Psychological Persuasion

by *Elliot Lipnowski & Laurent Mathevet*[*]

New York University[†]

PLEASE DO NOT CIRCULATE

February 24, 2015

Abstract

Should a well-intentioned advisor always tell the whole truth? In standard economics, the answer is yes (Blackwell (1953)), but in the world of psychological preferences (Geanakoplos, Pearce, and Stacchetti (1989))—where a listener's state of mind has a direct impact on his well-being—things are not so simple. In this paper, we study how a benevolent principal should disclose information to a psychological agent. After characterizing attitudes toward information, we study optimal information disclosure. Psychological information-aversion is of particular interest. In this case, we show that the principal can simply inform the agent by telling him what to do. Then, we study how the optimal policy changes with information-aversion. We also offer general tools of optimal disclosure. We apply our results to reputational concerns and cognitive dissonance; consumption-saving decisions with temptation problems (Gul and Pesendorfer (2001)); and doctor-patient relationships with equivocal information preferences.

[*]Email: `elliot.lipnowski@nyu.edu` and `lmath@nyu.edu`
[†]Address:

1 Introduction

The choices of information disclosure—what is revealed and what is concealed—affect what people know and, in turn, the decisions that they make. In many contexts, information disclosure is intended to improve people's welfare. In public policy, for instance, the regulator imposes legal rules of disclosure to this effect, such as required transparency by a seller of a good to a buyer. But should a well-intentioned advisor always tell the whole truth? In standard economics, the answer is yes, because a better informed person makes better decisions (Blackwell (1953)). However, parents do not always tell the whole truth to their children; people reveal information about their personal lives with great caution to their family elders; doctors do not always reveal to their patients all the details of their health, when dealing with depressive patients for instance; future parents do not always want to know the sex of their unborn child, even if it is pragmatic to know in advance; and many of us would not always like to know the caloric content of the food that we eat.

In this paper, we study how an informed principal should disclose information to an agent with psychological traits. It is hard to argue that our beliefs and our information affect our lives only through the actions that we take. Indeed, the mere fact of knowing, not knowing, or partially knowing can be a source of comfort or discomfort. Thus, how should information be disclosed to a person whose state of mind has a direct impact on his well-being? This question has important implications when we think of the principal as the informed party deciding what to say, but also when we think of the principal as an entity that, by law, can enforce what the informed party must disclose. Laws and regulations about public disclosure of information are widespread, ranging from administrative regulations, loans, credit cards, sales, and insurance contracts to self-storage facilities, restaurants, doctor-patient relationships and more (see Ben-shahar and Schneider (2014)).

In a paper entitled "Disclosure: Psychology Changes Everything," Loewenstein, Sunstein, and Goldman (2014) argue that psychology may lead one to rethink public disclosure policies. For example, it may be counter-productive to offer detailed information to a person with limited attention, as she may ignore critical information due to cognitive limitations. Likewise, if people are nonstandard in how they process information, i.e., if they are not always Bayesian, then the regulator should take into account their biased probability judgments in the choice of disclosure policies.

At a more personal level, a well-intentioned speaker may directly take into account the psychological well-being of her interlocutor when deciding what to say. Sometimes, older adults experience cognitive dissonance, an innate discomfort at having their beliefs shift from an existing position: how should information disclosure respond to this? As another example, there are many situations in which people would like to be fully informed when the news is good, but would rather not know how bad it is otherwise. A naive way of communicating with such a person would be to reveal the news if it is good and say nothing otherwise. Of course, it is not as simple, as the person would then know that the news is bad when she hears nothing.

In our model, an informed principal will know the realization of a state of nature and she must choose what to tell the agent in each contingency. After receiving his information from the principal, the agent updates his prior beliefs and then takes an action. The principal wants to maximize the agent's ex-ante expected utility by choosing the right disclosure policy.

To study the impact of psychology on information disclosure, we must decide how to model psychology and communication. We model the former by assuming that the agent's satisfaction depends not only on the physical outcome of the situation, but also directly on his updated beliefs. In doing so, we employ the framework of psychological preferences (Gilboa and Schmeidler (1988) and Geanakoplos, Pearce, and Stacchetti (1989)). This framework captures a wide range of emotions and situations and, in our model, it allows us to represent various psychological phenomena, including (forms of) purely psychological preferences, cognitive dissonance, temptation and self-control problems, multiplier preferences (Sargent and Hansen (2001)), and more.

We model communication by assuming that the principal can commit ex-ante, before the state is realized, to an information policy. This means that to every possible state of nature corresponds a message by the principal, leading to an updated belief of the agent. Since the state is distributed ex-ante according to some prior distribution, an information policy leads to a distribution over updated (or posterior) beliefs and (given how the agent responds to said information) over actions. This is the methodology of random posteriors introduced by Kamenica and Gentzkow (2011) (KG hereafter). Alternative methodologies are available, most notably the "Cheap Talk" model of Crawford and Sobel (1982). In that model, the principal cannot commit ex-ante. Here, we want to showcase the impact of psychological factors on information disclosure, which is more easily accomplished without the added complexity of strategic interactions between different types of the principal. Besides, a benevolent principal can be thought of as a third-party who regulates information disclosure between a sender and a receiver and who, by entering laws, creates exactly the type of commitment that we assume. Indeed, if someone is bound by law to reveal all the information that he knows, then he can decide which information to seek—one is (often) not accountable for not revealing what one does not know—and, in that, the informed party is committed ex-ante to revealing whatever information he will find.

Psychological preferences add a number of new considerations to information disclosure and call for new tools. First, it is important to understand the landscape of psychological preferences, for agents' relationship to information will guide disclosure. "Does the agent like information?" A classical agent is intrinsically indifferent to information. Holding his choices fixed, he derives no value from knowing more or less about the world. But if he can respond to information, then he can make better decisions and, thus, information is always a good thing (Blackwell (1953)). This observation leads us to distinguish between two general classes of information preference, each of them having two sub-classes. On one hand, an agent may be *psychologically* information-loving [or averse] if he likes [dislikes] information for its own sake. On the other, he may be *behaviorally* information-loving [or averse] if he likes [dislikes] information, taking its instrumental value into account. To see the distinction, imagine a person is giving a seminar, and consider the possibility of him finding out whether his presentation has any embarrassing typos right before he presents. If his laptop is elsewhere, he may prefer not to know about any typos, as knowing will distract him during the presentation. He is psychologically information-averse. But if he has a computer available to change the slides, he would like to know; any typos can then be changed, which is materially helpful. He is behaviorally information-loving.

We characterize these four classes of preferences in Theorem 1, and optimal information disclosure becomes easy to describe for three of them. Almost by definition, if the agent is behaviorally information-loving [-averse], it is best to tell him everything [nothing]. If an agent is psychologically information-loving, then the potential to make good use of any information only intensifies his preference for it; he is then behaviorally information-loving, so that giving full information is again optimal. More interesting is the case in which information causes some psychological distress, i.e., a psychologically information-averse agent. This case presents an intriguing dilemma, due to the tradeoff between the agent's need for information and his dislike for it. For such an agent, information is a liability that is only worth taking on if it brings enough instrumental value. A large part of the paper is devoted to information disclosure under information-aversion.

Under psychological information-aversion, and in psychological environments in general, the Revelation Principle no longer holds. In the presence of many potential disclosure policies, some of them being unreasonably cumbersome, this observation is disappointing. However, although the Revelation Principle fails in this environment, the basic message remains: Theorem 2 says that (when preferences display psychological information-aversion) it is sufficient to look at recommendation policies, that is, the principal can simply tell the agent what to do in an incentive-compatible fashion.[1] Recommendation policies are especially interesting because they are rather natural. In the doctor-patient example, the doctor may not tell the patient what the test statistically reveals about his illness, but she can simply tell him which action to take.

[1]To be sure, this result is nontrivial. Its proof is different from typical proofs of the Revelation Principle—a point on which we elaborate in the main text—and the result is false for general psychological preferences.

Many recommendation policies can a priori be optimal for a given agent with psychological information-aversion, including extreme ones such as full or no information. Can we say anything more for agents who are ranked by their information aversion? The question is important, for one because the optimal policy for an agent might just be found in the coarsening of another agent's policy. We propose an order on information-aversion that, by all standards of comparative *risk*-aversion, is very weak and yet obtain a strong conclusion. When two agents are ranked according to information aversion, their indirect utilities differ by the addition of a concave function. In this model, concavity is tantamount to disliking information everywhere, and hence one would think that a designer should respond by providing unambiguously less information to a more information-averse agent. Interestingly, this is true when the state of nature can take on two values, but it is true not in general due to psychological substitution effects.

We also provide general tools of optimal information disclosure to respond to the variety of psychological preferences, especially those that fall outside our classification. The concave envelope characterization of KG holds in our model (Theorem 3), and the benevolence of our design environment allows us to develop the method of "posterior covers." An agent may not be information-loving as a whole, but he can be locally information-loving at some beliefs. Accordingly, the principal knows to be as informative as possible in those regions. A posterior cover is a collection of convex sets of posterior beliefs over each of which the agent is information-loving. By Theorem 4, the search for an optimal policy can be limited to the extreme points of a posterior cover of the indirect utility. By Proposition 7, this problem can be reduced to that of finding posterior covers of primitives of the model. And by Proposition 8, there is a class of economic problems for which all these objects can be computed explicitly.

How should we talk to someone with cognitive dissonance? How should we talk to someone who has temptation and self-control problems? How should we talk to someone who likes to hear news only when it is good? We apply our results to answer those questions. In the case of cognitive dissonance,[2] the main message is that we should communicate with such an individual by an all-or-nothing policy. Either the person can bear information, in which case we should tell him the whole truth to enable good decision-making, or he cannot, in which case we should say nothing.

In the case of temptation and self-control, the optimal policy depends on the nature of information. We study a standard consumption-saving problem where an individual has im-

[2]In our model of cognitive dissonance, the agent pays a psychological penalty given by the (Euclidean) distance between his posterior belief and their prior position. To illustrate the range of applications of the model, we actually study the case of a politician with reputational concerns, but it is analytically equivalent to cognitive dissonance.

pulsive desires to consume rather than save/invest. When information is about the asset return, it only appeals to the agent's rational side. Hence, full information is optimal. But when information is about (say) vacation promotion, it only appeals to the agent's impulsive side. Yet, it does not follow that no information is optimal! Indeed, by revealing the promotion, it makes the agent consume and avoid all temptation.

Finally, we analyze a doctor-patient relationship in which the patient, who can be either healthy or ill, has equivocal information preferences. He likes to be informed about his health when the news is good, but when it is bad, he prefers not to know. So, he is neither information averse nor loving. Of course, it is suboptimal to reveal the news when it is good and say nothing otherwise, for the patient would infer that the news is bad when she hears nothing. It turns out that the doctor should only recommend 'no treatment' when the patient is healthy; but even when the patient is healthy, the doctor should at times recommend him to take the aggressive course of action (never smoke again, etc).

This paper is part of a burgeoning literature on information design. The most related work is KG. We adopt the same method as KG to study a different problem. They are interested in information disclosure when the principal and the agent have conflicting interests, and they study this problem with classical agents—their beliefs only affect their utility indirectly. We are interested in information disclosure with psychological agents, and we study this problem when the principal and the agent have aligned interests. At the intersection of both works is but one trivial configuration: a non-psychological agent with the same preferences as the principal. In this case, full disclosure is trivially optimal.

There is also a recent literature on information design in games, i.e. with one informed principal and many interacting agents (Bergemann and Morris (2013a,b) and Taneva (2014)), and on dynamic information design (Ely, Frankel, and Kamenica (2014)).

Our work does not belong to the cheap talk literature spanned by Crawford and Sobel (1982), because our principal can commit to a communication protocol. A notable contribution to that literature is Caplin and Leahy (2004) who study a detailed interaction between a doctor and a patient with quadratic psychological preferences. Unlike us, they allow for private information on the part of the agent. Our analysis imposes less specific structure but allows for a great range of applications.

Finally, there is a conceptual connection between psychological preferences and preferences over temporal resolution of uncertainty (e.g., Kreps and Porteus (1978)). In our framework, the latter can be reinterpreted as psychological preferences. As such, the mathematical arguments underlying our classification of attitudes toward information are similar to Grant, Kajii, and Polak (1998). However, information disclosure imposes additional constraints that are absent in that literature, such as Bayes-consistency. Indeed, the beliefs of a

Bayesian agent cannot be manipulated at will.[3] Moreover, our behavioral classification of psychological agents is not always relevant in existing studies.[4]

This paper is organized as follows. Section 2 introduces psychological preferences and the methodology of random posteriors. Section 3 presents the psychological agent, and studies and characterizes various attitudes towards information, namely psychological and behavioral information preferences. Section 4 concentrates on psychological information-aversion, highlighting information disclosure and proposing an order to compare information-aversion. Section 5 supplies the general tools of optimal information design. Section 6 applies these tools to various situations, including forms of cognitive dissonance and bounded rationality.

2 The Model

Consider an agent who must make a decision when the state of nature $\theta \in \Theta$ is uncertain.[5] Suppose that the agent has (full-support) prior $\mu \in \Delta\Theta$ and that he has access to additional information about the state. After receiving that information, the agent updates his prior beliefs and then makes a decision.

2.1 Psychological Preferences

In this paper, an *outcome* is an element (a, ν) of $A \times \Delta\Theta$ where a is the action taken by the agent and $\nu \in \Delta\Theta$ denotes the posterior belief of the agent at the moment when he must make his decision. We assume that the agent has some continuous utility function $u : A \times \Delta\Theta \longrightarrow \mathbb{R}$ over outcomes. The true state is excluded from the utility function without loss of generality. Indeed, given any true underlying preferences $\tilde{u} : A \times \Theta \times \Delta\Theta \longrightarrow \mathbb{R}$, we can define the reduced preferences u via $u(a, \nu) = \int_\Theta \tilde{u}(a, \cdot, \nu)\, \mathrm{d}\nu$. The reduced preferences are the only relevant information, both behaviorally and—assuming μ is empirically correct—from a welfare perspective.[6] Given posterior beliefs ν, the agent chooses an action $a \in A$ so

[3] On average, the posteriors of a Bayesian must equal the prior. This sometimes limits the type of tools that we can use, for example to compare information-aversion (see Section 6).

[4] Since the principal is interested in the agent's welfare after he acts, intrinsic preferences for information (i.e. before he acts), as studied in Grant, Kajii, and Polak (1998), cannot be the only relevant ones here.

[5] In this paper, all spaces are assumed nonempty compact metrizable spaces, while all maps are assumed Borel-measurable. For any space Y, we let $\Delta Y = \Delta(Y)$ denote the space of Borel probability measures on Y, endowed with the weak* topology, and so itself compact metrizable. Given $\pi \in \Delta Y$, let $\mathrm{supp}(\pi)$ denote the support of π, i.e. the smallest closed subset of Y of full π-measure.

[6] Later, when we consider questions of information design by a benevolent principal, we will also work with u rather than $\tilde{u}$. This remains without loss, but only because the designer is assumed to have full commitment power. This is in contrast to the setting of Caplin and Leahy (2004).

as to maximize $u(a, \nu)$.

In the *classical case*, $\tilde{u}$ does not depend on ν and so $u(a, \cdot)$ is affine for every a. We do not make this assumption here. In our environment, the agent's satisfaction depends not only on the physical outcome of the situation, but also on his posterior beliefs. In the literature, this is known as *psychological preferences* (Gilboa and Schmeidler (1988) and Geanakoplos, Pearce, and Stacchetti (1989)). In a game-theoretic context, a player's utility function could depend on the beliefs of the other players. Preferences concerning fashion and gossip are examples of this phenomenon. In a dynamic context, a player's utility at some time could depend on the relationship between his current beliefs and any of his past beliefs. Surprise and suspense are examples (Ely, Frankel, and Kamenica (2014)).

In a single-agent static environment, our formulation covers a wide range of phenomena. Consider the following examples where $\Theta = \{0, 1\}$ and $A = [0, 1]$, and note that we can write any posterior belief as $\nu = \text{Prob}(\{\theta = 1\})$:

(a) (Purely psychological agent) Let

$$u(a, \nu) = -\mathbb{V}(\nu)$$

represent an agent whose only satisfaction comes from his degree of certainty, represented by the variance $\mathbb{V}(\nu) = \nu(1 - \nu)$.

(b) (Psychological and material tensions) For $k \in \mathbb{R}$, let

$$u(a, \nu) = k\mathbb{V}(\nu) - \mathbb{E}_{\theta\sim\nu}\left[(\theta - a)^2\right]$$

This agent wants to guess the state but he has a psychological component based on the variance of his beliefs. When $k > 0$, the agent faces a tension between information, which he dislikes per se, and his need for it to guess the state accurately.

(c) (Cognitive dissonance) Let

$$u(a, \nu) = -|\nu - \mu| - \mathbb{E}_{\theta\sim\nu}\left[(\theta - a)^2\right]$$

represent an agent who wants to guess the state but experiences discomfort when information conflicts in any way with his prior beliefs. These preferences are familiar from Gabaix (2014)'s sparse-maximizer, though the economic interpretation is quite different: a distaste for dissonance, rather than a cost for considering information.

(d) (Temptation and self-control) Given two classical utility functions u_1 and u_2, and mixed actions A, let

$$u(a, \nu) = \mathbb{E}_{\theta\sim\nu}\left[u_1(a, \theta) + u_2(a, \theta) - \max_{b\in A} \mathbb{E}_{\theta\sim\nu}[u_2(b, \theta)]\right].$$

These preferences are directly inspired from Gul and Pesendorfer (2001). After choosing a menu A, the agent receives some information and then chooses an action. Her non-tempted self experiences utility u_1 from action a, while $\max_{b\in A}\mathbb{E}_{\theta\sim\nu}[u_2(b,\theta)-u_2(a,\theta)]$ is interpreted as the expected cost of self-control.

(e) (Multiplier preferences) For $k>0$, let

$$u(a,\nu)=\min_{\nu'}\int u_c(a,\theta)d\nu'(\theta)+kR(\nu'\|\nu)$$

where $R(\nu'\|\nu)=\sum_\theta \ln\left(\frac{\nu'(\theta)}{\nu(\theta)}\right)\nu'(\theta)$ is the relative entropy of ν' with respect to ν. These are the multiplier preferences proposed by Sargent and Hansen (2001). The agent has some best guess ν of the true distribution, but he does not fully trust it. Thus, he considers other probabilities ν' whose plausibility decreases as the distance to ν increases.

2.2 Signals and Random Posteriors

Before making his decision, the agent has access to additional information, which is given by a signal.

Definition 1. *A signal (S,σ) on (Θ,μ) is a space S equipped with a map $\sigma:\Theta\longrightarrow\Delta S$.*

This definition describes a technology that sends a random message to the agent in each state: if the state is θ, the realized message $s\in S$ is distributed according to $\sigma(\cdot|\theta)\in\Delta S$. The signal is the *only* information that the agent receives about the state.

Given the signal (S,σ) and the realized message s, the agent forms posterior beliefs about the state, denoted $\beta^{S,\sigma}:S\to\Delta\Theta$.[7] In each state, a signal sends a message (possibly random) to the agent, and upon receiving the message, the agent forms a posterior belief. Therefore, from an ex-ante perspective—when the state has not yet realized—a signal induces a distribution over posterior beliefs. We call this distribution, which is an element of $\Delta\Delta\Theta$, a *random posterior*.[8] The random posterior correspondence,

$$\begin{aligned}\mathcal{R}:\Delta\Theta &\rightrightarrows \Delta\Delta\Theta\\ \mu &\longmapsto \{p\in\Delta\Delta\Theta:\ \mathbb{E}_{\nu\sim p}\nu=\mu\}\\ &= \left\{p\in\Delta\Delta\Theta:\ \int_{\Delta\Theta}\nu(\hat\Theta)\,\mathrm{d}p(\nu)=\mu(\hat\Theta)\text{ for every Borel }\hat\Theta\subseteq\Theta\right\},\end{aligned}$$

[7] In the finite case, the posterior beliefs are given by $\beta^{S,\sigma}(\theta|s)=\sigma(s|\theta)\mu(\theta)/\sum_{\hat\theta}\sigma(s|\hat\theta)\mu(\hat\theta)$ for all $\theta\in\Theta$ and $s\in S$. In general, for any Borel $\hat\Theta\subseteq\Theta$, the map $\beta^{S,\sigma}(\hat\Theta|\cdot):S\longrightarrow\Delta\Theta$ is the *conditional expectation* of $1_{\hat\Theta}$ conditional on s.

[8] The map $(\theta,s)\mapsto\beta^{S,\sigma}(s)$ is a $\Delta\Theta$-valued random variable. The induced members of $\Delta\Delta\Theta$ are the distributions of this random variable.

maps every prior into the set of random posteriors whose mean equals the prior. As shown in Kamenica and Gentzkow (2011) and in the lemma below, Bayesian updating must yield random posterior beliefs in $\mathcal{R}(\mu)$, and conversely, all such random posteriors can be generated by some signal. This means that, while there is room for manipulation, the posterior beliefs of a Bayesian agent must on average be equal to his prior.

Lemma 1. *A signal* (S, σ) *induces a unique random posterior* $p^{S,\sigma} \in \mathcal{R}(\mu)$*. Conversely, given* $p \in \mathcal{R}(\mu)$*, there is a signal on* (Θ, μ) *inducing* p*.*

One could imagine many different signals that induce a given random posterior, but it is useful to consider the *direct signal* (S_p, σ_p) associated with p. The direct signal is one for which (*i*) every message is a posterior in $\Delta\Theta$ and (*ii*) when the agent hears message 's', his update yields a posterior belief equal to s. In short, the signal tells the agent what his posterior belief should be, and his beliefs abide. For this signal, the space of messages is $S_p \subseteq \Delta\Theta$ and, in the finite case, the signaling map is defined as $\sigma_p(s|\theta) = s(\theta)p(s)/\mu(\theta)$ for each $s \in S_p$ and θ.[9] One can easily verify that this signal induces posterior belief s (see Footnote 7) with probability $p(s)$.

3 The Psychological Agent

In this section, we describe the attitudes and the behavior of our psychological agent towards information. This leads us to classify a wide range of psychological agents into four classes of preferences: psychological information-averse [-loving] agents, and behaviorally information-averse [-loving] agents.

3.1 Information as Risk

Prospective information is a form of risk. Indeed, getting no information entails a deterministic posterior belief: the prior for sure remains the agent's only information. But getting some information about the state entails a random posterior belief, because the posterior then depends on the realized state, which is initially uncertain. In general, when the signal is not very informative, posterior beliefs will be concentrated around the prior, while when it is informative, posterior beliefs will be more dispersed. In this context where beliefs are payoff

[9] In the general case,

$$\sigma_p(\hat{S}|\theta) = \int_{\hat{S}} \frac{\mathrm{d}s}{\mathrm{d}\mu}(\theta)\, \mathrm{d}p(s) \text{ for every } \theta \in \Theta \text{ and Borel } \hat{S} \subseteq S_p.$$

relevant, it is appropriate to think of psychological information preferences as a particular instance of risk preferences.

Definition 2. *Given two random posteriors, $p, q \in \mathcal{R}(\mu)$, p is more (Blackwell-) informative than q, denoted $p \succeq_B^\mu q$, if (S_q, σ_q) is a garbling of (S_p, σ_p), i.e. if there is a map $g : S_p \longrightarrow \Delta(S_q)$ such that*

$$\sigma_q(\hat{S}|\theta) = \int_{S_p} g(\hat{S}|\cdot) d\sigma_p(\cdot|\theta) \tag{1}$$

for every $\theta \in \Theta$ and $\hat{S} \subset S_q$.

This definition of informativeness based on posteriors is equivalent to the traditional definition of informativeness based on signals. Usually, we say that a signal is more informative than another if the latter is a garbling of the former (see Blackwell (1953)). But if signal (S, σ) generates random posterior p, then it is easy to verify that (S, σ) and (S_p, σ_p) are Blackwell-equivalent in the usual sense. Therefore, a signal is more informative than another in the traditional sense if and only if it induces a more informative random posterior.

The next proposition formalizes the connection between informativeness and risk.[10]

Proposition 1. *Given two random posteriors $p, q \in \mathcal{R}(\mu)$, the ranking $p \succeq_B^\mu q$ holds if and only if there is a map $r : S_q \longrightarrow \Delta(S_p) \subseteq \Delta\Delta\Theta$ such that*

1. *For every Borel $S \subseteq \Delta\Theta$,*
$$p(S) = \int_{S_q} r(S|\cdot)\, dq.$$

2. *For every $t \in S_q$,*
$$r(\cdot|t) \in \mathcal{R}(t).$$

The proposition establishes a somewhat counterintuitive result. The random posterior p represents better information than q about θ if and only if it is a *mean-preserving spread* of q. That is, greater certainty about the state ex-post entails greater risk concerning one's own beliefs ex-ante. This will be important in our analysis, because beliefs (like the prizes of a lottery) directly enter the agent's utility, and so attitudes towards information (like attitudes towards risk) will play a critical role.

3.2 Love and Aversion for Information

The agent's welfare depends on his attitude toward information. In particular, we make a distinction between *psychological* and *behavioral* attitudes toward information.

[10]For finite-support random posteriors, the result follows from Grant, Kajii, and Polak (1998, Lemma A.1). Ganuza and Penalva (2010, Theorem 2) prove a related result in the special case of $\Theta \subseteq \mathbb{R}$.

Definition 3. *The agent is psychologically information-loving [resp. -averse, -neutral] if given any action $a \in A$ and any random posteriors $p, q \in \mathcal{R}(\mu)$ with $p \geq_B^\mu q$,*

$$\int_{\Delta\Theta} u(a,\cdot)\, dp \geq [resp.\ \ \leq, =] \int_{\Delta\Theta} u(a,\cdot)\, dq.$$

An agent is psychologically information-loving [-averse] if he likes [dislikes] informativeness for its own sake, in the hypothetical event that he cannot adapt his decisions to it. That is, information is intrinsically valuable [damaging], abstracting from its instrumental value.

In the next definition, the agent is assumed to respond optimally to information, which defines his behavioral attitude towards it.

Define the *posterior value function* as

$$\begin{aligned} U : \Delta\Theta &\longrightarrow \mathbb{R} \\ \nu &\longmapsto \max_{a \in A} u(a,\nu), \end{aligned} \tag{2}$$

i.e. $U(\nu)$ is the indirect utility associated with a posterior belief ν.

Definition 4. *The agent is behaviorally information-loving [resp. -averse, -neutral] if given any random posteriors $p, q \in \mathcal{R}(\mu)$ with $p \geq_B^\mu q$,*

$$\int_{\Delta\Theta} U\, dp \geq [resp.\ \ \leq, =] \int_{\Delta\Theta} U\, dq \tag{3}$$

To illustrate the distinction between psychological and behavioral attitudes toward information, consider the following simple story. A job market candidate has a job talk, and he may or may not find out five minutes before the presentation whether his slides have embarrassing typos. Knowing that his slides are fine makes the candidate comfortable, and learning that they are not makes him unhappy. Consider now the difference between having his laptop when he finds out the information and not having it. If he will not have it, the person might prefer to remain ignorant: the risk of hearing bad news (and being self-conscious during the presentation) outweighs the minor benefit of hearing good news. If he will have his laptop, however, he will have an opportunity to respond to the situation: there is time to fix the typos. Hence, the same person may like information when he can respond to it, while he may prefer to remain ignorant about that same information if he cannot respond to it. If so, this person intrinsically dislikes information (psychological information-aversion), but it is more than compensated by the benefit of being able to *use* it (behavioral information-loving). A formal example will be presented after the theorem.

Theorem 1. *Psychological and behavioral information preferences are closely linked.*

1. *The agent is psychologically information-loving [resp. -averse, -neutral] if and only if $u(a, \cdot)$ is convex [resp. concave, affine] for every $a \in A$.*

2. *The agent is behaviorally information-loving [resp. -averse, -neutral] if and only if U is convex [resp. concave, affine].*

3. *If the agent is psychologically information-loving, then he is behaviorally information-loving as well.*

An immediate consequence of the first two points is that a classical agent is psychologically information-neutral and—in an expression of the easy direction of Blackwell's theorem—behaviorally information-loving, because the linearity of u in ν implies that U is convex.

To Illustrate the theorem, consider Example (b) in Section 2.1. Given that $\mathbb{V}(\nu) = \nu(1-\nu)$ and the expectation is linear in ν, we have $\partial^2 u/\partial\nu^2 = -2k$. So, the theorem tells us that the agent is psychologically information-loving if $k < 0$, and psychologically information-averse if $k > 0$. Clearly, the agent's optimal action is $a^*(\nu) = \mathbb{E}_\nu[\theta] = \nu$, and so

$$U(\nu) = (k-1)\mathbb{V}(\nu) = (k-1)\nu(1-\nu).$$

Given $\partial^2 U/\partial\nu^2 = -2(k-1)$, the agent is behaviorally information-loving if $k < 1$, and behaviorally information-averse if $k > 1$. Therefore, for $k \in (0,1)$, the agent is psychologically information-averse, but behaviorally information-loving.

4 Designing Information

Imagine there is a designer who chooses what type of signal the agent will see, as in Kamenica and Gentzkow (2011). Assume that the designer's and the agent's preferences are perfectly aligned, in the sense that the designer wants to maximize the agent's ex-ante expected welfare.

Definition 5. *A random posterior $p \in \mathcal{R}(\mu)$ is an optimal policy if $\int_{\Delta\Theta} U\,dp \geq \int_{\Delta\Theta} U\,dq$ for all $q \in \mathcal{R}(\mu)$. A signal (S, σ) is optimal if it induces an optimal policy.*

For purposes of characterizing optimal direct signals, the designer need only consider the *design environment* $\langle \Theta, \mu, U \rangle$. As we shall see later, the extra data A and u can be relevant for constructing reasonable indirect signals.

An optimal policy is immediate for certain classes of agents. If the agent is behaviorally information-loving [-averse], then the most [least] informative policy is optimal: one should tell him everything [nothing]. Part 3 of Theorem 1 then tells us that fully revealing the state is also optimal for psychologically information-loving agents. For example, it is easy to show that agents with multiplier preferences (case (f) in Section 2.1) are psychologically information-loving,[11] so that they are behaviorally information-loving.

Things are not so simple for agents who exhibit psychological information-aversion, because they are often conflicted about the value of information. These agents are the focus of the following section.

5 Psychological Information-Aversion

Information-aversion presents an intriguing tradeoff between the need for informativeness, so that the agent can make the right decision, and the agent's intrinsic dislike for information. Given the relationship between risk and information, aversion to information is also perhaps a realistic assumption (Harrison and Rutstrom (2008)). In this section, we show that some natural policies are always optimal for information averse individuals.

So far, the designer has used unlimited tools to maximize welfare, with a focus on posterior policies. Those policies, called direct signals in Section 2.2, suppose that the designer sends entire posterior distributions to the agent. In a world with binary states, this can be plausible. For example, a doctor who uses a blood test to check for some disease may give a probability that the patient has the illness rather than not. When the state space is larger and more complicated, posterior policies are cumbersome and unrealistic. It is important to know if, instead of telling the patient what he should believe, a doctor could simply recommend some course of action. The Revelation Principle usually solves the problem, but as we will see, it does not hold in psychological environments.

5.1 Recommendation Policies

A *recommendation policy* is one that recommends an incentive-compatible action. These policies are appealing because they simplify the designer's communication protocol; they also ring true to casual observations. In the doctor-patient example, the doctor may not describe to the patient what the test statistically reveals about his illness, but she could simply tell him which action to take, such as diet and lifestyle.

[11] That $u(a, \cdot)$ is convex follows readily from the well-known fact that the relative entropy term $R(\nu' \| \nu)$ is convex in $(\nu', \nu) \in (\Delta\Theta)^2$. Part 1 of Theorem 1 then applies.

5.1.1 Failure of the Revelation Principle

All information policies admit a corresponding recommendation policy (with a smaller message space) in which the agent is simply told how he would have optimally responded to his posterior. But there is no guarantee a priori that the agent will want to follow this recommendation. The Revelation Principle (as in Aumann (1974), Myerson (1979), Kamenica and Gentzkow (2011)) is the result that usually guarantees it.

The Revelation Principle says that, for *any* signal (S, σ), the corresponding recommendation policy is incentive-compatible.

In psychological environments, however, the Revelation Principle (Proposition 1 in KG) does not hold. For example, take $\Theta = A = \{0, 1\}$ and $u(a, \nu) = a(\mathbb{V}(\nu) - 1/9)$ so that this agent is psychologically information-averse.[12] Let $\mu = .6$ and consider a signal that generates random posterior $p = \frac{3}{8}\delta_{.1} + \frac{5}{8}\delta_{.9}$. As shown in Figure 1, whether the posterior is .1 or .9, the agent wants to play action 0 because $u(0, \nu) > u(1, \nu)$ when $\nu \in \{.1, .9\}$.

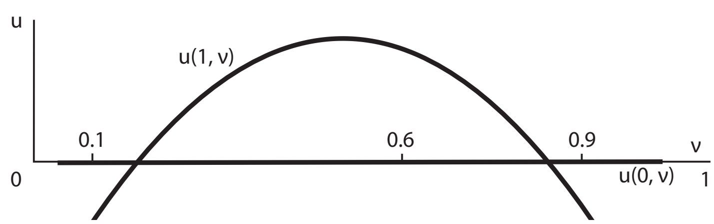


Figure 1: Here, $\{\nu : \ a = 0 \text{ is optimal at } \nu\}$ is not convex.

If, instead, the designer told the agent what he would have played, i.e., always sent the message "play $a = 0$," the agent would receive no information and his posterior would always be $\mu = .6$. By Figure 1, the agent would then want to play action 1 and the recommended action, $a = 0$, would not be incentive-compatible.

The reason why the Revelation Principle fails here is because the set of posterior beliefs for which a given action is optimal may not be convex. In classical environments, if an action is optimal for certain messages, then it is without loss to replace each of the messages with a direct recommendation of said action. By doing so, we "pool" these messages into one and instead send that pooled message to the agent, thereby also pooling the posteriors attached to these messages. In a psychological world, where the $u(a, \cdot)$'s are not linear, the action may not remain optimal under the averaged out posterior beliefs.

[12] Recall that a distribution is represented as the probability ν that state 1 occurs and hence $\mathbb{V}(\nu) = \nu(1 - \nu)$.

5.1.2 Result

In spite of the above, we are able to establish a result that is most useful for our purposes in the case of psychological information-aversion.

Theorem 2. *If the agent is psychologically information-averse, then some recommendation policy is optimal.*

It is surprising at first that a recommendation policy can serve the needs of all psychologically information-averse agents, even the behaviorally information-loving ones, because passing from a policy to its associated recommendations can only "lose" information. When full revelation is optimal and the agent is psychologically information-averse, this too can be implemented by a recommendation policy, if the optimal action reveals the state. The principal then recommends action $a = \theta$ with probability one.

The theorem relies on establishing that for any *optimal* signal (S, σ), the corresponding recommendation policy is incentive-compatible. The intuition is that if a policy maximizes the agent's utility, then playing what is recommended should be incentive-compatible. But maximizing the agent's utility is only relevant for benevolent design and, thus, this result would not hold if the principal and the agent had conflicting interests.

Similarly, the result breaks down for some information-loving agents. Consider $A = \{\bar{a}\}$, $\theta \in \{0, 1\}$, and $u(\bar{a}, \nu) = -\mathbb{V}(\nu)$. The only recommendation is $\bar{a}$ in all states, which gives the agent no information to update his prior, and so $\mathbb{V}(\nu) = \mathbb{V}(\mu)$. Clearly, it would be preferable to tell the agent which state has occurred, so that $\mathbb{V}(\nu) = 0$.

5.2 No Information

Two salient questions in the study of information design are when should one tell the agent everything, and when should one tell the agent nothing? In previous sections, we presented sufficient conditions for the first: if an agent is psychologically information-loving, then he is behaviorally information-loving, and so full revelation is optimal. The next result gives a sufficient condition on u for behavioral information-aversion.

In the special case in which u is concave, a constant recommendation policy—i.e. giving no information—is optimal.

Proposition 2. *Suppose $A \subset \mathbb{R}^k$ is convex.*[13] *If u is concave, then the agent is behaviorally information-averse. In particular, giving no information is optimal.*

[13] The same proposition can also be applied to the case of A finite, by extending the action space to include mixed actions.

Proof. Suppose u is concave. By the (convex) maximum theorem, U is concave. By Theorem 1, the agent is behaviorally information-averse. By Theorem 3, no information is optimal. □

The possibility to provide such sufficient conditions on primitives comes from our benevolent design setting. In an environment with conflicting interests, sufficient conditions must involve derived objects (such as KG's $\hat{v}$).

6 Comparative Information-Aversion

Many recommendation policies can be optimal for some psychologically information-averse agents, including extreme ones such as full or no information. Can we say anything more for agents who are ranked in terms of their information aversion? This is the question that we answer in this section. First, we propose an order on information-aversion and then explain how the optimal policy adapts to changes in information-aversion.

6.1 Ordering Relation

We begin by comparing attitudes towards information across agents. Consider two individuals whose indirect utilities (2) are denoted by U_1 and U_2.

The close relationship between risk and information (Section 3.1) suggests various ways of measuring information-aversion. We adopt an approach based on observed choices and say that an agent is more information-averse than another when the following applies:

Definition 6. *Agent 2 is more (behaviorally) information-averse than agent 1 if for any p and q such that $p \succeq_B^\mu q$,*

$$\int U_1\, dq \geq [resp.\ >] \int U_1\, dp \quad \Longrightarrow \quad \int U_2\, dq \geq [resp.\ >] \int U_2\, dp.$$

An agent is more information-averse than another, if any time the latter would prefer a less informative policy over a more informative one (with the same mean), so would the former agent.

This definition is qualitative rather than quantitative; it compares *when* two decision-makers prefer more/less information rather than the *degree* of such preference. This is quite different from the standard literature on comparative risk aversion (started by Pratt (1964) and Arrow (1971), including Kihlstrom and Mirman (1974), etc), in which risk aversion is measured by comparison to risk-free (i.e., degenerate) lotteries, called certainty equivalents.

Usually, an agent is said to be more risk averse than another, if all risky lotteries are worth no more to him than to the other individual, in risk-free terms. As such, certainty equivalents provide an objective "yardstick" to quantify risk aversion.

In the context of information disclosure, where random posteriors can be seen as lotteries over lotteries, Lemma 1 implies that the only available risk-free lottery is the no-information policy. Indeed, all our lotteries must have the same mean (the prior), hence the only risk-free lottery puts probability one on the prior. Without the tool of certainty equivalents—absent, since average posterior beliefs must always equal the prior—we have no objective yardstick. Of course, we could ask each agent which prior belief would hypothetically make him indifferent to any given policy, but this would seem like an unrealistic exercise.

Beyond definitional matters, adding motives for disliking information ought to lead to increasing information-aversion. Theorem 1 describes concavity as the expression of dislike for information and the next result confirms this intuition.

Proposition 3. *If $U_2 = \gamma U_1 + C$ where $\gamma > 0$ and $C : \Delta\Theta \to \mathbb{R}$ is a concave function, then agent 2 is more information-averse than agent 1.*

An interesting application of these ideas is to the mixture between any classical and any psychological preference. Consider a psychological utility u_1 that is independent of actions, a classical utility u_c, and define

$$u_\lambda(a,\nu) = \lambda u_1(\nu) + (1-\lambda)\int_\Theta u_c(a,\cdot)\,\mathrm{d}\nu, \tag{4}$$

where $\lambda \in [0,1]$ measures the departure from the classical case. Let U_λ be the indirect utility associated with u_λ, as defined in (2). For any $\lambda'' > \lambda'$ and letting $\gamma = \lambda''/\lambda'$, we have $U_{\lambda''} = \gamma U_{\lambda'} + (1-\gamma)U_0$. Since $(1-\gamma)U_0$ is concave,[14] the more psychological (4) becomes (i.e. the larger λ), the more information-averse he becomes. Interestingly, this is true regardless of the shape of the psychological component.

6.2 Comparative Statics of Optimal Policies

Almost all information policies are pairwise incomparable with respect to (Blackwell) informativeness (Proposition 9 in the Appendix). In light of this, our definition of comparative information-aversion appears to be extremely weak, as it only speaks to how two agents rank Blackwell comparable policies. Yet, this definition has strong implications for utility representation and, in two-state environments, this is enough to obtain comparative statics of optimal policies.

[14] By Theorem 1, U_0 is convex. Since $\lambda'' > \lambda'$, $(1-\lambda''/\lambda')U_0$ is concave.

Proposition 4. *Suppose agent 2 is more information-averse than agent 1, but 1 is not behaviorally information-loving. Then $U_2 = \gamma U_1 + C$ for some $\gamma \geq 0$ and concave $C : \Delta\Theta \to \mathbb{R}$.*

When two agents are ranked according to information aversion, their indirect utilities differ by the addition of a concave function. The proof of this result hinges on formulating comparative information-aversion as incomplete preference relations on the space of random posteriors, which then allows us to apply the representation uniqueness theorem from Dubra, Maccheroni, and Ok (2004).

This is one of the strongest conclusions that we could hope for, since it amounts to adding a motive for disliking information everywhere. From an information design perspective, one would think that a designer should respond by providing unambiguously less information to a more information-averse agent. But this is not optimal in general. In Section 9.6.2, we give a counterexample, the main idea of which is simple: two agents may be ranked by information aversion, but due to substitution effects, their optimal policies may not be Blackwell comparable. Indeed, suppose that the state is bi-dimensional, to be interpreted as two distinct sources of information. Suppose that both agents regard these sources as substitutes. If an agent will be informed by either source, then he prefers not to learn from the other, perhaps because he fears contradiction. Assume that agent 2 dislikes more than 1 the information coming from the first source. Then, it is optimal to provide 2 with more (less) information than 1 coming from the second (first) source. As a result, 2's optimal policy is not Blackwell comparable to 1's even though 2 is more information averse.

In the following proposition, we show that this multidimensionality issue is the only thing that comes in the way of comparative statics. In a world of binary uncertainty, a designer should provide a less informative policy to a more information-averse agent.

Proposition 5. *In a two-state environment, if agent 2 is more information-averse than agent 1, then there exist optimal policies p_1^* and p_2^* such that $p_1^* \succeq_B^{\mu} p_2^*$.*

In particular, if p_i^* is uniquely optimal for all i, then $p_1^* \succeq_B^{\mu} p_2^*$. This comparative statics result has real prescriptive power in that an agent who has information p_1^* is equipped to provide p_2^* to the other agent. For instance, if a doctor considers whether to tell parents the sex of their child, and the father is more information-averse than the mother, the doctor can simply provide optimal information p_1^* to the mother, who will then be equipped to inform the father. Likewise, if a child experiences only psychological consequences of discovery of some illness, while the parents care about the former and about making informed medical decisions, then a pediatrician may simply choose an optimal information policy for the parents, and leave the parents to tell their child about the diagnosis.

7 Optimal Policies: The General Case

So far, we have focused on optimal policies for special classes of psychological preferences. For psychologically information-loving agents, Theorem 1 tells us that full information is optimal. For psychologically information-averse agents, things are subtler. In some sense, Proposition 2 tells us that if the agent is extremely psychologically information-averse, then it is best to say nothing. More broadly, Theorem 2 delineates a family of optimal policies for this class, a large step toward characterizing optimal policies, but not a full solution. The general equivocal case—in which the agent is neither information averse nor loving—is subtler still. For such cases, we are in need of more general tools.

7.1 Optimal Welfare

Since any random posterior $p \in \mathcal{R}(\mu)$ can be induced by some signal, designing an optimal signal is equivalent to choosing the best random posterior in $\mathcal{R}(\mu)$. From there, we can determine the expected utility levels that an agent can reach with an appropriately chosen signal.

It follows from Lemma 1 that a distribution over outcomes $\Delta(A \times \Delta\Theta)$, or equivalently, over resulting utilities $\Delta(u(A \times \Delta\Theta))$, can be induced by a signal if and only if that distribution can be induced by a Bayes-consistent random posterior.

Proposition 6. *There exists a signal (S, σ) that induces distribution $\eta \in \Delta(u(A \times \Delta\Theta))$ if and only if there is some $p \in \mathcal{R}(\mu)$ such that $\eta = p \circ U^{-1}$.*

One of the most important results in Kamenica and Gentzkow (2011) is the full characterization of the maximal ex-ante expected utility. This characterization is in terms of the concave envelope of U, defined as

$$\bar{U}(\nu) = \min\{\phi(\nu) : \quad \phi : \Delta\Theta \to \mathbb{R} \text{ affine continuous}, \phi \geq U\}.$$

This result carries over to our environment for all psychological preferences:

Theorem 3. *There is a signal that induces expected value $\bar{U}(\mu)$, and no signal induces a higher value.*

7.2 Optimal Policies

The concave envelope is the general expression of optimality in the model (Theorem 3) as well as an encoding of the optimal policy. Unfortunately, it is a difficult task to characterize the concave envelope of a function in general, even more so to compute it. The main issue is that it should be solved for globally: evaluating it at one point can be as difficult as global maximization of the function itself (Tardella (2008)).

We approach the problem by reducing the support of the optimal policy based on local arguments, a method which will be successful in a class of problems that include many economic applications of interests.

Oftentimes, the designer can deduce from the primitives of the problem that the indirect utility U must be locally convex on various regions of $\Delta\Delta\Theta$. In every one of those regions, the agent likes (mean-preserving) spreads in beliefs, as they correspond to more informative policies. In consequence, an optimal policy need not employ beliefs in the interior of those regions, regardless of its general shape.

The central concept of our approach is that of posterior cover.

Definition 7. *Given a function $f : \Delta\Theta \to \mathbb{R}$, an f-(convex posterior) cover is a family $\mathcal{C}$ of closed convex subsets of $\Delta\Theta$ such that $\bigcup \mathcal{C} = \Delta\Theta$ and $f|_C$ is convex for every $C \in \mathcal{C}$.*

A posterior cover is a collection of sets of posterior beliefs over each of which some function is convex. Given a posterior cover $\mathcal{C}$, let

$$ext^*(\mathcal{C}) = \{\nu \in \Delta\Theta : \quad \forall C \in \mathcal{C} \text{ if } \nu \in C, \text{ then } \nu \in ext(C)\}$$

be the set of its extreme points.[15] The next theorem explains why posterior covers and their extreme points play an important role in finding an optimal policy.

Theorem 4. *If $\mathcal{C}$ is a countable U-cover, then there is an optimal policy p with $p(ext^*\mathcal{C}) = 1$.*

The search for an optimal policy can be limited to the extreme points of a U-cover. For this reason, we focus on posterior covers of the indirect utility. Clearly, Theorem 4 is only useful if three conditions are met. Given that the indirect utility is a derived object, it is important to tie the U-cover to primitives of the model. Afterwards, all calculations ought to be relatively easy, otherwise the result will not be useful in practice. Lastly, the set of extreme points must be small, so that solving for an optimal policy is tractable; in particular, if there are finitely many extreme points, computing the optimal policy and the concave envelope is a linear programming problem. We deal with these conditions one after the other.

First, we reduce the problem of finding a U-cover to that of finding posterior covers of primitives of the model.

[15] A point $\nu \in ext(C)$ if there is no nontrivial (of positive length) segment for which ν is an interior point.

Proposition 7. *1. If C_a is a $u(a,\cdot)$-cover for every $a \in A$, then*

$$C := \bigvee_{a\in A} C_a = \left\{ \bigcap_{a\in A} C_a : \quad C_a \in C_a \text{ for each } a \in A \right\}$$

is a U-cover.

2. If u takes the form

$$u(a,\nu) = u_p(\nu) + \int_\Theta u_c(a,\cdot)\, d\nu \tag{5}$$

and C is a u_p-cover, then C is a U-cover.

The benevolent feature of our design problem enables us to derive a U-cover from primitives of the model. By the first part of the proposition, if we can determine the posterior cover of $u(a,\cdot)$ for each a, then it is easy to name the U-cover. The second part of the proposition presents a class of problems for which the determination of every $u(a,\cdot)$-cover comes down to deriving the posterior cover of a single primitive function. This class of problems consists of all additive preferences between a psychological and a classical component. This includes all preferences in Section 2.1 from (a) to (d). For these problems, optimal design relies on finding the posterior cover of the psychological component alone.

After connecting the problem to posterior covers of primitives, it remains to elicit these posterior covers and find their extreme points. Doing so in complete generality is not practical, but we can point to a class of functions for which it can be done.

Proposition 8. *If the state space is finite and f is the pointwise minimum of a finite family of affine functions $\{f_i : \Delta\Theta \to \mathbb{R}\}_{i\in I}$, then $C := \{C_i : i \in I\}$ with*

$$C_i := \{\nu \in \Delta\Theta : f(\nu) = f_i(\nu)\}$$

is an f-cover.

The case described here, namely the minimum of affine functions, includes many economic situations of interest. For these problems, we can explicitly compute the U-cover. In the additive case (5) alone, it is common that $u_p(\nu) := -\max\{f_i(\nu)\} = \min\{-f_i(\nu)\}$ with affine f_i's, as in cognitive dissonance and temptation and self-control (Section 8). For these cases, u_p is piecewise linear; the pieces form a U-cover; and the extreme points are all those that do not live in any C_i without also being in $ext(C_i)$. The latter is essentially a reformulation of (7.2) and it leaves out how to go about computing $ext(C_i)$ for each i. In Section 9.10 of the Appendix, we offer a more hands-on characterization of the extreme points.

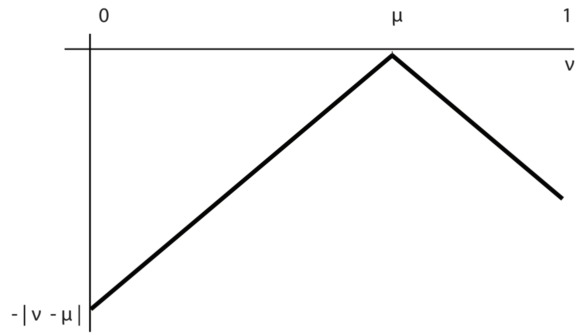


Figure 2: The Psychological Component: Concern for Reputation.

8 Applications

8.1 Reputation Concerns

The preferences given in Example (b) of Section 2.1 capture various sources of prior-bias, including cognitive dissonance, stubbornness and, as we study it here, reputation concerns. The analysis below applies to all.

Consider a ruling politician who must make a decision on behalf of her country—e.g., deciding whether or not to go to war—in the face of an uncertain state $\theta \in \{0, 1\}$. She will first seek whatever information she decides and then announce publicly the country's action $a \in A$ and (updated) belief ν that justifies it. The decision *must* serve the country's interests given whatever information is found (note that we cannot find what we do not seek).

On one hand, the politician wishes to make decisions that serve the country's motive $u_c(a, \theta)$. On the other, she campaigned on her deeply held beliefs $\mu \in (0, 1)$ and her political career will suffer if she is viewed as a 'flip-flopper'.[16] Publicly expressing a new belief ν entails a reputational cost, $u_p(\nu) = -\rho|\nu - \mu|$ for $\rho > 0$. When the politician decides which expertise to seek, she behaves as if she has psychological preferences

$$u(a, \nu) = u_p(\nu) + \mathbb{E}_{\theta\sim\nu}[u_c(\theta, a)].$$

Even without any information about u_c, we can infer that the optimal information policy takes a very special form: *the politician should either seek full information or no information at all!* She finds no tradeoff worth balancing.

[16]For example, John Kerry's perceived equivocation on the Iraq war damaged his 2004 campaign (for more details, see the September 19, 2004 and the June 23, 2008 issues of the New York Times).

The psychological component u_p is affine on $C := \{[0, \mu], [\mu, 1]\}$, hence C is a u_p-cover. Appealing to Proposition 7, it is also a U-cover. By Theorem 4, some optimal policy p^* is supported on $ext^*C = \{0, \mu, 1\}$. By Bayesian updating, it must be that $p^* = (1-\lambda)\delta_\mu + \lambda[(1-\mu)\delta_0 + \mu\delta_1]$ for some $\lambda \in [0, 1]$. That is, the politician is simply tossing a λ-coin, and either seeking full information or seeking none. But in this case, she must be indifferent between the two: either full information or no information must be optimal.

More can be said from our comparative static results. If we consider increasing ρ, then the politician becomes more information-averse by Proposition 3. Then, by Proposition 5, there is some cutoff $\rho^* \in \mathbb{R}_+$ such that, fixing μ, full information is optimal when $\rho < \rho^*$ and no information is optimal when $\rho > \rho^*$.

8.2 Temptation and Self-Control

Individuals often face temptations. Empirical work by psychologists (e.g., Baumeister (2002)) suggests the influence of temptation to consumer decision-making. In few contexts is the temptation and self-control problem as palpable as in consumption-saving decisions, as documented by Huang, Liu, and Zhu (2013). A recent literature in macroeconomics (e.g., Krusell, Kuruscu, and Smith (2010)) has asked how tax policy might be deliberately designed to alleviate the problem. In this section, we ask how information disclosure policy could do the same.

Consider a simple model of a consumer who decides whether to invest in a risky asset or to consume (later, the consumer will decide whether to buy an asset, sell an asset, or to consume). The state of the world, $\theta \in \{0, 1\}$, which we will interpret below in several ways, is unknown to him, distributed according to prior $\mu \in \Delta\Theta$. Before making his decision, the agent learns additional information from an advisor.

Investing draws its value from a higher consumption tomorrow (perhaps it also has intrinsic value) but it prevents today's consumption. When the agent deprives himself of consumption, it creates mental suffering. Only the commitment to not consuming could save the agent the cost of temptation and self-control, but such a commitment device is not available, hence investing comes with pain.

Assume that the consumer is a standard Gul and Pesendorfer (2001) agent (and expected utility maximizer). Given two classical utility functions u_1 and u_2, and action set A, the consumer gets a utility of

$$u(a, \nu) = \mathbb{E}_{\theta\sim\nu}\left\{u_1(a, \theta) + u_2(a, \theta) - \max_{b\in A} \mathbb{E}_{\hat{\theta}\sim\nu}[u_2(b, \hat{\theta})]\right\}, \tag{6}$$

when he chooses $a \in A$ at posterior belief ν. The usual interpretation is that the consumer has two "sides": a rational side u_1 and an impulsive side u_2, for now completely general. The rational side has expected utility $\mathbb{E}_{\theta\sim\nu}[u_1(a,\theta)]$ while the impulsive side bears the cost of self-control, $\max_{b\in A}\{\mathbb{E}_{\theta\sim\nu}[u_2(b,\theta)] - \mathbb{E}_{\theta\sim\nu}[u_2(a,\theta)]\}$. This cost is a psychological penalty measured by the utility loss between the chosen action and the action that the impulsive side would have chosen.

How should a well-intentioned financial advisor inform the consumer about θ when giving more information might induce more temptation? We first analyze two extreme cases that bring out the main trade-offs.

8.2.1 The State Is Irrelevant to One Side

Suppose first that u_2 is state-independent, which is a proxy for saying that the rational side is much more concerned with information than the impulsive side, the latter being too tempted to be affected by news. This scenario makes sense, for example, when θ is interpreted as the rate of return of the asset. In this case, information only enters via the expected value of u_1 and, thus, the agent is psychologically information-loving. By Theorem 1, the consumer is behaviorally information-loving, hence full information is optimal.

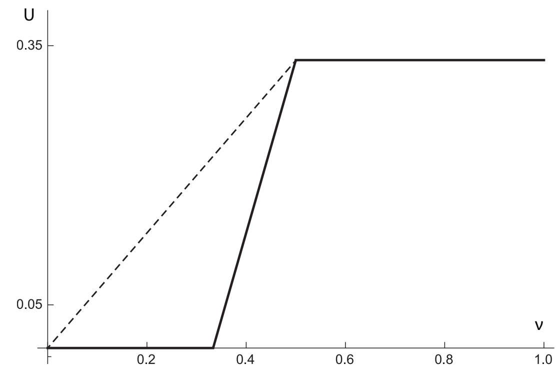
Suppose now that u_1 is state-independent, which makes sense, for example, when θ is interpreted as an airline promotion for a vacation. Given that utility $u(a,\nu) = u_1(a) + \mathbb{E}_{\theta\sim\nu}[u_2(a,\theta)] - U_2(\nu)$, the agent is easily seen to be psychologically information-averse by Theorem 1. Holding his choice fixed, it can only harm the consumer to find out whether or not there is a vacation promotion, as this exacerbates any temptation if his choice is to invest. Nonetheless, *a financial advisor may at times convey information about rationally irrelevant material (vacation promotion, car prices, etc.) so as to quell temptations*.

Consider an example where $A = \Theta = \{0,1\}$, $u_2(a,\theta) = -(a-\theta)^2$ and $u_1(a) = ka$ for some $k \geq 0$. If there will be a promotion for a vacation ($\theta = 0$), the consumer's tempted part would like to take it ($a = 0$), otherwise it prefers investing. To the rational side, any vacation promotions are irrelevant so that it always prefers investing. As usual, let $\nu = P(\{\theta = 1\})$. Then it is easy to compute $U_2(\nu) = \max\{2\nu - 1, 0\} - \nu$ and

$$u(a,\nu) = ka + \left(a - \mathbf{1}_{\nu\geq\frac{1}{2}}\right)(2\nu - 1),$$

so that when $k \in (0,1)$,

$$U(\nu) = \begin{cases} 0 & \text{if } 0 \le \nu \le \frac{1-k}{2} \\ 2\nu - (1-k) & \text{if } \frac{1-k}{2} < \nu \le \frac{1}{2} \\ k & \text{if } \frac{1}{2} < \nu \le 1 \end{cases}$$



Indirect Utility when $k = 1/3$

In particular, it is optimal to give information when the prior is below $\frac{1}{2}$. For example, suppose $\mu = .4$, so that there is 60% chance that some airline announces a promotion next month. If the advisor said nothing, the agent would invest and incur a big cost of self control, because there is a 60% chance of temptation occurring. Now, the advisor could do better with the following policy: say that there will be no promotion when it is the case and say that there will be a promotion only 1/3 of the time when there will actually be one. By doing so, the agent is never tempted when he receives news of a promotion (because he chooses to consume), and he is moderately tempted when he receives the other signal, as he invests but believes that there is a 50-50 chance of promotion.

8.2.2 The State Is Relevant to Both the Rational and the Tempted Side

In general, preferences (6) can be written as

$$u(a, \nu) = u_p(\nu) + \mathbb{E}_{\theta\sim\nu}[u_1(a, \theta) + u_2(a, \theta)],$$

where

$$u_p(\nu) = \min\left\{-\mathbb{E}_{\theta\sim\nu}[u_2(b, \theta)] : b \in A\right\}.$$

By linearity of expectation, u_p is a minimum of affine functions and, thus, the method of posterior covers developed in Section 7.2 is especially useful. Let us illustrate it in a concrete example.

Let $\theta \in \{\ell, h\}$ be the exchange rate of \$1 in euros tomorrow, where h (ℓ) stands for 'higher' ('lower') than today. Let $a \in \{a_1, a_2, a_3\}$ where a_i means "invest in firm i" for $i = 1, 2$, and a_3 means "buy a trip to Europe." Firm 1 uses domestic raw material to produce and mostly sells to European customers. Firm 2 imports foreign raw material from Europe and mostly sells domestically.

The utility functions are given by these matrices:

	ℓ	h
a_1	1	-1
a_2	-1	1
a_3	0	0

Function u_1

	ℓ	h
a_1	0	-1
a_2	-1	0
a_3	t	1

Function u_2

The rational side gets no value from consuming, but likes to invest in the firm that benefits from the exchange rate. The impulsive side gets no value from investing in the right firm, but has a disutility from investing in the wrong one. Moreover, it experiences great temptation when the exchange rate is high, and temptation governed by t when the rate is low.

When $t \geq 0$, temptation is strong in both states, consumption is a dominant action, and so $u_p(\nu) = -\mathbb{E}_{\theta\sim\nu}[u_2(a_3, \theta)]$. By Propositions 7 and 8, $C := \{[0, 1]\}$ is a U-cover, which immediately means that full information is optimal.

When $t \in (-1, 0)$, temptation is strong only when the exchange rate is high, otherwise traveling is worse than investing in firm 1 but still better than investing in firm 2. Overall, because investing in firm 2 is dominated by traveling, Propositions 7 and 8 imply that $C := \{[0, \nu^*], [\nu^*, 1]\}$ is a U-cover, where $\nu^* = \frac{t}{t-2}$. For now, we note $ext(C^*) = \{0, \nu^*, 1\}$.

Lastly, when $t < -2$.[17] It is easy to check that

$$C := \left\{\left[0, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{1+t}{t}\right], \left[\frac{1+t}{t}, 1\right]\right\}$$

is a U-cover with $ext(C^*) = \left\{0, \frac{1}{2}, \frac{1+t}{t}, 1\right\}$.

In each of the above cases, Theorem 4 reduces the search for an optimal policy to those supported on a given small set ext^*C. This reduction in hand, it is a straightforward exercise to compute an optimal policy.[18]

[17] When $t \in (-2, -1)$, the posterior cover once again has three extreme points and the argument is is similar to previous examples.

[18] It's a low-dimensional linear programming exercise. In particular, in the examples, the set of Bayes-consistent policies supported on ext^*C is either a line segment or a triangle, so that an optimal policy can be described as the best of two or three policies.

8.3 Equivocal Information Preferences

Consider a patient who is either healthy or very ill: $\Theta = \{0, 1\}$, with 0 corresponding to good health (μ and ν are probabilities to be ill). The patient is psychologically information-averse, but a medical choice can be made to best match the patient's state, as in Example (b) in Sections 2.1 and 3.2,[19]

$$u(a, \nu) = k\mathbb{V}(\nu) - \mathbb{E}_{\theta\sim\nu}\left[(\theta - a)^2\right]$$

for some $k \in (0, 1)$. If $A = [0, 1]$, we know the patient is behaviorally information-loving. But what if it is impossible to respond perfectly to posterior beliefs? Suppose that if the patient is likely enough to be ill, the disease is such that there is no appropriate action to take. Then, the inability to respond to information when the news is bad enough makes the agent equivocal: she likes to be informed when the news is good, but when it is bad, she prefers not to know.

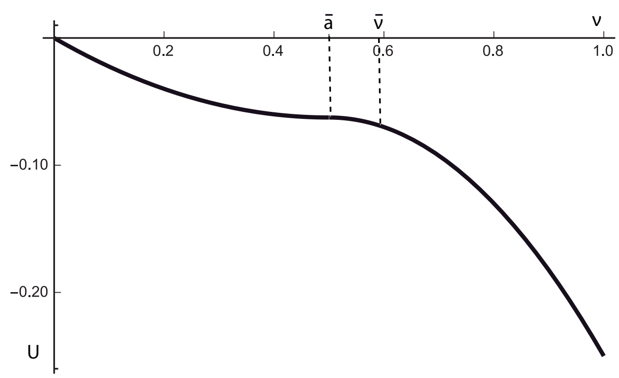


Figure 3: Equivocal Agent

We can model this by supposing $A = [0, \bar{a}]$ for some $\bar{a} \in (0, 1)$. Our intuition might tell us that (i) when $\mu < \bar{a}$, the doctor should give some information, since it can be useful (and such use outweighs the psychological discomfort when $k < 1$), and (ii) when $\mu > \bar{a}$, the doctor should say nothing, since there is no instrumental benefit to learning anything. The first intuition is correct, but the second leads us astray.

To solve for the optimal policy, our first tool is Theorem 2; some recommendation policy is optimal, because the agent is psychologically information-averse. If she could, the agent would choose action $E_\nu[\theta] = \nu$, but given the constraint, the optimal action is $a^*(\nu) = \min\{\nu, \bar{a}\}$. That is, the agent wants her action to match her belief if possible, or be

[19]The given $u(\nu, a)$ is not decreasing in beliefs, which does not match the 'illness' story. Of course, we could always subtract a constant multiple of ν to make it decreasing; this would have no design implications.

as close as possible. In this context, for the recommendation to be incentive-compatible, the patient's posterior belief upon hearing recommendation $a \in [0, \bar{a}]$ must be

$$\begin{cases} \nu = a & \text{if } a < \bar{a} \\ \nu \in [\bar{a}, 1] & \text{if } a = \bar{a}. \end{cases}$$

Moreover, the doctor would never use a policy that recommends actions $a \in (0, \bar{a})$, because it would always be better to give a little more information—run an additional test that will either give slightly good news or slightly bad news. This is because U is strictly convex at $a \in (0, \bar{a})$.[20] So, the doctor optimally makes a recommendation of either 'nothing specific required' ($a = 0$) or aggressive recommendation ($a = \bar{a}$).

For the 'no treatment' option to be incentive-compatible, she must *only* recommend it when the patient is healthy. Said differently, when the patient is ill, the doctor must always recommend the aggressive action. This leaves one remaining parameter: the probability λ of making the aggressive recommendation when the patient is healthy. The value of such a policy would be (assuming the aggressive action is incentive-compatible)

$$(1-\mu)(1-\lambda)u(0,0) + (\mu + (1-\mu)\lambda)u\left(\bar{a}, \frac{\mu}{\mu + (1-\mu)\lambda}\right) = 0 + \frac{\mu}{\nu}u(\bar{a}, \nu),$$

where $\nu = \frac{\mu}{\mu+(1-\mu)\lambda} \in [\mu, 1]$. Some algebra shows

$$\frac{\partial}{\partial \nu}\left[\frac{\mu}{\nu}u(\bar{a}, \nu)\right] = \frac{\mu}{\nu^2}(\bar{a}^2 - k\nu^2),$$

so that the patient's expected value of $\frac{\mu}{\nu}u(\bar{a}, \nu)$ is maximized at $\nu^* = \min\left\{1, \frac{\bar{a}}{\sqrt{k}}\right\}$. Thus, λ must be chosen so that $\nu = \nu^*$. Since $\nu^* > \bar{a}$, the described recommendation policy is indeed incentive-compatible.

The above work shows that the following is optimal:

- If the patient is healthy, the doctor suggests no treatment with probability $\frac{\nu^* - \mu}{\nu^*(1-\mu)}$. With complementary probability, she recommends the aggressive action.

- If he is ill, she recommends the aggressive action.

Notice that, even for some beliefs $\nu > \bar{a}$, at which the patient is (locally) behaviorally-information averse, the doctor optimally provides some information. This is because the potential to provide very good news with some probability—a clean bill of health, to which

[20] We could describe the patient as *locally* behaviorally information-loving at belief $a \in (0, \bar{a})$.

the patient responds with the no-treatment action—outweighs the damage of news to which she cannot respond.[21] Looking at ν^* above, this is particularly striking when $k \leq \bar{a}^2$, because in this case, the doctor optimally provides full information, even though the patient is not behaviorally information-loving. That is, full information is optimal, but poorly chosen partial information could be worse than saying nothing at all.

[21] Recall that giving no information when $\nu > \bar{a}$ results in action $\bar{a}$.

9 Appendix

9.1 Proof of Proposition 1

In both directions, the idea of the proof it to choose g (given r) or r (given g) to ensure

$$\mathrm{d}g(t|s)\,\mathrm{d}\sigma_p(s|\theta) = \mathrm{d}r(s|t)\,\mathrm{d}\sigma_q(t|\theta).$$

Everything else is just bookkeeping.

Proof. First suppose such r exists, and define $g : S_p \longrightarrow \Delta(S_q)$ via

$$g(T|s) := \int_T \frac{\mathrm{d}r(\cdot|t)}{\mathrm{d}p}(s)\,\mathrm{d}q(t),$$

for any $s \in S_p$ and Borel $T \subseteq S_q$. Then for $\theta \in \Theta$ and Borel $T \subseteq S_q$, it is straightforward to verify that $\int_{S_p} g(T|s)\,\mathrm{d}\sigma_p(s|\theta) = \sigma(T|\theta)$, so that g witnesses $p \succeq_B^\mu q$.

Conversely, suppose $p \succeq_B^\mu$ with the map $g : S_p \to \Delta(S_q)$ as in the definition of $\succeq_B^\mu$. Let

$$r(S|t) := \int_S \frac{\mathrm{d}g(\cdot|s)}{\mathrm{d}q}(t)\,\mathrm{d}p(s)$$

for $t \in S_q$ and Borel $S \subseteq S_p$. Then for Borel $S \subseteq S_p$, it is again straightforward to verify that $\int_{S_q} r(S|\cdot)\,\mathrm{d}q = p(S)$, which verifies the first condition.

Now, given Borel $T \subseteq S_q$ and $\theta \in \Theta$, one can check that

$$\int_T \frac{\mathrm{d}t}{\mathrm{d}\mu}(\theta)\,\mathrm{d}q(t) = \int_T \int_{\Delta\Theta} \frac{\mathrm{d}s}{\mathrm{d}\mu}(\theta)\,\mathrm{d}r(s|t)\,\mathrm{d}q(t),$$

so that $\frac{\mathrm{d}t}{\mathrm{d}\mu}(\theta) = \int_{\Delta\Theta} \frac{\mathrm{d}s}{\mathrm{d}\mu}(\theta)\,\mathrm{d}r(s|t)$ for a.e.-$\mu(\theta)q(t)$. From this we may show that, for any Borel $\hat{\Theta} \subseteq \Theta$,

$$\int_{\Delta\Theta} s(\hat{\Theta})\,\mathrm{d}r(s|t) = t(\hat{\Theta}),$$

which verifies the second condition. □

9.2 Proof of Theorem 1

Lemma 2. *The set $M := \{\gamma \in \Delta\Theta : \ \exists\epsilon > 0$ s.t. $\epsilon\gamma \leq \mu\}$ is dense*[22] *in $\Delta\Theta$.*

[22] Under the w^*-topology.

Proof. First, notice that $M = \left\{\gamma \in \Delta\Theta : \ \gamma << \mu \text{ and } \frac{\mathrm{d}\gamma}{\mathrm{d}\mu} \text{ is (essentially) bounded}\right\}$ is convex and extreme (i.e. a face of $\Delta\Theta$). Thus its w^*-closure $\bar{M}$ is closed (and so compact, by Banach-Alaoglu), convex, and extreme. Now, let E be the set of extreme points of $\bar{M}$. Because $\bar{M}$ is extreme, E is a subset of $ext(\Delta\Theta) = \{\delta_\theta\}_{\theta\in\Theta}$. So $E = \{\delta_\theta\}_{\theta\in\hat{\Theta}}$ for some $\hat{\Theta} \subseteq \Theta$. By Krein-Milman, $\bar{M} = \overline{co}E = \Delta(\Theta')$, where Θ' is the closure of $\hat{\Theta}$. Finally, notice that $\mu \in M$ implies $\Theta' \supseteq \text{supp}(\mu) = \Theta$. Thus $\bar{M} = \Delta\Theta$ as desired. □

Lemma 3. *Fix a continuous function $f : \Delta\Theta \to \mathbb{R}$. Then the following are equivalent (given μ is of full support):*

1. *For all $\nu \in \Delta\Theta$ and $p \in \mathcal{R}(\nu)$, we have $\int_{\Delta\Theta} f \, dp \geq f(\nu)$*
2. *For all $\mu' \in \Delta\Theta$ and $p, q \in \mathcal{R}(\mu')$ with $p \succeq_B^{\mu'} q$, we have $\int_{\Delta\Theta} f \, dp \geq \int_{\Delta\Theta} f \, dq$.*
3. *For all $p, q \in \mathcal{R}(\mu)$ with $p \succeq_B^{\mu} q$, we have $\int_{\Delta\Theta} f \, dp \geq \int_{\Delta\Theta} f \, dq$.*
4. *f is convex.*

Proof. Suppose (1) holds, and consider any $\mu' \in \Delta\Theta$ and $q \in \mathcal{R}(\mu')$. If $r : S_q \to \Delta\Delta\Theta$ satisfies $r(\cdot|\nu) \in \mathcal{R}(\nu)$ for every $\nu \in S_q$, (1) implies $\int_{S_q} \int_{\Delta\Theta} f \, \mathrm{d}r(\cdot|\nu) \, \mathrm{d}q(\nu) \geq \int_{S_q} f \, \mathrm{d}q$. Equivalently (by Proposition 1), any p more informative than q has $\int f \, \mathrm{d}p \geq \int f \, \mathrm{d}q$, which yields (2).

That (2) implies (3) is immediate.

Now, suppose (4) fails. That is, there exist $\gamma, \zeta, \eta \in \Delta\Theta$ and $\lambda \in (0, 1)$ such that

$$\begin{aligned} (1-\lambda)\zeta + \lambda\eta &= (1-\lambda)\zeta + \lambda\eta. \\ f((1-\lambda)\zeta + \lambda\eta) &< (1-\lambda)f(\zeta) + \lambda f(\eta). \end{aligned}$$

Now, we want to exploit the above to construct two information-ranked random posteriors such that f has higher expectation on the less informative of the two.

To start, let us show how to do it if $\epsilon\gamma \leq \mu$ for some $\epsilon \in (0, 1)$. In this case, let

$$\nu := \frac{1}{1-\epsilon}(\mu - \epsilon\gamma) \in \Delta\Theta, \quad p := (1-\epsilon)\delta_\nu + \epsilon(1-\lambda)\delta_\zeta + \epsilon\lambda\delta_\eta \in \mathcal{R}(\mu), \text{ and } q := (1-\epsilon)\delta_\nu + \epsilon\delta_\gamma \in \mathcal{R}(\mu).$$

Then $p \succeq_B^{\mu} q$, but

$$\int_{\Delta\Theta} f \, \mathrm{d}p - \int_{\Delta\Theta} f \, \mathrm{d}q = \epsilon \left[(1-\lambda)f(\zeta) + \lambda f(\eta) - f(\gamma)\right] < 0,$$

as desired.

Now, notice that we needn't assume that $\epsilon\gamma \leq \mu$ for some $\epsilon \in (0,1)$. Indeed, in light of Lemma 2 and continuity of f, we can always pick $\gamma, \zeta, \eta \in \Delta\Theta$ to ensure it. So given continuous nonconvex f, we can ensure existence of $p \geq_B^\mu q$ with $\int_{\Delta\Theta} f \, \mathrm{d}p < \int_{\Delta\Theta} f \, \mathrm{d}q$. That is, (3) fails too.

Finally, notice that (4) implies (1) by Jensen's inequality. □

Proof of Theorem 1.

1. This follows immediately from applying Lemma 3 to $u(a,\cdot)$ and $-u(a,\cdot)$ for each $a \in A$.

2. This follows immediately from applying Lemma 3 to U and $-U$.

3. This follows from the first two parts, and from the easy fact that a pointwise maximum of convex functions is convex.

4. By hypothesis, $U = u(a^*,\cdot)$, which is convex by psychological information-aversion.

□

9.3 Proof of Theorem 3

Below, we will prove the slightly stronger result. Fix any $\bar{p} \in \mathcal{R}(\mu)$, and let $S := \overline{co}(S_{\bar{p}})$ and $\Phi := \{\phi : S \to \mathbb{R} : \ \phi \text{ is affine continuous and } \phi \geq U|_S\}$. We will show that

$$\max_{p \in \Delta(S) \cap \mathcal{R}(\mu)} \int U \, \mathrm{d}p = cav_{\bar{p}} U(\mu),$$

where $cav_{\bar{p}} U(\nu) := \inf_{\phi \in \Phi} \phi(\nu)$.

Proof. Let $\beta : \Delta(S) \to S$ be the unique map such that $\mathbb{E}_{\nu \sim p} \nu = \beta(\nu)$. Such β is well-defined, continuous, affine, and surjective, as shown in Phelps (2001).

For every $p \in \Delta(S)$, define $\mathbb{E}U(p) = \int U \, \mathrm{d}p$. The map $\mathbb{E}U$ is then affine and continuous. For every $\nu \in S$, define $U^*(\nu) = \max_{p \in \Delta(S), \beta(p) = \nu} \mathbb{E}U(p)$, which (by Berge's theorem) is well-defined and upper-semicontinuous.

For any $\nu \in S$:

- $U^*(\nu) \geq \mathbb{E}U(\delta_\nu) = U(\nu)$. That is, an optimal policy for prior ν does at least as well as giving no information.

- For all $p \in \Delta(S)$ and all affine continuous $\phi : S \longrightarrow \mathbb{R}$ with $\phi \geq U|_S$

$$\mathbb{E}U(p) = \int U \, \mathrm{d}p \leq \int \phi(s) \, \mathrm{d}p(s) = \phi\left(\int_{\Delta\Theta} s \, \mathrm{d}p(s)\right) = \phi(\beta(p)).$$

  So, $U^*(\nu)$ can be no higher than $cav_{\bar{p}}U(\nu)$.

Moreover, if $\nu, \nu' \in S$ and $\lambda \in [0,1]$, then for all $p, q \in \Delta(S)$ with $\beta(p) = \nu$ and $\beta(q) = \nu'$, we know $\beta((1-\lambda)p + \lambda q) = (1-\lambda)\nu + \lambda\nu'$, so that

$$U^*((1-\lambda)\nu + \lambda\nu') \geq \mathbb{E}U((1-\lambda)p + \lambda q) = (1-\lambda)\mathbb{E}U(p) + \lambda\mathbb{E}U(q).$$

Optimizing over the right-hand side yields

$$U^*((1-\lambda)\nu + \lambda\nu') \geq (1-\lambda)U^*(\nu) + \lambda U^*(\nu').$$

That is, an optimal policy for prior $(1-\lambda)\nu + \lambda\nu'$ does at least as well as a signal inducing (interim) posteriors from $\{\nu, \nu'\}$ followed by an optimal signal for the induced interim belief.

So far, we have established that U^* is upper-semicontinuous and concave, and $U|_S \leq U^* \leq cav_{\bar{p}}U$. Since $cav_{\bar{p}}U$ is the pointwise-lowest u.s.c. concave function above $U|_S$, it must be that $cav_{\bar{p}}U \leq U^*$, and thus $U^* = cav_{\bar{p}}U$.

□

9.4 Proof of Theorem 2

Proof. Suppose the agent is psychologically information-averse.
Fix some measurable selection $a^* : \Delta\Theta \longrightarrow A$ of the best-response correspondence $\nu \mapsto \arg\max_{a\in A} u(a,\nu)$. In particular, given any $q \in \mathcal{R}(\mu)$, $a^*|_{S_q}$ is an optimal strategy for an agent with direct signal (S_q, σ_q).

Toward a proof of the theorem, we first verify the following claim.
Claim: Given any random posterior $p \in \mathcal{R}(\mu)$, we can construct a signal (A, α_p) such that:

1. The random posterior q_p induced by (A, α_p) is less informative than p.

2. An agent who follows the recommendations of α_p performs at least as well as an agent who receives signal (S_p, σ_p) and responds optimally, i.e.

$$\int_\Theta \int_A u\left(a, \beta^{A,\alpha_p}(\cdot|a)\right) \, \mathrm{d}\alpha_p(a|\theta) \, \mathrm{d}\mu(\theta) \geq \int_{\Delta\Theta} U \, \mathrm{d}p.$$

To verify the claim, fix any $p \in \mathcal{R}(\mu)$, and define the map

$$\begin{aligned} \alpha_p : \Theta &\longrightarrow \Delta(A) \\ \theta &\longmapsto \alpha_p(\cdot|\theta) = \sigma_p(\cdot|\theta) \circ a^{*-1}. \end{aligned}$$

Then (A, α_p) is a signal with

$$\alpha_p(\hat{A}|\theta) = \sigma_p\left(\left\{s \in S_p : \; a^*(s) \in \hat{A}\right\} \middle| \theta\right)$$

for every $\theta \in \Theta$ and Borel $\hat{A} \subseteq A$. The signal (A, α_p) is familiar: replace each message in (S_p, σ_p) with a recommendation of the action that would have been taken.

Let $q_p \in \mathcal{R}(\mu)$ denote the random posterior induced by signal (A, α_p). Now, let us show that q_p delivers at least as high an expected value as p.

By construction,[23] $p \succeq^\mu_B q$. Therefore, by Proposition 1, there is a map $r : S_q \longrightarrow \Delta(S_p)$ such that for every Borel $S \subseteq \Delta\Theta$, $p(S) = \int_{S_q} r(S|\cdot)\, \mathrm{d}q$, and for every $t \in S_q$, $r(\cdot|t) \in \mathcal{R}(t)$. Then, appealing to the definition of psychological information-aversion,

$$\begin{aligned} \int_A u\left(a, \beta^{A,\alpha_p}(\cdot|a)\right) \, \mathrm{d}\alpha_p(a|\theta) &= \int_{S_p} u\left(a^*(s), \beta^{A,\alpha_p}(\cdot|a^*(s))\right) \, \mathrm{d}\sigma_p(s|\theta) \\ &\geq \int_{S_p} \int_{S_p} u\left(a^*(s), \nu\right) \, \mathrm{d}r\left(\nu \middle| \beta^{A,\alpha_p}(\cdot|a^*(s))\right) \, \mathrm{d}\sigma_p(s|\theta) \\ &= \int_{S_p} \int_{S_p} U(\nu) \, \mathrm{d}r\left(\nu \middle| \beta^{A,\alpha_p}(\cdot|a^*(s))\right) \, \mathrm{d}\sigma_p(s|\theta) \\ &= \int_A \int_{S_p} U(\nu) \, \mathrm{d}r\left(\nu \middle| \beta^{A,\alpha_p}(\cdot|a)\right) \, \mathrm{d}\alpha_p(a|\theta). \end{aligned}$$

Therefore,

$$\begin{aligned} \int_\Theta \int_A u\left(a, \beta^{A,\alpha_p}(\cdot|a)\right) \, \mathrm{d}\alpha_p(a|\theta) \, \mathrm{d}\mu(\theta) &\geq \int_\Theta \int_A \int_{S_p} U(\nu) \, \mathrm{d}r\left(\nu \middle| \beta^{A,\alpha_p}(\cdot|a)\right) \, \mathrm{d}\alpha_p(a|\theta) \, \mathrm{d}\mu(\theta) \\ &= \int_{S_q} U(\nu) \, \mathrm{d}r(\nu|t) \, \mathrm{d}q(t) \\ &= \int_{S_p} U \, \mathrm{d}p, \end{aligned}$$

[23] Indeed we can define g in (1) via: $g(a|s) = 1$ if $a^*(s) = a$ and 0 otherwise.

which verifies the claim.

Now, fix some optimal policy $p^* \in \mathcal{R}(\mu)$, and let $\alpha = \alpha_{p^*}$ and $q = q_{p^*}$ be as delivered by the above claim. Let the measure $Q \in \Delta(A \times \Delta\Theta)$ over recommended actions and posterior beliefs be that induced by α_p. So

$$Q(\hat{A} \times \hat{S}) = \int_\Theta \int_{\hat{A}} 1_{\beta^{A,\alpha}(\cdot|a) \in S} \, \mathrm{d}\alpha(a|\theta) \, \mathrm{d}\mu(\theta)$$

for Borel $\hat{A} \subseteq A, \hat{S} \subseteq \Delta\Theta$.

Then,[24]

$$\int_{\Delta\Theta} U \, \mathrm{d}p \leq \int_{A \times \Delta\Theta} u \, \mathrm{d}Q \leq \int_{\Delta\Theta} U \, \mathrm{d}q \leq \int_{\Delta\Theta} U \, \mathrm{d}p,$$

so that:

$$\begin{aligned} \int_{\Delta\Theta} U \, \mathrm{d}q &= \int_{\Delta\Theta} U \, \mathrm{d}p, \text{ i.e. } q \text{ is optimal; and} \\ \int_{A \times \Delta\Theta} u(a, \nu) \, \mathrm{d}Q(a, \nu) &= \int_{\Delta\Theta} U \, \mathrm{d}q = \int_{A \times \Delta\Theta} \max_{\tilde{a} \in A} u(\tilde{a}, \nu) \, \mathrm{d}Q(a, \nu). \end{aligned}$$

The latter point implies that $a \in \operatorname{argmax}_{\tilde{a} \in A} u(\tilde{a}, \nu)$ a.s.-$Q(a, \nu)$. In other words, the recommendation (A, α) is incentive-compatible as well. This completes the proof. □

It seems worth noting that the claim in the above proof delivers something more than the result of Theorem 2. Indeed, given any finite-support random posterior p, the claim produces a constructive procedure to design an incentive-compatible recommendation policy which outperforms p. The reason is that (in the notation of the claim):

1. If $a^*|_{S_p}$ is injective, then $q_p = p$, so that (A, α_p) is an incentive-compatible recommendation policy inducing random posterior p itself.

2. Otherwise, $|S_{q_p}| < |S_p|$.

In the latter case, we can simply apply the claim to q_p. Iterating in this way—yielding a new, better policy at each stage—eventually (in fewer than $|S_p|$ stages) leads to a recommendation policy which is incentive-compatible and outperforms p.

[24]Indeed, the inequalities follow from the above claim, the definition of U along with the property $marg_{\Delta\Theta} Q = q$, and optimality of p, respectively.

9.5 Proof of Proposition 3

Proof. If C is a concave function and $p \geq_B^\mu q$, then by Proposition 1, $\int C\,\mathrm{d}q \geq \int C\,\mathrm{d}p$. Assuming $p \geq_B^\mu q$ and $\int U_1\,\mathrm{d}q \geq \int U_1\,\mathrm{d}p$, we obtain $\int \gamma U_1\,\mathrm{d}q \geq \int \gamma U_1\,\mathrm{d}p$ and, therefore, $\int U_2\,\mathrm{d}q \geq \int U_2\,\mathrm{d}p$. □

9.6 Comparative Statics: Proof of Proposition 4 and Example

9.6.1 Proof

Lemma 4. *Given $p, q \in \Delta\Delta\Theta$, the following are equivalent:*

1. *There is some $\nu \in \Delta\Theta$ such that $p, q \in \mathcal{R}(\nu)$ and $p \geq_B^\nu q$.*
2. *For every convex continuous $f : \Delta\Theta \to \mathbb{R}$, we have $\int_{\Delta\Theta} f\,dp \geq \int_{\Delta\Theta} f\,dq$.*

Proof. That (1) implies (2) follows from Lemma 3. Now, suppose (2) holds. The Theorem of Hardy-Littlewood-Polya-Blackwell-Stein-Sherman-Cartier reported in Phelps (2001, main Theorem, Section 15) then says that p is a mean-preserving spread of q. Then, Proposition 1 implies (1). □

Notation 1. *Given any $\mathcal{F} \cup \mathcal{G} \cup \{h\} \subseteq C(\Delta\Theta)$:*

- *Let $\succsim_{\mathcal{F}}$ be the (reflexive, transitive, continuous) binary relation on $\Delta\Delta\Theta$ given by*

$$p \succsim_{\mathcal{F}} q \iff \int_{\Delta\Theta} f\,dp \geq \int_{\Delta\Theta} f\,dq \ \ \forall f \in \mathcal{F}.$$

- *Let $\langle\mathcal{F}\rangle \subseteq C(\Delta\Theta)$ be the smallest closed convex cone in $C(\Delta\Theta)$ which contains $\mathcal{F}$ and all constant functions.*
- *Let $\mathcal{F} + \mathcal{G}$ denote the Minkowski sum $\{f + g : \ f \in \mathcal{F}, \ g \in \mathcal{G}\}$, which is a convex cone if $\mathcal{F}, \mathcal{G}$ are.*
- *Let $\mathbb{R}_+ h$ denote the ray $\{\alpha h : \ \alpha \in \mathbb{R}, \alpha \geq 0\}$.*

We now import the following representation uniqueness theorem to our setting.

Lemma 5 (Dubra, Maccheroni, and Ok (2004), Uniqueness Theorem, p. 124)**.** *Given any $\mathcal{F}, \mathcal{G} \subseteq C(\Delta\Theta)$,*

$$\succsim_{\mathcal{F}} \ = \ \succsim_{\mathcal{G}} \iff \langle\mathcal{F}\rangle \ = \ \langle\mathcal{G}\rangle.$$

As a consequence, we get the following:

Corollary 1. *Suppose $\mathcal{F} \cup \{g\} \subseteq C(\Delta\Theta)$. Then*

$$\succsim_{\mathcal{F}} \subseteq \succsim_{\{g\}} \text{ if and only if } g \in \langle \mathcal{F} \rangle.$$

Proof. If $g \in \langle \mathcal{F} \rangle$, then $\succsim_{\{g\}} \supseteq \succsim_{\langle \mathcal{F} \rangle}$, which is equal to $\succsim_{\mathcal{F}}$ by the uniqueness theorem. If $\succsim_{\mathcal{F}} \subseteq \succsim_{\{g\}}$, then $\succsim_{\mathcal{F}} \subseteq \succsim_{\mathcal{F}} \cap \succsim_{\{g\}} = \succsim_{\mathcal{F} \cup \{g\}} \subseteq \succsim_{\mathcal{F}}$. Therefore, $\succsim_{\mathcal{F} \cup \{g\}} = \succsim_{\mathcal{F}}$. By the uniqueness theorem, $\langle \mathcal{F} \cup \{g\} \rangle = \langle \mathcal{F} \rangle$, so that $g \in \langle \mathcal{F} \rangle$. □

Lemma 6. *Suppose $\mathcal{F} \subseteq C(\Delta\Theta)$ is a closed convex cone that contains the constants and $g \in C(\Delta\Theta)$. Then either $g \in -\mathcal{F}$ or $\langle \mathcal{F} \cup \{g\} \rangle = \mathcal{F} + \mathbb{R}_+ g$.*

Proof. Suppose $\mathcal{F}$ is as described and $-g \notin \mathcal{F}$.

Since a closed convex cone must be closed under sums and nonnegative scaling, it must be that $\langle \mathcal{F} \cup \{g\} \rangle \supseteq \mathcal{F} + \mathbb{R}_+ g$. Therefore, $\langle \mathcal{F} \cup \{g\} \rangle = \langle \mathcal{F} + \mathbb{R}_+ g \rangle$, which leaves us to show $\langle \mathcal{F} + \mathbb{R}_+ g \rangle = \mathcal{F} + \mathbb{R}_+ g$. As $\mathcal{F} + \mathbb{R}_+ g$ is a convex cone which contains the constants, we need only show it is closed.

To show it, consider sequences $\{\alpha_n\}_{n=1}^{\infty} \subseteq \mathbb{R}_+$ and $\{f_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ such that $(f_n + \alpha_n g)_{n=1}^{\infty}$ converges to some $h \in C(\Delta\Theta)$. We wish to show that $h \in \mathcal{F} + \mathbb{R}_+ g$. By dropping to a subsequence, we may assume $(\alpha_n)_n$ converges to some $\alpha \in \mathbb{R}_+ \cup \{\infty\}$.

If α were equal to ∞, then the sequence $\left(\frac{f_n}{\alpha_n}\right)_n$ from $\mathcal{F}$ would converge to $-g$, implying (since $\mathcal{F}$ is closed) $-g \in \mathcal{F}$. Thus α is finite, so that $(f_n)_n$ converges to $f = h - \alpha g$. Then, since $\mathcal{F}$ is closed, $h = f + \alpha g \in \mathcal{F} + \mathbb{R}_+ g$. □

Proof of Proposition 4

Proof. In light of Lemma 4, that 2 is more information-averse than 1 is equivalent to the pair of conditions:

$$\begin{aligned} \succsim_{C \cup \{U_2\}} &\subseteq \succsim_{\{U_1\}} \\ \succsim_{(-C) \cup \{U_1\}} &\subseteq \succsim_{\{U_2\}}, \end{aligned}$$

where C is the set of convex continuous functions on $\Delta\Theta$; notice that C is a closed convex cone which contains the constants.

Then, applying Corollary 1 and Lemma 6 twice tells us:

- Either U_2 is concave or $U_1 \in C + \mathbb{R}_+ U_2$.
- Either U_1 is convex or $U_2 \in -C + \mathbb{R}_+ U_1$.

By hypothesis[25], 1 is not behaviorally information-loving, i.e. U_1 is not convex. Then $U_2 \in -C + \mathbb{R}_+ U_1$, proving the proposition. $\square$

Notice: We cannot strengthen the above theorem to ensure $\gamma \in (0,1)$. Indeed, consider the $\Theta = \{0,1\}$ world with $U_1 = H$ (entropy) and $U_2 = \mathbb{V}$ (variance). Both are strictly concave, so that in particular U_2 is more information-averse than U_1. However, any sum of a concave function and a strictly positive multiple of U_1 is non-Lipschitz (as $H'(0) = \infty$), and so is not a multiple of U_2.

9.6.2 Example

Suppose that $\theta = (\theta_1, \theta_2) \in \{0,1\}^2$ is distributed uniformly. Let $x_1(\nu)$ and $x_2(\nu) \in [-1,1]$ be some well-behaved measures of information about θ_1, θ_2 respectively (the smaller x_i, the more dispersed the ith dimension of distribution ν).[26] Then define U_1, U_2 via

$$\begin{aligned} U_1(\nu) &:= -\frac{1}{2}\left(x_1(\nu)^2 + x_2(\nu)^2 + x_1(\nu)x_2(\nu)\right); \\ U_2(\nu) &:= U_1(\nu) - \frac{1}{5}x_1(\nu). \end{aligned}$$

Agent 2 receives extra utility from being less informed and so he is more information averse than 1. Meanwhile, we can easily verify (via first-order conditions to optimize (x_1, x_2)) that any optimal policy for 1 will give more information concerning θ_1, but less information concerning θ_2, than will any optimal policy for 2. Therefore, the optimal policies for 1 and 2 are Blackwell incomparable, although 2 is more information-averse than 1. Intuitively, the penalty for information about θ_1 induces a substitution effect toward information about θ_2. Technically, the issue is a lack of supermodularity.

9.7 Proof of Proposition 9

Proposition 9. *Blackwell-incomparability is generic, i.e. the set $N_\mu = \{(p,q) \in \mathcal{R}(\mu)^2 : p \not\geq_B^\mu q$ and $p \not\leq_B^\mu q\}$ is open and dense in $\mathcal{R}(\mu)^2$.*

Proof. To show that N_μ is open and dense, it suffices to show that $\not\geq_B^\mu$ is. Indeed, it would then be immediate that $\not\leq_B^\mu$ is open and dense (as switching coordinates is a homeomorphism), so that their intersection N_μ is dense too.

[25] Note: this is the first place we've used this hypothesis

[26] For instance, let $x_i(\nu) := 1 - 8\mathbb{V}(\mathrm{marg}_i \nu)$, $i = 1,2$. What we need is that $x_1(\nu), x_2(\nu)$ be convex $[-1,1]$-valued functions of $\mathrm{marg}_1 \nu, \mathrm{marg}_2 \nu$ respectively, taking the priors to -1 and atomistic beliefs to 1.

Given Lemma 1, it is straightforward[27] to express $\succeq_B^\mu \subseteq (\Delta\Theta)^2$ as the image of a continuous map with domain $\Delta\Delta\Delta\Theta$. Therefore $\succeq_B^\mu$ is compact, making its complement an open set. Take any $p, q \in \mathcal{R}(\mu)$. We will show existence of $\{p_\epsilon, q_\epsilon\}_{\epsilon\in(0,1)}$ such that $p_\epsilon \not\succeq_B^\mu q_\epsilon$ and $(p_\epsilon, q_\epsilon) \to (p, q)$ as $\epsilon \to 0$. For each fixed $\epsilon \in (0, 1)$, define

$$\begin{aligned} g_\epsilon : \Delta\Theta &\longrightarrow \Delta\Theta \\ \nu &\longmapsto (1-\epsilon)\nu + \epsilon\mu. \end{aligned}$$

Because g_ϵ is continuous and affine, its range G_ϵ is compact and convex. Define

$$p_\epsilon := p \circ g_\epsilon^{-1} \text{ and } q_\epsilon := (1-\epsilon)q + \epsilon f,$$

where $f \in \mathcal{R}(\mu)$ is the random posterior associated with full information.[28] It is immediate, say by direct computation with the Prokhorov metric and boundedness of Θ, that $(p_\epsilon, q_\epsilon) \to (p, q)$ as $\epsilon \to 0$. Moreover,

$$\overline{co}(S_{p_\epsilon}) \subseteq \overline{co}(G_\epsilon) = G_\epsilon \subsetneq \Delta\Theta = \overline{co}(S_{q_\epsilon}).$$

In particular, $\overline{co}(S_{p_\epsilon}) \not\supseteq \overline{co}(S_{p_\epsilon})$. Then, appealing to Lemma 1, $p_\epsilon \not\succeq_B^\mu q_\epsilon$ as desired. □

9.8 Toward Comparative Statics

A useful result for comparative statics is the following sharpening of Theorem 3 for extreme policies.

Lemma 7. *If $\bar{p} \in ext\mathcal{R}(\mu)$, then*

$$\max_{p\in\mathcal{R}(\mu):\ p\preceq_B^\mu \bar{p}} \int_{\Delta\Theta} U \, dp = cav_{\bar{p}} U(\mu),$$

Proof. Again, let $S := \overline{co}(S_{\bar{p}})$. Appealing to the proof of Theorem 3, we know $cav_{\bar{p}}U(\mu) = U^*(\mu) := \max_{p\in\mathcal{R}(\mu)\cap\Delta(S)} \mathbb{E}U(p)$.

Suppose that $U^*(\mu) < \max_{p\in\mathcal{R}(\mu):\ p\preceq_B^\mu \bar{p}} \int_{\Delta\Theta} U \, \mathrm{d}p$. Appealing to continuity of $\mathbb{E}U$, there is[29]

[27] Indeed, let $\beta_1 : \Delta\Delta\Theta \to \Delta\Theta$ and $\beta_2 : \Delta\Delta\Delta\Theta \to \Delta\Delta\Theta$ be the maps taking each measure to its barycenter. Then define $B : \Delta\Delta\Delta\Theta \to (\Delta\Delta\Theta)^2$ via $B(Q) = (Q \circ \beta_1^{-1}, \beta_2(Q))$. Then $\succeq_B^\mu = B((\beta_1 \circ \beta_2)^{-1}(\mu))$.

[28] i.e. $f\left(\{\delta_\theta : \theta \in \hat{\Theta}\}\right) = \mu(\hat{\Theta})$ for every Borel $\hat{\Theta} \subseteq \Theta$.

[29] TO-DO. Need to also use affinity to get from S to S_p.

some $p \in \mathcal{R}(\mu) \cap \Delta(S)$ and $\epsilon \in (0,1)$ such that $\epsilon p \leq \bar{p}$ and $U^*(\mu) < \int_{\Delta\Theta} U \, \mathrm{d}p$. Then

$$\bar{p} \in co\left\{\frac{1}{1-\epsilon}(\bar{p} - \epsilon p), p\right\},$$

so that $\bar{p}$ cannot be extreme in $\mathcal{R}(\mu)$. $\square$

Lemma 8. *Given μ is of full support and $\bar{p} \in \mathcal{R}(\mu)$, there is an affine continuous $\phi = \phi_{\bar{p}}$ on $S = \overline{co}(S_{\bar{p}})$ such that $\phi \geq U|_S$ and $cav_{\bar{p}}U(\mu) = \phi(\nu)$.*

Proof. There is some sequence $\{\phi_n\}_{n=1}^{\infty}$ of affine continuous functions on S which majorize $U|_S$, with $\phi_n(\mu) \to cav_{\bar{p}}U(\mu)$. As $cav_{\bar{p}}U$ is concave and continuous on the interior of its domain (and in particular in a neighborhood of μ), it is locally Lipschitz at μ, say of Lipschitz constant k. Therefore, we may assume without loss there is some constant[30] K such that $\phi_n \leq K$ for each $n \in \mathbb{N}$. Since $\{\phi \in \Phi : \ \phi \leq K\}$ is compact [TO-DO]

$\square$

Lemma 9. *Given $\bar{p} \in ext\mathcal{R}(\mu)$, let $\phi = \phi_{\bar{p}}$ be as delivered by Lemma 8. If $p^* \in \operatorname{argmax}_{p \preceq_B^{\mu} \bar{p}} \int U \, dp$, then $\phi|_{S_{p^*}} = U|_{S_{p^*}}$.*

Proof. Suppose otherwise. Then, by continuity, there is a nonempty open set $T \subseteq S$ and $\epsilon > 0$ such that $\phi|_T > \epsilon + U|_T$ and $p^*(T) > 0$. Then

$$\begin{aligned}
\int_{\Delta\Theta} U \, \mathrm{d}p^* &= \int_{\Delta\Theta \setminus T} U \, \mathrm{d}p^* + \int_T U \, \mathrm{d}p^* \\
&\leq \int_{\Delta\Theta \setminus T} \phi \, \mathrm{d}p^* + \int_T (\phi - \epsilon) \, \mathrm{d}p^* \\
&= \phi(\mu) + \epsilon p^*(T) \\
&> \phi(\mu),
\end{aligned}$$

so that (appealing to Lemma 8) p^* cannot be optimal. $\square$

The above lemmata imply the following:

Proposition 10. *Given $\bar{p} \in ext\mathcal{R}(\mu)$ and $p^* \preceq_B^{\mu} \bar{p}$, the following are equivalent.*

1. *$p^* \in \operatorname{argmax}_{p \precsim_B^{\mu} \bar{p}} \int U \, dp$.*
2. *There exists $\phi \in \Phi$ such that $\phi|_{S_{p^*}} = U|_{S_{p^*}}$.*

[30] For instance, we could take $K := 1U(\mu) + \max_{\nu \in \Delta(S_{\bar{p}})} k\|\nu - \mu\|$.

Proof. That (1) implies (2) is proven above. Now suppose that (2) holds. Then for any $p \preceq_B^\mu \bar{p}$,

$$\int U \, \mathrm{d}p \leq \int \phi \, \mathrm{d}p = \phi(\mu) = \int \phi \, \mathrm{d}p^* = \int U \, \mathrm{d}p^*,$$

as desired. □

9.9 Proof of Proposition 5

Proof. First, notice that we are done if either U_1 is information-loving (in which case p_1^* can just be full information) or U_2 is information-averse (in which case p_2^* can just be no information). Thus focus on the complementary case, where Proposition 4 guarantees $\gamma U_1 - U_2$ for some $\gamma > 0$.

Next, let us do some convenient normalizations:

- For $i \in \{1, 2\}$, let ϕ_i be the ϕ delivered by Lemma 8, when $U = U_i$ and $\bar{p}$ is full information. Replacing U_i with $U_i - \phi_i$, it is without loss to assume $U_i \leq 0$ and $\bar{U}_i(\mu) = 0$.

- By positively scaling if necessary, it is without loss to assume $f := U_1 - U_2$ is convex.

As $\mathbb{E}U$ is affine continuous on a compact convex domain, there some optimal policy p_2^* for U_2 which is extreme in $\mathcal{R}(\mu)$, and so has $|S_{p_2^*}| \leq 2$. Say $S_{p_2^*} = \{\nu_L, \nu_H\}$, where $\nu_L \leq \mu \leq \nu_H$. In view of the above normalization, Proposition 10 (applied to $\bar{p}$ = full information) guarantees $U_2(\nu_L) = U_2(\nu_H) = 0$. Then, for any $\lambda \in [0, 1]$,

$$\begin{aligned}
U_1((1-\lambda)\nu_L + \lambda\nu_H) &= U_2((1-\lambda)\nu_L + \lambda\nu_H) + f((1-\lambda)\nu_L + \lambda\nu_H) \\
&\leq f((1-\lambda)\nu_L + \lambda\nu_H) \\
&\leq (1-\lambda)f(\nu_L) + \lambda f(\nu_H) \\
&= (1-\lambda)(U_2 + f)(\nu_L) + \lambda(U_2 + f)(\nu_H) \\
&\leq (1-\lambda)U_1(\nu_L) + \lambda U_1(\nu_H).
\end{aligned}$$

Therefore, the map

$$\begin{aligned}
r : [0,1] &\longrightarrow \Delta[0,1] \\
\nu &\longmapsto \begin{cases} \delta_\nu & \text{if } \nu \notin (\nu_L, \nu_H) \\ (1-\lambda)\delta_{\nu_L} + \lambda\delta_{\nu_H} & \text{if } \nu = (1-\lambda)\nu_L + \lambda\nu_H \in [\nu_L, \nu_H] \end{cases}
\end{aligned}$$

is such that $\int U_1 \, \mathrm{d}r(\cdot|\nu) \geq U_1(\nu)$ for every $\nu \in [0, 1]$.

Now, take any optimal p_1 for U_1, and define $p_1^* \in \mathcal{R}(\mu)$ via

$$p_1^*(S) = \int_{\Delta\Theta} r(S|\cdot)\,\mathrm{d}p_1.$$

By the above, p_1^* is also optimal. It is also straightforward to show, given $p_1^*((\nu_L, \nu_H)) = 0$, that it is more informative than p_2^*, as desired.

□

9.10 On Optimal Policies: Proofs and Additional Results

Proof of Theorem 4

Proof. By continuity of Blackwell's order, there is a $\succeq_B^\mu$-maximal optimal policy $p \in \mathcal{R}(\mu)$.

For any $C \in \mathcal{C}$, it must be that $p(C \setminus extC) = 0$. Indeed, Phelps (2001, Theorem 11.4) provides a measurable map $r : C \to \Delta(extC)$ with $r(\cdot|\nu) \in \mathcal{R}(\nu)$ for every $\nu \in C$. Then we can define $p' \in \mathcal{R}(\mu)$ via $p'(S) = p(S \setminus C) + \int_C r(S|\cdot)\,\mathrm{d}p$ for each Borel $S \subseteq \Delta\Theta$. Then U-covering and Jensen's inequality imply $\int U\,\mathrm{d}p' \geq \int U\,\mathrm{d}p$, so that p' is optimal too. By construction, $p' \succeq_B^\mu p$, so that (given maximality of p) the two are equal. Therefore $p(C \setminus extC) = p'(C \setminus extC) = 0$. Then, since $\mathcal{C}$ is countable,

$$p(ext^*\mathcal{C}) = 1 - p\left(\bigcup_{C\in\mathcal{C}}[C \setminus extC]\right) = 1.$$

□

Given the above, it is useful to have results about how to find posterior covers. Below, we present two useful abstract propositions for doing just that.

Lemma 10.

1. *Suppose f is the pointwise supremum of a family of functions, $f = \sup_{i\in I} f_i$. If $\mathcal{C}_i$ is an f_i-cover for every $i \in I$, then*

$$\mathcal{C} := \bigvee_{i\in I} \mathcal{C}_i = \left\{\bigcap_{i\in I} C_i : \quad C_i \in \mathcal{C}_i \;\; \forall i \in I\right\}$$

   is an f-cover.

2. *Suppose f is the pointwise supremum of a family of functions, $f = \sup_{i\in I} f_i$. If $\mathcal{C}$ is an f_i-cover for every $i \in I$, then $\mathcal{C}$ is an f-cover.*

3. *If C is a g-cover and h is convex, then C is a $(g+h)$-cover.*

Proof of Proposition 8

Proof. By finiteness of I, the collection C covers $\Delta\Theta$. For each $i \in I$, note that $C_i = \bigcap_{j\in I}\{\nu \in \Delta\Theta : \ f_i(\nu) \geq f_j(\nu)\}$, an intersection of closed convex sets (since $\{f_j\}_{j\in I}$ are affine continuous), and so is itself closed convex. Restricted to C_i, f agrees with f_i and so is affine, and therefore convex. $\square$

Now, under the hypotheses of Proposition 8, we prove a claim that fully characterizes the set of extreme points of the f-cover:

Claim 1. *The f-cover $C = \{C_i : i \in I\}$ given by Proposition 8 satisfies $ext^*C = \{\nu^* \in \Delta\Theta : \{\nu^*\} = S(\nu^*)\}$ where*

$$S(\nu^*) := \{\nu \in \Delta\Theta : \ \textit{supp}(\nu) \subseteq \textit{supp}(\nu^*) \textit{ and } \exists J \subseteq I \textit{ s.t. } f_i(\nu) = f_j(\nu) = f(\nu) \ \ \forall i, j \in J\}. \quad (7)$$

Proof. Fix some $\nu^* \in \Delta\Theta$, for which we will show $\{\nu^*\} \neq S(\nu^*)$ if and only if $\nu^* \notin ext^*(C)$.

Let us begin by supposing $\{\nu^*\} \neq S(\nu^*)$; we have to show $\nu^* \notin ext^*C$. Since $\nu^* \in S(\nu^*)$ no matter what, there must then be some $\nu \in S(\nu^*)$ with $\nu \neq \nu^*$. We will show that $S(\nu^*)$ must then contain some line segment $co\{\nu, \nu'\}$ belonging to some C_i, in the interior of which lies ν^*; this will then imply $\nu^* \notin ext^*C$. Let $\hat{\Theta}$ be the support of ν^*, and let $J := \{i \in I : \ \nu^* \in C_i\}$. Given that $\nu \in S(\nu^*)$, we have $\nu \in \Delta\hat{\Theta}$ with $f_i(\nu) = f_j(\nu) = f(\nu)$ $\forall i, j \in J$. Now, for sufficiently small $\epsilon > 0$, we have $\epsilon(\nu - \nu^*) \leq \nu^*$.[31] Define $\nu' := \nu^* - \epsilon(\nu - \nu^*) \in \Delta\hat{\Theta}$. Then $f_i(\nu') = f_j(\nu') = f(\nu')$ $\forall i, j \in J$ too and, by definition of ν', we have $\nu^* \in co\{\nu, \nu'\}$. If $i \notin J$, then it implies $f_i(\nu^*) > f(\nu)$ because $\nu^* \notin C_i$. Therefore, by moving ν, ν' closer to ν^* if necessary, we can assume $f(\nu) = f_j(\nu) < f_i(\nu)$ and $f(\nu') = f_j(\nu') < f_i(\nu')$ for any $j \in J$ and $i \notin J$. In particular, fixing some $j \in J$ yields $\nu, \nu' \in C_j$, so that ν^* is not in ext^*C.

To complete the proof, let us suppose that $\nu^* \notin ext^*(C)$, or equivalently, $\nu^* \in C_i$ but $\nu^* \notin ext(C_i)$ for some $i \in I$. By definition of C_i, we have that $f_i(\nu^*) = f(\nu^*)$. The fact that $\nu^* \notin ext(C_i)$ implies that there is a non-trivial segment $L \subseteq C_i$ for which ν^* is an interior point. It must then be that $\text{supp}(\nu) \subseteq \text{supp}(\nu^*)$ and $f_i(\nu) = f(\nu)$ for all $\nu \in L$. As a result, $L \subseteq S(\nu^*)$ so that $\{\nu^*\} \neq S(\nu^*)$, completing the proof. $\square$

Corollary 2. *Suppose $\Theta = \{0, 1\}$; A is finite; and for each $a \in A$, $u(a, \cdot) = \min_{i\in I_a} f_{a,i}$, where $\{f_{a,i}\}_{i\in I_a}$ is a finite family of distinct affine functions for each a. Then, there exists an optimal policy that puts full probability on*

$$S := \{0, 1\} \cup \bigcup_{a\in A} \{\nu \in [0, 1] : \ f_{a,i}(\nu) = f_{a,j}(\nu) \textit{ for some distinct } i, j \in I_a\}.$$

[31] Here, $\leq$ is the usual component-wise order on $\mathbb{R}^{\hat{\Theta}}$.

References

Arrow, Kenneth J. 1971. *Essays in the Theory of Risk Bearing*. Chicago: Markham.

Aumann, Robert J. 1974. "Subjectivity and correlation in randomized strategies." *Journal of Mathematical Economics* 1 (1):67–96.

Baumeister, Roy F. 2002. "Yielding to Temptation: Self-Control Failure, Impulsive Purchasing, and Consumer Behavior." *Journal of Consumer Research* 28:670–676.

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.

Bergemann, Dirk and Stephen Morris. 2013a. "The Comparison of Information Structures in Games: Bayes Correlated Equilibrium and Individual Sufficiency." Princeton University.

———. 2013b. "Robust Predictions in Games With Incomplete Information." *Econometrica* 81 (4):1251–1308.

Blackwell, David. 1953. "Equivalent Comparison of Experiments." *Annals of Mathematics and Statistics* 32:265–272.

Caplin, Andrew and John Leahy. 2004. "The supply of information by a concerned expert." *Economic Journal* 114 (497):487–505.

Crawford, Vincent P. and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6):1431–51.

Dubra, J., F. Maccheroni, and E. Ok. 2004. "Expected utility theory without the completeness axiom." *Journal of Economic Theory* 115:118–133.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica. 2014. "Suspense and Surprise." *Quarterly Journal of Economics* :Forthcoming.

Gabaix, Xavier. 2014. "A Sparsity-Based Model of Bounded Rationality." *Quarterly Journal of Economics* :Forthcoming.

Ganuza, Juan-Jose and Jose S. Penalva. 2010. "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions." *Econometrica* 78 (3):1007–1030.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological games and sequential rationality." *Games and Economic Behavior* 1 (1):60–79.

Gilboa, Itzhak and David Schmeidler. 1988. "Information dependent games: Can common sense be common knowledge?" *Economics Letters* 27 (3):215–221.

Grant, Simon, Atsushi Kajii, and Ben Polak. 1998. "Intrinsic Preference for Information." *Journal of Economic Theory* 83 (2):233–259.

Gul, Faruk and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica* 69 (6):1403–1435.

Harrison, Glenn W. and E. Elisabeth Rutstrom. 2008. "Risk Aversion in the Laboratory." In *Risk Aversion in Experiments, Research in Experimental Economics*, vol. 12, edited by J. Cox Cox and Glenn W. Harrison. Bingley: Emerald, second ed., 41–196.

Huang, Kevin, Zheng Liu, and John Q. Zhu. 2013. "Temptation and self-control: some evidence and applications." Tech. Rep. 2013-23, Federal Reserve Bank of San Francisco.

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

Kihlstrom, Richard E. and Leonard Jay Mirman. 1974. "Risk aversion with many commodities." *Journal of Economic Theory* 8 (3):361–388.

Kreps, David M. and Evan L. Porteus. 1978. "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica* 46 (1):185–200.

Krusell, Per, Burhanettin Kuruscu, and Anthony A. Smith. 2010. "Temptation and Taxation." 78 (6):2063–2084.

Loewenstein, G., Cass R. Sunstein, and Russell Goldman. 2014. "Disclosure: Psychology Changes Everything." *Annual Economic Review* 6.

Myerson, Roger. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47 (1):61–74.

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

Pratt, J. W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32:122–136.

Sargent, Thomas J. and Lars Peter Hansen. 2001. "Robust Control and Model Uncertainty." *American Economic Review* 91 (2):60–66.

Taneva, Ina. 2014. "Information Design." Economics Department, Edinburgh University.

Tardella, Fabio. 2008. "On the existence of polyhedral convex envelopes." In *Frontiers In Global Optimization*, edited by C.A. Floudas and P.M. Pardalos. Boston: Kluwer, 149–188.