American Economic Journal: Microeconomics 2018, 10(4): 67–93
https://doi.org/10.1257/mic.20160247

Disclosure to a Psychological Audience†

By Elliot Lipnowski and Laurent Mathevet*

We study how a benevolent expert should disclose information to an agent with psychological concerns. We first provide a method to compute an optimal information policy for many psychological traits. The method suggests, for instance, that an agent suffering from temptation à la Gul and Pesendorfer (2001) should not know what he is missing, thereby explaining observed biases as an optimal reaction to costly self-control. We also show that simply recommending actions is optimal when the agent is intrinsically averse to information but has instrumental uses for it. This result, which circumvents the failure of the Revelation Principle in psychological environments, simplifies disclosure and informs the debate regarding mandated disclosure. (*JEL* D11, D82, D83, D91)

Disclosing or concealing information can influence a person's choices by affecting what he knows. In many contexts, information disclosure decisions aim to improve the welfare of the otherwise uninformed party: in public policy, for instance, regulators impose disclosure laws, such as required transparency by a seller of a good to a buyer, so that the buyer can make a better decision. But should a well-intentioned advisor always reveal everything? 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 conceal details about their personal lives from family elders, doctors do not reveal every minute detail of their patients' health, future parents do not always want to know the sex of their unborn children, and many do not want to know the caloric content of the food they eat.

We study how an informed expert should disclose information to an agent with psychological characteristics, i.e., someone whose state of mind has a direct impact on his well-being. This question is faced by many trusted advisors, from family and

* Lipnowski: Department of Economics, University of Chicago, 1126 E 59th Street, Chicago, IL 60637 (email: lipnowski@uchicago.edu); Mathevet: Department of Economics, New York University, 19 W 4th Street, New York, NY 10012 (email: lmath@nyu.edu). We thank Adam Brandenburger, Joyee Deb, Juan Dubra, Xavier Gabaix, Emir Kamenica, Alessandro Lizzeri, Alessandro Pavan, David Pearce, Debraj Ray, and Tomasz Strzalecki for their comments. We also thank seminar audiences at Penn State University, Princeton University, Indiana University, Harvard University and MIT, New York University, Midwest Theory (Kansas), and the Canadian Economic Theory Conference (UWO).

† Go to https://doi.org/10.1257/mic.20160247 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

friends to teachers and doctors. In medical science, the direct relevance of patients' beliefs is well-known, as Ubel (2001) reports:

> *Mildly but unrealistically positive beliefs can improve outcomes in patients with chronic or terminal diseases … Moreover, unrealistically optimistic views have been shown to improve quality of life.*

In choosing which medical tests to take, for example, many patients desire to be informed about their health only when the news is good, but prefer to know less if the news is bad. Even in this simple case, it is not obvious how best to disclose information. The first scheme that comes to mind—reveal the news when it is good and say nothing otherwise—accomplishes little here, since no news then reveals bad news. However, in line with Benoît and Dubra (2011), it is possible to leave many (but not all) patients with unrealistically positive beliefs, even if they process information in a Bayesian way.

In a regulatory context, where lawmakers often mandate disclosure with the uninformed party's welfare in mind, Loewenstein, Sunstein, and Golman (2014) argue that psychology may lead one to rethink public disclosure policies. For example, a comprehensive accounting of costs should include the psychological costs of dealing with the information. Think, for example, of giving information about foregone food to someone who is fasting or dieting. This might create temptation if he resists, and disappointment if he succumbs. These considerations are relevant to all industries with regulated disclosure, ranging from financial services and food services to medical services and more (see Ben-Shahar and Schneider 2011).

The Model in Brief.—In our model, an expert decides which information an agent will obtain about the state of nature. After receiving his information from the expert, the agent updates his beliefs and then acts. The expert chooses a disclosure policy to maximize the agent's ex ante expected utility. Depending on the application, the expert can be any of a well-intentioned advisor (e.g., a doctor) who decides which information to seek on behalf of the agent (e.g., which medical tests to perform); the agent himself, deciding which information to obtain (e.g., whether to read nutritional information for a planned meal); or a third party who regulates information disclosure between a sender (e.g., a seller) and the agent. In each of these cases, the expert sets the disclosure policy ex ante, in ignorance of the realized state, and thus faces no interim opportunity to lie. Accordingly, our principal enjoys the same commitment power as that of Kamenica and Gentzkow (2011).

We model psychology by assuming that the agent's welfare 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 (Geanakoplos, Pearce, and Stacchetti 1989), which captures varied behavioral and psychological phenomena including those listed below.

The Theory.—This paper proposes a method that simplifies the search for optimal information policies. The well-known concave envelope result in Kamenica and Gentzkow (2011) extends readily to our environment,[1] but in some applications it may not be enough—as deriving the concave envelope of a function can be demanding. In response, we develop a hands-on method for concavification and bring with it the natural economic structure amenable to this method.

Our key observation is that local information preferences can be used to simplify the search for a globally optimal disclosure policy. Though straightforward once formally expressed, our approach is surprisingly powerful when it can be applied without solving the agent's problem (i.e., without needing to derive the function that will be concavified). This is the case for many popular behavioral models, such as temptation and self-control problems (Gul and Pesendorfer 2001), belief distortion (Brunnermeier, Papakonstantinou, and Parker 2016), ego utility (Kőszegi 2006), shame (Dillenberger and Sadowski 2012), and reference-dependent choice (Kahneman and Tversky 1979, Kőszegi and Rabin 2009). In those models, the qualitative form of the psychological effect suggests many beliefs at which the agent will prefer more information, leading the expert to the heart of the underlying tradeoff.

The rest of the paper takes a more applied perspective. First, we illustrate our method in the above models, with a particular focus on temptation and self-control, and then we move on to issues of implementation.

Applications: Reference Dependence and Temptation.—How should we talk to someone who is loss-averse or suffers from temptation? When losses weigh on the agent at constant sensitivity, as studied in Kőszegi and Rabin (2009), there is no disappointment worth averaging. The agent is psychologically willing to incur maximal disappointment or none at all. When the same is true for gains, these maximal or null surprises correspond to 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 à la Gul and Pesendorfer (2001), the optimal policy depends on the nature of information. Focusing on a discrete consumption-saving problem, in which an individual has impulsive desires to consume rather than save, we demonstrate how the method of posterior covers can be brought to bear quite generally. The simple yet resonant lesson is that, as studied in psychology, a tempted agent *does not want to know what he is missing*. This desire to limit information is something one observes in real life. For example, someone who has decided to save this month might not sign up for a news feed about restaurant openings he will not be enjoying. By formalizing exactly in which way the agent does not want to know what he is missing, our analysis explains this desire as an optimal reaction to costly self-control. This phenomenon is an instance of strategic ignorance, related to Carrillo and Mariotti (2000) and Bénabou and Tirole (2002).[2]

[1] Concavification has a history in repeated games with incomplete information (Aumann and Maschler 1995), geometric moment theory (Kemperman 1968), optimal portfolio management (Anastassiou 2006), and more.

[2] In the Strotzian world studied by those papers, the consumer limits his own information today to alter his chosen action tomorrow; in the model of Gul and Pesendorfer (2001), even if information does not affect the agent's choice—which it might—it is still valuable to keep him ignorant about the value of more tempting choices that he resists. This way, the agent does not suffer as much from foregone opportunity.

Application: Giving Advice.—The existence of an optimal policy in theory does not tell one how to implement it in practice. In the two contexts of healthcare and campus safety, for instance, mandated disclosure requires an avalanche of information concerning all possible side effects of a medication and all recent on-campus crimes, respectively. Although there are reasons for wanting such transparency, professionals are questioning the benefits. As Ben-Shahar and Schneider (2011, 746) argue:

> *When we abandon the unreal world of mandated disclosure and ask how people really make decisions, we see that they generally seek—and that the market often supplies—not data, but* advice.

We demonstrate that a simple class of policies, advanced in the public debate, is in fact optimal for a broad class of agents: an expert can simply advise an agent by recommending an action. If an agent innately likes information, this minimalist class leaves something to be desired, but what if, to agents who like information only for its instrumental value (i.e., *psychologically information-averse agents*), we only recommend a behavior? A standard approach would use the Revelation Principle—that replacing every message with a recommendation preserves incentives. Unfortunately, the Revelation Principle fails in psychological settings. Despite this negative observation, Theorem 2 says that some incentive-compatible recommendation policy is optimal when preferences display psychological information aversion.

Related Literature.—This paper is part of a recently active literature on information design, as in Brocas and Carrillo (2007), Gentzkow and Kamenica (2014), Rayo and Segal (2010), and Gentzkow and Kamenica (2016).[3] The most related works, methodologically speaking, are Kamenica and Gentzkow (2011) and Gentzkow and Kamenica (2014), who also study information transmission by a sender with full commitment power. We adopt the same overall method (albeit in a different environment) with a practical mindset, which leads us to provide applied tools beyond concavification and to explore the subtleties of disclosure under various psychological attitudes.

The relevance of psychological preferences in information contexts is exemplified by Schweizer and Szech (2012),[4] who study optimal revelation of life-changing information between a sender and a receiver with anticipatory utility, and also by Ely, Frankel, and Kamenica (2015), who study dynamic information disclosure to an agent who gets satisfaction from being surprised or feeling suspense. The latter describe how a fixed amount of information should be parceled out over time to maximize surprise or suspense.

There is a conceptual connection between our classification of information preferences (Section III) and that of Grant, Kajii, and Polak (1998) in their study of intrinsic informational preferences (over lotteries of lotteries) absent the reduction axiom (à

[3] See also the broader literature on optimal information structures such as Ostrovsky and Schwarz (2010), Caplin and Eliaz (2003), and Hörner and Skrzypacz (2016).

[4] See also Caplin and Leahy (2004), who study a detailed interaction between a doctor and a patient with quadratic psychological preferences, with two-sided private information.

la Kreps and Porteus 1978). Our conceptual distinction between psychological and behavioral information preferences, absent in their world, is critical for information disclosure.

The paper is organized as follows. Section I introduces psychological preferences and the language of random posteriors. Section II presents the psychological agent and characterizes his possible attitudes toward information. Section III develops the method of posterior covers. Section IV applies the method to analyze disclosure to a tempted agent. Section V delineates when good advice suffices for optimal information. Section VI concludes.

I. The Environment

An agent must make a decision when the state of nature $\theta \in \Theta$ is uncertain.[5] The agent has (full support) prior $\mu \in \Delta\Theta$, and receives additional information about the state from the principal. After receiving said information, the agent forms a posterior belief by updating his prior, and then he makes a decision.

A. *Psychological Preferences*

An *outcome* is an element $(a,\theta,\nu) \in A \times \Theta \times \Delta\Theta$, where a is the action taken by the agent, $\theta \in \Theta$ is the true state of the world, and $\nu \in \Delta\Theta$ denotes the agent's posterior belief at the moment when he makes his decision. We assume that the agent has a utility function $u : A \times \Theta \times \Delta\Theta \to \mathbb{R}$ over outcomes. In a slight abuse of notation, we can define $u(a,\nu) := \int_\Theta u(a,\theta,\nu)\, d\nu(\theta)$. For a benevolent sender with commitment (as studied here), these "reduced preferences" $u : A \times \Delta\Theta \to \mathbb{R}$ are the only relevant feature of the agent's preferences.[6] Given posterior beliefs ν, the agent chooses an action $a \in A$ to maximize $u(a,\nu)$. We assume either that u is continuous, or that u is upper-semicontinuous and the state space Θ is finite. For every posterior belief $\nu \in \Delta\Theta$, define the **indirect utility** associated with ν as $U(\nu) = \max_{a\in A} u(a,\nu)$.

In the *classical* case, the agent's belief does not enter his welfare, that is, $u \equiv u(a,\theta)$, so that $\mathbb{E}_{\theta\sim\nu} u(a,\theta)$ is affine in beliefs ν for every a. In our environment, the agent's satisfaction depends not only on the physical outcome, but also on his posterior beliefs. In the literature, such an agent is said to have *psychological preferences* (Geanakoplos, Pearce, and Stacchetti 1989).

This formulation covers a wide range of phenomena. Below is a list of psychological models, familiar from behavioral economics and behavioral finance, which can be readily accommodated by our forthcoming method (see Section A in the Appendix for a more extensive list, along with optimal disclosure prescriptions

[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 and metrizable. Given $\pi \in \Delta Y$, let $\operatorname{supp}(\pi)$ denote the support of π (i.e., the smallest closed subset of Y of full π-measure).

[6] This contrasts with the cheap talk setting of Caplin and Leahy (2004), wherein the realized state affects the hypothetical gains of a deviating sender.

for each). For simplicity, let $\Theta = \{0,1\}$, and write any posterior belief as $\nu = \Pr(\{\theta = 1\})$:

(**a**) (**Purely psychological agent**): Let

$$u(a,\theta,\nu) = -\mathrm{var}_{\theta\sim\nu}(\theta) = -\nu(1-\nu)$$

represent an agent whose only satisfaction comes from his degree of certainty, as quantified by the variance $\mathrm{var}_{\theta\sim\nu}(\theta) = \mathbb{E}_{\theta\sim\nu}(\theta^2) - (\mathbb{E}_{\theta\sim\nu}\theta)^2$.

(**b**) (**Stubbornness/prior-bias**): For $\rho > 0$ and classical motive u_C, let

$$u(a,\theta,\nu) = u_C(a,\theta) - \rho|\nu - \mu|$$

represent an agent who wants to make good choices but experiences discomfort when information conflicts in any way with his prior beliefs.

(**c**) (**Temptation and self-control**): Given two classical utility functions u_R (rational) and u_T (tempted), let

$$u(a,\theta,\nu) = u_R(a,\theta) - \max_{b\in A}\mathbb{E}_{\hat{\theta}\sim\nu}\left[u_T(b,\hat{\theta}) - u_T(a,\hat{\theta})\right],$$

in the spirit of Gul and Pesendorfer (2001). The agent receives his information and then chooses an action in a finite menu A.[7] His non-tempted self experiences utility $u_R(a,\theta)$ from action a, while the forgone value $\max_{b\in A}\mathbb{E}_{\hat{\theta}\sim\nu}\left[u_T(b,\hat{\theta}) - u_T(a,\hat{\theta})\right]$ is a cost of self-control faced by the tempted side.

B. *Signals and Random Posteriors*

The principal discloses information about the state to the agent by choosing a signal (S,σ), which consists of a space S of messages and a map $\sigma : \Theta \to \Delta S$. In any state θ, the agent sees a message $s \in S$ drawn according to $\sigma(\cdot|\theta) \in \Delta S$, and then forms a posterior belief via Bayesian updating. Since the state is ex ante uncertain, a signal induces a distribution over the agent's posterior beliefs. From a welfare perspective, only this distribution over posterior beliefs matters. Choosing a signal is then equivalent to choosing an information policy, defined below.

[7] In that model, preferences over menus are used to identify the impact of unchosen alternatives on the decision maker. Our interest is not in *identifying* temptation, but rather in studying its implications for optimal disclosure. Accordingly, we consider an agent facing a fixed menu of alternatives, and we adapt the functional form of Gul and Pesendorfer (2001) to accommodate lotteries over lotteries (by assuming independence of "early resolution" lotteries, as well as recursivity, as defined in Grant, Kajii, and Polak 1998). Even with a fixed action set, the functional form is suitable to study the effect of information under temptation. As we will see, providing information to a tempted agent can exacerbate his (expected) self-control costs.

DEFINITION 1: *An information policy (given prior μ) is an element of*

$$\mathcal{R}(\mu) := \{p \in \Delta\Delta\Theta : \mathbb{E}_{\nu\sim p}[\nu] = \mu\}$$
$$= \left\{p \in \Delta\Delta\Theta : \int_{\Delta\Theta}\nu(\hat{\Theta})\,dp(\nu) = \mu(\hat{\Theta}) \text{ for every Borel } \hat{\Theta} \subseteq \Theta\right\}.$$

Although there is room for manipulation, the posterior beliefs of a Bayesian agent must, on average, equal his prior. All signals generate an information policy, and, as described in Benoît and Dubra (2011) and Kamenica and Gentzkow (2011), all information policies can be generated by some signal. In particular, one could employ a **direct signal** (also known as a **posterior policy**) (S_p, σ_p) to produce p, i.e., a signal for which every message is a posterior, $S_p := \mathrm{supp}(p) \subseteq \Delta\Theta$, and when the agent hears message "s," his update yields a posterior belief equal to s. This signal tells the agent what his posterior belief should be, and his beliefs conform.[8]

II. The Psychological Agent

Some agents prefer to be informed, and others do not (see Frankl, Oye, and Bellamy 1989 and Oster, Shoulson, and Dorsey 2013 on patients' attitudes toward information concerning life support and Huntington's disease, respectively). Attitudes toward information, by driving the agent's welfare, are a critical aspect of optimal information disclosure.

We begin by distinguishing *psychological* and *behavioral* attitudes toward information. Given two information policies, $p, q \in \mathcal{R}(\mu)$, p is **more informative** than q, denoted $p \succeq_B^\mu q$, if p is a mean-preserving spread of q.[9] So a more informative signal induces a lottery over agent beliefs that is more correlated with the uncertain state, and, hence, riskier. This definition is known to be equivalent to the usual Blackwell (1953) garbling definition (Lemma 1, in the Appendix).

DEFINITION 2: *The agent is psychologically information-loving [resp. -averse, -neutral] if, given any action $a \in A$ and any information policies $p, q \in \mathcal{R}(\mu)$ with $p \succeq_B^\mu q$,*

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

The agent is behaviorally information-loving [resp. -averse, -neutral] if, given any information policies $p, q \in \mathcal{R}(\mu)$ with $p \succeq_B^\mu q$,

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

[8] When $s \ll \mu$ a.s.-$p(s)$, the signal with this law works: $\sigma_p(\hat{S}|\theta) = \int_{\hat{S}}\frac{ds}{d\mu}(\theta)\,dp(s)$ for every $\theta \in \Theta$ and Borel $\hat{S} \subseteq S_p$.

[9] That is, if there is a map $r : S_q \to \Delta(S_p) \subseteq \Delta\Delta\Theta$ such that (i) for every Borel $S \subseteq \Delta\Theta$, $p(S) = \int_{S_q} r(S|\cdot)\,dq$ and (ii) for every $t \in S_q$, $r(\cdot|t) \in \mathcal{R}(t)$.

When an agent likes [dislikes] more information for its own sake, in the hypothetical event that he cannot adapt his decisions to it, he is psychologically information-loving [-averse]. For such an agent, information is intrinsically valuable [damaging], abstracting from its instrumental value.[10] In contrast, behavioral attitudes toward information assume that the agent can respond optimally to information. Imagine a person whose imminent seminar presentation may have some errors. If his laptop is elsewhere, he may prefer not to know about any typos, since knowing will distract him. He is psychologically information-averse. But if he has a computer available to change the slides, he may want to know, as such information is instrumentally helpful. He is behaviorally information-loving.

Put differently, there are two effects of information: psychological and instrumental. These two forces act in the same direction when the agent is psychologically information-loving, and in opposite directions when he is psychologically information-averse.

PROPOSITION 1:

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

(*ii*) *If the agent is psychologically information-loving, then he is behaviorally information-loving. If (*$A \subseteq \mathbb{R}^k$ *is convex and)* u *is jointly concave, then the agent is behaviorally information-averse.*

The first part of the proposition, which is not new,[11] shows that psychological and behavioral information preferences are closely linked, while the second part tells us exactly how. Consider a binary-state world with $A = [0,1]$ and $u_k(a,\theta,\nu) = k\text{var}(\nu) - (\theta - a)^2$. Psychological information preferences depend on whether $k > 0$, while behavioral preferences depend on whether $k > 1$. In particular, if the agent likes information per se $(k < 0)$, he also likes it when he can use it $(k < 1)$. However, he might dislike information from a psychological perspective $(k > 0)$ but have this effect overwhelmed by the value of making good decisions $(k < 1)$. This conflict disappears when the agent's utility is *jointly* concave in (a,ν), a condition that both imposes psychological aversion to information and limits its instrumental use.[12]

[10] Grant, Kajii, and Polak (1998) define what it means to be *single-action information-loving* for an agent who makes no decision. In their model, there is no distinction between psychological and behavioral information preferences.

[11] This result is familiar to readers of the literature on temporal resolution of uncertainty, as in Kreps and Porteus (1978), and has been substantially generalized by Grant, Kajii, and Polak (1998, Proposition 1) in the case in which no action is taken, ($A = \{a\}$, so that $u = U$).

[12] This concavity condition is essentially never satisfied for classical preferences, as bilinear functions have saddle points.

Perhaps the most economically interesting case is psychological information-aversion. For such agents, the primary question is whether the instrumental effect dominates the psychological effect. When the first-order approach is valid and the environment smooth, the answer can be computed and the tradeoff formalized. At any belief ν, the local value of information can be expressed as

$$U''(\nu) = u_{\nu\nu}(a,\nu) + \frac{u_{a\nu}(a,\nu)^2}{-u_{aa}(a,\nu)},$$

where the right-hand side is evaluated at the optimal action. When information is instrumentally useful ($u_{a\nu} \neq 0, -u_{aa} > 0$) but psychologically damaging ($u_{\nu\nu} < 0$), the agent faces a genuine tradeoff.

III. Information Disclosure: Theory

Consider information disclosure by a benevolent expert who is a trusted advisor, the agent himself, or a public entity that serves the agent's interests. The expert's goal is to maximize the agent's ex ante expected welfare. An information policy p is **optimal** 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 generates an optimal information policy. An optimal policy is already transparent for some classes of agents. If the agent is information-loving, tell him everything; if he is behaviorally information-averse, tell him nothing. For agents with some aversion to information, but who are not behaviorally information-averse, things are subtler.

The concavification result of Kamenica and Gentzkow (2011) extends readily to all psychological agents, providing an abstract characterization of the optimal value. For any prior μ, an optimal information policy exists and induces expected indirect utility

$$\bar{U}(\mu) = \inf\{\phi(\mu) \mid \phi : \Delta\Theta \to \mathbb{R} \text{ affine continuous}, \phi \geq U\},$$

i.e., the concave envelope of U evaluated at the prior. In applications, however, concavification is not a trivial affair. Unless the designer can compute the envelope or somehow derive qualitative properties of it, which is known to be difficult beyond specific examples (see Tardella 2003), the applied lessons from concavification are elusive. In response, we develop a method to simplify the computation of an optimal policy in models of interest and deliver prescriptions.

Method of Posterior Covers.—We approach the expert's problem by reducing the support of the optimal policy based on local arguments.

In some situations, the designer can deduce from the primitives $\langle A,u\rangle$ that the indirect utility U must be locally convex on various regions of $\Delta\Theta$. In every such region, the agent likes (mean-preserving) spreads in beliefs, which correspond to more informative policies. Consequently, an optimal policy need not employ beliefs inside of those regions, regardless of its other features. The central concept of our approach is the posterior cover, a collection of such regions.

DEFINITION 3: *Given* $f : \Delta\Theta \to \mathbb{R}$, *an f-(posterior) cover is a finite family* $\mathcal{C}$ *of closed convex subsets of* $\Delta\Theta$ *such that* $f|_C$ *is convex for every* $C \in \mathcal{C}$.

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

$$out(\mathcal{C}) = \{\nu \in \Delta\Theta : \nu \in ext(C) \text{ whenever } \nu \in C \in \mathcal{C}\}$$

be its set of **outer points**. That is, outer points are those posterior beliefs that are extreme in any member of $\mathcal{C}$ to which they belong. In particular, any point outside $\cup\mathcal{C}$ is an outer point, as is any deterministic belief.

The next theorem develops the method of posterior covers in three parts.

THEOREM 1 (Method of Posterior Covers):

(OPTIMAL SUPPORT).—*If* $\mathcal{C}$ *is a U-cover, then* $p(out(\mathcal{C})) = 1$ *for some optimal policy p.*

(REDUCTION TO PRIMITIVES).—*If* $\mathcal{C}$ *is a* $u(a,\cdot)$*-cover* $\forall a \in A$, *then* $\mathcal{C}$ *is a U-cover. In particular, if u takes the form* $u(a,\theta,\nu) = u_C(a,\theta) + u_P(\nu)$, *then any* u_P*-cover is a U-cover.*

(CONCRETE DERIVATION).—*If f is the pointwise minimum of finitely many affine functions* $\{f_i : i \in I\}$, *and* ϕ *is convex, then* $\mathcal{C} := \{\{\nu : f(\nu) = f_i(\nu)\} : i \in I\}$ *is a* $(\phi \circ f)$*-cover.*

The first part restricts the search for an optimal policy to the outer points of a *U*-cover. Since the agent is locally information-loving within each member of a *U*-cover, an optimal policy need never leave him with non-extremal beliefs in such a set. Having an appropriate *U*-cover in hand thus simplifies optimal disclosure, but how does one find such a cover? One potential obstacle is that the indirect utility is a *derived* object. To make the first part useful, then, it is important to tie posterior covers of *U* to primitive features of the model. This is the next step.

The second part simplifies the search for a *U*-cover by analyzing the form of *u*, a step made simpler by the assumption of aligned interests. Any $\mathcal{C}$ that is a $u(a,\cdot)$-cover for all $a \in A$ is guaranteed to be a *U*-cover, making the first part applicable. This observation comes from applying the logic of Proposition 1 separately for each element of the cover. Note that, starting from posterior covers, one for each action, one can construct such a simultaneous posterior cover if *A* is finite.[13]

[13] If $\mathcal{C}_a$ is a $u(a,\cdot)$-cover for every $a \in A$, then $\mathcal{C} := \{\bigcap_{a\in A} C_a : C_a \in \mathcal{C}_a \, \forall a \in A\}$ is a *U*-cover.

Moreover, for agents who can separate their psychological effects from their classical motive, the determination of a U-cover reduces to finding a posterior cover of the psychological component alone. This includes all preferences (a)-(f) in Appendix Section A.

Having connected the problem to primitives, it remains to compute a useful cover of those primitive functions. Generally, this task depends on the nature of the function, and can be difficult. The third part points to a flexible, economically relevant class of preferences—the minimum of affine functions (or a convex transformation thereof)—for which such a computation is possible. For this class, a useful U-cover and its (finite) set of outer points can be computed explicitly. This class includes economic situations of interest.

A remarkable, though easy to overlook, aspect of the theorem is that the expert need not solve the agent's optimization problem at every posterior belief in order to derive an optimal policy. Said technically, our method enables the concavification of U without needing to derive U, if the primitives $\langle A, u \rangle$ meet some conditions. In this case, an optimal policy can be characterized while computing U *only* on the outer points of the posterior cover. Without our method, one would have to first solve the agent's problem for every possible posterior belief, and then concavify the induced indirect utility over its whole domain. Instead, after deducing a small set of outer points, one can characterize an optimal policy by solving the agent's problem at each of these posterior beliefs and solving the restricted disclosure problem. In various behavioral models, this is feasible because of the specific structure of the psychological effect: it is piecewise (weakly) convex. This structure is simple but rich, and very common in applied behavioral work (again, see Section A in the Appendix). The resulting simplification, also enabled by our restriction to aligned interests, considerably reduces the task of finding an optimal policy.

IV. Application: Temptation and Information

Economists and psychologists have presented evidence that temptation can be especially strong in consumption decisions (Gruber and Kőszegi 2001 and Baumeister 2002). Motivated by this evidence, a recent literature asks how tax policies can be designed to alleviate the effects of temptation (O'Donoghue and Rabin 2003; Gruber and Kőszegi 2004; Krusell, Kuruşçu, and Smith 2010). In this section, we examine a discrete choice model with tempting options, and instead ask how information disclosure policies can alleviate the effects of temptation. As we will see, a tempted agent *does not want to know what he is missing*. This maxim has been studied in psychology, for example, by Otto and Love (2010). The analysis below formalizes it and explains it as an optimal reaction to costly self-control.

Consider a consumer who decides how to spend a sum of money. The consumer has various options $A = \{a_1, \ldots, a_n\}$ that may include different types of investments and even immediate consumption. The unknown state of the world, $\theta \in \Theta$, is distributed according to prior μ. The agent has preferences à la Gul and Pesendorfer

(2001): given A and two classical utility functions u_R and u_T, the agent faces a welfare of

$$u(a,\theta,\nu) \;=\; u_R(a,\theta) - \max_{b\in A} \mathbb{E}_{\hat{\theta}\sim\nu}\Big[u_T\big(b,\hat{\theta}\big) - u_T\big(a,\hat{\theta}\big)\Big],$$

when he chooses action a at posterior belief ν. With this functional form, it is as if the agent has a "rational side" u_R and "tempted side" u_T. The rational side has an expected value of $\mathbb{E}_{\theta\sim\nu}\{u_R(a,\theta)\}$, but exerting self-control entails a personal cost of $\max_b \mathbb{E}_{\hat{\theta}\sim\nu}\Big\{u_T\big(b,\hat{\theta}\big) - u_T\big(a,\hat{\theta}\big)\Big\}$. This psychological penalty is the value forgone by the tempted side when consuming a.

Letting $u_P(\nu) \coloneqq \min_{b\in A}\big\{-\mathbb{E}_{\theta\sim\nu}\big[u_T(b,\theta)\big]\big\}$, the agent's reduced preferences can be written as

$$(1) \qquad u(a,\nu) \;=\; \mathbb{E}_{\theta\sim\nu}\Big\{u_R(a,\theta) - \max_{b\in A}\mathbb{E}_{\hat{\theta}\sim\nu}\Big[u_T\big(b,\hat{\theta}\big) - u_T\big(a,\hat{\theta}\big)\Big]\Big\}$$

$$= \mathbb{E}_{\theta\sim\nu}\Big\{-\max_{b\in A}\mathbb{E}_{\hat{\theta}\sim\nu}u_T\big(b,\hat{\theta}\big)\Big\} + \mathbb{E}_{\theta\sim\nu}u_R(a,\theta) + \mathbb{E}_{\hat{\theta}\sim\nu}u_T\big(a,\hat{\theta}\big)$$

$$= u_P(\nu) + \mathbb{E}_{\theta\sim\nu}\big[u_R(a,\theta) + u_T(a,\theta)\big].$$

By linearity of expectation, u_P is a minimum of affine functions, so the full strength of Theorem 1 applies. It is now straightforward to find a u_P-cover with a small set of outer points, and then to name an optimal policy.

Section VA offers prescriptions of the optimal information policy.

Not Knowing What One Is Missing.—In the model of Gul and Pesendorfer (2001), information is critical for rational decision-making, $\max_a \mathbb{E}[u_R + u_T]$, but it also induces more temptation since information increases the value to the impulsive side, $\max_b \mathbb{E}[u_T]$. Optimal disclosure balances these two forces.

The general advice from the method of posterior covers is that an optimal information policy is supported on beliefs at which the tempted self has, in some sense, as little instrumental information as possible. Knowing more exacerbates the appeal of forgone choices, for no instrumental benefit. Mathematically, say that the agent *does not know what he is missing* at belief ν if, for all $\nu' \neq \nu$ supported by some $q \in \mathcal{R}(\nu)$,

$$(2) \qquad \arg\max_{b\in A}\mathbb{E}_{\theta\sim\nu'}\big[u_T(b,\theta)\big] \;\not\supseteq\; \arg\max_{b\in A}\mathbb{E}_{\theta\sim\nu}\big[u_T(b,\theta)\big].$$

When an agent does not know what he is missing, more information $\big(q \in \mathcal{R}(\nu)\big)$ would only serve to eliminate some temptations from the agent's mind, and give him a better idea of precisely what choices he is missing.

PROPOSITION 2: *There is an optimal policy p for the tempted agent whereby a.s.-$p(\nu)$ he does not know what he is missing at ν.*

To illustrate the proposition's maxim, suppose a consumer has two options, $A = \{\text{save}, \text{consume}\}$ $(= \{0,1\})$; that the state is binary, $\theta \in \{0,1\}$, and $\mu \equiv \mu\{\theta = 1\} = 1/2$; and that information affects only one side. If u_T is state-independent, the agent is psychologically, and hence behaviorally, information-loving by Proposition 1.

More interesting is the case of state-independent u_R, say if θ is the unknown release date of a music album or a new technology. The consumer then exhibits psychological information aversion because u_P is concave (see equation (1)). For example, let $u_T(a,\theta) = 2\theta a + (1-a)(1-\theta)$ and $u_R(a) = -a/3$. The consumer is tempted to consume only when $\theta = 1$, while his rational side prefers not to consume. Despite information aversion, information can make the consumer better off. Indeed, if he receives no information, he always succumbs to temptation and never saves. But the following policy is an improvement: if the state is tempting—say if the new iPhone has just come out—then with probability $1/2$, tell the consumer to buy it, and with complementary probability or when the state is not tempting, tell him to save. This policy, whose advice the agent optimally follows, ensures that the agent splurges only 25 percent of the time, while investing and mitigating temptation the remaining 75 percent of the time. When the agent is told to consume, he suffers no cost of self-control, and when told to save, faces a reduced cost, believing the state likely to be bad. At this belief, the tempted side is indifferent between saving and consuming, so the consumer does not know what he is missing.

V. Application: Giving Advice

Even knowing that a policy is optimal in theory need not make it easy to implement it in practice. In reality, many professionals are concerned about standards of disclosure required by law. Ben-Shahar and Schneider (2011, 665) write:

> *The great paradox of the Disclosure Empire is that even as it grows, so also grows the evidence that mandated disclosure repeatedly fails to accomplish its ends.*

Among mandated disclosure laws figure both Informed Consent and the Clery Act. The former demands that a doctor provide a detailed statistical description of all possible side effects of a treatment or drug, and the latter requires institutions of higher education to publish an annual campus security report about all crimes from the past three years. In both cases, the requirements are akin to a posterior policy (also called direct signals in Section IIB), demanding that the principal send entire probability distributions to the agent. In a binary-state context, this can be plausible. For example, a doctor who uses a blood test to check for a disease might report the updated probability that the patient has the illness rather than not. When the state space is more complicated, however, posterior policies are costly to implement, and the listener might not know how best to interpret what he learns.

As Ben-Shahar and Schneider (2011, 746) argue,

> *When we … ask how people really make decisions, we see that they generally seek … not data, but* advice.

Consequently, we study whether this simple form of communication—giving advice—delivers the same benefit as other policies do. In psychological contexts, an agent might want information for its own sake, and so be well served by extraneous information. Information-averse agents, however, have no such desire. For them, we might want to employ a **recommendation policy**, i.e., a signal of the form (A, σ). For example, a doctor could simply recommend special dietary choices if her analysis reveals a pathology, or police could recommend avoiding walking in certain areas if they notice potential dangers during their patrols. While a posterior policy gives the agent all the information he might want, a recommendation policy gives him only the information that he *needs*.

As with any recommendation, the problem is that the person might not follow the advice. This is captured here by the *failure of the Revelation Principle*. Every information policy admits a corresponding recommendation policy whereby the agent is simply told how he would have optimally responded to his posterior. What are the incentive and welfare consequences of this new policy? In the classical world, the Revelation Principle (as in Myerson 1979, Aumann 1987, and Kamenica and Gentzkow 2011) says that, for any signal, the corresponding recommendation policy is incentive-compatible and welfare-equivalent.[14]

Under psychological information aversion, and a fortiori in general psychological environments, this is no longer true. For example, take $\Theta = A = \{0,1\}$ and

$$u(a,\nu) = a\left[\operatorname{var}_{\theta\sim\nu}(\theta) - \tfrac{1}{9}\right] = a\left[\nu(1-\nu) - \tfrac{1}{9}\right],$$

so that the agent is psychologically information-averse. Let $\mu = 0.6$ and consider a signal that generates information policy $p = (3/8)\delta_{0.1} + (5/8)\delta_{0.9}$. As shown in Figure 1, whether the realized posterior belief is 0.1 or 0.9, the agent plays action 0. If the designer were to replace each message with a recommendation of the action the agent would have played, then she would be recommending action $a = 0$ with probability 1, independent of the state. This would convey no information whatsoever to the agent, and so his posterior belief would equal his prior belief (0.6). He would then optimally play action 1: the recommended action, $a = 0$, would not be incentive-compatible. At a technical level, the Revelation Principle fails under psychological preferences because the set of posterior beliefs for which a given action is optimal might not be convex, and so "pooling" messages from such a region may render the action no longer optimal.

Recommendation policies are practical, but they cannot generate all joint distributions over actions and beliefs that posterior policies can attain, given the failure of the Revelation Principle. Even so, our next theorem says that this class is always optimal for psychologically information-averse agents.

[14] In its standard form, the Revelation Principle (Myerson 1991, 260) says that (given principal commitment power) all attainable outcomes can be attained by having players report all private information to the principal, while the principal simply makes incentive-compatible action recommendations to players. In the present context, with the agent having no private information, we focus on failure of the latter.

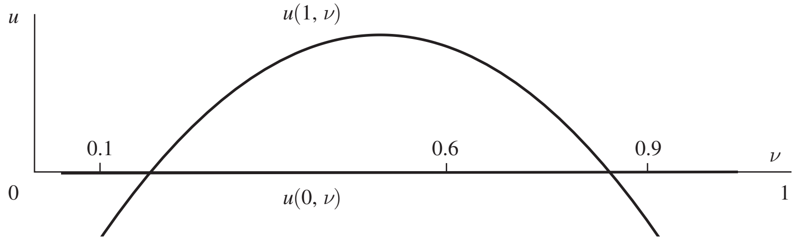

Figure 1. The Set $\{\nu : a = 0 \text{ is optimal at } \nu\}$ Is Not Convex

THEOREM 2 (Optimality of Advice): *If the agent is psychologically information-averse, then there exists an incentive-compatible recommendation policy which is optimal.*

A recommendation is the coarsest information policy that leads to a given state-action mapping. If an information-averse agent complies with the recommendation, then giving him only instrumental information is, of course, beneficial. The nontrivial content of the theorem, then, is that such a coarse policy can be used without disrupting the agent's incentives (i.e., retaining incentive-compatibility).

Although Theorem 2 tells us that giving advice is best in a broad set of circumstances, this reasoning has limits. The result no longer holds if the agent likes information for its own sake, or if the expert and agent have conflicting interests. To see the former, consider $A = \{\bar{a}\}$, $\Theta = \{0, 1\}$, and $u(\bar{a}, \nu) = -\text{var}(\nu)$. The sole recommendation policy reveals nothing, and so is suboptimal for this information-loving agent. To see the latter, consider the above agent of Figure 1 (at prior $\mu = 0.6$) paired with a principal who only wants the agent to choose action 0. Any incentive-compatible recommendation policy must sometimes recommend action 1, but full information would persuade the agent to always choose 0.

VI. Conclusion

This paper proposes a method for computing concave envelopes (hence, optimal policies) and gives a natural domain of application for these methods in the world of psychological preferences. This domain includes a broad range of behavioral models for which concrete prescriptions become possible. Information disclosure to a tempted agent is of particular interest.

In a recent note, Lipnowski and Mathevet (2017), we have adapted and specialized our method of posterior covers to standard Bayesian persuasion, where the principal and the agent are expected utility maximizers but have conflicting preferences.[15] In contrast to the classical world, psychological phenomena make disclosure a subtle issue, even without conflicting interests. Although the assumption of aligned preferences is well-suited to many situations, conflicting preferences are

[15] We thank an anonymous referee for encouraging us to pursue this direction.

also reasonable in a psychological world. The first and third parts of the method of posterior covers still apply to conflicting preferences, but the second part is (in its current form) special to benevolence. In the expected utility world, the principal is information-neutral over beliefs at which the agent's optimal action does not change, making the second part again applicable. Extending the tools herein, to study misaligned preferences with psychological concerns, is an exciting avenue for future work.

Appendix A. A List of Models and Their Optimal Policies

Below is a list of models, accommodated by our methods, followed by descriptions of their respective optimal policies:

(**a**) (**Purely psychological agent**): Defined in the main text.

(**b**) (**Stubbornness/prior-bias**): Defined in the main text.

(**c**) (**Temptation and self-control**): Defined in the main text.

(**d**) (**Ego utility**): Given a classical utility $\tilde{u}$ and a psychological utility u_E (ego utility), let

$$u(a,\theta,\nu) \ = \ u_C(a,\theta) + u_E(\nu).$$

This is a version of Kőszegi (2006).[16] The agent has poor or great ability, $\theta \in \{0,1\}$. He does not know θ, but has a prior notion μ of how good he is. He must choose a task σ that will reveal any desired amount of information about his ability and then choose $a \in A$. From these choices, he gets the expected value of u_C, but also derives "ego utility" u_E from his belief ν about his ability. Following Kőszegi (2006, 679), "a step-function ego utility captures the qualitative features of such preferences," so, for example, assume $u_E(\nu)$ is proportional to $\mathbf{1}_{\nu \geq \bar{e}}$, where $\bar{e} \in (0,1)$ is the minimum belief that satisfies the agent's ego:

(**e**) (**Shame**): Given a finite set A, two classical utility functions u_S (selfish) and u_N (normative), and a function g increasing in its second argument,[17] let

$$u(a,\theta,\nu) \ = \ u_S(a,\theta) - g\Big(a, \max_{b \in A} \mathbb{E}_{\hat{\theta}\sim\nu}\big[u_N(b,\hat{\theta})\big]\Big),$$

following Dillenberger and Sadowski (2012). The agent chooses an action a that affects both himself and some other recipient. Think of deciding how much to tip at a restaurant. The utility u_S captures the agent's selfish enjoyment of a, while u_N is

[16] Also see Dal Bó and Terviö (2013) for an application of the same to an agent's accrued "moral capital."

[17] For brevity, we have omitted restrictions on A, u_S, u_N, g implied by Dillenberger and Sadowski's (2012) representation theorem.

his normative utility. The function g represents the shame from choosing a instead of the normatively right action. Of particular interest to us is the case of diminishing marginal shame, with g concave in its second argument.

(**f**) (**Reference-dependent choice/Loss aversion**): Given a classical utility u_C, let

$$u(a,\theta,\nu) = u_C(a,\theta) + g\Big(\mathbb{E}_{\hat{\theta}\sim\nu}\big[u_C(a,\hat{\theta})\big] - \mathbb{E}_{\hat{\theta}\sim\mu}\big[u_C(a,\hat{\theta})\big]\Big),$$

where $g : \mathbb{R} \to \mathbb{R}$. The agent evaluates an action a not only based on its expected utility, but also based on its "gain-loss" utility, the second component of u. This model is a two-period version of that of Kőszegi and Rabin (2009),[18] also related to Kahneman and Tversky (1979), according to which people evaluate alternatives from a reference point. The reference point is what the chosen action would yield absent any information, $\mathbb{E}_{\hat{\theta}\sim\mu}\big[u_C(a,\hat{\theta})\big]$. Then, if $\mathbb{E}_{\hat{\theta}\sim\nu}\big[u_C(a,\hat{\theta})\big] > \mathbb{E}_{\hat{\theta}\sim\mu}\big[u_C(a,\hat{\theta})\big]$, the agent is positively surprised by the information. Kőszegi and Rabin (2009) focus especially on the case where $g(x)$ is proportional to $(1 + \epsilon\mathbf{1}_{x<0})x$ for $\epsilon > 0$, so that relative losses loom larger than gains:[19]

(**g**) (**Belief distortion**): For one-period utility $\tilde{u}(a_t) = -\frac{1}{2}a_t^2$ and discount factor $\delta \in [0,1]$, let

$$u(a,\theta,\nu) = \mathbb{E}_{\hat{\theta}\sim\nu^*(\nu)}\Big[\tilde{u}(a) + \delta\tilde{u}\big(1+\hat{\theta}-a\big)\Big],$$

where $\nu^* : [0,1] \to [0,1]$ is weakly increasing and piecewise linear, with

$$\nu^*(\nu) = \begin{cases} 0 & \text{if } \nu \leq \underline{\nu} \\ \alpha(\nu - \underline{\nu}) & \text{if } \underline{\nu} \leq \nu \leq \bar{\nu}. \\ \alpha(\bar{\nu} - \underline{\nu}) & \text{if } \bar{\nu} \leq \nu \end{cases}$$

This model is the planning model of Brunnermeier, Papakonstantinou, and Parker (2016) in which an agent has two periods to complete a task whose difficulty (i.e., required total effort) $1 + \theta$ is unknown until the second period. The agent chooses his effort a_t in period t such that $a \coloneqq a_1 \leq 1$ and $a_1 + a_2 = 1 + \theta$. If all "objective" information would leave our agent with beliefs ν, our agent optimally distorts his beliefs (à la Brunnermeier and Parker 2005), which yields the functional form $\nu^*(\nu)$.[20] When ν^* is also optimistic, i.e., $\nu^*(\nu) \leq \nu$, the agent is prone to the so-called planning fallacy, hence he procrastinates.

[18] Information can exacerbate loss aversion because we take ν-expectations inside the argument of g (if ν-expectation were taken outside, $u(a,\nu)$ would be linear in ν and full information immediately optimal). This functional form is different from the baseline model of Kőszegi and Rabin (2007) and Kőszegi and Rabin (2009), but these authors allude to the "alternative [model in which] the decision maker compares the means of her new and old beliefs" and mention that most of their findings apply to it.

[19] See Assumption A3′, section 2, and footnote 10 in Kőszegi and Rabin (2009).

[20] Brunnermeier, Papakonstantinou, and Parker (2016) provide explicit formulas for $\alpha, \underline{\nu}$, and $\bar{\nu}$. Crucially, the optimal belief distortion does not depend on the entire distribution of posterior expectations, but only on the realized posterior expectation itself.

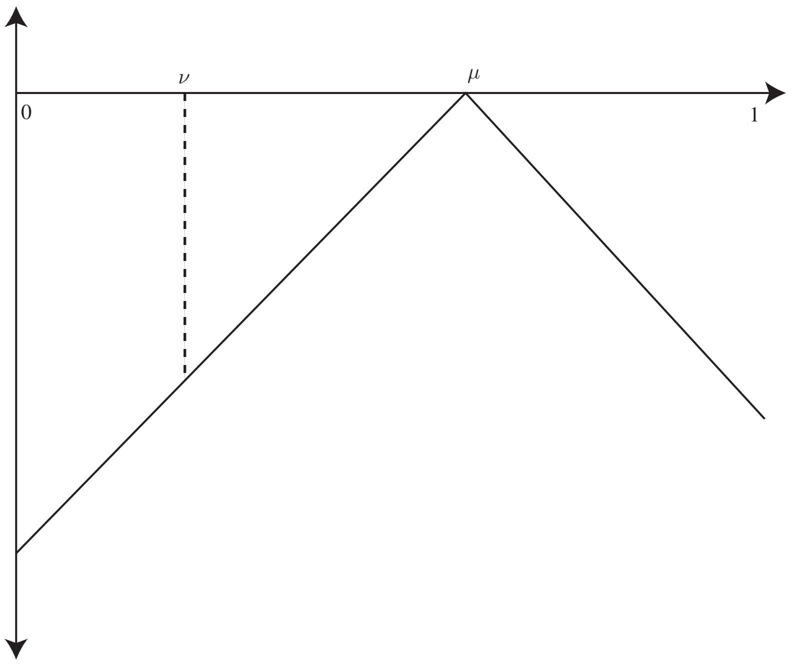

FIGURE 2. FUNCTION $u_P(\nu) = -|\nu - \mu|$ (*prior-bias*) IS AFFINE ON $[0,\mu]$ AND $[\mu,1]$

Below are the corresponding prescriptions from the method of posterior covers:

(**a**) Full information is trivially optimal.

(**b**) Either full or no information is optimal. Indeed, since $u_P(\nu) = -\rho|\nu - \mu|$ is affine in ν on each member of $\mathcal{C} := \{[0,\mu],[\mu,1]\}$, $\mathcal{C}$ is a u_P-cover (see Figure 2). By Theorem 1, it is also a U-cover, and, thus, some optimal policy p^* is supported on $out(\mathcal{C}) = \{0,\mu,1\}$. Bayes-consistency then requires that $p^* = (1-\lambda)\delta_\mu + \lambda[(1-\mu)\delta_0 + \mu\delta_1]$ for some $\lambda \in [0,1]$. That is, we should give the agent no information or full information.

(**c**) See Proposition 2.

(**d**) It is optimal to either give the agent full information or keep his posterior belief at $\bar{e}$ with the largest feasible probability. To see why, let $u_E(\nu) = \mathbf{1}_{\nu \geq \bar{e}}$, as suggested above, and focus on the case of $\mu \geq \bar{e}$, so that an ignorant agent is confident. There, $u_E(\nu)$ is weakly convex in ν on each member of $\mathcal{C} := \{[0,\bar{e}],[\bar{e},1]\}$, so that $\mathcal{C}$ is a u_E-cover and also a U-cover by the theorem. Thus, some optimal policy is supported on $out(\mathcal{C}) = \{0,\bar{e},1\}$. The agent then faces the following trade-off: either making a good decision matters most, in which case he should seek full information, or he should limit his research to protect his ego (so that his posterior belief is either $\bar{e}$ or 1).

(**e**) Under diminishing marginal shame, it is optimal to keep the agent in moral ambiguity. Given the functional form, the result is formally related to Proposition 2. The agent may like information to inform his decision-making, but in a way that will always leave him morally indifferent to the course of action.[21]

(**f**) Either full or no information is optimal. Since $u_P(a,\nu)$ is affine in ν on each member of $\mathcal{C} := \{[0,\mu],[\mu,1]\}$ for all a, Theorem 1 says $\mathcal{C}$ is a u_P-cover for all a and, hence, a U-cover. The argument follows as in (b).

(**g**) By definition of ν^*, $u(a,\theta,\nu)$ is convex (at least, weakly) in ν on each member of $\mathcal{C} := \{[0,\underline{\nu}],[\underline{\nu},\bar{\nu}],[\bar{\nu},1]\}$ for all a, hence, $\mathcal{C}$ is a U-cover. Thus, $out(\mathcal{C}) = \{0,\underline{\nu},\bar{\nu},1\}$. Since ν^* is weakly increasing and bounded above by $\alpha(\bar{\nu}-\underline{\nu})$, an optimal p is supported on $\{0,\underline{\nu},\bar{\nu}\}$.

Appendix B. Proofs

Throughout the rest of the Appendix, we let $u : A \times \Delta\Theta \to \mathbb{R}$ refer only to the reduced preferences of the agent.

We begin by stating a well-known connection between our "mean-preserving spread" definition of informativeness and the usual garbling definition.

LEMMA 1: *Given two information policies $p,q \in \mathcal{R}(\mu)$, the ranking $p \succeq_B^\mu q$ holds if and only if (S_q,σ_q) is a garbling of (S_p,σ_p), i.e., if there is a map $g : S_p \to \Delta(S_q)$ such that*

$$\sigma_q(\hat{S}|\theta) = \int_{S_p} g(\hat{S}|\cdot)\, d\sigma_p(\cdot|\theta) \tag{3}$$

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

A. *Proof of Proposition 1*

LEMMA 2: *The set $M := \{\gamma \in \Delta\Theta : \exists\epsilon > 0$ such that $\epsilon\gamma \leq \mu\}$ is w^*-dense in $\Delta\Theta$.*

PROOF:

First, notice that $M = \{\gamma \in \Delta\Theta : \gamma \ll \mu$ and $d\gamma/d\mu$ is (essentially) bounded$\}$ is convex and extreme (i.e., a face of $\Delta\Theta$). Thus, its w^*-closure $\bar{M}$ is closed (and so compact, by the Banach-Alaoglu theorem (Aliprantis and Border 1999, Theorem 6.21)), 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 the Krein-Milman theorem (Aliprantis and Border 1999, Theorem 7.68), $\bar{M} = \overline{co}E = \Delta(\Theta')$, 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. ∎

[21] That is, he should be left at posterior beliefs $\nu \in \Delta\Theta$ such that, for all $\nu' \neq \nu$ supported by some $q \in \mathcal{R}(\nu)$, $\arg\max_b \mathbb{E}_{\theta\sim\nu'}[u_N(b,\theta)] \not\supseteq \arg\max_b \mathbb{E}_{\theta\sim\nu}[u_N(b,\theta)]$.

LEMMA 3: *Fix a bounded measurable function* $f : \Delta\Theta \to \mathbb{R}$, *and suppose either that* Θ *is finite or that* f *is continuous. Then, the following are equivalent* (*given* μ *is of full support*):

(*i*) *For all* $\nu \in \Delta\Theta$ *and* $p \in \mathcal{R}(\nu)$, *we have* $\int_{\Delta\Theta} f dp \geq f(\nu)$

(*ii*) *For all* $\mu' \in \Delta\Theta$ *and* $p, q \in \mathcal{R}(\mu')$ *with* $p \succeq_B^{\mu'} q$, *we have* $\int_{\Delta\Theta} f dp \geq \int_{\Delta\Theta} f dq$.

(*iii*) *For all* $p, q \in \mathcal{R}(\mu)$ *with* $p \succeq_B^{\mu} q$, *we have* $\int_{\Delta\Theta} f dp \geq \int_{\Delta\Theta} f dq$.

(*iv*) f *is convex.*

PROOF:

Suppose (i) 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 $\nu \in S_q$, (i) implies $\int_{S_q}\int_{\Delta\Theta} f dr(\cdot|\nu)\, dq(\nu) \geq \int_{S_q} f dq$. Equivalently (by definition of the Blackwell order), any p more informative than q has $\int f dp \geq \int f dq$, which yields (ii).

That (ii) implies (iii) is immediate.

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

$$\gamma = (1-\lambda)\zeta + \lambda\eta;$$

$$f(\gamma) < (1-\lambda) f(\zeta) + \lambda f(\eta).$$

Moreover, we can always pick $\gamma, \zeta, \eta \in \Delta\Theta$, and $\lambda \in (0,1)$ to ensure that $\epsilon\gamma \leq \mu$ for some $\epsilon \in (0,1)$. Indeed, Lemma 2 guarantees this if f is continuous, and it comes for free—we can simply let $\epsilon := \min_{\theta\in\Theta} \mu(\theta)$—if Θ is finite.

Now, we can exploit the above to construct two information-ranked information policies such that f has higher expectation on the less informative of the two. Let

$$\nu := \frac{1}{1-\epsilon}(\mu - \epsilon\gamma) \in \Delta\Theta,$$

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

$$q := (1-\epsilon)\,\delta_\nu + \epsilon\,\delta_\gamma \in \mathcal{R}(\mu).$$

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

$$\int_{\Delta\Theta} f dp - \int_{\Delta\Theta} f dq = \epsilon\big[(1-\lambda) f(\zeta) + \lambda f(\eta) - f(\gamma)\big] < 0,$$

as desired.

Finally, notice that (iv) implies (i) by Jensen's inequality. ∎

PROOF OF PROPOSITION 1:

The first part follows immediately from applying Lemma 3 to $u(a,\cdot)$ and $-u(a,\cdot)$ for each $a \in A$, and to U and $-U$. That a psychologically information-loving agent is behaviorally information-loving follows from the first part, and from the easy fact that a pointwise maximum of convex functions is convex.

Lastly, suppose u is concave. It is upper-semicontinuous by hypothesis and lower semicontinuous by Rockafellar (1970, Theorem 10.2), thus continuous. The convex maximum theorem then applies, so that U is concave too. Then, the first part yields behavioral information aversion. ∎

B. *Proof of Theorem 1*

Proof of the First Part.—First, notice that, for any $C \in \mathcal{C}$, the extreme points of C are a Borel set. Therefore, $out(\mathcal{C}) = \Delta\Theta \backslash \bigcup_{C\in\mathcal{C}} [C \backslash ext(C)]$ is Borel, and $p(out(\mathcal{C}))$ is in fact well-defined for $p \in \mathcal{R}(\mu)$.

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

$$p(out(\mathcal{C})) = 1 - p\Big(\bigcup_{C\in\mathcal{C}} [C\backslash ext(C)]\Big) = 1. \blacksquare$$

Proof of the Second Part.—As a sum or supremum of convex functions is convex, the following are immediate:

(i) Suppose f is the pointwise supremum of a finite 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.

(ii) If $\mathcal{C}$ is a g-cover and h is convex, then $\mathcal{C}$ is a $(g+h)$-cover.

From there, we need only notice that $U = \sup_{a\in A} u(a,\cdot)$, and that in the separable case $U = u_P + U_C$ for convex U_C. ∎

Proof of the Third Part.—By finiteness of I, the collection $\mathcal{C}$ 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 and convex. Restricted to C_i, f agrees with f_i and so is affine, and therefore $\phi \circ f$ is convex. ∎

C. *Proof of Proposition 2*

First, under the hypotheses of the third part of Theorem 1, we prove a claim that fully characterizes the set of outer points of the *f*-cover, reducing their computation to linear algebra.

LEMMA 4: *Let Θ be finite and let f, $\mathcal{C} = \{C_i\}_{i\in I}$ be as given in the third part of Theorem* 1. *Then, the f-cover $\mathcal{C}$ satisfies $out(\mathcal{C}) = \{\nu^* \in \Delta\Theta : \{\nu^*\} = S(\nu^*)\}$, where*

$$S(\nu^*) \coloneqq \left\{\nu \in \Delta\Theta : \mathrm{supp}(\nu) \subseteq \mathrm{supp}(\nu^*) \text{ and } f_j(\nu) = f(\nu)\ \forall j \in \arg\min_{i\in I} f_i(\nu^*)\right\}$$
$$\subseteq \left\{\nu \in \Delta\Theta : \mathrm{supp}(\nu) \subseteq \mathrm{supp}(\nu^*) \text{ and } f_j(\nu) = f_k(\nu)\ \forall j,k \in \arg\min_{i\in I} f_i(\nu^*)\right\}.$$

PROOF:

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

Let us begin by supposing $\{\nu^*\} \neq S(\nu^*)$; we have to show $\nu^* \notin out(\mathcal{C})$. 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 $\nu^* \notin out(\mathcal{C})$. Let $\hat{\Theta}$ be the support of ν^*, and let $J \coloneqq \arg\min_{i\in I} f_i(\nu^*)$. 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 $\epsilon(\nu - \nu^*) \leq \nu^*$.[22] Define $\nu' \coloneqq \nu^* - \epsilon(\nu - \nu^*) \in \Delta\hat{\Theta}$. Then, $f_i(\nu') = f_j(\nu')\ \forall i,j \in J$ too (by affinity) and, by definition of ν', we have $\nu^* \in co\{\nu,\nu'\}$. If $i \notin J$, then $f_i(\nu^*) > f(\nu)$ by definition. 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 $out(\mathcal{C})$.

To complete the proof, let us suppose that $\nu^* \notin out(\mathcal{C})$, 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 $\mathrm{supp}(\nu) \subseteq \mathrm{supp}(\nu^*)$ 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. ∎

In passing, an immediate consequence of the above Lemma is the following simple reduction in the two-state world.

COROLLARY 1: *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*

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

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

PROOF OF PROPOSITION 2:

For each $a \in A$, let $C_a := \{\nu \in \Delta\Theta : a \in \arg\max_b \mathbb{E}_{\theta\sim\nu}[u_T(b,\theta)]\}$. By Theorem 1, $\mathcal{C} := \{C_a\}_{a\in A}$ is a U-cover, and so some optimal policy p is supported on $out(\mathcal{C})$. By Lemma 4, we then know $S(\nu^*) = \{\nu^*\}$—where $S(\nu^*)$ is as in the statement of Lemma 4 (with $I := A$ and $f_a(\nu) := -\mathbb{E}_{\theta\sim\nu}u_T(a,\theta)$)—for every $\nu^* \in S_p$. It remains to show that, at each $\nu^* \in S_p$, the agent does not know what he is missing. To that end, consider any $q \in \mathcal{R}(\nu^*)$, $\nu \in S_q \backslash \{\nu^*\}$. Because $q \in \mathcal{R}(\nu^*)$, it must be that $\text{supp}(\nu) \subseteq \text{supp}(\nu^*)$. Therefore, $\nu \notin S(\nu^*)$ means it cannot be that $f_a(\nu) = \arg\max_{b\in A} f_b(\nu)\ \ \forall a \in \arg\max_{b\in A} f_b(\nu^*)$. That is, $\arg\max_b \mathbb{E}_{\theta\sim\nu^*}[u_T(b,\theta)] \not\subseteq \arg\max_b \mathbb{E}_{\theta\sim\nu}[u_T(b,\theta)]$. So the agent does not know what he is missing at ν^*.

D. *Proof of Theorem 2*

PROOF:

Suppose the agent is psychologically information-averse.

Fix some measurable selection[23] $a^* : \Delta\Theta \to A$ 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 1: *Given any information policy* $p \in \mathcal{R}(\mu)$, *we can construct a signal* (A, α_p) *such that:*

(*i*) *The information policy* q_p *induced by* (A, α_p) *is less informative than* p.

(*ii*) *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.,*

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

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

$$\alpha_p : \Theta \to \Delta(A)$$
$$\theta \mapsto \alpha_p(\cdot|\theta) = \sigma_p(\cdot|\theta) \circ a^{*-1}.$$

Then, (A, α_p) is a signal with

$$\alpha_p(\hat{A}|\theta) = \sigma_p\big(\{s \in S_p : a^*(s) \in \hat{A}\}\big|\theta\big),$$

[23] One exists, by Aliprantis and Border (1999, Theorem 8.13).

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 information policy 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,[24] q is a garbling of p. Therefore, by Lemma 1, $p \succeq_B^\mu q$, so that there is a map $r: S_q \to \Delta(S_p)$ such that for every Borel $S \subseteq \Delta\Theta$, $p(S) = \int_{S_q} r(S|\cdot)\, dq$, and for every $t \in S_q$, $r(\cdot|t) \in \mathcal{R}(t)$. Then, appealing to the definition of psychological information aversion,

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

Therefore,

$$\begin{aligned}
&\int_\Theta\int_A u\left(a, \beta^{A,\alpha_p}(\cdot|a)\right) d\alpha_p(a|\theta)\, d\mu(\theta) \\
&\quad\geq \int_\Theta\int_A\int_{S_p} U(\nu)\, dr\left(\nu|\beta^{A,\alpha_p}(\cdot|a)\right) d\alpha_p(a|\theta)\, d\mu(\theta) \\
&\quad= \int_{S_q} U(\nu)\, dr(\nu|t)\, dq(t) \\
&\quad= \int_{S_p} U\, dp, \quad \text{which verifies the claim.}
\end{aligned}$$

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

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

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

Then,[25]

$$\int_{\Delta\Theta} U\, dp \leq \int_{A\times\Delta\Theta} u\, dQ \leq \int_{\Delta\Theta} U\, dq \leq \int_{\Delta\Theta} U\, dp,$$

so that:

$$\int_{\Delta\Theta} U\, dq = \int_{\Delta\Theta} U\, dp, \quad \text{i.e., } q \text{ is optimal;}$$

[24] Indeed, we can define g in (3) via: $g(a|s) = 1$ if $a^*(s) = a$ and 0 otherwise.

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

and

$$\int_{A\times\Delta\Theta} u(a,\nu)\, dQ(a,\nu) \;=\; \int_{\Delta\Theta} U\, dq \;=\; \int_{A\times\Delta\Theta} \max_{\tilde{a}\in A} u(\tilde{a},\nu)\, dQ(a,\nu).$$

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

We note that the claim in the above proof delivers something more than the statement of Theorem 2. Indeed, given any finite-support information policy 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):

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

(ii) 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.

REFERENCES

Aliprantis, Charalambos D., and Kim Border. 1999. *Infinite Dimensional Analysis*. Berlin: Springer.

Anastassiou, G. A. 2006. "Applications of geometric moment theory related to optimal portfolio management." *Computers and Mathematics with Applications* 51 (9–10): 1405–30.

Aumann, Robert J. 1987. "Correlated Equilibrium as an Expression of Bayesian Rationality." *Econometrica* 55 (1): 1–18.

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

Baumeister, Roy F. 2002. "Yielding to Temptation: Self-Control Failure, Impulsive Purchasing, and Consumer Behavior." *Journal of Consumer Research* 28 (4): 670–76.

Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117 (3): 871–915.

Benoît, Jean-Pierre, and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica* 79 (5): 1591–1625.

Ben-Shahar, Omri, and Carl E. Schneider. 2011. *The Failure of Mandated Disclosure: More Than You Wanted to Know*. Princeton: Princeton University Press.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *Annals of Mathematics and Statistics* 24 (2): 265–72.

Brocas, Isabelle, and Juan D. Carrillo. 2007. "Influence through ignorance." *RAND Journal of Economics* 38 (4): 931–47.

Brunnermeier, Marcus K., Filippos Papakonstantinou, and Jonathan A. Parker. 2016. "Optimal Time-inconsistent Beliefs: Misplanning, Procrastination, and Commitment." *Management Science* 63 (5): 1318–40.

Brunnermeier, Markus K., and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118.

Caplin, Andrew, and Kfir Eliaz. 2003. "AIDS Policy and Psychology: A Mechanism-Design Approach." *RAND Journal of Economics* 34 (4): 631–46.

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

Carrillo, Juan D., and Thomas Mariotti. 2000. "Strategic Ignorance as a Self-Disciplining Device." *Review of Economic Studies* 67 (3): 529–44.

Dal Bó, Ernesto, and Marko Terviö. 2013. "Self-Esteem, Moral Capital, and Wrongdoing." *Journal of the European Economic Association* 11 (3): 599–663.

Dillenberger, David, and Philipp Sadowski. 2012. "Ashamed to be selfish." *Theoretical Economics* 7 (1): 99–124.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica. 2015. "Suspense and Surprise." *Journal of Political Economy* 123 (1): 215–60.

Frankl, David, Robert K. Oye, and Paule E. Bellamy. 1989. "Attitudes of hospitalized patients toward life support: A survey of 200 medical inpatients." *American Journal of Medicine* 86 (6): 645–48.

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

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

Gentzkow, Matthew, and Emir Kamenica. 2016. "A Rothschild-Stiglitz Approach to Bayesian Persuasion." *American Economic Review* 106 (5): 597–601.

Grant, Simon, Atsushi Kajii, and Ben Polak. 1998. "Intrinsic Preference for Information." *Journal of Economic Theory* 83 (2): 233–59.

Gruber, Jonathan, and Botond Kőszegi. 2001. "Is Addiction 'Rational'? Theory And Evidence." *Quarterly Journal of Economics* 116 (4): 1261–1303.

Gruber, Jonathan, and Botond Kőszegi. 2004. "Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes." *Journal of Public Economics* 88 (9–10): 1959–87.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica* 69 (6): 1403–35.

Hörner, Johannes, and Andrzej Skrzypacz. 2016. "Selling Information." *Journal of Political Economy* 124 (6): 1515–62.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–92.

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

Kemperman, J. H. B. 1968. "The General Moment Problem: A Geometric Approach." *Annals of Mathematical Statistics* 39 (1): 93–122.

Kőszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review* 97 (4): 1047–73.

Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review* 99 (3): 909–36.

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

Krusell, Per, Burhanettin Kuruşçu, and Anthony A. Smith, Jr. 2010. "Temptation and Taxation." *Econometrica* 78 (6): 2063–84.

Lipnowski, Elliot, and Laurent Mathevet. 2017. "Simplifying Bayesian Persuasion." http://www.laurentmathevet.com/wp-content/uploads/2017/03/BPCover.pdf.

Loewenstein, George, Cass R. Sunstein, and Russell Golman. 2014. "Disclosure: Psychology Changes Everything." *Annual Review of Economics* 6 (1): 391–419.

Myerson, Roger B. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47 (1): 61–73.

Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.

O'Donoghue, Ted, and Matthew Rabin. 2003. "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes." *American Economic Review* 93 (2): 186–91.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review* 103 (2): 804–30.

Ostrovsky, Michael, and Michael Schwarz. 2010. "Information Disclosure and Unraveling in Matching Markets." *American Economic Journal: Microeconomics* 2 (2): 34–63.

Otto, A. Ross, and Bradley C. Love. 2010. "You don't want to know what you're missing: When information about forgone rewards impedes dynamic decision making." *Judgment and Decision Making* 5 (1): 1–10.

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

Rayo, Luis, and Ilya Segal. 2010. "Optimal Information Disclosure." *Journal of Political Economy* 118 (5): 949–87.

Rockafellar, R. Tyrrell. 1970. *Convex Analysis*. Princeton: Princeton University Press.

Schweizer, Nikolaus, and Nora Szech. 2012. "Optimal Revelation of Life-Changing Information." https://web.stanford.edu/group/SITE/archive/SITE_2012/2012_segment_6/2012_segment_6_papers/szech.pdf.

Tardella, Fabio. 2003. "On the existence of polyhedral convex envelopes." In *Frontiers In Global Optimization*, edited by C. A. Floudas and P. M. Pardalos, 563–73. Berlin: Springer.

Ubel, Peter A. 2001. "Truth in the Most Optimistic Way." *Annals of Internal Medicine* 134 (12): 1142–43.