m) := \int_{\Delta X} \gamma(\hat{X}) dm(\gamma)$, $\forall m \in \Delta\Delta X$, Borel $\hat{X} \subseteq X$. In other words, $\beta_X(m) = \mathbb{E}_{\nu \sim m}(\nu)$ for all $m \in \Delta\Delta X$. Note that $\mathscr{R}(\mu) = \beta_\Theta^{-1}(\mu)$, by definition.

Define $\Phi : \Delta\Delta\Delta\Theta \to (\Delta\Delta\Theta)^2$ by $\Phi(\mathbb{P}) = \left(\beta_{\Delta\Theta}(\mathbb{P}), \mathbb{P} \circ \beta_\Theta^{-1}\right)$. While we offer no specific interpretation to this map, it will be of use in deriving required properties of the Blackwell order.

Define the garbling correspondence $G : \Delta\Delta\Theta \rightrightarrows \Delta\Delta\Theta$ by

$$G(p) := \left\{ q \in \Delta\Delta\Theta : p \succeq^B q \right\}.$$

We can view the principal's problem as a delegation problem in which she offers the agent a delegation set $\hat{G} \in \{G(p)\}_{p \in \mathscr{R}(\mu)}$, and the agent makes a selection $q \in \hat{G}$. Recall, the agent's optimal garbling correspondence $G^* : \Delta\Delta\Theta \rightrightarrows \Delta\Delta\Theta$ is given by

$$G^*(p) := \operatorname*{argmax}_{q \in G(p)} \int_{\Delta\Theta} U_A \, \mathrm{d}q.$$

Claim 1. β_X is continuous for every compact metrizable space X.

Proof. This follows from Phelps (2001, Proposition 1.1). ■

Claim 2. Φ is continuous.

Proof. Suppose $\{\mathbb{P}_n\}_n \subseteq \Delta\Delta\Delta\Theta$ converges to $\mathbb{P}$. Since $\Delta\Theta$ is compact metrizable, $\beta_{\Delta\Theta}(\mathbb{P}_n) \to \beta_{\Delta\Theta}(\mathbb{P})$, by Claim 1. To show $\mathbb{P}_n \circ \beta_\Theta^{-1} \to \mathbb{P} \circ \beta_\Theta^{-1}$, take any continuous and bounded function $f : \Delta \to \mathbb{R}$. Continuity of β_Θ implies that $f \circ \beta_\Theta$ is continuous. Then,

$$\begin{aligned}
\int_{\Delta\Theta} f \, \mathrm{d}\left(\mathbb{P}_n \circ \beta_\Theta^{-1}\right) &= \int_{\Delta\Delta\Theta} f \circ \beta_\Theta \, \mathrm{d}\mathbb{P}_n \\
&\to \int_{\Delta\Delta\Theta} f \circ \beta_\Theta \, \mathrm{d}\mathbb{P} \\
&= \int_{\Delta\Theta} f \, \mathrm{d}\left(\mathbb{P} \circ \beta_\Theta^{-1}\right)
\end{aligned}$$

where the second line follows from the weak convergence of $\mathbb{P}_n$ to $\mathbb{P}$. ■

Claim 3. The partial order $\succeq^B$ is given by $\succeq^B = \Phi(\Delta\Delta\Delta\Theta)$.

Proof. First, take any $p \succeq^B q$ witnessed by mean-preserving spread $r : \Delta\Theta \to \Delta\Delta\Theta$ as in footnote 5. Define $\mathbb{P} := q \circ r^{-1} \in \Delta\Delta\Delta\Theta$. We now show that $\Phi(\mathbb{P}) = (p, q)$. Notice that $\mathscr{R}(v) \cap \mathscr{R}(v') = \emptyset$ for $v \neq v'$. Therefore, any $v \in \Delta\Theta$ satisfies $\beta_\Theta^{-1}(v) \cap r(\Delta\Theta) = r(v)$. As a result, for any Borel $S \subseteq \Delta\Theta$,

$$\mathbb{P} \circ \beta_\Theta^{-1}(S) = q \circ r^{-1}\left(\beta_\Theta^{-1}(S)\right) = q \circ r^{-1}(r(S)) = q(S),$$

and

$$\beta_{\Delta\Theta}(S|\mathbb{P}) = \int_{\Delta\Delta\Theta} \tilde{p}(S)\, \mathrm{d}\mathbb{P}(\tilde{p}) = \int_{\Delta\Delta\Theta} \tilde{p}(S)\, \mathrm{d}\left[q \circ r^{-1}\right](\tilde{p}) = \int_{\Delta\Theta} r(S|\tilde{p})\, \mathrm{d}q(\tilde{p}) = p(S).$$

Therefore, $(p, q) = \Phi(\mathbb{P})$.

Next, take any $\mathbb{P} \in \Delta\Delta\Delta\Theta$ and let $(\bar{p}, \bar{q}) := \Phi(\mathbb{P})$. We want to show that $\bar{p} \succeq^B \bar{q}$. Notice that we can view β_Θ as a $(\Delta\Theta)$-valued random variable on the probability space $(\Delta\Delta\Theta, \mathscr{B}(\Delta\Delta\Theta), \mathbb{P})$. Let $\gamma : \Delta\Delta\Theta \to \Delta\Delta\Theta$ be a conditional expectation $\gamma = \mathbb{E}_{q\sim\mathbb{P}}\left[q|\beta_\Theta(q)\right]$, which exists by Chatterji (1960, Theorem 1). So γ is β_Θ-measurable, and $\forall$ Borel $S \subseteq \Delta\Theta$, we have

$$\int_{\Delta\Delta\Theta} q(S)\, \mathrm{d}\mathbb{P}(q) = \int_{\Delta\Delta\Theta} \gamma(S|\cdot)\, \mathrm{d}\mathbb{P}.$$

By Doob's theorem (Kallenberg, 2006, Lemma 1.13), there exists a measurable $r : \Delta\Theta \to \Delta\Delta\Theta$ such that $\gamma = r \circ \beta_\Theta$. Then, $\forall$ Borel $S \subseteq \Delta\Theta$,

$$\int_{\Delta\Theta} r(S|\cdot)\, \mathrm{d}\bar{q} = \int_{\Delta\Delta\Theta} \left(r \circ \beta_\Theta\right)(S|\cdot)\, \mathrm{d}\mathbb{P} = \int_{\Delta\Delta\Theta} \gamma(S|\cdot)\, \mathrm{d}\mathbb{P} = \int_{\Delta\Delta\Theta} q(S)\, \mathrm{d}\mathbb{P}(q) = \beta_{\Delta\Theta}(S|\mathbb{P}) = \bar{p}(S).$$

Now, that β_Θ is affine and continuous implies

$$\beta_\Theta \circ \gamma = \mathbb{E}\left[\beta_\Theta \circ \mathrm{id}_{\Delta\Delta\Theta} | \beta_\Theta\right],$$

which is $\mathbb{P}$-a.s. equal to β_Θ. That is, $\beta_\Theta \circ r \circ \beta_\Theta = \mathrm{id}_{\Delta\Theta} \circ \beta_\Theta$, a.s.-$\mathbb{P}$. Equivalently, $\beta_\Theta \circ r = \mathrm{id}_{\Delta\Theta}$, a.s.-$\bar{q}$. The measurable function

$$\begin{aligned} \bar{r} : \Delta\Theta \quad &\to \quad \Delta\Delta\Theta \\ v \quad &\mapsto \quad \begin{cases} r(v) & : r(v) \in \mathscr{R}(v) \\ \delta_v & : r(v) \notin \mathscr{R}(v) \end{cases} \end{aligned}$$

is then $\bar{q}$-a.s. equal to r and satisfies $\beta_\Theta \circ \bar{r} = \mathrm{id}_{\Delta\Theta}$. Thus, $\bar{r}$ is a mean-preserving spread witnessing $\bar{p} \succeq^B \bar{q}$. ■

CLAIM 4. $\succeq^B$ is a continuous partial order, i.e. $\succeq^B \subseteq (\Delta\Delta\Theta)^2$ is closed.

Proof. This follows from Claims 2 and 3, because the continuous image of a compact set is compact. ■

CLAIM 5. The garbling correspondence G is continuous and nonempty-compact-valued.

Proof. It is nonempty-valued because $\succeq^B$ is reflexive, and upper hemicontinuous and compact-valued by Claim 4. Toward showing G is lower hemicontinuous, fix some open $D \subseteq \Delta\Delta\Theta$. Then,

$$\begin{aligned}\{p \in \Delta\Delta\Theta : G(p) \cap D \neq \emptyset\} &= \left\{p \in \Delta\Delta\Theta :\ p \succeq^B q,\ q \in D\right\} \\ &= \{p :\ (p,q) \in \Phi(\Delta\Delta\Delta\Theta),\ q \in D\} \\ &= \Phi_1 \circ \Phi_2^{-1}(D) \\ &= \beta_{\Delta\Theta}\left(\Phi_2^{-1}(D)\right)\end{aligned}$$

where the second line follows from Claim 3, and the last line follows from the definition of Φ_1. By Claim 2, since D is open, so is $\Phi_2^{-1}(D)$. In addition, $\beta_{\Delta\Theta}$ is an open map by O'Brien (1976, Corollary 1). So $\beta_{\Delta\Theta}\left(\Phi_2^{-1}(D)\right)$ is open, implying that G is lower hemicontinuous. ■

CLAIM 6. The optimal garbling correspondence G^* is upper hemicontinuous and nonempty-compact-valued.

Proof. As the indirect utility function U_A is (by Berge's theorem) continuous, so is $q \mapsto \int_{\Delta\Theta} U_A \, \mathrm{d}q$. The result then follows from Claim 5 and Berge's theorem. ■

CLAIM 7. If $q^* \in \mathscr{R}(\mu)$ is such that (q^*, q^*) solves the principal's problem in (2), then there is a set $\mathscr{P} \subseteq \operatorname{ext}\left[\mathscr{R}(\mu)\right]$ such that $q^* \in \overline{\operatorname{co}}\mathscr{P}$ and (p^*, p^*) solves the principal's problem for every $p^* \in \mathscr{P}$.

Proof. By Choquet's theorem, $\exists \mathbb{Q} \in \Delta\left[\mathscr{R}(\mu)\right]$ such that:

$$\begin{aligned}\mathbb{Q}\left[\operatorname{ext}\mathscr{R}(\mu)\right] &= 1, \\ \beta_{\Delta\Theta}(\mathbb{Q}) &= q^*.\end{aligned}$$

By Claim 6 and the Kuratowski-Ryll-Nardzewski Selection Theorem Aliprantis and Border (2006, Theorem 18.13), which applies here by Aliprantis and Border (2006, Theorem 18.10), there is some measurable selector g of G^*. The random posterior $q_g := \beta_{\Delta\Theta}(\mathbb{Q} \circ g^{-1})$ is then a garbling of q^*. Moreover, that $q^* \in G^*(q^*)$ implies

$$\begin{aligned}0 &\leq \int_{\Delta\Theta} U_A \, \mathrm{d}q^* - \int_{\Delta\Theta} U_A \, \mathrm{d}q_g \\ &= \int_{\operatorname{ext}\mathscr{R}(\mu)} \left[\int_{\Delta\Theta} U_A \, \mathrm{d}q - \max_{\tilde{q} \in G(q)} \int_{\Delta\Theta} U_A \, \mathrm{d}\tilde{q}\right] \mathrm{d}\mathbb{Q}(q).\end{aligned}$$

Since the latter integrand is everywhere nonpositive and the integral is nonnegative, it must be that the integrand is almost everywhere zero. That is, $q \in G^*(q)$ for $\mathbb{Q}$-

almost every q. Then, by Claim 6, $q \in G^*(q)$ for every $q \in \text{supp}(\mathbb{Q})$. Therefore, $\mathscr{P} := \text{supp}(\mathbb{Q}) \cap \text{ext}\mathscr{R}(\mu)$ is as desired. ■

CLAIM 8. There is some $p^* \in \text{ext}\left[\mathscr{R}(\mu)\right]$ such that (p^*, p^*) solves the principal's problem in (2).

Proof. The principal's objective can be formulated as a mapping $\text{Graph}(G^*) \to \mathbb{R}$ with $(p,q) \mapsto \int_{\Delta\Theta} U_P \, \mathrm{d}q$. It is upper semicontinuous and, by Claim 6, has compact domain. Therefore, there is some solution $(\hat{p}, q^*)$ to (2). As $G(q^*) \subseteq G(\hat{p})$, it is immediate that $q^* \in G^*(q^*)$; that is, q^* is IC. Letting $\mathscr{P}$ be as delivered by Claim 7, and taking any $p^* \in \mathscr{P}$ completes the claim. ■

CLAIM 9. If $|\Theta| < \infty$, then: $p \in \text{ext}\left[\mathscr{R}(\mu)\right]$ if and only if $\text{supp}(p)$ is affinely independent.

Proof. First, we prove the "only if" direction. Take any $p \in \mathscr{R}(\mu)$. Then $\mu \in \overline{\text{co}}[\text{supp}(p)] = \text{co}[\text{supp}(p)]$, where the equality follows from Θ being finite. By Carathéodory's theorem, there exists an affinely independent $S \subseteq \text{supp}(p)$ such that $\mu \in \text{co}(S)$; without loss, let S be a smallest such set. Since Θ is finite, $S \subset \mathbb{R}^{|\Theta|}$, so affine independence implies that S is finite. Therefore, $\exists N : S \rightrightarrows \Delta\Theta$ such that, $\forall \nu \in S$, the set $N(\nu)$ is a closed convex neighborhood of ν with $S \cap N(\nu) = \{\nu\}$. Making $\{N(\nu)\}_{\nu \in S}$ smaller, we may assume for all selectors η of N, $\{\eta(\nu)\}_{\nu \in S}$ is affinely independent.

Now define a specific selector $\eta : S \to \Delta\Theta$ by:

$$\eta(\nu) = \beta_\Theta\left(\frac{p\left(N(\nu) \cap \cdot\right)}{p\left(N(\nu)\right)}\right).$$ [11]

Since $\mu \in \text{co}(S)$, $\exists w \in \Delta S$ such that $\sum_{\nu \in S} w(\nu)\eta(\nu) = \mu$, and ($S$ being minimal) $w(\nu) > 0$ for all $\nu \in S$. Let

$$q := \sum_{\nu \in S} w(\nu) \frac{p\left(N(\nu) \cap \cdot\right)}{p\left(N(\nu)\right)}$$

$$\varepsilon := \min_{\nu \in S} \frac{w(\nu)}{p(N(\nu))}$$

Note that $q \in \mathscr{R}(\mu)$. Therefore, $\frac{p - \varepsilon q}{1 - \varepsilon} \in \mathscr{R}(\mu)$ and $p \in \text{co}\left\{q, \frac{p - \varepsilon q}{1 - \varepsilon}\right\}$.

Now, if $p \in \text{ext}\left[\mathscr{R}(\mu)\right]$, then it must be that $q = p$, even if we make each neighborhood in $\{N(\nu)\}_{\nu \in S}$ smaller, for otherwise $p \in \text{co}\left\{q, \frac{p - \varepsilon q}{1 - \varepsilon}\right\}$ contradicts $p \in \text{ext}\left[\mathscr{R}(\mu)\right]$. But then, $\text{supp}(p) = S$, and since S is affinely independent, so is $\text{supp}(p)$.

Now, we prove the "if" direction. Suppose $p \in \mathscr{R}(\mu)$ has affinely independent support S. Suppose $q, q' \in \mathscr{R}(\mu)$ have $p = (1 - \lambda)q + \lambda q'$ for some $\lambda \in (0,1)$. Then the support of q

11. Note that $p(N(\nu)) > 0$ for every $\nu \in S \subseteq \text{supp}(p)$, so that $\eta(\nu)$ is well-defined. That $N(\nu)$ is closed and convex for every $\nu \in S$ implies η is a selector of N.

must be contained in S. However, q is Bayes-plausible:

$$\sum_{\nu \in S} q(\nu)\nu = \mu = \sum_{\nu \in S} p(\nu)\nu.$$

But S is affinely independent, implying that $q(\nu) = p(\nu)$ for all $\nu \in S$. That is, $q = p$. As q, q', λ were arbitrary, it must be that p is an extreme point. ■

Proof of Lemma 1. By Claim 8, a solution to (2) exists. By Claims 8 and 9, (2) admits some optimal solution, (q^*, q^*), where $\text{supp}(q^*)$ is affinely independent if Θ is finite. This implies that $q^* \in G^*(q^*)$. Finally, notice that the optimal value of the problem in (3) is no larger than that of (2), since the former is a relaxation of the latter. So (q^*, q^*) is also a solution to (3). ■

REMARK 1. In the above work, the only properties of U_A and U_P that we use are that the former is continuous and the latter upper semicontinuous. For this reason, Lemma 1 applies without change to environments in which the principal and the agent have different material motives, to settings in which the principal partially internalizes the agent's attention costs, and more.

A.2 *Theorem 1: Two States*

We first prove a result that equivalently characterizes the Blackwell order, specialized to the case where the more informative information policy has affinely independent support. This characterization is important in proving both directions of Theorem 1.

LEMMA 2. Suppose $|\Theta| < \infty$. $\forall p, q \in \mathcal{R}(\mu)$ such that $\text{supp}(p)$ is affinely independent,[12]

$$p \succeq^B q \iff \text{supp}(q) \subseteq \text{co}[\text{supp}(p)].$$

The special case of this lemma with both policies being finite support is the same as Wu (2018, Theorem 5). We include this slightly more general, nearly identical proof for the sake of completeness.

Proof. Take any $p, q \in \mathcal{R}(\mu)$ with p nonredundant. Since Θ is finite, affine independence implies that $\text{supp}(p)$ is finite.

"If" part: Suppose $\text{supp}(q) \subseteq \text{co}[\text{supp}(p)]$. Since $\text{supp}(p)$ is affinely independent, we can find a unique $r : \text{supp}(q) \to \Delta(\text{supp}(p))$ such that $r(\cdot|\nu_q) \in \mathcal{R}(\nu_q)$ for all $\nu_q \in \text{supp}(q)$; that is,

$$\nu_q = \sum_{\nu_p \in \text{supp}(p)} \nu_p r(\nu_p|\nu_q), \forall \nu_q \in \text{supp}(q). \tag{6}$$

12. i.e., if no posterior belief in the support can be expressed as a finite linear combination of the other posterior beliefs with coefficients adding up to one.

Moreover, we have

$$\begin{aligned}\sum_{\nu_p \in \mathrm{supp}(p)} p(\nu_p)\nu_p &= \mu \\ &= \int_{\Delta\Theta} \nu_q \,\mathrm{d}q(\nu_q) \\ &= \int_{\Delta\Theta} \left[\sum_{\nu_p \in \mathrm{supp}(p)} r(\nu_p|\nu_q)\nu_p \right] \mathrm{d}q(\nu_q) \\ &= \sum_{\nu_p \in \mathrm{supp}(p)} \left[\int_{\Delta\Theta} r(\nu_p|\cdot)\,\mathrm{d}q \right] \nu_p,\end{aligned}$$

where the first two equalities follow from $p,q \in \mathscr{R}(\mu)$, the third equality follows from (6), and the last equality comes from changing the order of summation. Since $\mathrm{supp}(p)$ is affinely independent, the weights under which the average of the supported beliefs is μ is unique. Therefore, we have

$$p(\nu_p) = \int_{\Delta\Theta} r(\nu_p|\cdot)\,\mathrm{d}q, \ \forall \nu_p \in \mathrm{supp}(p). \tag{7}$$

From (6) and (7), we know that p is a mean-preserving spread of q (witnessed by r), thus $p \succeq^B q$.

"Only if" part: Suppose $p \succeq^B q$ is witnessed by $r : \mathrm{supp}(q) \to \Delta(\mathrm{supp}(p))$ such that (6) and (7) hold. By (6), we directly know that $\nu_q \in \mathrm{co}[\mathrm{supp}(p)], \forall \nu_q \in \mathrm{supp}(q)$, thus $\mathrm{supp}(q) \subseteq \mathrm{co}[\mathrm{supp}(p)]$, as desired. ■

We now specialize to binary environments and establish one direction of Theorem 1.

CLAIM 10. If $|\Theta| = 2$, then full disclosure is optimal for every $\langle A, u, c, \mu \rangle$.

Proof. Fix any $\langle A, u, c, \mu \rangle$. Let $p^F := \mu \circ (\delta_{(\cdot)})^{-1} \in \mathscr{R}(\mu)$, the full disclosure policy. Lemma 1 delivers an optimal IC policy $q^* \in \mathscr{R}(\mu)$ supported on at most two beliefs. If we could find some $q^F \in G^*(p^F)$ with $q^F \succeq^B q^*$, then we could prove the claim. Indeed, convexity of U_P would imply that $\int_{\Delta\Theta} U_P \,\mathrm{d}q^F \geq \int_{\Delta\Theta} U_P \,\mathrm{d}q^*$; and optimality of (p^F, q^F) would then follow from optimality of (q^*, q^*).[13]

If $|\mathrm{supp}(q^*)| = 1$, then any $q^F \in G^*(p^F)$ has $q^F \succeq^B q^*$. Now, focus on the complementary case of $|\mathrm{supp}(q^*)| = 2$. Identifying $\Delta\Theta$ with $[0,1]$, say $\mathrm{supp}(q^*) = \{\nu_0, \nu_1\}$, where

13. Under a different preference specification with U_P not convex, the same conclusion would obtain if $U_P - U_A$ were convex. Then we could deduce that

$$\int_{\Delta\Theta} U_P \,\mathrm{d}(q^F - q^*) \geq \int_{\Delta\Theta} U_A \,\mathrm{d}(q^F - q^*) \geq 0,$$

where the first inequality following from Jensen's inequality, and the second following from $G(q^*) \subseteq G(p^F)$.

$0 \le \nu_0 < \mu < \nu_1 \le 1$.

For any $\lambda \in (0,1)$, there is some $\epsilon \in (0,1)$ such that $\epsilon(1-\lambda,\ \lambda) \le (q^*(\nu_0),\ q^*(\nu_1))$. Therefore, $p_\lambda := q^* - \epsilon\left[(1-\lambda)\delta_{\nu_0} + \lambda\delta_{\nu_1}\right] + \epsilon\delta_{(1-\lambda)\nu_0+\lambda\nu_1} \in \mathscr{R}(\mu)$ too. As $q^* \in G^*(q^*)$ and $p_\lambda \preceq^B q^*$, it must be that

$$0 \le \int_{\Delta\Theta} U_A \ \mathrm{d}q^* - \int_{\Delta\Theta} U_A \ \mathrm{d}p_\lambda = \epsilon\left[(1-\lambda)U_A(\nu_0) + \lambda U_A(\nu_1) - U_A\left((1-\lambda)\nu_0 + \lambda\nu_1\right)\right].$$

So, defining

$$\begin{aligned} r : \Delta\Theta &\to \Delta\Delta\Theta \\ \nu &\mapsto \begin{cases} (1-\lambda)\delta_{\nu_0} + \lambda\delta_{\nu_1} & : \ \nu = (1-\lambda)\nu_0 + \lambda\nu_1 \text{ for some } \lambda \in (0,1), \\ \delta_\nu & : \ \text{otherwise}, \end{cases} \end{aligned}$$

r is a mean-preserving spread with $\int_{\Delta\Theta} U_A \ \mathrm{d}r(\cdot|\nu) \ge U_A(\nu) \ \forall \nu \in \Delta\Theta$.

Now, take any $\tilde{q} \in G^*(p^F)$, and define $q^F := \int_{\Delta\Theta} r \ \mathrm{d}\tilde{q} \in \mathscr{R}(\mu)$. As

$$\int_{\Delta\Theta} U_A \ \mathrm{d}q^F - \int_{\Delta\Theta} U_A \ \mathrm{d}\tilde{q} = \int_{\Delta\Theta} \left[\int_{\Delta\Theta} U_A \ \mathrm{d}r(\cdot|\nu) - U_A(\nu)\right] \ \mathrm{d}\tilde{q}(\nu) \ge 0$$

and $\tilde{q} \in G^*(p^F)$, it follows that $q^F \in G^*(p^F)$ too. Moreover, by construction, $q^F\left([0,\nu_0] \cup [\nu_1,1]\right) = 1$, so that $q^F \succeq^B q^*$. The claim follows. ■

A.3 Theorem 1: Three States

Throughout this subsection, we specialize to the instance of our model described in Section 3.2. Recall that the state space and action space are ternary, the prior symmetric, the material loss quadratic, and the attention cost proportional to entropy reduction.

A.3.1 Mathematical preliminaries

We find it convenient to represent beliefs as two-dimensional vectors. For any $\ell, r \in [0,1]$ such that $\ell + r \le 1$, identify (ℓ, r) with the belief $\nu_{\ell,r} := \ell\delta_{-1} + (1-\ell-r)\delta_0 + r\delta_1 \in \Delta\Theta$. Clearly, the map $(\ell, r) \mapsto \nu_{\ell,r}$ is bijective and affine.

Now, let $x := e^{-1/\kappa} \in (0,1)$, which is in one-to-one correspondence with the marginal cost of attention. We will typically not make the dependence of x on κ notationally explicit.

A.3.2 Optimal agent behavior with restricted choice sets

As a stepping stone to studying the agent's behavior under various principal choices of information, we consider the agent's optimal behavior given various restricted choice sets over actions. To that end, define the restricted choice sets $B_1 := \{0\}$, $B_2 := \{-1,1\}$, and $B_3 := \{-1,0,1\} = A$.

We now document an assumption which will ensure (it is in fact equivalent, but the converse is not important for our purposes, so we do not prove it) that the agent optimally chooses all three actions with positive probability given full disclosure.

ASSUMPTION 1. $x\bar{\mu}_0 < \mu_0 < \bar{\mu}_0$, where $\bar{\mu}_0 := \frac{1-x-x^2-x^3}{(1-x)(1+x)^2}$.

NOTATION 1. Define the following elements of $\mathscr{R}(\mu)$:[14][15]

$$\begin{aligned} q_1 &= \delta_\mu = \delta_{\left(\frac{1-\mu_0}{2},\frac{1-\mu_0}{2}\right)} \\ q_2 &= \frac{1}{2}\delta_{\left(\frac{x^4(1-\mu_0)}{1+x^4},\frac{1-\mu_0}{1+x^4}\right)} + \frac{1}{2}\delta_{\left(\frac{1-\mu_0}{1+x^4},\frac{x^4(1-\mu_0)}{1+x^4}\right)} \\ q_3 &= \alpha\delta_{\left(\frac{x(1-x)}{(1-x^2)^2},\frac{x(1-x)}{(1-x^2)^2}\right)} + (1-\alpha)\left(\frac{1}{2}\delta_{\left(\frac{1-x}{(1-x^2)^2},\frac{x^4(1-x)}{(1-x^2)^2}\right)} + \frac{1}{2}\delta_{\left(\frac{x^4(1-x)}{(1-x^2)^2},\frac{1-x}{(1-x^2)^2}\right)}\right), \end{aligned}$$

where $\alpha := \frac{\frac{(1-x^2)^2}{1-2x+x^4}\mu_0 - x}{1-x}$.

For a graphical illustration of these policies, see Figure III where q_2 and q_3 are depicted in green and red, respectively.

CLAIM 11. For each $k \in \{1,2,3\}$, q_k is the unique solution to the restricted-action agent problem

$$\max_{q\in\mathscr{R}(\mu)} \int_{\Delta\Theta} \max_{a\in B_k} \left\{-\mathbb{E}_{\theta\sim\nu}\left[(a-\theta)^2\right] - \kappa\left[\mathbf{H}(\mu) - \mathbf{H}(\nu)\right]\right\} \, \mathrm{d}q(\nu)$$

if Assumption 1 holds.

Proof. Agent optimality of q_k follows directly from verifying the sufficient conditions provided in Caplin and Dean (2013, Theorem 1).[16]

Toward uniqueness, direct computation shows that the matrix

$$\begin{pmatrix} 1 & x & x^4 \\ x & 1 & x \\ x^4 & x & 1 \end{pmatrix}$$

is of full rank (recall $0 < x < 1$). In particular, any subset of its columns is affinely independent. Uniqueness then follows from Caplin and Dean (2013, Theorem 2). ■

14. Recall, we identify an ordered pair (ℓ, r) with the measure $\ell\delta_{-1} + (1-\ell-r)\delta_0 + r\delta_1$.
15. One can easily verify that $\mathscr{R}(\mu)$ contains q_1, q_2, and (under Assumption 1) q_3.
16. Also see Csiszár (1974).

CLAIM 12. The attention outcomes from the previous claim satisfy:

$$\begin{aligned}
\int_{\Delta\Theta} U_A \, \mathrm{d}q_1 &= -(1-\mu_0) \\
\int_{\Delta\Theta} U_A \, \mathrm{d}q_2 &= -\left(\frac{4x^4}{1+x^4} + \frac{1-3x^4}{1+x^4}\mu_0\right) - \kappa(1-\mu_0)\left[h\left(\frac{1}{2}\right) - h\left(\frac{1}{1+x^4}\right)\right] \\
\int_{\Delta\Theta} U_P \, \mathrm{d}q_2 &= -\left(\frac{4x^4}{1+x^4} + \frac{1-3x^4}{1+x^4}\mu_0\right) \\
\int_{\Delta\Theta} U_P \, \mathrm{d}q_3 &= -\left[\frac{x(1-3x)}{1-x^2} + \frac{x(1+2x+3x^2)}{1-x-x^2-x^3}\mu_0\right],
\end{aligned}$$

where $h(p) := -p\log p - (1-p)\log(1-p)$ for all $p \in (0,1)$.

Proof. The claim follows from direct computation using the definitions of U_A, U_P, q_1, q_2, and q_3. ■

CLAIM 13. There is a nonempty open region of values $(\mu_0, \kappa) \in (0,1) \times (0,\infty)$ satisfying Assumption 1 such that:

1. $\int_\Theta U_P \, \mathrm{d}q_2 > \int_\Theta U_P \, \mathrm{d}q_3$;
2. $\int_\Theta U_A \, \mathrm{d}q_1 < \int_\Theta U_A \, \mathrm{d}q_2$.

Proof. First, as the finite collection of inequalities we need to satisfy are all strict inequalities between continuous functions of (μ_0, κ), the range where they are all satisfied is open. It therefore suffices to find a single pair of (μ_0, κ) with the desired properties.[17]

Now, define the polynomial $M(x) \equiv 1 - 3x - 4x^3 + x^4 + x^5$, and recall that $\bar{\mu}_0(x) = \frac{1-x-x^2-x^3}{(1-x)(1+x)^2}$. Direct computation shows that $\frac{\mathrm{d}[x\bar{\mu}_0(x)]}{\mathrm{d}x} = \frac{M(x)}{(1-x)^2(1+x)^3}$, where the denominator is always strictly positive over $(0,1)$. That $M(0) > 0 > M(1)$ then implies that $x \mapsto x\bar{\mu}_0(x)$ is strictly increasing [resp. decreasing] in a neighborhood to the right [left] of 0 [1]. Maximizing the continuous function over a large enough compact subinterval of $(0,1)$, there is some $x^* \in \operatorname{argmax}_{x\in(0,1)} x\bar{\mu}_0(x)$.

A rational, non-affine function that attains an interior maximum will have zero derivative at the maximizer and be strictly concave in a neighborhood of the maximizer. Therefore, $M(x) < M(x^*) = 0$ for sufficiently small $x > x^*$.

Since $\bar{\mu}_0(x) \to 1$ as $x \to 0$, it follows that $x\bar{\mu}_0(x) > 0$ when x is sufficiently small; the maximum value $x^*\bar{\mu}_0(x^*)$ is therefore strictly positive. As a consequence, $0 < x\bar{\mu}_0(x) < \bar{\mu}_0(x)$ for x near enough to x^*. Moreover expanding the denominator defining $\bar{\mu}_0$ shows that $\bar{\mu}_0 < 1$ for any $x \in (0,1)$.

So fix $x \in (x^*, 1)$ small enough to ensure that $M(x) < 0$ and $\bar{\mu}_0(x) > 0$—which the above work shows exists—and take $\kappa := \frac{-1}{\log x}$. Then for any μ_0 in the nonempty interval

17. One could proceed numerically to see that $(\mu_0, \kappa) = (.2, 1)$ satisfies the desired properties. We provide an analytical argument below for the sake of completeness.

$(x\bar{\mu}_0, \bar{\mu}_0) \subseteq (0,1)$, Assumption 1 will be satisfied. It remains to show that such μ_0 can be taken to satisfy the two desired payoff rankings.

For any $\mu_0 \in (x\bar{\mu}_0, \bar{\mu}_0)$, the computations of Claim 12 show:

$$\begin{aligned}
&\int_\Theta U_P \, \mathrm{d}q_2 - \int_\Theta U_P \, \mathrm{d}q_3 \\
= \;& -\left(\frac{4x^4}{1+x^4} + \frac{1-3x^4}{1+x^4}\mu_0\right) + \left[\frac{x(1-3x)}{1-x^2} + \frac{x(1+2x+3x^2)}{1-x-x^2-x^3}\mu_0\right] \\
= \;& \frac{x(x^5+x^4-4x^3-3x+1)}{(1-x^2)(1+x^4)} - \frac{\mu_0(1-x^2)(x^5+x^4-4x^3-3x+1)}{(1-2x+x^4)(1+x^4)} \\
= \;& \frac{M(x)\left(x - \frac{\mu_0}{\bar{\mu}_0}\right)}{(1-x^2)(1+x^4)} \\
> \;& 0,
\end{aligned}$$

where the second-to-last line follows from $\bar{\mu}_0 = \frac{1-x-x^2-x^3}{(1-x)(1+x)^2} = \frac{1-2x+x^4}{(1-x^2)^2}$, and the inequality follows from $M(x) < 0$ and $\mu_0 > x\bar{\mu}_0$. Thus, the principal payoff ranking (1) is as required.

Finally, we show that the desired agent payoff ranking (2) can be ensured. When $\mu_0 = x\bar{\mu}_0$ exactly, an appeal to Caplin and Dean (2013, Theorems 1 and 2) just as in the proof of Claim 11 shows that q_2 is the agent's unique best response to full information; in particular, $\int_\Theta U_A \, \mathrm{d}q_1 < \int_\Theta U_A \, \mathrm{d}q_2$ at such parameter values. By continuity, the same ranking will hold for sufficiently small $\mu_0 \in (x\bar{\mu}_0, \bar{\mu}_0)$. ■

A.3.3 Optimal agent behavior with restricted information

Below, we show that, for appropriate parameters, our solutions to the agent's problem under restricted choice can be imported to the agent's problem under restricted information.

Claim 14. If $\int_\Theta U_A \, \mathrm{d}q_1 < \int_\Theta U_A \, \mathrm{d}q_2$ and Assumption 1 holds, then $G^*(p^F) = \{q_3\}$ and $G^*(p^O) = \{q_2\}$.

Proof. That $G^*(p^F) = \{q_3\}$ follows directly from Claim 11. We now turn to characterizing the agent's best responses to p^O. To that end, let $\bar{u} := \int_\Theta U_A \, \mathrm{d}q_2$, let $S := \mathrm{co}[\mathrm{supp}(p^O)]$, and define $f : S \times A \to \mathbb{R}$ by letting $f(\nu, a) := \int_\Theta u_A(a, \cdot) \, \mathrm{d}\nu$.

By Lemma 2, we know $q_2 \preceq^B p^O$, so that any element of $G^*(p)$ must generate a value of at least $\bar{u}$ to the agent.

Let $\{\nu^1, \nu^{-1}\} = \mathrm{supp}(q^2)$ where $\nu^1(1) > \nu^{-1}(1)$. Consider any $\nu \in S$. As $f(\cdot, 1)$ is strictly concave and (as direct computation shows) has zero derivative (in S) at ν^1, it follows that $f(\nu, 1) \le \bar{u}$—strictly so if $\nu \notin \mathrm{supp}(q_2)$. Similarly, $f(\nu, -1) \le \bar{u}$, strictly so if $\nu \notin \mathrm{supp}(q_2)$. Finally, as $f(\cdot, 0)$ is concave and has zero derivative at μ, and $f(\mu, 0) = \int_\Theta U_A \, \mathrm{d}q_1 < \bar{u}$, we learn that $f(\nu, 0) < \bar{u}$. Maximizing over $a \in A$ tells us $U_A(\nu) \le \bar{u}$, strictly so if $\nu \notin \mathrm{supp}(q_2)$.

Take any $q \in G(p^O) \setminus \Delta[\text{supp}(q_2)]$. Lemma 2 tells us $q(D) = 1$, so that $U_A|_{\text{supp}(q)} \le \bar{u}$, but U_A is not q-a.s. equal to $\bar{u}$. Therefore $\int U_A \, \mathrm{d}q < \bar{u} = \int U_A \, \mathrm{d}q_2$, so that q cannot be a best response for the agent.

We thus know that $G^*(p^O) \subseteq \mathscr{R}(\mu) \cap \Delta[\text{supp}(q_2)]$. Unique optimality of q_2 then follows because $\mathscr{R}(\mu) \cap \Delta[\text{supp}(q_2)] = \{q_2\}$. ■

A.3.4 Payoff gains from attention management

Now, we are ready to show that the principal strictly prefers to limit the agent's information.

Claim 15. There is a nonempty open region of values $(\mu_0, \kappa) \in (0,1) \times (0,\infty)$ such that:

1. There is a unique $q^F \in G^*(p^F)$ and a unique $q^O \in G^*(p^O)$.
2. $\int_\Theta U_P \, \mathrm{d}q^F < \int_\Theta U_P \, \mathrm{d}q^O$.

Proof. Take values for μ_0 and κ as delivered by Claim 13. Then, Claim 13(2) and Claim 14 tell us that $q^F = q_3$ and $q^O = q_2$ are unique agent best responses to p^F and p^O, respectively. Finally, Claim 13(1) delivers the desired payoff comparison for the principal. ■

A.4 Proof of the Main Theorem

Finally, we collect and adapt existing results to prove Theorem 1.

Proof. That 2 implies 1 follows directly from Claim 10.

To see that 1 implies 2, suppose that $|\Theta| > 2$. Then, relabeling states, we may assume without loss that $\Theta \supseteq \Theta_3 := \{-1,0,1\}$. Let $A := \{-1,0,1\}$. Let $u : A \times \Theta \to \mathbb{R}$ be some continuous function such that $u(a,\theta) = -(a-\theta)^2$ for every $a \in A$ and $\theta \in \Theta_3$; it exists by the Tietze extension theorem. Let the cost function $c : \Delta\Theta \to \mathbb{R}$ be given by

$$c(\nu) = \kappa \sum_{\theta \in \{-1,0,1\}:\ \nu(\theta)>0} \left[\nu(\theta)\log\nu(\theta) - \mu(\theta)\log\mu(\theta)\right]$$

for some $\kappa > 0$; it is convex and continuous. Finally, let $\mu \in \Delta\Theta$ be the prior with $\mu(\Theta \setminus \Theta_3) = 0$, $\mu(0) = \mu_0$, and $\mu(-1) = \mu(1) = \frac{1-\mu_0}{2}$ for some $\mu_0 \in (0,1)$. Choosing (μ_0, κ) to be as delivered by Claim 15 then completes the proof. ■