Attention Management*

ELLIOT LIPNOWSKI† LAURENT MATHEVET‡ DONG WEI§

University of Chicago New York University UC Berkeley

April 15, 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

> "...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) foreshadows the potential hazards of detailed communication: "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 they deem superfluous. In the standard paradigm of rational inattention (Sims, 1998, 2003; Caplin and Dean, 2015; etc.), discarded information is ignored at *no cost*: the agent pays only for the informative content of what he chooses to heed. Information being free-disposal, one may suspect that disclosing more of it can never hinder decision making. But there is a risk—not that some provided information will be superfluous, but rather that it will be "inferior" and crowd out other, more useful, information.

Absent attention concerns, if a principal (she) and an agent (he) disagree on the best state-contingent choice, information filtering can be used for instrumental reasons: restricting the agent to partial information can *persuade* him to choose an action the principal prefers (e.g., Kamenica and Gentzkow, 2011). To isolate optimal filtering for attention purposes, we separate it from its persuasive purpose by assuming that the principal and the agent have the same material motives—hence, agree on the best action in all states.

In our model, a principal provides information to a rationally inattentive agent about some underlying state of the world. The agent cares about his material benefit, but also about the cost of processing information. The cost could be direct, because processing information is mentally taxing, or an opportunity cost, because people and organizations have limited information processing capacity (Sims, 1998, 2003). The principal, however, is motivated only by the agent's material benefit, as a teacher is motivated by her student's educational outcomes or a doctor by the fitness of her patient's medical decisions. In this paternalistic, benevolent design problem, the principal does not internalize the agent's attention cost. Given such costs, the agent decides how informed he wants to be, taking whatever information the principal provides as an upper bound. Choosing that upper bound optimally 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. The principal never has a reason to provide more or less information than q^F, as more would be ignored, by revealed preference, 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.

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 attention costs) if and only if the state is binary. The intuition for the positive result is that, with only two states, information can never be misused, because its sole use is to separate one state from the other. No piece of information being "inferior", full information is optimal for the principal, even though the agent will typically discard some of it. By contrast, with three states or more (call them -1, 0 and 1), information can be used in multiple ways, such as separating 1 from -1, 0 from $\{-1,1\}$, etc. Call each of these an "issue." Left to his own devices, the agent may focus on the wrong issues.

Conversely, we show how information filtering dominates full disclosure in a canonical example with three states, Shannon attention cost, and a motive for matching the state. With all information available, the agent learns to discern the moderate state 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 a high attention cost. Information about the moderate state crowds out information about the extreme states, even though the latter could forestall harmful mistakes. This phenomenon is supported by strong evidence, despite its stylized incarnation herein. Ben-Shahar and Schneider (2014) present many 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 disclosees can receive, only so much information, mandated disclosures effectively keep disclosees from acquiring other information." By excluding cheaper information from consideration, filtering can encourage the agent to pay more attention to issues he might otherwise neglect.

Related Literature. Our paper bridges two literatures: persuasion 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 costly parallel information acquisition by the agent (Matyskova, 2018). Other works study persuasion with departures from "classical" preferences, including psychological preferences (Lipnowski and Mathevet, 2018), ambiguity aversion (Beauchene, Li, and Li, 2017), and heterogenous beliefs (Alonso and Câmara, 2016; Galperti, 2017). In the above papers, the receiver is a passive learner, automatically processing whatever information is revealed by the sender. In our paper, the receiver actively filters his own information to limit information processing costs.

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 face an attention cost (or attention constraint) and 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.

Our paper contributes to the literature on costly information acquisition under moral hazard (e.g., Dewatripont and Tirole, 1999; Li, 2001). In particular, in a setting of delegated decision-making, Szalay (2005) illustrates that eliminating "safe" actions from the agent's choice set can sharpen incentives to seek information, which may be worthwhile 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).

Other papers featuring information transmission under some form of inattention are those of Bloedel and Segal (2018), Lester, Persico, and Visschers (2012), and Wei (2018). In contemporaneous 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,[1] their work 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

1. 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 attention cost 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.

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. Our paper studies the same tradeoff in a flexible information-choice framework. Finally, Wei (2018) extends our framework to misaligned preferences, studying a buyer-seller relationship.

2. The Attention Management Problem

2.1 Our Model

Let Θ and A be compact metrizable spaces 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 prior $\mu \in \Delta\Theta$. When he chooses a in state θ, his material payoff is $u(a,\theta)$, where $u : A \times \Theta \to \mathbb{R}$ is continuous. The principal's payoff is equal to the agent's material utility.[2]

In addition to his material utility, the agent incurs an attention cost, the utility loss from processing information. To define it, first let

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

be the set of **(information) policies**, which are the distributions over agent beliefs that average to the prior. It is well-known (e.g., Kamenica and Gentzkow, 2011) that signal structures and information policies are equivalent formalisms. Our agent's attention cost function is a mapping $C : \Delta\Delta\Theta \to \mathbb{R}_+$ given by[3]

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

for some convex continuous $c : \Delta\Theta \to \mathbb{R}_+$. Jensen's inequality tells us that an agent who processes more information, in the sense of obtaining a policy q' that is more (Blackwell) informative than q, denoted $q' \succeq^B q$,[4]

2. For reasons outside the model, the agent must make the decision and cannot cede responsibility to the principal.

3. Caplin, Dean, and Leahy (2018) call such a cost functional "posterior-separable".

4. For any $q, q' \in \mathcal{R}(\mu)$, $q' \succeq^B q$ if q' is a mean-preserving spread of q, that is, there is $r : \Delta\Theta \to \Delta\Delta\Theta$

incurs a higher attention cost for q' than for q.

The timing of the game is as follows:

- The principal first chooses 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 $q \preceq^B p$. Such a policy q is called an **(attention) outcome**.
- Nature draws an agent belief $\nu \in \Delta\Theta$ via q.
- The agent chooses an action $a \in A$.
- Nature chooses a state $\theta \in \Theta$ via ν.

We study principal-preferred subgame perfect equilibrium of this game.

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

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

such that (i) $q'(S) = \int_{\Delta\Theta} r(S|\cdot)\, \mathrm{d}q, \forall$ Borel $S \subseteq \Delta\Theta$ and (ii) $r(\cdot|\nu) \in \mathscr{R}(\nu), \forall \nu \in \Delta\Theta$.

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 q^*. The corresponding q^* is an **optimal (attention) outcome**.

Notice that the policy p chosen by the principal appears only in the constraint and does not *directly* affect payoffs. In deciding which information to make available, the principal proposes a menu of information policies, from which the agent chooses.

2.2 Existence

As a preliminary result, we prove that a solution to the principal's problem exists. Say that 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^*.

Existence follows from a compactness argument, after establishing that the garbling correspondence is continuous. That IC policies are without loss, analogous to the revelation principle, relies on revealed preference reasoning: if q is an optimal attention outcome, then it must be an optimal garbling of itself, $q \in G^*(q)$.

3. FILTERING AND THE NUMBER OF ISSUES

The only preference divergence in our model 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.

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 is made available. In this fashion, information filtering can 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 $\delta_\theta \in \Delta\Theta$ assigns probability 1 to state θ.

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.

We illustrate below the two directions of the theorem. First, we present an intuitive argument for why full disclosure is optimal with two states. Second, we demonstrate that full disclosure can be suboptimal with three states or more by counterexample. See the Appendix for the formal arguments.

3.1 Full Disclosure in Binary-State Environments

The principal never has a reason to provide more or less information than what the agent would acquire given full information (denoted q^F). 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, for intuition, consider the case in which there is a unique optimal attention outcome and a unique agent best response to full information, each with binary support. The Blackwell ranking of policies with binary support enjoys a simple characterization

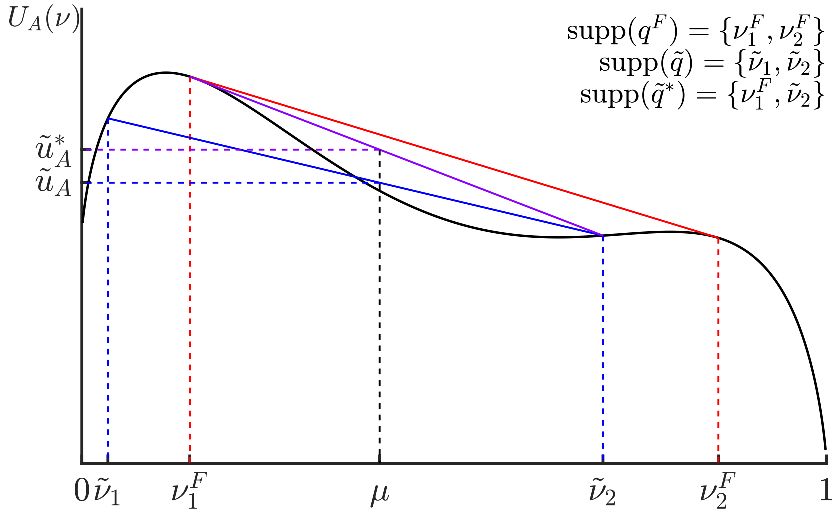

FIGURE I: OPTIMALITY OF FULL INFORMATION WHEN $|\Theta| = 2$

(see Lemma 2): for any $p, q \in \mathscr{R}(\mu)$ with $\text{supp}(p) = \{\nu_1^p, \nu_2^p\}$ and $\text{supp}(q) = \{\nu_1^q, \nu_2^q\}$,

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

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 q^F by providing full information.

3.2 Information Filtering in Multi-State Environments

Consider now a canonical ternary-state specialization of our model, readily extended to more states, such that 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 $\Theta = A = \{-1, 0, 1\}$; the prior be $\mu = \left(\frac{1-\mu_0}{2}, \mu_0, \frac{1-\mu_0}{2}\right)$ for some $\mu_0 \in (0,1)$; the material utility be $u(\theta, a) = -(a-\theta)^2$; and the attention cost be

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

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, 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 claim for this auxiliary problem: for some parameter values (μ_0, κ)

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

(‡) 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 (†) 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 be-come 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 (‡) 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 (‡) holds, then the agent's optimal behavior 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 *endogenously* restricts the agent's choice set by (‡), which serves the principal's objective by (†).

What information filtering strategy might the principal want to adopt? 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 $p^O \in \mathcal{R}(\mu)$ 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.

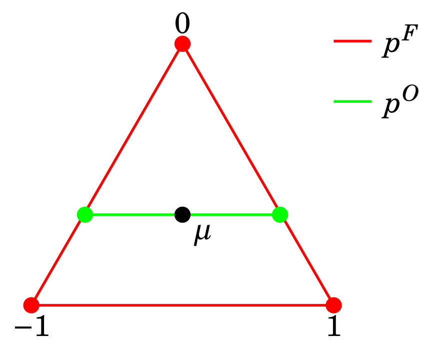

FIGURE II: p^F AND p^O

Observe how, given (‡), the agent's behavior under restricted information p^O coincides exactly with his behavior under restricted actions $\{-1, 1\}$. Indeed, as p^O has binary support S, Lemma 2 implies the agent's optimal value given p^O is simply the concave envelope of:

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

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.[5] Therefore, if the agent strictly prefers restriction $\{-1,1\}$ to restriction $\{0\}$, his behavior under p^O perfectly coincides with that under restriction $\{-1,1\}$.

Beyond conceptually tying attention management to 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, the situation is described by Figure III. As foreshadowed, the induced attention outcomes q^O and q^F capture different "issues," i.e., dimensions of uncertainty. Specifically, they are Blackwell-incomparable (see Lemma 2).[6] When choosing q^F, the agent pays greater 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

5. 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).

6. In contrast to the binary-state world, this environment exhibits information policies (for example, q^O) which are both IC *and* incomparable to q^F.

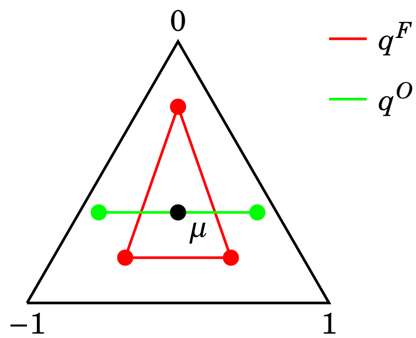

FIGURE III: q^F AND q^O

when unsupervised, whereas the restriction p^O (inducing 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 a rationally 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.

A. APPENDIX: PROOFS

A.1 Theorem 1: Two States

We first record a geometric characterization of the Blackwell order under affine independence. The proof, nearly identical to Wu (2018, Theorem 5), is omitted.

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

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

We next show that binary-support policies are without loss in the principal's problem.

CLAIM 1. If $|\Theta| = 2$, then some optimal IC policy $q^* \in \mathcal{R}(\mu)$ exists with $|\text{supp}(q^*)| \leq 2$.

Proof. Lemma 1 delivers an optimal IC policy $q \in \mathcal{R}(\mu)$. Identifying $\Delta\Theta$ with $[0,1]$, for each $\lambda \in [0,1]$, let $\nu_\lambda := (1-\lambda)\min[\text{supp}(q)] + \lambda \max[\text{supp}(q)]$. If $\nu_0 = \nu_1$, then $q^* = q$ works, so we focus on the case that $\nu_0 < \nu_1$. That $q \in \mathcal{R}(\mu)$ tells us $\nu_0 < \mu < \nu_1$, so that there is a unique $q^* \in \mathcal{R}(\mu) \cap \Delta\{\nu_0, \nu_1\}$. Clearly, $q^* \succeq^B q$ by Lemma 2.

Fix any $\lambda \in (0,1)$. For each $\epsilon \in (0, \frac{\nu_1 - \nu_0}{2})$, that $\nu_0, \nu_1 \in \text{supp}(q)$ implies that there exist $m_0^\epsilon \in \Delta[\nu_0, \nu_0 + \epsilon]$, $m_1^\epsilon \in \Delta[\nu_1 - \epsilon, \nu_1]$, and $\gamma^\epsilon > 0$ such that $\gamma^\epsilon[(1-\lambda)m_0^\epsilon + \lambda m_1^\epsilon] \leq q$. That q is IC then implies

$$\int U_A \, \mathrm{d}[(1-\lambda)m_0^\epsilon + \lambda m_1^\epsilon] \geq U_A\left(\int \nu \, \mathrm{d}[(1-\lambda)m_0^\epsilon + \lambda m_1^\epsilon](\nu)\right),$$

for otherwise the agent could profitably pool the mass from $\gamma^\epsilon[(1-\lambda)m_0^\epsilon + \lambda m_1^\epsilon]$. But then, taking limits as $\epsilon \to 0$ and appealing to continuity of U_A tells us $(1-\lambda)U_A(\nu_0) + \lambda U_A(\nu_1) \geq U_A(\nu_\lambda)$.

Consider any $\hat{q} \in \mathcal{R}(\mu)$ such that $\hat{q} \preceq^B q^*$—which Lemma 2 tells us is equivalent to $\hat{q}\{\nu_\lambda : 0 \leq \lambda \leq 1\} = 1$. That $(1-\lambda)U_A(\nu_0) + \lambda U_A(\nu_1) \geq U_A(\nu_\lambda)$ for any $\lambda \in [0,1]$ immediately implies $\int U_A \, \mathrm{d}q^* \geq \int U_A \, \mathrm{d}\hat{q}$. As $\hat{q}$ was arbitrary, it follows that q^* is IC. But $q^* \succeq^B q$, so that $\int U_P \, \mathrm{d}q^* \geq \int U_P \, \mathrm{d}q$, meaning q^* too is an optimal IC policy. ■

Now, we establish one direction of Theorem 1.

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

Proof. Fix any $\langle A, u, c, \mu \rangle$. Claim 1 delivers an optimal IC policy $q^* \in \mathcal{R}(\mu)$ 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^*).

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

For any $\lambda \in (0,1)$, there is some $\epsilon \in (0,1)$ with $\epsilon(1-\lambda,\ \lambda) \leq (q^*(\nu_0),\ q^*(\nu_1))$. 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 $p_\lambda \preceq^B q^*$, we learn

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

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 $\int_{\Delta\Theta} U_A \, \mathrm{d}r(\cdot|\nu) \geq U_A(\nu)\ \forall \nu \in \Delta\Theta$.

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

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

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

A.2 Theorem 1: Three States

Consider the model from Section 3.2, parametrized by (κ, μ_0). Let $x := e^{-1/\kappa} \in (0,1)$.

For any $\ell, r \in [0,1]$ such that $\ell + r \leq 1$, identify (ℓ, r) with the belief $\nu_{\ell,r} := \ell\delta_{-1} + (1-\ell-r)\delta_0 + r\delta_1 \in \Delta\Theta$. The map $(\ell, r) \mapsto \nu_{\ell,r}$ is bijective and affine.

We consider the agent's optimal behavior given various restricted action sets: $B_1 := \{0\}$, $B_2 := \{-1,1\}$, and $B_3 := \{-1,0,1\} = A$.

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. Let:[7]

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

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

CLAIM 3. For each $k \in \{1,2,3\}$, q_k is the unique solution to

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

if Assumption 1 holds.

Proof. Optimality of q_k and uniqueness, respectively, follow directly from Caplin and Dean (2013, Theorems 1 and 2).[8] ■

CLAIM 4. There are values $(\mu_0,\kappa) \in (0,1)\times(0,\infty)$ satisfying Assumption 1 with:

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. 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

7. One easily verifies that $\mathscr{R}(\mu)$ contains q_1, q_2, and (under Assumption 1) q_3.
8. Also see Csiszár (1974).

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 $(x\bar{\mu}_0, \bar{\mu}_0) \subseteq (0,1)$, 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 direct computation shows:

$$\int_\Theta U_P \,\mathrm{d}q_2 - \int_\Theta U_P \,\mathrm{d}q_3 = \frac{M(x)\left(x - \frac{\mu_0}{\bar{\mu}_0}\right)}{(1-x^2)(1+x^4)} > 0.$$

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) 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 $\mu_0 \in (x\bar{\mu}_0, \bar{\mu}_0)$. ■

CLAIM 5. 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 from Claim 3. We now characterize the agent's best responses to p^O. Let $\bar{u} := \int_\Theta U_A \,\mathrm{d}q_2$, $S := \mathrm{co}[\mathrm{supp}(p^O)]$, and define $f : S \times A \to \mathbb{R}$ by $f(\nu, a) := \int_\Theta u(a,\cdot)\,\mathrm{d}\nu - c(\nu)$.

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 $\{\nu^1, \nu^{-1}\} = \text{supp}(q_2)$ where $\nu^1(1) > \nu^{-1}(1)$. Consider any $\nu \in S$. As $f(\cdot, 1)$ is strictly concave and has zero derivative (in S) at ν^1, it follows that $f(\nu, 1) \le \bar{u}$—strictly so if $\nu \notin \text{supp}(q_2)$. Similarly, $f(\nu, -1) \le \bar{u}$, strictly so if $\nu \notin \text{supp}(q_2)$. 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 $\nu \notin \text{supp}(q_2)$.

Take any $q \in G(p^O) \setminus \Delta[\text{supp}(q_2)]$. Lemma 2 implies $q(S) = 1$, 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 follows because $\mathscr{R}(\mu) \cap \Delta[\text{supp}(q_2)] = \{q_2\}$. ■

CLAIM 6. Some $(\mu_0, \kappa) \in (0,1) \times (0, \infty)$ are 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 4. Then, Claims 4(2) and 5 imply that $q^F = q_3$ and $q^O = q_2$ are unique agent best responses to p^F and p^O, respectively. Claim 4(1) delivers the desired payoff comparison for the principal. ■

A.3 Proof of The Main Theorem

Proof. That 2 implies 1 is Claim 2.

To see that 1 implies 2, suppose $|\Theta| > 2$. Then, relabeling states, 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 continuous with $u(a, \theta) = -(a - \theta)^2$ for every $a \in A$ and $\theta \in \Theta_3$; it exists by the Tietze extension theorem. Define the cost function $c : \Delta\Theta \to \mathbb{R}$ 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 $\mu_0 \in (0,1)$. Taking (μ_0, κ) as delivered by Claim 6 completes the proof. ■

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.

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, Mark Dean, and John Leahy. 2018. "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy." Tech. rep., National Bureau of Economic Research.

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

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

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.

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

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.

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.

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: Existence Proof

In this supplementary appendix, we provide a formal proof of Lemma 1. In fact, we prove the slightly stronger result, that an optimum exists to the program of Lemma 1 that, if $|\Theta|$ is finite, has affinely independent support. This strengthening of the lemma is not invoked in our paper,[9] but may be of use to future users of the Attention Management framework.[10]

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

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

Claim 7. β_Θ, $\beta_{\Delta\Theta}$ are continuous.

9. The strengthened result implies Claim 1, but we instead provide an independent, elementary proof in the main appendix.

10. Given results proven in this online appendix, one could employ results of Harris (1985) to establish existence. We instead prove the result directly, enabling us to strengthen the lemma.

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

CLAIM 8. 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 7. 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 9. 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 4. Define $\mathbb{P} := q \circ r^{-1} \in \Delta\Delta\Delta\Theta$. We now show that $\Phi(\mathbb{P}) = (p, q)$. Notice that $\mathscr{R}(\nu) \cap \mathscr{R}(\nu') = \emptyset$ for $\nu \neq \nu'$. Therefore, any $\nu \in \Delta\Theta$ satisfies $\beta_\Theta^{-1}(\nu) \cap r(\Delta\Theta) = r(\nu)$. 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

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

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

Proof. This follows from Claims 8 and 9, because the continuous image of a compact set is compact. ■

CLAIM 11. 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 10. 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 9, and the last line follows from the definition of Φ_1. By Claim 8, 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 12. 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 11 and Berge's theorem. ∎

CLAIM 13. If $q^* \in \mathscr{R}(\mu)$ is such that (q^*, q^*) solves the principal's problem in (2), then there is a set $\mathscr{P} \subseteq \text{ext}\left[\mathscr{R}(\mu)\right]$ such that $q^* \in \overline{\text{co}}\mathscr{P}$ 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:

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

By Claim 12 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

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

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 12, $q \in G^*(q)$ 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 14. 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 12, 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 13, and taking any $p^* \in \mathscr{P}$ completes the claim. ■

CLAIM 15. 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:

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

$$q := \sum_{\nu \in S} w(\nu) \frac{p\,(N(\nu) \cap \cdot)}{p\,(N(\nu))}$$
$$\varepsilon := \min_{\nu \in S} \frac{w(\nu)}{p(N(\nu))}$$

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

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.

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 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 14, a solution to (2) exists. By Claims 14 and 15, (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.