Bayesian Persuasion

Emir Kamenica and Matthew Gentzkow*
University of Chicago

September 2009

Abstract

When is it possible for one person to persuade another to change her action? We take a mechanism design approach to this question. Taking preferences and initial beliefs as given, we introduce the notion of a *persuasion mechanism*: a game between Sender and Receiver defined by an information structure and a message technology. We derive necessary and sufficient conditions for the existence of a persuasion mechanism that strictly benefits Sender. We characterize the optimal mechanism. Finally, we analyze several examples that illustrate the applicability of our results.

JEL classification: *D83, K41, L15, M37*
Keywords: strategic communication, disclosure, signalling

*We thank Richard Holden for many important contributions to this paper. We would also like to thank Eric Budish, Navin Kartik, Canice Prendergast, Maxwell Stinchcombe, Lars Stole and participants at seminars at University of Mannheim, Duke/Northwestern/Texas IO Theory Conference, Stanford GSB, Simon Fraser University, University of British Columbia, and University of Chicago. This work is supported by the Initiative on Global Markets, the George J. Stigler Center for the Study of the Economy and the State, the James S. Kemper Foundation Faculty Research Fund, the Centel Foundation / Robert P. Reuss Faculty Research Fund, and the Neubauer Family Foundation, all at the University of Chicago Booth School of Business. E-mail: emir.kamenica@chicagoBooth.edu; matthew.gentzkow@chicagoBooth.edu.

1 Introduction

Suppose one person, call him Sender, wishes to persuade another, call her Receiver, to change her action. If Receiver is a rational Bayesian, can Sender persuade her to take an action *he* would prefer over the action she was originally going to take? If Receiver understands that Sender chose what information to convey with the intent of manipulating her action for his own benefit, can Sender still gain from persuasion? If so, what is the optimal way to persuade?

These questions are of substantial economic importance. As McCloskey and Klamer (1995) emphasize, attempts at persuasion command a sizeable share of our resources. Persuasion, as we will define it below, plays an important role in advertising, courts, lobbying, financial disclosure, and political campaigns, among many other economic activities.

Consider the example of a prosecutor trying to convince a judge that a defendant is guilty. When the defendant is indeed guilty, revealing the facts of the case will tend to help the prosecutor's case. When the defendant is innocent, revealing facts will tend to hurt the prosecutor's case. Can the prosecutor structure his arguments, selection of evidence, *etc.* so as to increase the probability of conviction by a rational judge *on average*? Perhaps surprisingly, the answer to this question is yes. Bayes' Law restricts the expectation of posterior beliefs but puts no other constraints on their distribution. Therefore, so long as the judge's action is not linear in her beliefs, the prosecutor may benefit from persuasion.

To make this concrete, suppose the judge (Receiver) must choose one of two actions: to *acquit* or *convict* a defendant. There are two states of the world: the defendant is either *guilty* or *innocent*. The judge gets utility 1 for choosing the just action (convict when guilty and acquit when innocent) and utility 0 for choosing the unjust action (convict when innocent and acquit when guilty). The prosecutor (Sender) gets utility 1 if the judge convicts and utility 0 if the judge acquits, regardless of the state. The prosecutor and the judge share a prior belief $\Pr(guilty) = 0.3$.

The prosecutor conducts an investigation and is required by law to report its full outcome. We can think of the choice of the investigation as consisting of the decisions on whom to subpoena, what forensic tests to conduct, what question to ask an expert witness, *etc.* We formalize an investigation as distributions $\pi(\cdot|guilty)$ and $\pi(\cdot|innocent)$ on some set of signal realizations. The prosecutor chooses π and must honestly report the signal realization to the judge. Importantly, we assume

that the prosecutor can choose any π whatsoever, i.e., that the space of possible investigations is arbitrarily rich.

If there is no communication (or equivalently, if π is completely uninformative), the judge always acquits because guilt is less likely than innocence under her prior. If the prosecutor chooses a fully informative investigation, one that leaves no uncertainty about the state, the judge convicts 30 percent of the time. The prosecutor can do better, however. His uniquely optimal investigation is a binary signal

$$\begin{aligned} \pi\left(i|innocent\right) &= \tfrac{4}{7} \quad \pi\left(i|guilty\right) = 0 \\ \pi\left(g|innocent\right) &= \tfrac{3}{7} \quad \pi\left(g|guilty\right) = 1. \end{aligned} \tag{1}$$

This leads the judge to convict with probability 60 percent. Note that the judge knows 70 percent of defendants are innocent, yet she convicts 60 percent of them! She does so even though she is fully aware that the investigation was designed to maximize the probability of conviction.

In this paper, we study the general problem of persuading a rational agent. Our approach follows the literature on mechanism design. We consider a setting with an arbitrary state space and action space, and with arbitrary state-dependent preferences for both Sender and Receiver. We introduce a broad class of "persuasion mechanisms" that encompasses cheap talk games (e.g., Crawford and Sobel 1982), persuasion games (e.g., Milgrom and Roberts 1986), and signalling games (e.g., Spence 1973), among many others. The key distinguishing feature of a persuasion mechanism is that Sender can affect Receiver's action only by changing Receiver's *beliefs*. We do not allow Sender to make transfers or affect Receiver's *payoffs* in any way. In contrast to most other papers on strategic communication, we allow for mechanisms where Sender can fully commit on two counts: to fully disclose all he knows *and* to limit the extent of his private information. Given this definition, we focus on two questions: (i) when does there exist a persuasion mechanism that strictly benefits Sender, and (ii) what is an optimal mechanism from Sender's perspective?

We begin by establishing some results that simplify our analysis. We show that, without loss of generality, we can restrict attention to mechanisms where Sender learns a recommended action for Receiver, reports it truthfully, and then Receiver chooses the recommended action. In the example above, we can think of i as a recommendation to *acquit* and g as a recommendation to *convict*. We then show that we can re-express the problem of choosing such a mechanism as a

search over distributions of posteriors subject to the constraint that the expected posterior is equal to the prior.

When does there exist a persuasion mechanism that strictly benefits Sender? Consider why the prosecutor in the example benefits from the opportunity to provide information to the judge. Since the judge is rational, providing information must sometimes make her more convinced and sometimes less convinced that the defendant is guilty. The former will strictly improve the prosecutor's payoff if the information is strong enough to induce conviction. The latter, however, will not reduce the prosecutor's payoff, since the judge already acquits the defendant by default. The net effect is to increase the prosecutor's payoff in expectation. We show that in general Sender benefits from persuasion whenever (i) Receiver does not take Sender's preferred action by default (in a sense we make precise below) and (ii) Receiver's action is constant in some neighborhood of beliefs around the prior. When these conditions hold, Sender can benefit by sending a signal that induces a better action with positive probability and balances this with a worse belief that leaves Receiver's action unchanged. We also show that whether Sender benefits from persuasion depends in a natural way on the concavity or convexity of Sender's payoff as a function of Receiver's beliefs.

We next turn to studying optimal mechanisms. We use tools from convex analysis to show that an optimal mechanism exists and to characterize it for any given set of preferences and initial beliefs. We show that no disclosure of information is optimal when Sender's payoff is concave in Receiver's beliefs, and full disclosure is optimal when Sender's payoff is convex in Receiver's beliefs. We also establish that an optimal mechanism need never induce more actions in equilibrium than there are states.

We then generalize three important properties of the optimal mechanism in the example above. Notice, first, that when the judge chooses the prosecutor's least-preferred action (*acquit*), she is certain of the state. That is, she never acquits guilty defendants. Otherwise, we would have $\pi(i|guilty) > 0$. But then the prosecutor could increase his payoff by decreasing $\pi(i|guilty)$ and increasing $\pi(g|guilty)$; this would strictly increase the probability of g and would only increase the willingness of the judge to convict when she sees g. We establish that, in general, whenever Receiver takes Sender's least-preferred action, she knows with certainty that the state is one where this action is optimal.

Second, notice that when the judge convicts, she is exactly indifferent between convicting and acquitting. If she strictly preferred to convict upon seeing g, the prosecutor could increase his payoff by slightly decreasing $\pi(i|innocent)$ and increasing $\pi(g|innocent)$; this would increase the probability of g and leave the judge's optimal action given the message unchanged, thus increasing the probability of conviction. We show that, in general, whenever Receiver has an interior posterior, she is effectively indifferent among two actions.

Finally, notice that because the prosecutor's payoff is (weakly) increasing in the judge's posterior belief that the state is *guilty*, it is meaningful to talk about beliefs that place more weight on *innocent* as being "worse" from the prosecutor's perspective. A different way to look at the last two results is that the prosecutor chooses an investigation that induces the worst possible belief consistent with a given action by the judge—certainty of innocence when the action is *acquit*, and indifference when the action is *convict*. We show that in general when Sender's payoffs are monotonic in Receiver's beliefs, Sender typically induces the worst belief consistent with a given action.

We next apply our results to three examples. Our first example examines what type of feedback a university should provide to an assistant professor whose research effort depends on her beliefs about the chance that she will get tenure. The second example studies how preference disagreement between Sender and Receiver impacts information transmission under an optimal mechanism. Lastly, we analyze the optimal structure of informative advertisements in a setting with unit demand. These examples illustrate both the breadth of situations captured by our model and the practical applicability of our propositions. Finally, we discuss extensions of our results to dynamic mechanisms, incomplete information on the part of Receiver, multiple Receivers, multiple Senders, limited messaging technologies, and limited commitment.

The observation that Bayesian updating only restricts the expectation of posteriors has been made before and has been utilized in a variety of contexts.[1] The work most closely related to our

[1]The formal methods employed in our analysis are very close to Aumann and Maschler's (1995) analysis of repeated games of incomplete information. They study the value to a player of knowing which game is being played when the other player lacks this knowledge, a fixed zero-sum game is repeated *ad infinitum*, players maximize their long-run non-discounted average payoffs, and payoffs are not observed. The fact that the informed player's initial actions have no impact on his long-run average payoffs (and can thus be treated as just a signal) combined with a focus on Nash equilibria (which implicitly allow for commitment) makes Aumann and Maschler's problem mathematically analogous to ours.

paper is Brocas and Carrillo (2007). They analyze the gain to Sender from controlling the flow of public information in a setting with a binary state space and information that consists of a sequence of symmetric binary signals. Lewis and Sappington (1994) and Johnson and Myatt (2006) consider how much information a monopolist would want to provide to his potential customers. Carillo and Mariotti (2000), Bodner and Prelec (2003), and Bénabou and Tirole (2002, 2003, 2004) employ a form of Bayesian persuasion to study self-signaling and self-regulation. Caillaud and Tirole (2007) rely on a similar mechanism to study persuasion in group settings. Lazear (2006) applies a closely-related intuition to examine when providing information about a test increases learning. In contrast to these papers, we derive results that apply to arbitrary state spaces, information structures, preferences and initial beliefs.[2]

This paper also relates to a broader literature on optimal information structures. Prendergast (1992) studies the assignment of individuals into groups (and the resulting information about their types) when individuals are risk-averse over the realization of their type. Ostrovsky and Schwarz (2008) examine the equilibrium design of grade transcripts (and the resulting information about quality of students) when schools compete to place their students in good jobs. Rayo and Segal's (2008) concurrent work characterizes the optimal disclosure policy under specific assumptions about preferences and about Receiver's outside option.

Our results also contribute to the literature on contract theory. An important aspect of our setting is that Receiver's action is not contractible. Most work in contract theory examines two remedies for such non-contractibility: payment for outcomes correlated with the action (e.g., Holmstrom 1979, Grossman and Hart 1983) and suitable allocation of property rights (e.g., Grossman and Hart 1986, Hart and Moore 1990). Our results highlight another instrument for implementing a second-best outcome, namely the control of the agent's informational environment.[3] Our example on how to optimally structure midterm review of tenure-track faculty so as to induce second-best effort illustrates this interpetation of our results.

Finally, past work has studied related questions in contexts where Receivers are not perfect

[2]Glazer and Rubinstein (2004, 2006) study related problems where the communication technology effectively limits the set of signals Sender can convey. They focus on Receiver's part of the problem, however, and their approach differs markedly from that in all of the aforementioned papers.

[3]Taub (1997) analyzes the impact of information provision on incentives in a dynamic framework.

Bayesians (Mullainathan, Schwartzstein, and Shleifer 2008, Ettinger and Jehiel forthcoming)[4]. While persuasive activities may reflect such failures of rationality, assessing the relevant evidence requires a more complete understanding of when and how persuading a fully rational Bayesian is possible.

2 A model of persuasion

Receiver has a continuous utility function $u(a, \omega)$ that depends on her action $a \in A$ and the state of the world $\omega \in \Omega$. Sender has a continuous utility function $v(a, \omega)$ that depends on Receiver's action and the state of the world. Sender and Receiver share a prior $\mu_0 \in int(\Delta(\Omega))$.[5] Let $a^*(\mu)$ to be the set of actions that maximize Receiver's expected utility given her belief is μ. We assume that that there are at least two actions in A and that for any action a there exists a μ s.t. $a^*(\mu) = \{a\}$. The action space A is compact and the state space Ω is finite. The latter assumption is mainly for ease of exposition: Appendix B demonstrates that our central characterization result extends to the case where Ω is any compact metric space.

A special case of particular interest is where ω is a real-valued random variable, Receiver's action depends only on the expectation $E_\mu[\omega]$, rather than the entire distribution μ, and Sender's preferences over Receiver's actions do not depend on ω. This holds, for example, if $u(a, \omega) = -(a - \omega)^2$ and $v(a, \omega) = a$. When these conditions are satisfied, we will say that *payoffs depend only on the expected state.*

We define a *persuasion mechanism* (π, c) to be a combination of a signal and a message technology. Sender's private *signal* π consists of a finite realization space S and a family of distributions $\{\pi(\cdot|\omega)\}_{\omega \in \Omega}$ over S. A *message technology* c consists of a finite message space M and a family of functions $c(\cdot|s) : M \to \overline{\mathbb{R}}_+$; $c(m|s)$ denotes the cost to Sender of sending message m after receiving signal realization s.[6] The assumptions that S and M are finite are without loss of generality (cf. Proposition 9) and are used solely for notational convenience.

A persuasion mechanism defines a game. The timing is as follows. First, nature selects ω from

[4] Cain, Loewenstein, and Moore (2005) provide experimental results on susceptibility to persuasion.

[5] $int(X)$ denotes the interior of set X and $\Delta(X)$ the set of all probability distributions on X.

[6] $\overline{\mathbb{R}}_+$ denotes the affinely extended non-negative real numbers: $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{\infty\}$. Allowing c to take on the value of ∞ is useful for characterizing the cases where Sender cannot lie and cases where he must reveal all his information.

Ω according to μ_0. Neither Sender nor Receiver observe nature's move. Then, Sender privately observes a realization $s \in S$ from $\pi(\cdot|\omega)$ and chooses a message $m \in M$. Finally, Receiver observes m and chooses an action $a \in A$. Sender's payoff is $v(a,\omega) - c(m|s)$ and Receiver's payoff is $u(a,\omega)$. We represent the Sender's and Receiver's (possibly stochastic) strategies by σ and ρ, respectively. We use $\mu(\omega|m)$ to denote Receiver's posterior belief that the state is ω after observing m.

A perfect Bayesian equilibrium of a persuasion mechanism is a triplet $(\sigma^*, \rho^*, \mu^*)$ satisfying the usual conditions. We also apply an additional equilibrium selection criterion: we focus on Sender-preferred equilibria, i.e., equilibria where the expectation of $v(a,\omega) - c(m|s)$ is the greatest. The focus on Sender-preferred equilibria provides a consistent comparison across mechanisms which prevents us from generating benefits of persuasion simply through equilibrium selection. Moreover, this particular comparison, unlike say comparing equilibria worst for Sender, ensures the existence of an optimal mechanism (cf. proof of Proposition 7). In the remainder of the paper, we use the term "equilibrium" to mean a Sender-preferred perfect Bayesian equilibrium of a persuasion mechanism.

Motivated by this definition of equilibria, we let $\hat{a}(\mu)$ denote an element of $a^*(\mu)$ that maximizes Sender's expected utility at belief μ. If there is more than one such action, we let $\hat{a}(\mu)$ be an arbitrary element from this set.[7] We refer to $\hat{a}(\mu_0)$, as the *default action*.

We define the *value* of a mechanism to be the equilibrium expectation of $v(a,\omega) - c(m|s)$. The *gain* from a mechanism is the difference between its value and the equilibrium expectation of $v(a,\omega)$ when Receiver obtains no information. *Sender benefits from persuasion* if there is a mechanism with a strictly positive gain. A mechanism is *optimal* if no other mechanism has higher value.

2.1 Varieties of Persuasion Mechanisms

A few examples help clarify the varieties of games that are captured by the definition of a persuasion mechanism. If π is perfectly informative and c is constant, the mechanism is a cheap talk game as in Crawford and Sobel (1982). If π is arbitrary and c is constant, the mechanism coincides with the information-transmission game of Green and Stokey (2007). If π is perfectly informative and $c(m|s) = -(m-s)^2$, the mechanism is a communication game with lying costs developed

[7] This allows us to use convenient notation such as $v(\hat{a}(\mu), \omega)$.

in Kartik (forthcoming). If π is perfectly informative, $M = \mathcal{P}(\Omega)$, and $c(m|s) = \begin{cases} 0 \text{ if } s \in m \\ \infty \text{ if } s \notin m \end{cases}$, the mechanism is a persuasion game as in Grossman (1981) and Milgrom (1981).[8] If π is perfectly informative, $M = \mathbb{R}_+$, and $c(m|s) = m/s$, the mechanism is Spence's (1973) education signalling game. The model can also be easily re-interpreted to allow for multiple receivers (as in Lewis and Sappington 1994), for Receiver to be uncertain about what information Sender has (as in Shin 2003), or for Seller to have discretion over which costly information to acquire (as in Jovanovic 1982). We consider extensions to our model in Section 7 below, where we also discuss in more detail what types of games our definition rules out.

2.2 Honest Mechanisms

A particularly important type of a persuasion mechanism is one where $M = S$ and $c(m|s) = \begin{cases} k \text{ if } s=m \\ \infty \text{ if } s \neq m \end{cases}$ for some $k \in \mathbb{R}_+$. We call such mechanisms *honest*. In contrast to Grossman (1981) and Milgrom (1981), where Sender is simply not allowed to tell an explicit lie, under an honest mechanism Sender must tell the truth, *the whole truth*, and nothing but the truth. In other words, he has committed to fully disclose all his private information. Much of our analysis depends on allowing for the possibility that Sender can commit in this way.

Note that an honest mechanism can be interpreted as either (i) a choice of a signal π on S given an honest messaging technology $c(m|s) = \begin{cases} k \text{ if } s=m \\ \infty \text{ if } s \neq m \end{cases}$, or (ii) a disclosure rule $\pi : \Omega \to \Delta(S)$, as in Rayo and Segal (2008). While these two interpretations are formally equivalent, one or the other may be more natural in particular settings.

Examples where the first interpretation makes sense include a prosecutor requesting a forensic test or a firm conducting a public celebrity taste test against a rival product. In these settings, Sender has an ability to commit on two counts: he can choose to remain imperfectly informed *and* he can commit to fully disclose anything he learns. The latter commitment may arise either because Receiver directly observes the revelation of the information (as in the taste test example) or through an institutional structure that requires Sender to always report all the tests he has conducted and what their outcomes were (as in the prosecutor example).

Examples where the second interpretation is more natural include a school or a rating agency

[8] $\mathcal{P}(X)$ denotes the set of all subsets of X.

choosing a coarse grading policy. Here, Sender might be unable to avoid learning the full information about the state but can plausibly commit to an *ex ante*, potentially stochastic, disclosure rule that is not fully revealing.

3 Simplifying the Problem

The class of persuasion mechanisms defined above is large, including cases where Sender's strategy might involve complex messages, signaling, lying, and so on. In this section, we show that to determine whether Sender benefits from persuasion and what an optimal mechanism is, it suffices to consider a much simpler problem.

The key intuition is the following. An equilibrium of any persuasion mechanism induces a particular distribution of Receiver's beliefs. This distribution of beliefs in turn determines a distribution over Receiver's actions. From Sender's perspective, any equilibrium that induces the same distribution of actions conditional on states must have the same value. To determine whether there exists a persuasion mechanism with some given value, therefore, it is sufficient to ask whether there exists a distribution of Receiver's beliefs that is compatible with Bayes' rule and generates expected utility for Sender equal to that value.

Let a *distribution of posteriors* τ be a distribution on $\Delta(\Omega)$. A persuasion mechanism *induces* τ if there exists an equilibrium of the mechanism such that $Supp(\tau) = \{\mu_m\}_{m \in M}$ and

$$\begin{aligned} \text{(i) } \mu_m(\cdot) &= \mu^*(\cdot|m) \\ \text{(ii) } \tau(\mu_m) &= \sum_{\omega \in \Omega} \sum_{s \in S} \sigma^*(m|s)\, \pi(s|\omega)\, \mu_0(\omega). \end{aligned}$$

A belief μ is induced by a mechanism if τ is induced by the mechanism and $\tau(\mu) > 0$. A distribution of posteriors is *Bayes-plausible* if the expected posterior probability of each state equals its prior probability:

$$\int \mu d\tau(\mu) = \mu_0.$$

Bayesian rationality requires that any equilibrium distribution of Receiver's beliefs be Bayes-plausible. Our first Proposition below shows that this is the *only* restriction imposed by Bayesian

rationality. That is, for any Bayes-plausible distribution of posteriors there is a persuasion mechanism that induces this distribution in equilibrium.

Now, let

$$\hat{v}(\mu) \equiv \sum_{\omega \in \Omega} v(\hat{a}(\mu), \omega) \mu(\omega).$$

This denotes Sender's expected utility when both he and Receiver hold belief μ.

Sender's utility (gross of messaging costs) in any mechanism which induces τ is simply the expectation of $\hat{v}$ under τ, $E_\tau \hat{v}(\mu)$. At terminal nodes of the game, Sender and Receiver may hold different beliefs, say μ_S and μ_R. For example, Sender may have observed a highly informative signal but chosen to send a message that reveals no information. Sender's payoff at such a node is neither $\hat{v}(\mu_S)$ nor $\hat{v}(\mu_R)$, but rather $\sum_{\omega \in \Omega} v(\hat{a}(\mu_R), \omega) \mu_S(\omega)$. A more obvious statement, then, would have been that the Sender's utility in a mechanism is the expectation of $\sum_{\omega \in \Omega} v(\hat{a}(\mu_R), \omega) \mu_S(\omega)$ over the *joint* distribution of Sender's and Receiver's beliefs. What allows us to collapse this potentially complicated expression to $E_\tau \hat{v}(\mu)$ is the following observation. Because Receiver's beliefs satisfy the equilibrium condition, it must be the case that from the *ex ante* perspective, before he has obtained any private information, Sender's belief conditional on learning that Receiver will have belief μ must also be μ. Hence, his *ex ante* expected utility from inducing μ is $\hat{v}(\mu)$.

Another reason why we do not need to worry about the joint distribution of Sender's and Receiver's beliefs is that we can restrict our attention, without loss of generality, to honest mechanisms where Sender's and Receiver's beliefs always coincide. In fact, we can restrict our attention even further, to a particular type of honest mechanism. Say that a mechanism is *straightforward* if it is honest, $S \subset A$, and Receiver's equilibrium action equals the message. In other words, in straightforward mechanisms, the signal produces a "recommended action" for Receiver, Sender reports the recommendation honestly, and Receiver takes the action recommended. Given a distribution of actions induced by any mechanism, there exists a straightforward mechanism that induces the same distribution of actions. This result is closely analogous to the revelation principle (e.g., Myerson 1979). Of course, the revelation principle applies to problems where players' information is a given, while our problem is that of designing the informational environment.

This leads us to the proposition that greatly simplifies our problem.

Proposition 1 *The following are equivalent:*

1. *There exists a persuasion mechanism with value v^*;*

2. *There exists a straightforward mechanism with value v^*;*

3. *There exists a Bayes-plausible distribution of posteriors τ such that $E_\tau \hat{v}(\mu) = v^*$.*

Detailed proofs of all propositions are in Appendix A. We sketch the basic argument here. That (2) implies (1) and (3) is immediate. To see that (1) implies (2), let $\alpha(\cdot|\omega)$ be the distribution of actions in an equilibrium of any mechanism. Consider the honest mechanism with $S = A$ and $\pi(a|\omega) = \alpha(a|\omega)$. We need to show that in an equilibrium of this mechanism $\rho^*(a|m) = \begin{cases} 1 \text{ if } a=m \\ 0 \text{ if } a \neq m \end{cases}$. This follows from two observations: (i) the belief induced by sending message a is a convex combination of beliefs that induced a in the original equilibrium; (ii) if an action is optimal for a set of beliefs, it is optimal for a belief that is in the convex hull of that set. Finally, that (3) implies (1) is equivalent to the claim that Bayes-plausability is the only restriction on the equilibrium distribution of posteriors. This part of our argument is closely related to Shmaya and Yariv's (2009) concurrent work that identifies which sequences of distributions of posteriors are consistent with Bayesian rationality. Given any Bayes-plausible τ, let S index $Supp(\tau)$ and consider a signal $\pi(s|\omega) = \frac{\mu_s(\omega)\tau(\mu_s)}{\mu_0(\omega)}$. The honest mechanism with signal π induces τ.

The key implication of Proposition 1 is that to evaluate whether Sender benefits from persuasion and to determine the value of an optimal mechanism we need only ask how $E_\tau \hat{v}(\mu)$ varies over the space of Bayes-plausible distributions of posteriors.

Corollary 1 *Sender benefits from persuasion if and only if there exists a Bayes-plausible distribution of posteriors such that*

$$E_\tau \hat{v}(\mu) > \hat{v}(\mu_0).$$

The value of an optimal mechanism is

$$\max_\tau E_\tau \hat{v}(\mu)$$
$$s.t. \int \mu d\tau(\mu) = \mu_0.$$

Note that Corollary 1 does not by itself tell us that an optimal mechanism exists. As we will show later, however, this is indeed always the case.

Proposition 1 implies that we can restrict our attention to mechanisms where Sender is compelled to report all he knows truthfully. Consequently, none of our results depend on the interpretation of v as Sender's utility. We could let v denote a social welfare function, for example. Our results would then identify socially-optimal rather than Sender-optimal mechanisms. Or, we could let v denote a weighted combination of Sender's and Receiver's utilities resulting from *ex ante* bargaining over which mechanism to use. Throughout the paper we will refer to v as Sender's utility, but one should keep in mind that our results apply for any objective function over Receiver's action and the state.

We introduce a final definition that will be useful in the analysis that follows. Let V be the *concave closure* of $\hat{v}$:

$$V(\mu) \equiv \sup \{z | (\mu, z) \in co(\hat{v})\},$$

where $co(\hat{v})$ denotes the convex hull of the graph of $\hat{v}$. Note that V is concave by construction. In fact, it is the smallest concave function which is everywhere weakly greater than $\hat{v}$.[9] Figure 1 shows an example of the construction of V. In this figure, as in all figures in the paper, we identify a distribution μ with a point in $\mathbb{R}^{n-1}$, where n is the number of states. So in Figure 1, if $\Omega = \{\omega_L, \omega_R\}$, the μ on the x-axis is the probability of one of the states, say ω_L. Specifying this probability of course uniquely pins down the distribution μ.

To see why V is a useful construct, observe that if $(\mu', z) \in co(\hat{v})$, then there exists a distribution of posteriors τ such that $E_\tau \mu = \mu'$ and $E_\tau \hat{v}(\mu) = z$. Thus, by Proposition 1, $co(\hat{v})$ is the set of (μ, z) such that if the prior is μ, there exists a mechanism with value z. Hence, $V(\mu)$ is the largest payoff Sender can achieve with any mechanism when the prior is μ.

[9] Our definition of concave closure is closely related to the notion of a *biconjugate* function in convex analysis (Hiriart-Urruty and Lemaréchal 2004). Note that we can alternatively express V as

$$V(\mu) = \inf_{s,r} \left\{ \langle s, \mu \rangle - r | \langle s, \mu' \rangle - r \geq \hat{v}(\mu') \ \ \forall \mu' \right\}$$

where $\langle \cdot, \cdot \rangle$ denotes inner product. Hence, V is a "concave version" of the (convex) biconjugate function defined by

$$\hat{v}^{**}(\mu) \equiv \sup_{s,r} \left\{ \langle s, \mu \rangle - r | \langle s, \mu' \rangle - r \leq \hat{v}(\mu') \ \ \forall \mu' \right\}.$$

Specifically, $V = -((-\hat{v})^{**})$. Aumann and Maschler (1995) refer to V as the concavification of $\hat{v}$.

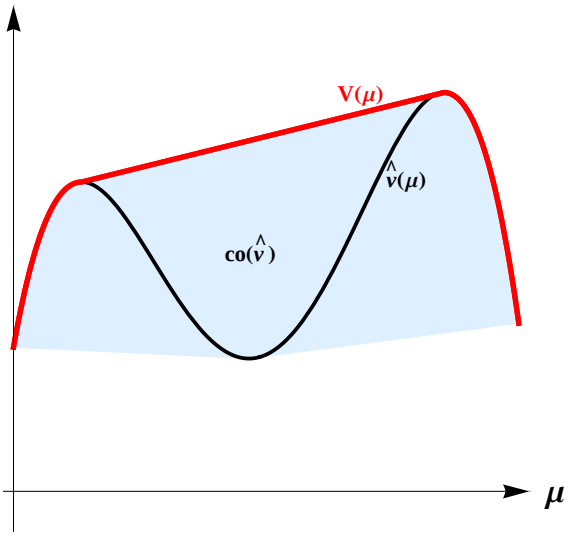


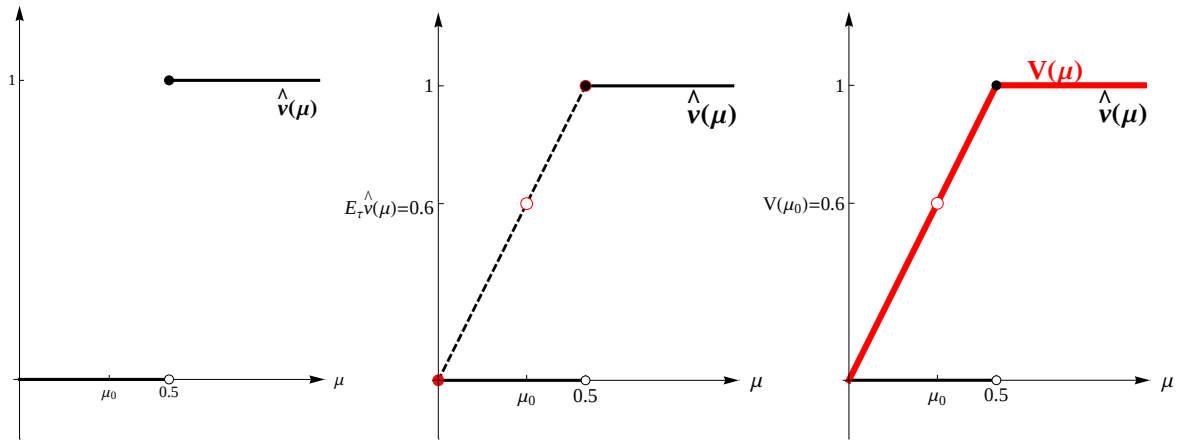
Figure 1: an illustration of concave closure

Corollary 2 *The value of an optimal mechanism is* $V(\mu_0)$. *Sender benefits from persuasion if and only if* $V(\mu_0) > \hat{v}(\mu_0)$.

Figure 2 shows the function $\hat{v}$, the optimal mechanism, and the concave closure V in the motivating example from the introduction. In the figure, μ denotes the probability that the state is *guilty*. As panel (a) shows, $\hat{v}$ is a step function: the prosecutor's expected payoff is 0 whenever μ is less than 0.5 (since the judge will choose *acquit)* and 1 whenever μ is greater than or equal to .5 (since the judge will choose *convict*). As panel (b) shows, the optimal signal induces two posterior beliefs. When the judge observes i, her posterior belief is $\mu = 0$ and $\hat{v}(0) = 0$. When the judge observes g, her posterior belief is $\mu = .5$ and $\hat{v}(.5) = 1$. The distribution τ over these beliefs places probability .4 on $\mu = 0$ and probability .6 on $\mu = .5$. Hence, the prosecutor's expected utility is $E_\tau \hat{v}(\mu) = .6$. The distribution τ is Bayes plausible since $\mu_0 = .3 = .4(0) + .6(.5)$. As panel (c) shows, the concave closure V is equal to 2μ when $\mu \leq 0.5$ and constant at 1 when $\mu > 0.5$. It is clear that $V(\mu_0) > \hat{v}(\mu_0)$ and that the value of the optimal mechanism is $V(\mu_0)$.

4 When does Sender benefit from persuasion?

Corollary 2 provides a necessary and sufficient condition for Sender to benefit from persuasion in terms of the concave closure V. In any problem where we can graph the function $\hat{v}$, it is



(a) function $\hat{v}$ (b) optimal mechanism (c) function V

Figure 2: the motivating example

straightforward to construct V and determine the prior beliefs, if any, at which $V(\mu_0) > \hat{v}(\mu_0)$. In Figure 3, for example Sender benefits from persuasion for any $\mu_0 \in (\mu_l, \mu_h)$, and does not benefit from persuasion for any $\mu_0 \leq \mu_l$ or $\mu_0 \geq \mu_h$. In Figure 2, Sender benefits from persuasion for any $\mu_0 \in (0, .5)$—i.e. at any prior belief at which the judge does not convict by default. In this section, we characterize the conditions when $V(\mu_0) > \hat{v}(\mu_0)$ holds, first in terms of the properties of $\hat{v}$, and then in terms of the primitives of our model, namely Sender and Receiver's preferences and initial beliefs.

Corollaries 1 and 2 tell us that Sender benefits from persuasion if and only if there exists a τ such that $E_\tau(\hat{v}(\mu)) > \hat{v}(E_\tau(\mu))$. Whether this is the case is naturally tied to the concavity or convexity of $\hat{v}$. Note that since $\hat{v}$ is not necessarily differentiable, we cannot speak of its convexity or concavity "at a point." The analogue for a potentially non-differentiable function to being weakly convex everywhere and strictly convex somewhere is that it be convex and not concave.

Proposition 2 *If $\hat{v}$ is concave, Sender does not benefit from persuasion for any prior. If $\hat{v}$ is convex and not concave, Sender benefits from persuasion for every prior.*

Observe that in the simple case where Sender's payoff does not depend on the state, $\hat{v}(\mu) = v(\hat{a}(\mu))$. The concavity or convexity of $\hat{v}$ then depends on just two things: whether Receiver's action $\hat{a}(\mu)$ is concave or convex in μ, and whether Sender's payoff $v(a)$ is concave or convex in a.

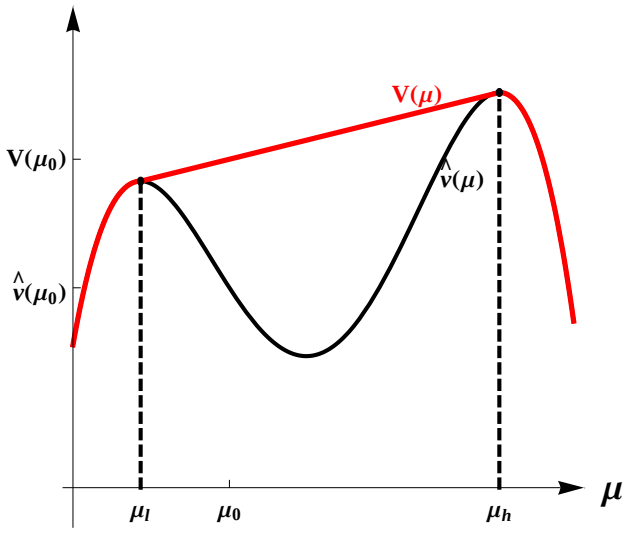


Figure 3: an illustration with an arbitrary $\hat{v}$

If both $\hat{a}$ and v are concave, Sender does not benefit from persuasion. If both $\hat{a}$ and v are convex and at least one of them is not concave, Sender benefits from persuasion.

Consider a simple example where $u(a, \omega) = -(a-\omega)^2$ and $v(a, \omega) = a$. In this case, both $\hat{a}$ and v are linear, and hence concave, so Sender does not benefit from persuasion. Moreover, since $\hat{v}(\mu) = E_\mu[\omega]$ is linear in μ, Sender is completely indifferent about the information Receiver obtains. This is not because he does not care about Receiver's belief: his utility is strictly increasing in the expectation of μ. Rather, it is because he knows that in any informational environment, Receiver's beliefs will on average equal the prior, so his expected utility from Receiver's action is fixed at $E_{\mu_0}[\omega]$.[10]

There is also a sense in which the concavity or convexity of $\hat{v}$ depends on the extent to which Sender and Receiver's preferences are aligned. At one extreme, if Sender and Receiver's preferences are *perfectly* aligned ($v = u$), we have $\hat{v}(\mu) \equiv \max_a \sum_\omega u(a, \omega)\mu(\omega)$. Since the maximand is linear in μ, $\hat{v}$ is convex. Moreover, since $\hat{a}(\mu)$ is not constant, $\hat{v}$ is not concave. Hence, Sender benefits from persuasion. By the same logic, if Sender and Receiver's preferences are perfectly *misaligned* ($v = -u$), $\hat{v}$ is concave and Sender can never benefit from persuasion. In Subsection 6.2, we explore alignment of preferences in more detail.

[10]This does not mean Sender is indifferent across all mechanisms. Some mechanisms, such as a signalling game, would induce messaging costs in equilibrium and thus lead to a lower overall utility.

Often, $\hat{v}$ will be neither convex nor concave. This is true, for example, in our motivating example as shown in Figure 2. As we discussed earlier, the fact that Sender benefits from persuasion in that example hinges on (i) the fact that Receiver does not take Sender's preferred action by default, and (ii) the fact that Receiver's action is constant in a neighborhood around the prior. We now show that these two conditions, suitably generalized, play a crucial role more broadly. Specifically, the generalization of (i) is necessary, while generalizations of (i) and (ii) are jointly sufficient, for Sender to benefit from persuasion.

To generalize (i), say *there is information Sender would share* if $\exists \mu$ s.t.

$$\hat{v}(\mu) > \sum_{\omega} v(\hat{a}(\mu_0), \omega)\, \mu(\omega). \tag{2}$$

In other words, there must exist a μ such that, if Sender had private information that led him to believe μ, he would prefer to share this information with Receiver rather than have Receiver act based on μ_0. Note that when v does not depend on ω, there is information Sender would share as long the default action is not dominant, i.e., $v(\hat{a}(\mu_0)) < v(a)$ for some $a \in A$. This is the sense in which equation (2) generalizes condition (i).

When there is no information Sender would share, Sender cannot benefit from persuasion since there is no informative message he would ever wish Receiver to see.

Proposition 3 *If there is no information Sender would share, Sender does not benefit from persuasion.*

Now, to generalize (ii), we say Receiver's *preference is discrete* at belief μ if Receiver's expected utility from her preferred action $\hat{a}(\mu)$ is bounded away from her expected utility from any other action, i.e., if there is an $\varepsilon > 0$ s.t. s.t. $\forall a \neq \hat{a}(\mu)$,

$$\sum u(\hat{a}(\mu), \omega)\, \mu(\omega) > \sum u(a, \omega)\, \mu(\omega) + \varepsilon.$$

The following Proposition is the main result of this section; it demonstrates that the generalizations of (i) and (ii) are sufficient for Sender to benefit from persuasion.

Proposition 4 *If there is information Sender would share and Receiver's preference is discrete at the prior, Sender benefits from persuasion.*

The intuition for the proof is as follows. First, because there is information that Sender would share we can find a belief μ_h such that $\hat{v}(\mu_h) > \sum_{\omega} v(\hat{a}(\mu_0), \omega)\mu_h(\omega)$. Second, the discreteness of Receiver's preference implies that there is a belief near the prior, say μ_l, such that $\hat{a}(\mu_l)$ is equal to Receiver's default action and μ_0 is on the segment between μ_l and μ_h. That mixing point μ_l and μ_h produces a strictly positive gain is obvious in a case like the motivating example where Sender's payoff v does not depend on the state. The argument is more subtle when v does depend on the state. The key observation is that for any *given* action by Receiver, Sender's utility is linear in μ. In particular, $\sum_{\omega} v(\hat{a}(\mu_0), \omega)\mu(\omega)$ is linear in μ. This implies that mixing μ_l with μ_h yields a strictly positive gain.

Proposition 4 is particularly useful when the action space is finite. In that case, Receiver's preference is generically discrete at the prior in the sense that the set of beliefs at which Receiver's preference is not discrete is Lebesgue measure-zero in $\Delta(\Omega)$. To see why this is the case, note that with a finite action space, Receiver's preference can be non-discrete at μ only if Receiver is exactly indifferent between two distinct actions at μ. Such indifference is a knife-edge case that generically does not hold at the prior.

Proposition 5 *If A is finite, Receiver's preference is discrete at the prior generically.*

The key implication of Proposition 5 is that when the action space is finite, we should expect Sender to benefit from persuasion if and only if there is information Sender would share. Note, however, that this result is not meant to suggest that there is some form of discontinuity in Sender's benefit from persuasion as we move from large finite choice sets to infinite ones. As the action space becomes large, the gain from persuasion may become arbitrarily small.

We now turn to the case where payoffs depend only on the expected state. We have shown that when $\hat{v}$ can be graphed, inspection of the graph can show directly whether Sender benefits from persuasion. Remember, however, that the domain of $\hat{v}$ is $\Delta(\Omega)$. This means that it is only possible to easily depict $\hat{v}$ when there are two or three states. When there are more states, our Propositions still apply, but one cannot approach the problem by simply studying the graph of $\hat{v}$.

When payoffs depend only on the expected state, however, a natural conjecture is that we could learn about Sender's gain from persuasion by graphing Sender's expected payoff as a function of the expected state $E_\mu[\omega]$ rather than as a function of μ directly. If so, we would have a simple two-dimensional representation of this subclass of problems regardless of the size of the state space.

When payoffs depend only on the expected state, there exists a $\widetilde{v}:\mathbb{R}\to\mathbb{R}$ such that $\widetilde{v}(E_\mu[\omega]) = \hat{v}(\mu)$. Let $\widetilde{V}$ be the concave closure of $\widetilde{v}$. The following proposition shows that the conjecture above is correct: we can determine whether Sender benefits from persuasion simply by inspecting $\tilde{V}$ and $\tilde{v}$. In Section 6 below, we provide an example of how this result can greatly simplify the analysis of problems with a large state space.

Proposition 6 *Sender benefits from persuasion if and only if* $\widetilde{V}\left(E_{\mu_0}[\omega]\right) > \widetilde{v}\left(E_{\mu_0}[\omega]\right)$.

To see that $\widetilde{V}\left(E_{\mu_0}[\omega]\right) \leq \widetilde{v}\left(E_{\mu_0}[\omega]\right)$ implies Sender cannot benefit from persuasion, we need only note that for any Bayes-plausible τ,

$$E_\tau[\hat{v}(\mu)] = E_\tau[\widetilde{v}(E_\mu[\omega])] \leq \widetilde{V}(E_\tau[E_\mu[\omega]]) = \widetilde{V}\left(E_{\mu_0}[\omega]\right).$$

The proof of the converse is more involved. If $\widetilde{V}\left(E_{\mu_0}[\omega]\right) > \widetilde{v}\left(E_{\mu_0}[\omega]\right)$, we know there is a τ s.t. $E_\tau[E_\mu[\omega]] = E_{\mu_0}[\omega]$ and $E_\tau[\hat{v}(\mu)] > \hat{v}(\mu_0)$. If this τ were Bayes-plausible, we could construct an honest mechanism that induces it and we would be done. The trouble is that $E_\tau[E_\mu[\omega]] = E_{\mu_0}[\omega]$ does not guarantee that τ is Bayes-plausible. To construct a persuasion mechanism with a strictly positive gain, we show that it is always possible to find a belief μ' such that $E_{\mu'}[\omega] = E_{\mu_0}[\omega]$ and a Bayes-plausible τ' that is a mixture of τ and μ'. Since $E_\tau[\hat{v}(\mu)] > \hat{v}(\mu_0)$ and $\hat{v}(\mu') = \hat{v}(\mu_0)$, we know that $E_{\tau'}[\hat{v}(\mu)] > \hat{v}(\mu_0)$.[11]

5 Optimal mechanisms

Corollary 2 shows that the value of an optimal mechanism is $V(\mu_0)$. When the problem is simple enough that it is possible to graph $\hat{v}$ and V, we can read $V(\mu_0)$ and the gain $V(\mu_0) - \hat{v}(\mu_0)$ off

[11] As the detailed proof in Appendix A shows, we can also establish a result somewhat stronger than Proposition 6. Suppose there exists a linear $T:\Delta(\Omega)\to\mathbb{R}^k$ and a $\widetilde{v}:\mathbb{R}^k\to\mathbb{R}$ s.t. $\hat{v}(\mu) = \widetilde{v}(T\mu)$. Then, Sender benefits from persuasion if and only if $\widetilde{v}$ is below its concave closure at $T\mu_0$. We focus on the case where $T\mu = E_\mu[\omega]$ only so we can simplify the exposition of the result.

the graph directly. Figure 3 illustrates this for the arbitrary $\hat{v}$ shown earlier in Figure 1. Similarly, in Figure 2, we can easily see that the value of the optimal mechanism in our motivating example must be $V(\mu_0) = .6$, and the gain is $V(\mu_0) - \hat{v}(\mu_0) = .6 - 0 = .6$.

The graph of $\hat{v}$ and V also identifies the beliefs induced by the optimal mechanism—these are the points on $\hat{v}$ whose convex combination yields value $V(\mu_0)$. In Figure 3, we see that the optimal mechanism induces beliefs μ_l and μ_h. It is clear from this figure that the optimal mechanism is unique, since there are no beliefs other than μ_l and μ_h that could be combined to produce value $V(\mu_0)$. Similarly, we can easily see in Figure 2 that the optimal mechanism must induce beliefs $\mu = 0$ and $\mu = .5$.

From here, it is straightforward to construct the optimal mechanism. The condition that $\int \mu d\tau(\mu) = \mu_0$ allows us to compute the distribution τ. For the motivating example in Figure 2, we must have $0\tau(0) + .5\tau(.5) = .3$, so $\tau(0) = .4$ and $\tau(.5) = .6$. Following the proof of Proposition 1, we then set:

$$\begin{aligned} \pi(s|\omega) &= \frac{\mu_s(\omega)\,\tau(\mu_s)}{\mu_0(\omega)} \\ c(m|s) &= \begin{cases} 0 \text{ if } s = m \\ \infty \text{ if } s \neq m \end{cases}. \end{aligned}$$

For the motivating example, this yields the π in Equation 1.

These steps rely on being able to visualize the graphs of $\hat{v}$ and V, which is only possible when the number of states is small. More generally, $\hat{v}$ is upper semicontinuous[12], which implies that any element of the graph of V can be expressed as a convex combination of elements of the graph of $\hat{v}$.[13] In particular, there exists a Bayes-plausible distribution of posteriors τ such that $E_\tau \hat{v}(\mu) = V(\mu_0)$. From τ and the prior μ_0, we can construct $\pi(s|\omega)$ and $c(m|s)$ from the expressions above. Then, (π, c) is an optimal mechanism.

Thus:

Proposition 7 *An optimal mechanism exists.*

We first characterize the optimal mechanism in terms of the convexity or concavity of $\hat{v}$. Say

[12] This is a consequence of our focus on Sender-preferred equilibria. We show this in Lemma 2.

[13] We establish this implication in Lemma 4.

that no disclosure is optimal if $\mu^*(\cdot|m) = \mu_0$ for all m sent in equilibrium of an optimal mechanism.[14] If, at the other extreme, $\mu^*(\cdot|m)$ is degenerate for all m sent in equilibrium of an optimal mechanism, say that full disclosure is optimal.[15] Finally, if $\mu^*(\cdot|m)$ is at the boundary of $\Delta(\Omega)$ for all m sent in equilibrium of an optimal mechanism, we say that *strong disclosure* is optimal.

Proposition 8 *For any prior,*

- *if $\hat{v}$ is (strictly) concave, no disclosure is (uniquely) optimal;*
- *if $\hat{v}$ is (strictly) convex, full disclosure is (uniquely) optimal;*
- *if $\hat{v}$ is convex and not concave, strong disclosure is uniquely optimal.*

The first two parts of Proposition 8 follow directly from the definition of convexity and concavity. To see why the third part holds, first note that for any μ induced by an optimal mechanism, it must be the case that $V(\mu) = \hat{v}(\mu)$. This observation will also be quite useful for several other Propositions. Now, if $\hat{v}$ is not concave, there is some belief μ_l s.t. $V(\mu_l) > \hat{v}(\mu_l)$. If an optimal mechanism induces an interior belief, this belief can be expressed as a convex combination of μ_l and some other belief. But then, $V(\mu) = \hat{v}(\mu)$ and $V(\mu_l) > \hat{v}(\mu_l)$ coupled with concavity of V imply that $\hat{v}$ cannot be convex.

Next, we show that in an optimal mechanism, the number of actions Sender needs to induce in equilibrium is limited by the number of states.[16]

Proposition 9 *There exists an optimal straightforward mechanism in which the number of actions Receiver takes with positive probability in equilibrium is at most $|\Omega|$.*

Note, first, that this proposition is clearly true for our motivating example, where there are two states and the optimal mechanism induces exactly two actions. It is also true for the example of Figure 3, where there are also two states and the optimal signal induces two beliefs μ_l and μ_h; since we have assumed no belief induces more than one action in equilibrium, the optimal mechanism induces no more than two actions.

[14] Note that in contrast to Sender, who can be hurt by revelation of information, Receiver is always made weakly better off by any mechanism.

[15] We say μ is degenerate if there is an ω s.t. $\mu(\omega) = 1$.

[16] In fact, as the proof of the proposition shows, we establish a somewhat stronger result: there exists an optimal straightforward mechanism that induces a τ whose support has cardinality no greater than that of Ω.

Formally, the proof of Proposition 9 hinges on showing that any point on the graph of V can be written as a convex combination of at most $|\Omega|$ points on the graph of $\hat{v}$. This is closely related to Caratheodory's theorem, which shows that if a point p is in the convex hull of a set of points P in $\mathbb{R}^N$, it is possible to write p as a convex combination of $N+1$ or fewer of the points in P. This would be sufficient to prove a version of Proposition 9 where the bound is $|\Omega|+1$ rather than $|\Omega|$. To prove that only $|\Omega|$ points are required, we rely on a related result from convex analysis.[17]

The intuition for Proposition 9 is best seen graphically. Referring back to Figure 3, it is easy to see that when there are two states, V will always be made up of points on $\hat{v}$ and line segments whose endpoints are on $\hat{v}$. This means we can write any point on the graph of V as a convex combination of at most two points of the graph of $\hat{v}$. More generally, V will always be made up of points on $\hat{v}$ and the $(|\Omega|-1)$-dimensional faces whose vertices are on $\hat{v}$. Any point on such a face can be written as a convex combination of at most $|\Omega|$ points of $\hat{v}$.

The bound provided by Proposition 9 is tight - one can easily construct examples where the optimal mechanism cannot be achieved with a signal that has fewer than $|\Omega|$ possible realizations. For instance, whenever full disclosure is uniquely optimal, no mechanism with $|S|<|\Omega|$ can be optimal. It is also worthwhile to note that this is a property specifically of *optimal* mechanisms. One can construct an example in which there exists a mechanism with value v^* but there is no mechanism with $|\Omega|$ or fewer actions induced in equilibrium that also gives value v^*. Proposition 9, however, implies that such a v^* must be strictly less than the value of an optimal mechanism.

We now turn to generalizing three features of the optimal mechanism in our motivating example: (i) whenever the judge chooses the prosecutor's least-preferred action (*acquit*), the judge is certain of the state; (ii) whenever the judge chooses an action that is not the prosecutor's least-preferred (*convict*), she is indifferent between that action and a worse one; (iii) the prosecutor always induces the worst belief consistent with a given action by the judge.

To generalize (i), we show that if there exists an action $\underline{a}$ which is Sender's least-preferred regardless of the state, then at any μ induced by an optimal mechanism that leads Receiver to choose $\underline{a}$, Receiver is certain that $\underline{a}$ is her optimal action.

[17] Caratheodory's Theorem guarantees that given any mechanism with value v^* (optimal or otherwise), there is a mechanism that induces no more than $|\Omega|+1$ actions which also yields v^*. Fenchel and Bunt's strengthening of Caratheodory's theorem implies that given an *optimal* mechanism, we need no more than $|\Omega|$ actions.

Proposition 10 *Suppose that* $v(\underline{a},\omega) < v(a,\omega)$ *for all* ω *and all* $a \neq \underline{a}$. *Suppose that* μ *is induced by an optimal mechanism and* $\hat{a}(\mu) = \underline{a}$. *Then, for any* ω *s.t.* $\{\underline{a}\} \neq \max u(a,\omega)$, *we have* $\mu(\omega) = 0$.

Intuitively, suppose that in a straightforward mechanism there is a belief which induces $\underline{a}$ but puts positive probability on a state ω where $\underline{a}$ is not optimal. For this to be true, Sender must report message $\underline{a}$ with positive probability in state ω. Sender's payoff would be strictly higher in a mechanism that simply revealed that the true state is ω in all such cases, because this would induce Receiver to choose an action Sender likes strictly better than $\underline{a}$ and would leave Receiver's actions given all other messages unchanged.

To generalize (ii), we establish that at any interior μ induced by an optimal mechanism, Receiver's preference for $\hat{a}(\mu)$ cannot be discrete. To show this result, it is necessary to rule out a pathological case where there can be a default action such that there is no information Sender would share, yet there is some other action which yields *exactly* the same utility to Sender as the default action.

Assumption 1 *There exists no action* a *s.t. (i)* $\forall \mu$, $\hat{v}(\mu) \leq \sum_{\omega} v(a,\omega)\mu(\omega)$ *and (ii)* $\exists \mu$ *s.t.* $a \neq \hat{a}(\mu)$ *and* $\hat{v}(\mu) = \sum_{\omega} v(a,\omega)\mu(\omega)$.

Since this assumption rules out only a particular type of indifference, we conjecture the set of Sender's and Receiver's preferences that violate Assumption 1 has Lebesgue measure zero.

Proposition 11 *Suppose Assumption 1 holds. If Sender benefits from persuasion, Receiver's preference is not discrete at any interior* μ *induced by an optimal mechanism.*

For an intuition for this result, note that if Receiver's preference at μ were discrete, then if μ were the prior Sender could benefit from persuasion unless there was no information Sender would share (by Proposition 4). So, since we know $V(\mu) = \hat{v}(\mu)$, it must be the case that if μ were the prior, there would be no information Sender would share. That would mean that $\hat{a}(\mu)$ is an extremely desirable action for Sender, so he would want to maximize the chance of inducing it. But, if μ were discrete, there would be another belief, close to μ, which would yield the same action and which Sender could induce more often than μ. Hence the Proposition above must hold.

The fact that Receiver's preference at μ is not discrete means that, in at least one direction, Receiver's action changes at μ. Note that Sender must weakly dislike at least one such change; otherwise he would be better off providing more information. Moreover, when the action space is finite, Proposition 11 implies that at any interior belief induced by an optimal mechanism, Receiver is indifferent between two actions. Hence, in that case, the optimal mechanism necessarily brings Receiver as close as possible to taking some action less desirable than the equilibrium one.

To generalize (iii), say that $\hat{v}$ is monotonic if for any μ, μ', $\hat{v}\left(\gamma\mu + (1-\gamma)\mu'\right)$ is monotonic in γ. When $\hat{v}$ is monotonic in μ, it is meaningful to think about beliefs that are better or worse from Sender's perspective. The simplest definition would be that μ is worse than μ' if $\hat{v}(\mu) \leq \hat{v}(\mu')$. Note, however, that because $v(a,\omega)$ depends on ω directly, whether μ is worse in this sense depends both on how Receiver's action changes at μ and how μ affects Sender's expected utility directly. It turns out that for our result we need a definition of worse that isolates the way beliefs affect Receiver's actions.

When $\hat{v}$ is monotonic, there is a rational relation on A defined by $a \succsim a'$ if $\hat{v}(\mu) \geq \hat{v}(\mu')$ whenever $a = \hat{a}(\mu)$ and $a' = \hat{a}(\mu')$. This relation on A implies a partial order on $\Delta(\Omega)$: say that $\mu \rhd \mu'$ if

$$E_{\mu}u(a,\omega) - E_{\mu}u\left(a',\omega\right) > E_{\mu'}u(a,\omega) - E_{\mu'}u\left(a',\omega\right)$$

for any $a \succsim a'$. In other words, a belief is higher in this partial order if it makes better actions (from Sender's perspective) more desirable for Receiver. The order is partial since a belief might make both a better and a worse action more desirable for Receiver. We say that μ is a *worst belief inducing* $\hat{a}(\mu)$ if there is no $\mu' \lhd \mu$ s.t. $\hat{a}(\mu) = \hat{a}(\mu')$. We then have the following:

Proposition 12 *Suppose Assumption 1 holds. If $\hat{v}$ is monotonic, A is finite, and Sender benefits from persuasion, then for any interior belief μ induced by an optimal mechanism either: (i) μ is a worst belief inducing $\hat{a}(\mu)$, or (ii) both Sender and Receiver are indifferent between two actions at μ.*

We have already discussed the basic intuition behind this Proposition: the expected posterior must equal the prior; hence more undesirable beliefs that induce a given action increase the probability of beliefs that induce a more desirable action. Proposition 12 is the reason why, in our

initial example, when the judge convicts she is barely willing to do so. Proposition 12 shows that the force behind this result applies more broadly.

Case (ii) is necessary to deal with the possibility of a belief which could be interpreted both as a worst or a best belief inducing $\hat{a}(\mu)$: if both Sender and Receiver are indifferent between, say a and a' at μ, the choice of $\hat{a}(\mu)$ from $\{a, a'\}$ is entirely arbitrary. However, for generic preferences, there will be no belief where both Sender and Receiver are indifferent between two actions at a same belief.

Finally, we consider the case where payoffs depend only on the expected state. Recall that we established earlier that in this case there is a function $\widetilde{v}$ s.t. $\widetilde{v}(E_{\mu}[\omega]) = \hat{v}(\mu)$, and that Sender benefits from persuasion if and only if $\widetilde{V}\left(E_{\mu_0}[\omega]\right) > \widetilde{v}\left(E_{\mu_0}[\omega]\right)$. We might conjecture that the value of an optimal mechanism is $\widetilde{V}\left(E_{\mu_0}[\omega]\right)$. This conjecture turns out to be false. Recall from the discussion of Proposition 6 that even though we know there is always a τ s.t. $E_{\tau}[E_{\mu}[\omega]] = E_{\mu_0}[\omega]$ and $E_{\tau}[\hat{v}(\mu)] = \widetilde{V}\left(E_{\mu_0}[\omega]\right)$, such τ need not be Bayes-plausible. In order to show that Sender could benefit from persuasion, we had to mix the beliefs in the support of τ with another belief μ' such that $E_{\mu'}[\omega] = E_{\mu_0}[\omega]$. This reduces the value of the mechanism strictly below $\widetilde{V}\left(E_{\mu_0}[\omega]\right)$.

For a concrete example, suppose that $A = [0,1]$, $\Omega = \{-1, 0, 1\}$, $u(a, \omega) = -(a-\omega)^2$, and $v(a, \omega) = a^2$. In this case, $\hat{v}(\mu) = \widetilde{v}\left(E_{\mu_0}[\omega]\right) = (E_{\mu}[\omega])^2$. Hence, $\widetilde{V}$ is constant at 1 and in particular $\widetilde{V}\left(E_{\mu_0}[\omega]\right) = 1$. Yet, whenever the prior puts any weight on $\omega = 0$, the value of any mechanism is strictly less than 1. Specifically, suppose that the prior μ_0 is $(\varepsilon/2, 1-\varepsilon, \varepsilon/2)$. In that case, when ε is close to 0, the value of an optimal mechanism is close to 0. Hence, when payoffs depend only on the expected state, we can use $\widetilde{v}$ to determine whether Sender benefits from persuasion, but *not* to determine an optimal mechanism or its value. To do that, we need to analyze $\hat{v}$ directly or derive the properties of an optimal mechanism from Propositions 11, 12, and 10.

6 Examples

In this section we develop several examples that are meant to demonstrate the breadth of settings captured by our model and illustrate the ways in which the results developed in the previous two sections can be applied.

6.1 Tenure-track midterm review

We begin with an example that illustrates the applicability of our results to problems from contract theory. We examine what type of information a university should provide to an assistant professor whose willingness to exert effort depends on her likelihood of getting tenure. This setting is one where structuring information provision may be a particularly useful way to induce effort because other instruments are less potent than usually. Paying the professor based on the quality of her research may be infeasible as this quality is likely to be non-verifiable, while the institutional structure of universities makes it difficult to motivate untenured faculty by suitably allocating property rights.

An assistant professor chooses an effort level $a \in [0,1]$ to exert before coming up for tenure. There are two types of individuals denoted by $\omega \in \{1,2\}$. The university and the professor share a prior μ_0 over the professor's type. The quality of research produced by an individual of type ω who exerts effort a is $a\omega$. At the end of the tenure clock, the individual is tenured if the quality of her research is above some cutoff level, say $3/2$.[18] If she is tenured, she receives utility $a\omega - a^2$, receiving the recognition for the quality of her research ($a\omega$), but suffering a disutility a^2 from her effort. If she is not tenured, she leaves academia and receives no recognition for her research but still suffers the sunk cost of effort, i.e. her utility is $-a^2$. The university wants to maximize the expected quality of the research produced by the professor. It conducts a midterm review which results in a signal $\pi : \{1,2\} \to \Delta(S)$. What type of a midterm review process maximizes the university's objective?

We begin by computing $\hat{v}$, denoting $\Pr(\omega = 2)$ by μ. Simple algebra reveals that the professor's

[18] Note that such a rule is feasible even if quality of research is non-verifiable if the university *wants* to give tenure when the quality exceeds this threshold.

optimal effort is:

$$\hat{a}(\mu) = \begin{cases} 0 & \text{if } \mu < 3/8 \\ 3/4 & \text{if } 3/8 \leq \mu < 3/4 \\ \mu & \text{if } \mu \geq 3/4. \end{cases}$$

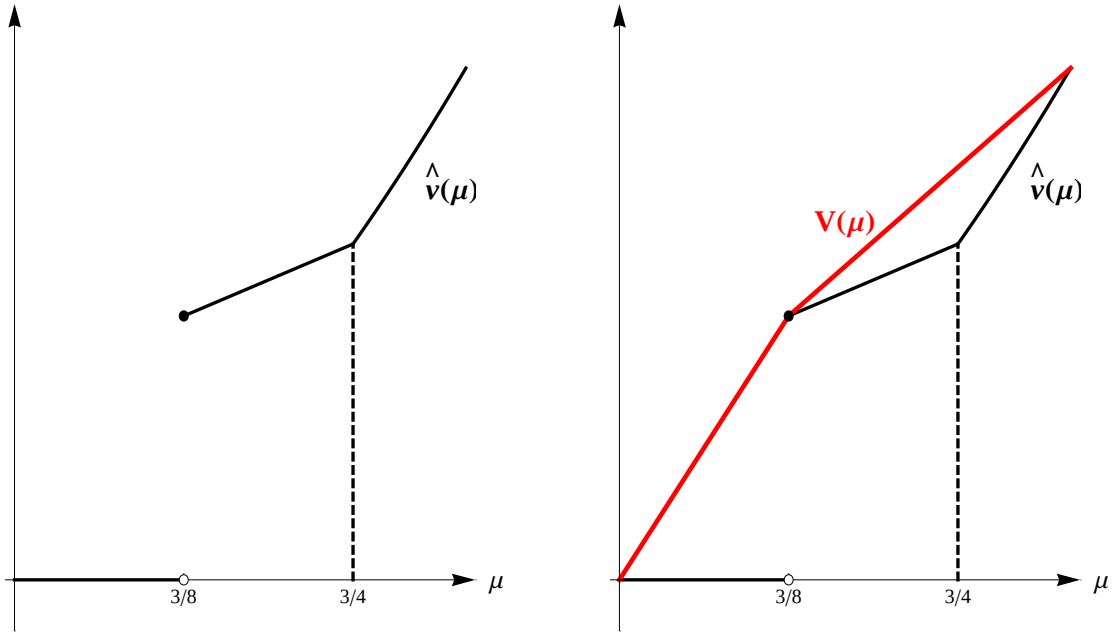
Hence, the university's expected utility given the professor's belief is:

$$\hat{v}(\mu) = \begin{cases} 0 & \text{if } \mu < 3/8 \\ 3/4\,(1+\mu) & \text{if } 3/8 \leq \mu < 3/4 \\ \mu + \mu^2 & \text{if } \mu \geq 3/4. \end{cases}$$

What do our Propositions tell us about this example? Propositions 2 and 8 do not apply since $\hat{v}$ is neither convex nor concave. Also, since there is information the university would share, Proposition 3 does not rule out the possibility that the university might benefit from persuasion. In fact, Proposition 4 directly tells us that, at least if $\mu_0 < 3/8$, the university will benefit from persuasion. Proposition 9 tells us that the optimal midterm review need not induce more than two distinct effort levels. Finally, Proposition 10 implies that whenever the professor exerts no effort, she is completely certain that she is a low type. The reason for this feature of the optimal review is that any posterior weight on the high type is "wasted" when the posterior is below 3/8 since that weight is still insufficient to lift effort below zero and yet it reduces the probability of a more favorable, effort-inducing, opinion the professor can have about herself.

Because of the simplicity of the state space in this example, we can also easily depict $\hat{v}$ and its concave closure (Figure 4). The figure makes it clear that the university generically benefits from persuasion.[19] Moreover, it shows that the optimal structure of the performance review depends on the prior. When the initial probability that the professor is a high type is below 3/8, the optimal review induces posteriors $\mu = 0$ and $\mu = 3/8$. When the prior is above 3/8, then the optimal review induces $\mu = 3/8$ and $\mu = 1$. Note that regardless of what the prior is, it is never optimal to fully reveal the type. When the prior is below 3/8, revealing that the type is high with certainty is very costly because this realization happens too rarely relative to the effort it induces. When the

[19] No disclosure is optimal only when the prior is exactly 3/8.



(a) function $\hat{v}$ (b) function V

Figure 4: optimal midterm review

prior is above 3/8, revealing that the type is low with certainty is very costly because effort drops discontinuously when prospect of tenure becomes too dim (i.e., when μ falls below 3/8).

This simple model illustrates the way in which the tools developed in this paper can be used to study optimal feedback when actions are not fully contractible. Moreover, as we mentioned earlier, by redefining v we can use the same approach to find mechanisms that maximize social welfare rather than the university's preferences.

6.2 Preference disagreement

Here we consider the question of how certain types of preference disagreement between Sender and Receiver affect the amount of information transmitted under the optimal persuasion mechanism. Let $u = -(a-\omega)^2$, $v = -(a-(b_1+b_2\omega))^2$, and $A = \Omega = [0,1]$.[20] Parameters b_1 and b_2 capture two types of preference disagreement: values of b_1 away from 0 indicate that Sender and Receiver disagree about the best average level of a while values of b_2 away from 1 indicates that they disagree about how the action should vary with the state.

[20]Recall that Appendix B shows our approach extends to compact metric state spaces.

In order to determine how information transmission depends on b_1 and b_2 it will suffice to compute $\hat{v}$. Since $u = -(a-\omega)^2$ we know that

$$\hat{a}(\mu) = E_\mu[\omega].$$

Given this $\hat{a}$, simple algebra reveals that

$$\hat{v}(\mu) = -b_1^2 + 2b_1(1-b_2)E_\mu[\omega] - b_2^2 E_\mu\left[\omega^2\right] + (2b_2-1)(E_\mu[\omega])^2.$$

Hence, by Corollary 1, Sender solves

$$\begin{aligned} &\max_\tau \; E_\tau\left[-b_1^2 + 2b_1(1-b_2)E_\mu[\omega] - b_2^2 E_\mu\left[\omega^2\right] + (2b_2-1)(E_\mu[\omega])^2\right] \\ &\text{s.t.} \int \mu d\tau(\mu) = \mu_0. \end{aligned}$$

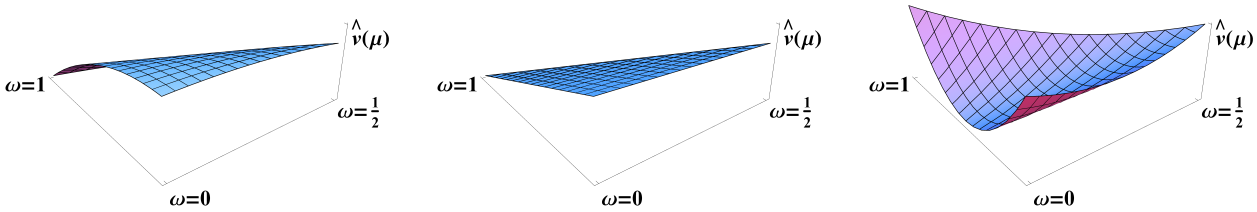
Since the $-b_1^2 + 2b_1(1-b_2)E_\mu[\omega] - b_2^2 E_\mu\left[\omega^2\right]$ term is constant across all Bayes-plausible τ's, this maximization problem simplifies to

$$\begin{aligned} &\max_\tau \; E_\tau\left[(2b_2-1)(E_\mu[\omega])^2\right] \\ &\text{s.t.} \int \mu d\tau(\mu) = \mu_0. \end{aligned}$$

Hence, b_1 does not affect Sender's problem. Moreover, $\hat{v}$ is linear when $b_2 = \frac{1}{2}$, strictly convex when $b_2 > \frac{1}{2}$, and strictly concave when $b_2 < \frac{1}{2}$. Therefore, by Proposition 8, no disclosure is uniquely optimal when $b_2 < \frac{1}{2}$ and full disclosure is uniquely optimal when $b_2 > \frac{1}{2}$. When $b_2 = \frac{1}{2}$ all mechanisms yield the same value.

If we consider a simpler state space, we can develop a geometric intuition for this example. Keeping utility functions and the action space the same, suppose that $\Omega = \left\{0, \frac{1}{2}, 1\right\}$.[21] Figure 5 depicts $\hat{v}$ for a few values of b_2. Recall that, for the reasons we just discussed, the shape of $\hat{v}$ does not depend on b_1; this parameter only affects the vertical location $\hat{v}$. Figure 5 clearly demonstrates that b_2 changes the concavity of $\hat{v}$. While these graphs provide some geometric intuition, it is

[21] Note that the expression for $\hat{v}$ above still applies with this simplified state space.



(a) $b_2 = 0$ (b) $b_2 = \frac{1}{2}$ (c) $b_2 = 1$

Figure 5: shape of $\hat{v}$ depends on b_2

worthwhile to note that the algebra above provides an illustration of the practical usefulness of formulating Sender's problem in terms of $\hat{v}$ even when the state space is too rich for this function to be depicted.

Now, what do the results above tell us about the impact of preference disagreement on information transmitted in an optimal mechanism? First, note that b_1 has no impact on the relative value of any two mechanisms. This stands in sharp contrast to the results from cheap talk games. Crawford and Sobel's (1982) main example considers preferences of the form $u = -(a-\omega)^2$, $v = -(a-\omega-b_1)^2$, which are equivalent to our example when $b_2 = 1$. They show that as soon as $|b_1| > \frac{1}{4}$, no communication is possible in any equilibrium of the cheap talk game. By comparison, the optimal persuasion mechanism induces full disclosure for any value of b_1. The reason for this difference is that Sender's commitment frees him from the temptation to try to increase the *average* level of Receiver's action, the temptation which is so costly both for him and for Receiver in the equilibrium of the cheap talk game. Under any mechanism, Receiver's expected action is a given. Moreover, conditional on Receiver's expected action, Sender and Receiver completely agree on which actions are more desirable in which state. Hence, providing Receiver with anything short of a fully informative signal only induces an additional cost to Sender.

The other parameter, b_2, measures disagreement over how the action should vary with the state. When $b_2 = 1$, Sender and Receiver completely agree on how action should vary with the state. Hence, full disclosure is uniquely optimal. When $b_2 > 1$, Receiver's actions are insufficiently sensitive to the state from Sender's perspective. Because Receiver *under*-reacts to information, however, Sender still strictly prefers to fully disclose everything, even if preference disagreement is very large, because anything short of that would only exacerbate the problem of

under-reaction. When $0 < b_2 < 1$, Receiver's actions are overly sensitive to the state, though the action always moves in the right direction. When the level of disagreement is sufficiently small $\left(b_2 > \frac{1}{2}\right)$, Sender still prefers to reveal all information, but when Receiver's over-reaction becomes too severe $\left(b_2 < \frac{1}{2}\right)$, Sender cuts all information flow to Receiver. Finally, when $b_2 < 0$, Receiver reacts to any information in a way opposite to the one Sender wishes, so he reduces her information as much as possible.

6.3 Informative advertisements with unit demand

Now consider an example where a firm faces a continuum of *ex ante* identical consumers, each of whom decides whether to buy one unit of the firm's product or not. The firm's product has quality $\omega \in [0,1]$ and the consumers' utility from purchasing the product is $\omega - p$, where p is an exogenous price, also in the unit interval. The consumers and the firm share the prior μ_0 on the quality of the product. The assumption of symmetric information is most palatable if we conceptualize quality as driven by the match between consumers' uncertain tastes and the products' uncertain characteristics. Since consumers are risk-neutral, each will buy the product if and only if $E_\mu[\omega] \geq p$ where μ is their posterior belief about the quality of the product. To make things interesting, we suppose that consumers's default action is not to buy, i.e., $E_{\mu_0}[\omega] < p$.

The firm chooses an advertising campaign which provides a signal about quality $\pi : [0,1] \to \Delta(S)$. All consumers see the advertisement. We assume that the firm is risk-neutral and only cares about the expected revenue, so it does not matter whether each consumer gets an independent realization of the signal (which would mean they end up with heterogeneous posteriors) or all consumers get the same realization (which means they would have identical posteriors). Because consumers are *ex ante* homogeneous, our results apply directly to this example even though we have multiple Receivers.[22]

We again begin by computing $\hat{v}$. Denoting the decision to buy with 1 and the alternative with

[22]This example is closely related to the analysis in Lewis and Sappington (1994).

0, we have

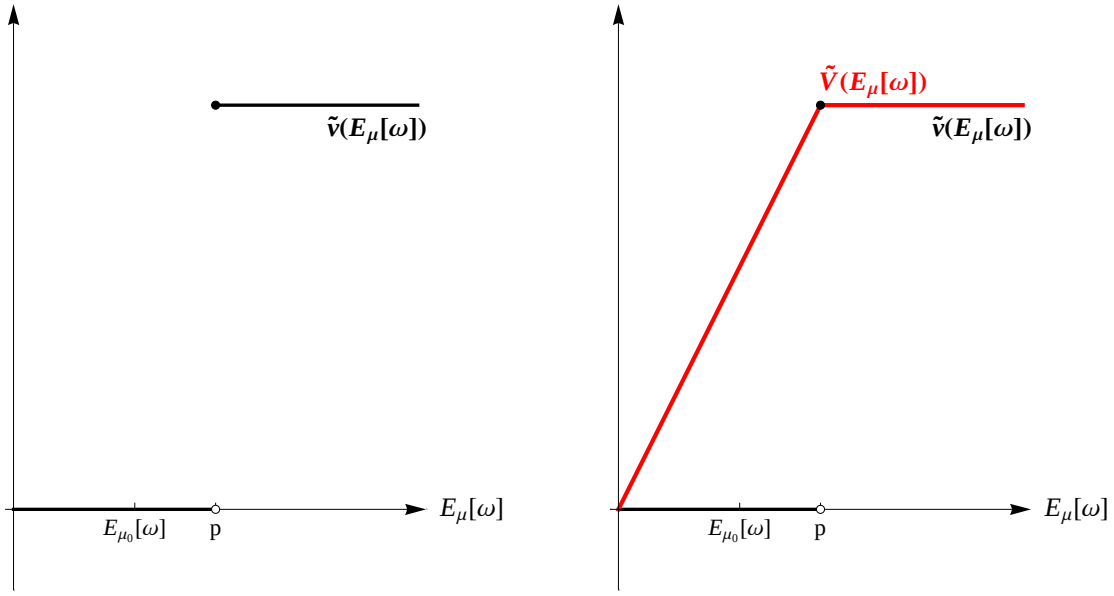
$$\hat{a}(\mu) = \begin{cases} 0 & \text{if } E_\mu[\omega] < p \\ 1 & E_\mu[\omega] \geq p \end{cases}$$

$$\hat{v}(\mu) = \begin{cases} 0 & \text{if } E_\mu[\omega] < p \\ 1 & E_\mu[\omega] \geq p \end{cases}.$$

Moreover, since payoffs depend only on the expected state, we can define $\widetilde{v}(E_\mu[\omega]) = \hat{v}(\mu)$.

What do our Propositions tell us about this example? Propositions 2 and 8 do not apply since $\hat{v}$ is neither convex nor concave. Because $E_{\mu_0}[\omega] < p$, there is information the firm would share. Hence, Proposition 3 does not exclude the possibility that the firm might benefit from persuasion. In fact, since consumers' preference for not buying is discrete at the prior, Proposition 4 implies that the firm does benefit from persuasion. Moreover, we know that any consumer who buys the product will have the posterior μ s.t. $E_\mu[\omega] = p$ (Proposition 12), while any consumer who does not buy will have a posterior that puts zero weight on the possibility that $\omega \geq p$ (Proposition 10). This tells us that the optimal advertising campaign induces two types of reactions in consumers: some consumers become completely convinced that the product is not worth its price, while others buy the product, albeit reluctantly. The basic intuition for this stems from the fact that worsening the opinion of the product by those who are already not buying it is costless to the firm, and yet it increases the probability of more favorable reactions by others (because of Bayes-plausibility). Hence, those who do not buy the product are driven to complete certainty that it is not worth buying. On the other hand, improving the opinion of the product by those who are already willing to buy it does not generate further sales, and yet it decreases the likelihood of such favorable impressions. Hence, the optimal advertising campaign makes all buyers marginal.

In this example, $\hat{v}$ is difficult to visualize but since payoffs depend only on the expected state, we can depict $\widetilde{v}$ and $\widetilde{V}$ (Figure 6). Since $\widetilde{V}\left(E_{\mu_0}[\omega]\right) > \widetilde{v}\left(E_{\mu_0}[\omega]\right)$, Proposition 6 tells us there exists an advertising campaign that increases firm's revenue. Also, as we discussed at the end of Section 5, while we cannot determine the optimal campaign by examining $\widetilde{v}$, we know that $\widetilde{V}\left(E_{\mu_0}[\omega]\right)$ is an upper bound on the market share that can be achieved by any advertising campaign.



(a) function $\hat{v}$ (b) function V

Figure 6: advertising to increase sales

7 Extensions and Limitations

7.1 Dynamic Mechanisms

We have restricted our attention to a class of persuasion mechanisms where there is a single stage of communication. Sender sees only one draw of private information, and sends only one message to Receiver. In reality, however, persuasion often happens over multiple periods. Firms may send multiple ads, a prosecutor may go through successive rounds of gathering and reporting information, and so forth.

Our framework can easily accommodate one class of dynamic mechanisms. Consider an extended definition of a persuasion mechanism that has T periods with possibly different signals and messaging technologies (π_t, c_t) at each stage. Sender privately observes realization $s_1 \in S_1$ from $\pi_1(\cdot|\omega)$ and chooses a message $m_1 \in M_1$ which Receiver observes. Sender then privately observes realization $s_2 \in S_2$, sends message $m_2 \in M_2$, observes realization $s_3 \in S_3$ and so on up to period T. After Receiver observes the final message m_T, she chooses her action $a \in A$. We define an equilibrium of such a mechanism exactly as before.

Any equilibrium of such a dynamic mechanism must still induce a distribution of posteriors τ

at the point that Receiver chooses her action, and the expected value of v will still be equal across all mechanisms that induce the same distribution τ. Therefore, the argument that was used to prove Proposition 1 implies that if there exists some dynamic mechanism with value v^* there also exists a straightforward static mechanism with value v^*. This means that all our core results on optimal mechanisms and the situations where Sender benefits from persuasion apply directly to this dynamic case as well.

A different class of dynamic mechanisms is one where Receiver chooses an action in each period. Suppose, for example, that we modify the definition of a dynamic mechanism above and assume that after observing m_1 Receiver chooses action $a_1 \in A$; after observing m_2 Receiver chooses action $a_2 \in A$; and so on. Sender's and Receiver's payoffs are $v(a_1, ..., a_T, \omega)$ and $u(a_1, ..., a_T, \omega)$, respectively.

This case is much more complicated. Sender's expected payoff now depends not only on the distribution of Receiver's beliefs at time T, but also on the distribution of her beliefs at each stage along the way. In fact, it depends on the *joint* distribution of beliefs in each period, $\tau \in \Delta(\Omega^T)$. We conjecture that we can still express Sender's expected utility as $E_\tau \hat{v}(\mu_1, ..., \mu_T)$, and moreover, that it is still possible to restrict attention to something analogous to honest mechanisms, provided that Sender's private signals can be conditioned on Receiver's past actions. However, the construction of $\hat{v}$ for any given problem is now more complex, and the Bayes-plausibility constraint for sequences of beliefs is more complicated than for a single distribution of beliefs. We have not attempted to extend either our geometric intuitions or our analytical results to this case.

7.2 Receiver's Private Information

Extending our analysis to situations where Receiver has private information is straightforward. Given a mechanism (π, c), suppose that the timing of the game is as follows. First, nature selects ω from Ω according to μ_0. Neither Sender nor Receiver observe nature's move. Then, Sender privately observes a realization $s \in S$ from $\pi(\cdot|\omega)$ and chooses a message $m \in M$. Then, Receiver privately observes a realization $r \in R$ from some signal $\chi(\cdot|\omega)$. Finally, Receiver observes m and chooses an action $a \in A$. Note that if Receiver instead observes her private information before Sender observes his signal or before he sends his message, the game can be re-formulated with the

timing above without loss of generality because Sender's action is independent of Receiver's private information (since he does not observe it) and Receiver always chooses her action after observing m. We still assume that Sender and Receiver share a prior μ_0 at the outset of the game.

The only way in which Receiver's private information changes our analysis is that we can no longer construct a deterministic function $\hat{a}(\mu)$ which gives Receiver's action at any belief. Rather, for any μ, Receiver's optimal action $\hat{a}(\mu, r)$ depends on the realization of her private signal and so is stochastic from Sender's perspective. When his posterior is μ, Sender assigns probability $\chi(r|\omega)\mu(\omega)$ to the event that Receiver's signal is r and the state is ω. Hence, Sender's expected utility when he induces belief μ is:

$$\hat{v}(\mu) = \sum_{\omega \in \Omega} \sum_{r \in R} v(\hat{a}(\mu, r), \omega) \chi(r|\omega) \mu(\omega).$$

Once we reformulate $\hat{v}$ this way, our approach applies directly. In particular, our key simplifying results—Proposition 1, Corollary 1, Corollary 2—still hold. Aside from the fact that constructing $\hat{v}$ is slightly more complicated, the analysis of the problem in terms of the properties of $\hat{v}$ and its concave closure V proceeds exactly as before. That said, some of our characterization results, such as that Receiver's preference is never discrete at any interior μ induced by an optimal mechanism, will no longer hold.

A different type of situation that involves private information is when Receiver's preferences depend on some parameter $\theta \in \Theta$ which is unrelated to ω. This distinction matters because we still assume that Sender's private signal is informative only about ω. Hence, no mechanism provides any additional information to Sender about θ. For example, in our motivating example, the judge might have unobservable, idiosyncratic aversion toward convicting an innocent defendant, which would affect the posterior cutoff at which she is willing to convict.[23] The prosecutor's investigation, however, would not provide any information about the judge's preferences.

Despite this distinction, we can handle this situation in the same way as the one before. Again, the only impact of private information is that Receiver's optimal action $\hat{a}(\mu, \theta)$ is stochastic from Sender's perspective. Letting ϕ denote the distribution of θ, Sender's expected utility when he

[23] Another example of this type of private information is the value of Receiver's outside option in Rayo and Segal (2008).

induces belief μ is:

$$\hat{v}(\mu) = \sum_{\omega \in \Omega} \int_{\Theta} v(\hat{a}(\mu, \theta), \omega) \, d\phi(\theta) \, \mu(\omega).$$

Here again, once we thus reformulate $\hat{v}$, we can analyze the problem in terms of $\hat{v}$ and V in exactly the same way as before.

7.3 Multiple Receivers

In many settings of interest—politicians persuading voters, firms advertising to consumers, auctions—our assumption that there is a single Receiver is unrealistic. Suppose there are n receivers. For ease of exposition we maintain our common prior assumption, which in this setting means that Sender and all receivers share a prior μ_0 over Ω.[24] Sender's utility is now a function of each receiver's action: $v(a_1, ..., a_n, \omega)$. There are two classes of multiple-receiver models where our results can be extended quite easily.

The first class is one where Sender sends separate (possibly correlated) messages to each receiver, Sender's utility is separable across receivers' actions,[25] and each receiver cares only about her own action. In this case, we can simply apply our approach separately to Sender's problem *vis-à-vis* each receiver. Since Sender's utility is separable, each receiver sees only her own message, and no receiver cares about what other receivers are doing, we basically have n copies of our standard problem with a single Receiver. In the special case where all receivers have the same utility function, the optimal mechanism will of course be the same for each receiver, so the analysis collapses to a single problem identical to the one we have analyzed before, as in the example of section 6.3.

The second class of models is where Sender can only persuade by revealing *public* information. That is, any message from Sender is observed by all receivers. In this case, our approach applies no matter whether receivers then choose their individual actions simultaneously, in sequence, or according to some other game. Moreover, Sender's utility need not be separable across receivers' actions, receivers might care about each other's actions, and they might have heterogeneous util-

[24] There are no additional complications from having both multiple receivers and private information on their part. The approach from the previous Subsection for dealing with private information applies equally well to the case with multiple receivers.

[25] When v is not separable across the a_i's the problem is similar to that of dynamic mechanisms where Receiver chooses an action in each period.

ity functions. An example of a setting like this is Milgrom and Weber's (1982) model of public information revelation in auctions with a common-value component.[26]

For simplicity, consider the case where the equilibrium of the post-message game is in pure strategies.[27] If the post-message game does not have a unique equilibrium, we focus on an equilibrium which yields the highest payoff to Sender, analogously to our earlier equilibrium selection rule. Let $\hat{a}_i(\mu)$ represent the ith receiver's equilibrium action when she has belief μ. We can then define $\hat{v}$ as a function of receivers' shared posterior μ:

$$\hat{v}(\mu) \equiv \sum_{\omega \in \Omega} v(\hat{a}_1(\mu), ..., \hat{a}_n(\mu), \omega) \mu(\omega).$$

With this reformulation of $\hat{v}$, our basic approach again applies. Proposition 1, Corollary 1, and Corollary 2 all still hold. Constructing $\hat{v}$ is potentially much more complicated here since it involves solving for the equilibria of the post-message game, but the analysis of the problem in terms of the properties of $\hat{v}$ and V is exactly the same as before. Of course, characterization results which are stated in terms of Receiver's preferences, such as Proposition 4, would have to be reinterpreted.

There is an important third class of multiple-receiver models where our results do not extend easily: those where the receivers care about each other's actions and Sender can send private signals to individual receivers. The crucial problem with this case is that for a given set of beliefs that receivers hold after observing their messages, the receivers' actions may vary as a function of the *mechanism* that produced those beliefs. In a common value auction for example, a bidder with a given belief will behave differently if she believes that other bidders are receiving highly informative signals than if she believes they are receiving uninformative signals. This means that the key simplifying step in our analysis—reducing the problem of finding an optimal mechanism to one of maximizing over distributions of posterior beliefs—does not apply.

[26] Milgrom and Weber (1982) allow for bidders to have private information. As we mentioned earlier, the previous Subsection provides a way of incorporating that possibility.

[27] If the equilibrium is in mixed strategies, the only additional complication is that actions are stochastic for a given belief, but we have already shown that this poses no problems for our approach.

7.4 Multiple Senders

Our model can also be used to think about settings with multiple senders. Suppose there is a single Receiver who receives messages from multiple senders. Receiver observes all the messages and then takes a single action. All senders and Receiver share a common prior and each sender's utility depends on Receiver's action and the state of the world.

If we simply wish to know whether there is a an informational environment that increases some weighted function of senders' utilities, our previous results apply directly since they do not depend on any particular interpretation of the objective function $v(a, \omega)$. Hence, we could simply let v be any weighted function of senders' utilities. A more interesting question is what happens if multiple senders play a non-cooperative game. Specifically, consider a game where each sender i simultaneously[28] chooses a $\pi_i : \Omega \to \Delta(S_i)$, Receiver then observes all s_i's and takes her action.[29] Taking other senders' choices as given, each sender's problem is identical to one in Subsection 7.2 where he is the only Sender and Receiver has some private information about ω. The private information here is simply the set of signal realizations from all the other senders. Hence, our tools for finding an optimal mechanism provide a way to compute the best-response functions. Unless one can solve for an optimal mechanism analytically, however, solving for the equilibria of this game might be challenging. Another issue is that we can no longer speak of Receiver taking a "Sender-preferred" action from $a^*(\mu)$. Consequently, an optimal mechanism, and hence a best-response function, might not always exist. To guarantee the existence of an equilibrium of this game we would need to assume there is a uniquely optimal action for Receiver at each belief.

7.5 Limited set of signals

Throughout the paper we have assumed that the mechanism designer can choose any signal π whatsoever. In many settings, this might be an unreasonable assumption. What can we say about the case where only some subset of potential signals, say Π, is feasible? We can still formulate our

[28] The analysis of games where senders move sequentially is very similar.

[29] For the same reason that we can restrict our attention to honest mechanisms in our main model, this game is isomorphic to one where each sender chooses a (π, c) pair, observes a signal realization s from π, and then sends a message m to Receiver.

problem as a search over distributions of posteriors. The value of an optimal mechanism is

$$\begin{aligned} &\max_{\tau} E_{\tau}\hat{v}(\mu) \\ &\text{s.t.} \int \mu d\tau(\mu) = \mu_0 \\ &\text{and } \tau \in \Gamma \end{aligned}$$

where Γ denotes distributions of posteriors induced by honest mechanisms with a signal in Π. When this additional constraint binds, however, we will not necessarily be able to use our geometric approach. Knowing that $V(\mu_0) > \hat{v}(\mu_0)$ does not tell us whether there is a Bayes-plausible $\tau \in \Gamma$ s.t. $E_{\tau}\hat{v}(\mu) > \hat{v}(\mu_0)$.

There is one particular type of limitation on Π, however, which our approach handles easily. Suppose that we can express the state space as $\Omega \times \Theta$, denoting a particular state by (ω, θ). Moreover, suppose that the set of potential signals Π consists of all signals that provide no information about θ. Then, the situation is closely analogous to one in Subsection 7.2 where the space of signals is arbitrarily rich, but Receiver's preferences depend on some parameter $\theta \in \Theta$ which is unrelated to ω. The only difference is that Receiver's action no longer depends on θ, but Sender's utility $v(a, (\omega, \theta))$ now does. This poses no further complications. Sender's utility when he induces a belief μ on Ω is simply:

$$\hat{v}(\mu) = \sum_{\omega \in \Omega} \int_{\Theta} v(\hat{a}(\mu), (\omega, \theta)) \, d\phi(\theta) \, \mu(\omega),$$

where ϕ denotes his prior on Θ. Again, with this reformulation of $\hat{v}$, Proposition 1, Corollary 1, and Corollary 2 all still hold.

Finally, the possibility that Γ might not always include all Bayes-plausible τ's has important implications. Consider a modification of our initial example. Suppose there are two types of prosecutors: a high-ability prosecutor who can structure his investigation so as to generate any signal π, and a low-ability prosecutor who has access to a limited set of signals. If this set does not include the optimal signal, the high-ability prosecutor will convict a higher percentage of defendants than the low-ability one even if the judge is fully aware of the difference in prosecutors' abilities. The rationality of the judge does not imply that she will somehow compensate for the prosecutor's

type. More broadly, the benefit to Sender from expanding Γ might partly explain the observed large expenditures on persuasive activities.

7.6 Limited Commitment

What can we say about settings where Sender is unable to commit to an honest mechanism?

The first thing to note is that our results provide an upper bound on gains from communication in any persuasion mechanism. By Proposition 1, we know that the value to Sender from being able to communicate with Receiver can only be weakly lower when he cannot commit to an honest mechanism. This observation has important implications even in models without commitment. Consider, for example, Spence's (1973) signalling model. Since the worker's wage in that model is the expectation of his type, we know that $\hat{v}$ is linear, so the worker (Sender) cannot benefit from persuasion. Hence, even without solving for the equilibria of this signalling game, we know that in any equilibrium the average worker would be weakly better off if a government policy outlawed education.[30]

More broadly, in any game captured by our definition of a persuasion mechanism, Sender values his ability to communicate with Receiver no more than $V(\mu_0) - \hat{v}(\mu_0)$. However, this provides only an upper bound to the value of communication in games without commitment. In some settings, $V(\mu_0) - \hat{v}(\mu_0)$ might be large and yet Sender might not benefit at all from an opportunity to communicate with Receiver.

In the remainder of this section we analyze more directly what the gains from communication are when Sender cannot commit to an honest messaging technology but retains his choice of what information to gather. Specifically, we analyze the choice of an optimal signal when the message technology c is constant. We call such mechanisms *commitment-free*. If there is a commitment-free mechanism with a strictly positive gain, we will say that *beneficial persuasion without commitment is possible*.

The analysis of commitment-free mechanisms is closely related to the questions in Green and Stokey (2007). Assuming cheap talk messaging, they also consider the benefits that might arise from reducing the informativeness of Sender's signal.[31] In contrast to their paper, which focuses on

[30] In fact, in any equilibrium with positive signalling costs, he would be strictly better off under such policy.

[31] Note that Crawford and Sobel (1982) implicitly assume that Sender receives a fully informative signal. They

local improvements in informativeness, we here derive bounds on the benefit from persuasion when Sender has a choice over an arbitrary signal. The analysis here is also related to Ivanov (2008). He, however, focuses on the question of when restricting Sender's information benefits Receiver, while we explore when Sender can be made better off by being less informed. Throughout the analysis we assume that Receiver observes what information Sender has. In a different game, where Sender chose his signal covertly, the only equilibrium would be for Sender to choose the fully informative signal and the set of equilibria would be isomorphic to those of Crawford and Sobel (1982).

As in all cheap talk settings, incentive compatibility is the key obstacle to communication in commitment-free mechanisms. Sender, however, can often choose to gather information which mitigates the incentive compatibility problem: he can choose to learn information with the property that *ex post* his utility is maximized by truthfully reporting what he learns. When the chosen signal has this property, Receiver can accept Sender's messages at face value.

Consider the case where Sender's preferences are independent of the state. In this case, Sender can credibly convey messages if and only if he is completely indifferent between the actions they induce. For example, suppose $\Omega = \{0,1\}$, $A = [0,1]$, $\mu_0 = \frac{1}{4}$, $u = -(a-\omega)^2$, and $v(a,\omega) = \left(a - \frac{1}{4}\right)^2$. Here, Sender's optimal commitment-free mechanism is a signal π on $\{l,h\}$ with

$$\begin{aligned} \pi(l|0) &= \tfrac{2}{3} \quad \pi(l|1) = 0 \\ \pi(h|0) &= \tfrac{1}{3} \quad \pi(h|1) = 1. \end{aligned}$$

This signal induces posteriors $\mu_l(1) = 0$ and $\mu_h(1) = \frac{1}{2}$, which yield exactly the same value of $\hat{v}$. Hence, Sender can truthfully reveal the signal realization. By contrast, if Sender were fully informed, Receiver would not believe his reports that $\mu(1) = 1$ and thus Sender would not benefit from persuasion.

This example is an instance of a more general result that the gain which Sender can obtain when his utility is independent of the state is equal to the distance between $\hat{v}(\mu_0)$ and the greatest

assume that Sender observes the value of an arbitrary random variable, but since both Sender's and Receiver's utilities are defined solely in terms of this random variable, the state space is effectively the set of realizations of the random variable and Sender is thus perfectly informed of the state.

value of $\hat{v}$ that is constant across beliefs whose convex hull includes the prior. Let

$$E = \sup \{\overline{v} | \exists M \subset \Delta(\Omega) \text{ s.t. } \hat{v}(\mu) = \overline{v} \ \forall \mu \in M \text{ and } \mu_0 \in co(M)\} - \hat{v}(\mu_0).$$

Because $\hat{v}$ is not necessarily continuous, the sup in the expression above may not be obtained, and hence an optimal commitment-free mechanisms might not always exist. However, we can always find a commitment-free mechanism whose value is arbitrarily close to E.

Proposition 13 *Suppose v is independent of ω. The gain from any commitment-free mechanism is weakly less than E. For any $\varepsilon > 0$, there exists a commitment-free mechanism whose gain exceeds $E - \varepsilon$. Beneficial persuasion without commitment is possible if and only if $E > 0$.*

Often, E will be equal to zero and beneficial persuasion without commitment will be possible only if v depends on the state. When v varies with ω, Sender can mitigate the incentive compatibility constraint by choosing a signal that induces posteriors at which his preferences are aligned with those of Receiver. Consider the following example. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $A = \{a_1, a_2\}$. The beliefs where an agent is indifferent between the two actions form a straight line in the probability triangle. Suppose preferences are such that these lines are as in Figure 7. The line where Receiver is indifferent between the actions is the darker, steeper one; she prefers a_1 when her beliefs are in the Southeast portion of the simplex. The line where Sender is indifferent is the one that is less dark and less steep; he prefers a_1 in the Northwest portion of the simplex.

For most beliefs, Sender and Receiver disagree on which action is preferable. The shaded areas indicate the small regions of agreement. Let Z denote the union of these two areas. Because they disagree on the appropriate action at the degenerate beliefs, no communication would be possible if Sender were perfectly informed. If Receiver takes both a_1 and a_2 in equilibrium, Sender will always send a message that induces his preferred action. But Receiver knows that, for any degenerate belief that Sender has, she prefers the other action. Thus, an equilibrium where both actions are taken cannot exist.

If, however, Sender limits his information, both Receiver and Sender can benefit from communication. In particular whenever μ_0 belongs to the convex hull of Z, but not to Z itself, beneficial persuasion without commitment is possible. For instance, Figure 7 illustrates that if Sender selects

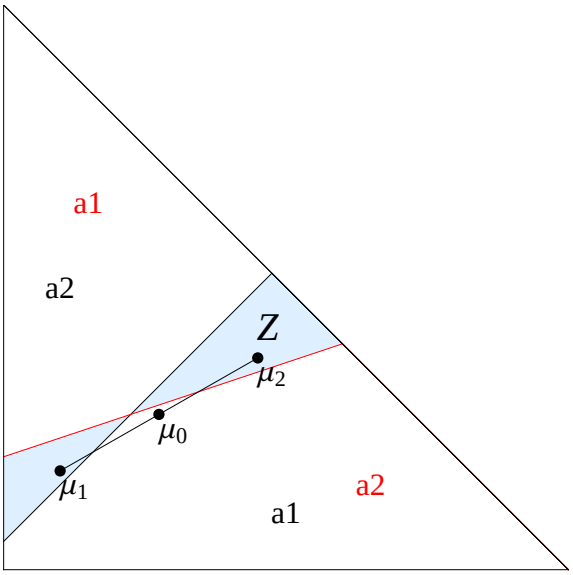


Figure 7: persuasion without committment

π which induces μ_1 and μ_2, he will have the incentive to truthfully reveal the signal realization. Receiver will be aware of this and both will benefit from persuasion without commitment.

The gain that Sender can achieve from persuasion due to such alignment of preferences is bounded by the extent to which his preferences vary with the state. Let

$$D = \max_{a \in A} \left(\max_{\omega} v(a, \omega) - \min_{\omega} v(a, \omega) \right)$$

denote a measure of the sensitivity of v to ω. When $\hat{v}(\mu)$ is monotonic,[32] Sender can achieve incentive compatibility only by selecting posteriors at which his preferences are aligned with Receiver's, and the gain from any commitment-free mechanism cannot exceed D.

Proposition 14 *If $\hat{v}$ is monotonic, the gain from any commitment-free mechanism is at most D.*

Since E is necessarily equal to zero when $\hat{v}$ is monotonic, both Proposition 13 and Proposition 13 yield the following Corollary.

Corollary 3 *If $\hat{v}$ is monotonic and v is state-independent, beneficial persuasion without commitment is not possible.*

[32] A condition weaker than monotonicity will also suffice. Say $\hat{v}$ is *without troughs* if there do not exist μ_a, μ_b, μ_c and $\gamma \in (0,1)$ s.t. $\mu_b = \gamma \mu_a + (1-\gamma)\mu_c$ and $\min\{\hat{v}(\mu_a), \hat{v}(\mu_c)\} > \hat{v}(\mu_b)$. As the proof in Appendix A shows, the subsequent Proposition and Corollary both hold if we assume $\hat{v}$ is without troughs rather than monotonic.

8 Conclusion

There are two ways to induce a person to do something. One is to incentivize her, whether by increasing her reward from taking a particular action, or by increasing her punishment for doing something else. Such incentive schemes can be blunt, as when you pay someone for performing a task or coerce her into doing it, or more subtle, as when you increase the supply of goods complementary to an activity. All such schemes, however, rely on changing the individual's marginal preferences. The other way to induce a person to do something is to change her beliefs. Changes in beliefs can change the expected action for several reasons. One is that the person being persuaded may fail to fully account for the motives of the persuader. Another is that overwhelming an individual with information may make it too difficult to process all of it an appropriate way. Yet another, which is the focus of this paper, is that even a perfectly rational Bayesian can often be persuaded.

9 Appendix A: Proofs

9.1 Proof of Proposition 1

As we mentioned in the text, (2) immediately implies (1) and (3). We first show that (1) implies (2).

Consider an equilibrium $\varepsilon^o = (\sigma^o, \rho^o, \mu^o)$ of some mechanism (π^o, c^o) with value v^*. For any a, let M^a be the set of messages which induce a: $M^a = \{m|\hat{a}(\mu) = m\}$ in this equilibrium. Now, consider a straightforward mechanism with $S = A$ and

$$\pi(a|\omega) = \sum_{m \in M^a} \sum_{s \in S} \sigma^o(m|s)\, \pi^o(s|\omega).$$

In other words, in the proposed mechanism Sender "replaces" each message with a recommendation of the action that the message induced. Since a was an optimal response to each $m \in M^a$ in ε^o, it must also be an optimal response to the message a in the proposed straightforward mechanism. Hence, the distribution of Receiver's actions conditional on the state is the same as in ε^o. Therefore, if we set k equal to the messaging costs in ε^o, i.e., $k = \sum_m \sum_s \sum_\omega c(m|s)\, \sigma^o(m|s)\, \pi^o(s|\omega)$, the

value of the straightforward mechanism is exactly v^*. Hence, (1) implies (2).

It remains to show that (3) implies (1). In other words, we need to show that given any Bayes-plausible distribution of posteriors, there exists a mechanism that induces it. We will show a stronger claim that there exists an honest mechanism that induces it. An honest mechanism with signal π induces τ if $Supp(\tau) = \{\mu_s\}_{s \in S}$ and

$$\begin{aligned} \text{(i) } \mu_s(\omega) &= \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega')} \text{ for all } s \text{ and } \omega \\ \text{(ii) } \tau(\mu_s) &= \sum_{\omega \in \Omega} \pi(s|\omega)\mu_0(\omega) \text{ for all } s. \end{aligned}$$

Given a Bayes-plausible τ, let

$$\pi(s|\omega) = \frac{\mu_s(\omega)\tau(\mu_s)}{\mu_0(\omega)}.$$

Now,

$$\begin{aligned} \pi(s|\omega) &= \frac{\mu_s(\omega)\tau(\mu_s)}{\mu_0(\omega)} \Rightarrow \\ \pi(s|\omega)\mu_0(\omega) &= \mu_s(\omega)\tau(\mu_s) \Rightarrow \\ \sum_{\omega} \pi(s|\omega)\mu_0(\omega) &= \sum_{\omega} \mu_s(\omega)\tau(\mu_s) \Rightarrow \\ \sum_{\omega} \pi(s|\omega)\mu_0(\omega) &= \tau(\mu_s). \end{aligned}$$

which establishes (ii). Moreover,

$$\begin{aligned} \pi(s|\omega) &= \frac{\mu_s(\omega)\tau(\mu_s)}{\mu_0(\omega)} \Rightarrow \\ \mu_s(\omega) &= \frac{\pi(s|\omega)\mu_0(\omega)}{\tau(\mu_s)} \Rightarrow \\ \mu_s(\omega) &= \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega')}, \end{aligned}$$

which establishes (i). Hence, π induces τ.

9.2 Proof of Proposition 2

Suppose that $\hat{v}$ is concave. That means that for any τ, $E_\tau\left(\hat{v}\left(\mu\right)\right) \leq \hat{v}\left(E_\tau\left(\mu\right)\right)$. Hence, for any Bayes-plausible τ, $E_\tau\left(\hat{v}\left(\mu\right)\right) \leq \hat{v}\left(\mu_0\right)$. Hence, by Corollary 1, Sender does not benefit from persuasion. Now, suppose that $\hat{v}$ is convex and not concave. The fact that it is not concave means that there exists some μ_a and μ_b and $\lambda \in (0,1)$ s.t. $\hat{v}\left(\lambda\mu_a + (1-\lambda)\mu_b\right) < \lambda\hat{v}\left(\mu_a\right) + (1-\lambda)\hat{v}\left(\mu_b\right)$. Now, consider the belief $\mu_c = \lambda\mu_a + (1-\lambda)\mu_b$. Since μ_0 is not on the boundary of $\Delta\left(\Omega\right)$, there exists a belief μ_d and a $\gamma \in (0,1)$ s.t. $\mu_0 = \gamma\mu_c + (1-\gamma)\mu_d$. Moreover, since $\hat{v}$ is convex, we know that

$$\begin{aligned}
\hat{v}\left(\mu_0\right) &= \hat{v}\left(\gamma\mu_c + (1-\gamma)\mu_d\right) \\
&\leq \gamma\hat{v}\left(\mu_c\right) + (1-\gamma)\hat{v}\left(\mu_d\right) \\
&= \gamma\hat{v}\left(\lambda\mu_a + (1-\lambda)\mu_b\right) + (1-\gamma)\hat{v}\left(\mu_d\right) \\
&< \gamma\left(\lambda\hat{v}\left(\mu_a\right) + (1-\lambda)\hat{v}\left(\mu_b\right)\right) + (1-\gamma)\hat{v}\left(\mu_d\right).
\end{aligned}$$

Now, let τ be the distribution of posteriors that puts probability $\gamma\lambda$ on μ_a, $\gamma(1-\lambda)$ on μ_b and $(1-\gamma)$ on μ_d. By construction, τ is Bayes-plausible and $E_\tau\hat{v}\left(\mu\right) > \hat{v}\left(\mu_0\right)$.

9.3 Proof of Proposition 3

Suppose that $\forall\mu$, $\hat{v}\left(\mu\right) \leq \sum_\omega v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu\left(\omega\right)$. Given an equilibrium (σ^*,ρ^*,μ^*) of some mechanism, let τ_m denote the probability of message m:

$$\tau_m = \sum_\omega \sum_s \sigma^*\left(m|s\right)\pi\left(s|\omega\right)\mu_0\left(\omega\right).$$

The value of the mechanism is at most:

$$\begin{aligned}
& \sum_{m\in M}\tau_m\hat{v}\left(\mu^*\left(\cdot|m\right)\right)\\
\leq & \sum_{m\in M}\tau_m\sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu^*\left(\omega|m\right)\\
= & \sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\sum_{m\in M}\tau_m\mu^*\left(\omega|m\right)\\
= & \sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu_0\left(\omega\right)\\
= & \hat{v}\left(\mu_0\right).
\end{aligned}$$

9.4 Proof of Proposition 4

Suppose there is information Sender would share and Receiver's preference is discrete at the prior. Since u is continuous in ω, $\sum u\left(\hat{a}\left(\mu_0\right),\omega\right)\mu\left(\omega\right)$ is continuous in μ. Therefore, since Receiver's preference is discrete at the prior, $\exists\delta>0$ s.t. for all μ in an δ-ball around μ_0, $\hat{a}\left(\mu\right)=\hat{a}\left(\mu_0\right)$. Denote this ball by B_δ. Since there is information Sender would share, $\exists\mu_h$ s.t. $\hat{v}\left(\mu_h\right)>\sum_\omega v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu_h\left(\omega\right)$. Consider a ray from μ_h through μ_0. Since μ_0 is not on the boundary of $\Delta\left(\Omega\right)$, there exists a belief on that ray, μ_l s.t. $\mu_l\in B_\delta$ and $\mu_0=\gamma\mu_l+\left(1-\gamma\right)\mu_h$ for some $\gamma\in(0,1)$. Now, consider the distribution of posteriors $\tau\left(\mu_l\right)=\gamma$, $\tau\left(\mu_h\right)=1-\gamma$. Since $\hat{a}\left(\mu_0\right)=\hat{a}\left(\mu_l\right)$, $\hat{v}\left(\mu_l\right)=\sum_\omega v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu_l\left(\omega\right)$. Hence,

$$\begin{aligned}
& \gamma\hat{v}\left(\mu_l\right)+\left(1-\gamma\right)\hat{v}\left(\mu_h\right)\\
> & \gamma\sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu_l\left(\omega\right)+\left(1-\gamma\right)\sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu_h\left(\omega\right)\\
= & \sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\left(\gamma\mu_l\left(\omega\right)+\left(1-\gamma\right)\mu_h\left(\omega\right)\right)\\
= & \sum_{\omega}v\left(\hat{a}\left(\mu_0\right),\omega\right)\mu_0\left(\omega\right)\\
= & \hat{v}\left(\mu_0\right)
\end{aligned}$$

Since τ is Bayes-plausible, Sender benefits from persuasion.

9.5 Proof of Proposition 5

Suppose A is finite. We begin with the following Lemma:

Lemma 1 *If Receiver's preference at a belief μ is not discrete, there must be an action $a \neq \hat{a}(\mu)$ such that*

$$\sum u(\hat{a}(\mu), \omega)\mu(\omega) = \sum u(a, \omega)\mu(\omega).$$

Proof. Suppose there is no such action. Then, we can define an $\varepsilon > 0$ by

$$\varepsilon = \frac{1}{2} \min_{a \neq \hat{a}(\mu)} \left\{ \sum u(\hat{a}(\mu), \omega)\mu(\omega) - \sum u(a, \omega)\mu(\omega) \right\}.$$

Since A is finite, the minimum is obtained. But then, $\sum u(\hat{a}(\mu), \omega)\mu(\omega) > \sum u(a, \omega)\mu(\omega) + \varepsilon$ $\forall a \neq \hat{a}(\mu)$, which means that Receiver's preference is discrete at μ. ■

Given this Lemma, it will suffice to show that the set

$$\left\{ \mu | \exists a \neq \hat{a}(\mu) \text{ s.t. } \sum u(\hat{a}(\mu), \omega)\mu(\omega) = \sum u(a, \omega)\mu(\omega) \right\}$$

has measure zero. Note that this set is a subset of

$$\left\{ \mu | \exists a, a' \text{ s.t. } a \neq a', \sum u(a, \omega)\mu(\omega) = \sum u(a', \omega)\mu(\omega) \right\}.$$

Hence, it will suffice to show that the latter set has measure zero. In fact, since there are only finitely many pairs of actions a, a', and since the union of a finite number of measure-zero sets has measure zero, it will suffice to show that given any distinct a and a' the set

$$\left\{ \mu | \sum u(a, \omega)\mu(\omega) = \sum u(a', \omega)\mu(\omega) \right\}$$

has measure zero.

Given any distinct a and a', index states by i and let $\beta_i = u(a, \omega_i) - u(a', \omega_i)$. Let $\beta = [\beta_1, ..., \beta_n]$ and $\mu = [\mu(\omega_1), ..., \mu(\omega_n)]$. We need to show that the set $\{\mu | \beta'\mu = 0\}$ has measure zero. Recall that for any action a there exits a μ s.t. $a^*(\mu) = \{a\}$. That means that there is necessarily an ω s.t. $u(a, \omega) \neq u(a', \omega)$. Hence, there is at least one $\beta_i \neq 0$. Therefore, β is

a linear transformation of rank 1. Hence, the kernel of β is a vector space of dimension $n-1$. Therefore, $\{\mu|\beta'\mu=0\}$ is measure zero with respect to the Lebesgue measure on $\mathbb{R}^n$.

9.6 Proof of Proposition 6

As we mentioned in footnote 11, we will establish a somewhat stronger proposition which implies Proposition 6. Suppose that there exists a linear transformation $T:\Delta(\Omega)\to\mathbb{R}^k$ s.t. $\hat{v}(\mu)=\widetilde{v}(T\mu)$. Let $\widetilde{V}$ denote the concave closure of $\widetilde{v}$. Then,

Proposition 15 *Sender benefits from persuasion if and only if* $\widetilde{V}(T\mu_0)>\widetilde{v}(T\mu_0)$.

Proof. Suppose $\widetilde{V}(T\mu_0)>\widetilde{v}(T\mu_0)$. That implies there exists a z s.t. $z>\widetilde{v}(T\mu_0)$ and $(T\mu_0,z)\in co(\widetilde{v})$. Hence, there exists a set $(t_i)_{i=1}^{k+1}$ w/ $t_i\in\text{Image}(T)$ and weights $\gamma\in\Delta^{k+1}$ s.t. $\sum_i\gamma_i t_i=T\mu_0$ and $\sum_i\gamma_i\widetilde{v}(t_i)>\widetilde{v}(T\mu_0)$. For each i, select any μ_i from $T^{-1}t_i$. Let $\mu_a=\sum_{i=1}^{k+1}\gamma_i\mu_i$. Note that since T is linear $T\mu_a=T\sum_i\gamma_i\mu_i=\sum_i\gamma_iT\mu_i=\sum_i\gamma_it_i=T\mu_0$. Since μ_0 is not on the boundary of Δ^n, there exists a belief μ_b and a $\lambda\in(0,1)$ s.t. $\lambda\mu_a+(1-\lambda)\mu_b=\mu_0$. Since T is linear, $T\mu_b=\frac{1}{1-\lambda}(T\mu_0-\lambda T\mu_a)$. Therefore, since $T\mu_a=T\mu_0$, we have $T\mu_b=T\mu_0$. Hence, $\widetilde{v}(T\mu_0)=\widetilde{v}(T\mu_b)$. Now, consider a mechanism that induces the distribution of posteriors

$$\begin{aligned}\tau(\mu_i) &= \lambda\gamma_i \text{ for } i=1,...,k+1\\ \tau(\mu_b) &= 1-\lambda\end{aligned}$$

Since $\mu_a=\sum_{i=1}^{k+1}\gamma_i\mu_i$ and $\lambda\mu_a+(1-\lambda)\mu_b=\mu_0$, this τ is Bayes-plausible. The value of the mechanism that induces this τ is

$$\begin{aligned}&\sum_i\lambda\gamma_i\hat{v}(\mu_i)+(1-\lambda)\hat{v}(\mu_b)\\ =\;&\lambda\sum_i\gamma_i\widetilde{v}(T\mu_i)+(1-\lambda)\widetilde{v}(T\mu_b)\\ =\;&\lambda\sum_i\gamma_i\widetilde{v}(t_i)+(1-\lambda)\widetilde{v}(T\mu_0)\\ >\;&\lambda\widetilde{v}(T\mu_0)+(1-\lambda)\widetilde{v}(T\mu_0)\\ =\;&\widetilde{v}(T\mu_0)\\ =\;&\hat{v}(\mu_0).\end{aligned}$$

Hence, Sender benefits from persuasion. Now suppose $\widetilde{V}(T\mu_0) \leq \widetilde{v}(T\mu_0)$. For any Bayes-plausible distribution of posteriors τ, $E_\tau[\mu] = \mu_0$ implies $E_\tau[T\mu] = T\mu_0$, so $E_\tau[\hat{v}(\mu)] = E_\tau[\widetilde{v}(T\mu)] \leq \widetilde{V}(T\mu_0) \leq \widetilde{v}(T\mu_0) = \hat{v}(\mu_0)$. Hence, by Corollary 1 Sender does not benefit from persuasion. ■

9.7 Proof of Proposition 7

We begin the proof by showing that selection of Sender-preferred equilibria implies that $\hat{v}$ is upper semicontinuous.

Lemma 2 *$\hat{v}$ is upper semicontinuous.*

Proof. Suppose that $\hat{v}$ is discontinuous at some μ. Since $\hat{v}(\mu) = \sum_\omega v(\hat{a}(\mu), \omega)\mu(\omega)$ and v is continuous, it must be that $\hat{a}(\mu)$ is discontinuous at μ. Since u is continuous, by Berge's Maximum Theorem this means that Receiver must be indifferent between a set of actions at μ, i.e., $a^*(\mu)$ is not a singleton. By definition, however, $\hat{v}(\mu) \equiv \max_{a \in a^*(\mu)} \sum_\omega v(a, \omega)\mu(\omega)$. Hence, $\hat{v}$ is upper semicontinuous. ■

Now, let $hyp(\hat{v})$ denote the hypograph of $\hat{v}$, i.e., the set of points lying on or below the graph. Because, unlike the graph of $\hat{v}$, $hyp(\hat{v})$ is closed and connected, it will be easier construct to work with.

Lemma 3 *Given μ and $S \subset hyp(\hat{v})$, if $(\mu, V(\mu))$ is in the convex hull of S, it is also in the convex hull of the intersection of S and graph of $\hat{v}$.*

Proof. Given μ and $S = \{(\mu_i, z_i)\}_{i \in I} \subset hyp(\hat{v})$, suppose $(\mu, V(\mu)) = \sum_{i \in I} \gamma_i (\mu_i, z_i)$ with $\sum_{i \in I} \gamma_i = 1$ and $\gamma_i \in [0,1]$ $\forall i \in I$. We need to show that for any $\gamma_i > 0$, it must be the case that $z_i = \hat{v}(\mu_i)$. Suppose to the contrary that $\exists j \in I$ s.t. $\gamma_j > 0$ and $z_j < \hat{v}(\mu_j)$. Consider the set $\{(\mu_i, z_i)\}_{i \in I \setminus \{j\}} \cup \{(\mu_j, \hat{v}(\mu_j))\} \subset hyp(\hat{v})$. Since $\gamma_j \mu_j + \sum_{i \in I \setminus \{j\}} \gamma_i \mu_i = \mu$ and $\gamma_j \hat{v}(\mu_j) + \sum_{i \in I \setminus \{j\}} \gamma_i z_i > \sum_{i \in I} \gamma_i z_i = V(\mu)$, we have that $V(\mu) < \sup\{z | (\mu, z) \in co(hyp(\hat{v}))\} = \sup\{z | (\mu, z) \in co(\hat{v})\}$. Hence, we've reached a contradiction. ■

Combining the two Lemmas above, we obtain the following:

Lemma 4 *Any element of the graph of V can be expressed as a convex combination of elements of the graph of $\hat{v}$.*

Proof. Since $\hat{v}$ is upper semicontinuous, $hyp\,\hat{v}$ is closed. Let $H = \left\{(\mu, z) \in hyp\ \hat{v} | z \geq \inf_{\mu'} \hat{v}(\mu')\right\}$. Since v is continuous and A is compact, $\hat{v}$ is bounded so H is bounded. Hence, since $hyp\ \hat{v}$ is closed, H is compact. Hence, $co(H)$ is compact which implies $co(hyp(\hat{v}))$ is closed. Hence, $hyp(V) = co(hyp(\hat{v}))$. Hence, any element of the graph of V can be expressed a convex combination of elements of $hyp(\hat{v})$. But then, by Lemma 3, it can also be expressed as a convex combination of elements of the graph of $\hat{v}$. ■

The remainder of the proof of Proposition 7 is straightforward. By Corollary 2, there can be no mechanism with value strictly greater than $V(\mu_0)$. By Lemma 4, $(\mu_0, V(\mu_0)) \in co(\hat{v})$. Hence, there exists an optimal mechanism with value $V(\mu_0)$.

9.8 Proof of Proposition 8

It follows directly from definition of convexity and concavity that if $\hat{v}$ is (strictly) concave, no disclosure is (uniquely) optimal, and if it is (strictly) convex, full disclosure is (uniquely) optimal. To establish the last part of Proposition 8, we begin with a key Lemma which will also be useful for establishing several other Propositions.

Lemma 5 *If μ' is induced by an optimal mechanism, $V(\mu') = \hat{v}(\mu')$.*

Proof. Suppose that an optimal mechanism induces τ and there is some μ' s.t. $\tau(\mu') > 0$ and $V(\mu') > \hat{v}(\mu')$. Since $(\mu', V(\mu')) \in co(\hat{v})$, there exists a distribution of posteriors τ' such that $E_{\tau'}\mu = \mu'$ and $E_{\tau'}\hat{v}(\mu) = V(\mu')$. But then we can then take all the weight from μ' and place it on τ' which would yield higher value while preserving Bayes-plausability. Formally, consider the distribution of posteriors

$$\tau^*(\mu) = \begin{cases} \tau(\mu')\,\tau'(\mu) & \text{if } \mu \in Supp(\tau') \setminus Supp(\tau) \\ \tau(\mu) + \tau(\mu')\,\tau'(\mu) & \text{if } \mu \in Supp(\tau') \cap Supp(\tau) \\ \tau(\mu) & \text{if } \mu \in Supp(\tau) \setminus (Supp(\tau') \cup \{\mu'\}). \end{cases}$$

By construction, τ^* is plausible and yields a higher value than τ does. ■

Suppose $\hat{v}$ is not concave and an optimal mechanism induces an interior μ_m. By Lemma 5, we know $V(\mu_m) = \hat{v}(\mu_m)$. Since $\hat{v}$ is not concave, there is some belief μ_l s.t. $V(\mu_l) > \hat{v}(\mu_l)$.

Because μ_m is interior, we know there is a μ_r and $\gamma \in (0,1)$ s.t. $\mu_m = \gamma\mu_l + (1-\gamma)\mu_r$. Now,

$$\begin{aligned} \hat{v}(\mu_m) = V(\mu_m) &\geq \gamma V(\mu_l) + (1-\gamma) V(\mu_r) \qquad \text{(by concavity of } V\text{)} \\ &\geq \gamma V(\mu_l) + (1-\gamma)\hat{v}(\mu_r) \\ &> \gamma\hat{v}(\mu_l) + (1-\gamma)\hat{v}(\mu_r) \end{aligned}$$

which means that $\hat{v}$ is not convex.

9.9 Proof of Proposition 9

Since $\hat{v}$ is bounded, $hyp(\hat{v})$ is path-connected. Therefore, it is connected. The Fenchel-Bunt Theorem (Hiriart-Urruty and Lemaréchal 2004, Thm 1.3.7) states that if $S \subset \mathbb{R}^n$ has no more than n connected components (in particular, if S is connected), then any $x \in co(S)$ can be expressed as a convex combination of n elements of S. Hence, since $hyp(\hat{v}) \subset \mathbb{R}^{|\Omega|}$, any element of $co(hyp(\hat{v}))$ can be expressed as a convex combination of $|\Omega|$ elements of $hyp(\hat{v})$. In particular, $(\mu_0, V(\mu_0)) \in co(hyp(\hat{v}))$ can be expressed a convex combination of $|\Omega|$ elements of $hyp(\hat{v})$. But then, by Lemma 3 this further implies that $(\mu_0, V(\mu_0))$ can be expressed as a convex combination of $|\Omega|$ elements of the graph of $\hat{v}$. Hence, there exists an optimal straightforward mechanism which induces a distribution of posteriors whose support has no more than $|\Omega|$ elements. Since Receiver takes only a single action at any of her beliefs, this implies there exists an optimal straightforward mechanism in which the number of actions Receiver takes with positive probability is at most $|\Omega|$.

9.10 Proof of Proposition 10

Suppose that $v(\underline{a}, \omega) < v(a, \omega)$ for all ω and all $a \neq \underline{a}$. For any ω, let μ_ω denote the distribution s.t. $\mu_\omega(\omega) = 1$. Let Ω^- be the set of states where $\underline{a}$ is the uniquely optimal action for Receiver, i.e., $\Omega^- = \{\omega | \hat{a}(\mu_\omega) = \underline{a}\}$. Let Ω^+ be the complement of Ω^-.

Now, suppose contrary to Proposition 10 that an optimal mechanism induces τ and there is a

belief μ' s.t. $\tau(\mu') > 0$, $\hat{a}(\mu') = \underline{a}$ and $\exists \omega \in \Omega^+$ s.t. $\mu'(\omega) > 0$. Consider the following two beliefs

$$\mu^-(\omega) = \begin{cases} \frac{\mu'(\omega)}{\sum_{\omega' \in \Omega^-} \mu'(\omega')} & \text{if } \omega \in \Omega^- \\ 0 & \text{if } \omega \in \Omega^+ \end{cases}$$

$$\mu^+(\omega) = \begin{cases} 0 & \text{if } \omega \in \Omega^- \\ \frac{\mu'(\omega)}{\sum_{\omega' \in \Omega^+} \mu'(\omega')} & \text{if } \omega \in \Omega^+. \end{cases}$$

It is easy to see that μ' is a convex combination of μ^- and μ^+. Hence, we can "replace" μ' with μ^- and μ^+, and since $\hat{a}(\mu^-) = \hat{a}(\mu')$ while μ^+ induces an action Sender prefers over $\underline{a}$, this will yield a higher value. Formally, consider the following alternative distribution of beliefs:

$$\begin{aligned} \tau^*(\mu^-) &= \left(\sum_{\omega' \in \Omega^-} \mu'(\omega') \right) \tau(\mu') + \tau(\mu^-) \\ \tau^*(\mu^+) &= \left(\sum_{\omega' \in \Omega^+} \mu'(\omega') \right) \tau(\mu') + \tau(\mu^+) \\ \tau^*(\mu) &= \tau(\mu) \text{ if } \mu \notin \{\mu', \mu^-, \mu^+\}. \end{aligned}$$

Simple algebra reveals that τ^* is Bayes-plausible and yields a higher value than τ does.

9.11 Proof of Proposition 11

We first prove a preliminary lemma.

Lemma 6 *Suppose μ_l and μ_r are induced by an optimal mechanism and $\mu_m = \gamma\mu_l + (1-\gamma)\mu_r$ for some $\gamma \in [0,1]$. Then, $\hat{v}(\mu_m) \leq \gamma\hat{v}(\mu_l) + (1-\gamma)\hat{v}(\mu_r)$.*

Proof. Suppose to the contrary that τ is induced by an optimal mechanism, $\tau(\mu_l)$, $\tau(\mu_r) > 0$, and $\hat{v}(\mu_m) > \gamma\hat{v}(\mu_l) + (1-\gamma)\hat{v}(\mu_r)$. Then we can take some weight from μ_l and μ_r and place it on μ_m which would yield higher value while preserving Bayes-plausability. Formally, pick any

$\varepsilon \in (0,1)$. Let $\varepsilon' = \varepsilon\tau(\mu_l)/\tau(\mu_r)$. Consider an alternative τ^* defined by:

$$\begin{aligned}
\tau^*(\mu_l) &= (1-\gamma\varepsilon)\,\tau(\mu_l) \\
\tau^*(\mu_r) &= \left(1-(1-\gamma)\,\varepsilon'\right)\tau(\mu_r) \\
\tau^*(\mu_m) &= \tau(\mu_m) + \varepsilon\tau(\mu_l) \\
\tau^*(\mu) &= \tau(\mu) \text{ if } \mu \notin \{\mu_l, \mu_m, \mu_r\}.
\end{aligned}$$

Simple algebra reveals that τ^* is Bayes-plausible and yields a higher value than τ does. ■

Say that action a is induced-dominant if $\forall\mu$, $\hat{v}(\mu) \leq \sum_{\omega} v(a,\omega)\,\mu(\omega)$. Say that a is strictly induced-dominant if $\forall\mu$ s.t. $a \neq \hat{a}(\mu)$, $\hat{v}(\mu) < \sum_{\omega} v(a,\omega)\,\mu(\omega)$. Say that a is weakly but not strictly dominant (wnsd) if it is induced-dominant and $\exists\mu$ s.t. $a \neq \hat{a}(\mu)$ and $\hat{v}(\mu) = \sum_{\omega} v(a,\omega)\,\mu(\omega)$. Note that there is information Sender would share if and only if $\hat{a}(\mu_0)$ is not induced-dominant, and that Assumption 1 states that there are no wnsd actions.

We now prove the proposition in two steps.

Lemma 7 *Suppose that Assumption 1 holds. Let μ be an interior belief induced by an optimal mechanism. Then, either: (i) Receiver's preference at μ is not discrete, or (ii) $\hat{a}(\mu)$ is strictly induced-dominant.*

Proof. Suppose that Assumption 1 holds and μ is an interior belief induced by an optimal mechanism. Now, suppose Receiver's preference at μ is discrete. By Proposition 4, we know that if μ were the prior either: (i) there would be no information Sender would want to share, i.e., $\hat{a}(\mu)$ is induced dominant; or (ii) Sender would benefit from persuasion. But, Sender would not benefit from persuasion if μ were the prior because by Lemma 5 we know $V(\mu) = \hat{v}(\mu)$. Thus, $\hat{a}(\mu)$ is induced-dominant so by Assumption 1 it is strictly induced-dominant. ■

Lemma 8 *Suppose Sender benefits from persuasion, μ is an interior belief induced by an optimal mechanism, and $\hat{a}(\mu)$ is strictly induced-dominant. Then, Receiver's preference at μ is not discrete.*

Proof. Suppose Sender benefits from persuasion, μ is an interior belief induced by an optimal mechanism, and $\hat{a}(\mu)$ is strictly induced-dominant. First note that the set of beliefs that induces

any particular action is necessarily convex. Hence, when Sender benefits from persuasion, any optimal mechanism must induce at least two distinct actions. Therefore, there must be a μ' induced by the mechanism at which $\hat{a}(\mu) \neq \hat{a}(\mu')$. Now, suppose contrary to the Lemma that Receiver's preference at μ is discrete. Then, there is an $\varepsilon > 0$ s.t. $\hat{a}(\varepsilon\mu' + (1-\varepsilon)\mu) = \hat{a}(\mu)$. Let $\mu_m = \varepsilon\mu' + (1-\varepsilon)\mu$. Since both μ and μ' are induced by an optimal mechanism, Lemma 6 tells us that

$$\hat{v}(\mu_m) \leq \varepsilon\hat{v}(\mu') + (1-\varepsilon)\hat{v}(\mu). \tag{3}$$

But,

$$\begin{aligned}
\hat{v}(\mu_m) &= \sum_{\omega} v(\hat{a}(\mu_m), \omega)\mu_m(\omega) \\
&= \sum_{\omega} v(\hat{a}(\mu), \omega)\mu_m(\omega) \\
&= \varepsilon\sum_{\omega} v(\hat{a}(\mu), \omega)\mu'(\omega) + (1-\varepsilon)\sum_{\omega} v(\hat{a}(\mu), \omega)\mu(\omega) \\
&= \varepsilon\sum_{\omega} v(\hat{a}(\mu), \omega)\mu'(\omega) + (1-\varepsilon)\hat{v}(\mu)
\end{aligned}$$

Hence, Equation (3) is equivalent to

$$\sum_{\omega} v(\hat{a}(\mu), \omega)\mu'(\omega) \leq \hat{v}(\mu'),$$

which means $\hat{a}(\mu)$ is not strictly induced-dominant. ■

Combining these two lemmata, we know that if Assumption 1 holds, Sender benefits from persuasion, and μ is an interior belief induced by an optimal mechanism, Receiver's preference at μ is not discrete.

9.12 Proof of Proposition 12

Suppose Assumption 1 holds, $\hat{v}$ is monotonic, A is finite, and Sender benefits from persuasion. Now, suppose μ is an interior belief induced by an optimal mechanism. Since Assumption 1 holds and Sender benefits from persuasion, Receiver's preference at μ is not discrete by Proposition 11. Therefore, Lemma 1 tells us $\exists a$ such that $E_\mu u(\hat{a}(\mu), \omega) = E_\mu u(a, \omega)$. If (ii) does not hold, we

know $E_\mu v(\hat{a}(\mu),\omega) > E_\mu v(a,\omega)$. Therefore, $\hat{a}(\mu) \succsim a$. Hence, given any $\mu' \lhd \mu$,

$$0 = E_\mu u(\hat{a}(\mu),\omega) - E_\mu u(a,\omega) > E_{\mu'} u(\hat{a}(\mu),\omega) - E_{\mu'} u(a,\omega).$$

Since $E_{\mu'} u(a,\omega) > E_{\mu'} u(\hat{a}(\mu),\omega)$, we know that $\hat{a}(\mu)$ is not Receiver's optimal action when her beliefs are μ. Hence, for any $\mu' \lhd \mu$, $\hat{a}(\mu') \neq \hat{a}(\mu)$, which means that μ is a worst belief inducing $\hat{a}(\mu)$.

9.13 Proof of Proposition 13

We first show that no commitment-free mechanism can have a gain greater than E. Suppose the contrary. That means that in equilibrium of a mechanism, Sender conveys two messages with positive probability m and m' s.t. $\hat{v}(\mu^*(\cdot|m)) > \hat{v}(\mu^*(\cdot|m'))$. But then, since v is independent of ω, it must be the case that Sender would profit by a deviation that always sends m instead of m'.

Next, we show that for any $\varepsilon > 0$, there is a commitment-free mechanism whose gain exceeds $E-\varepsilon$. Given ε, choose $\overline{v} > E - \varepsilon + \hat{v}(\mu_0)$ s.t. $\exists M \subset \Delta(\Omega)$ s.t. $\hat{v}(\mu) = \overline{v}\ \forall \mu \in M$ and $\mu_0 \in co(M)$. By definition of E, such a $\overline{v}$ exists. Now, since $\mu_0 \in co(M)$, there exists a Bayes-plausible τ s.t. $supp(\tau) \subset M$. Consider a commitment-free mechanism with the signal π that induces this τ. Since $\hat{v}(\mu) = \overline{v}$ and v is independent of ω, incentive compatibility does not bind. Hence, the value of this mechanism is $\overline{v}$ so its gain is strictly greater than $E - \varepsilon$.

9.14 Proof of Proposition 14

Even though it is not feasible for Sender to commit to an honest mechanism, he still may tell truth in equilibrium. We say an equilibrium of a commitment-free mechanism is truthful if $\sigma^*(m|s) = \begin{cases} 1 \text{ if } m=s \\ 0 \text{ if } m\neq s \end{cases}$. Given any equilibrium of any commitment-free mechanism, there exists a truthful equilibrium of some commitment-free mechanism that generates the same distribution of actions and beliefs. To see this, note that if some commitment-free mechanism π^o has an equilibrium with a message strategy σ^o, then $\sigma(m|s) = \begin{cases} 1 \text{ if } m=s \\ 0 \text{ if } m\neq s \end{cases}$ is an equilibrium strategy if $\pi(m|\omega) = \sum_s \sigma^o(m|s)\,\pi^o(s|\omega)$. Moreover, this π and σ generate the same conditional distribution of messages, and thus of actions and beliefs, as π^o and σ^o. Hence, we can restrict our attention

to truthful equilibria without loss of generality.

With that observation, we establish the following Lemma:

Lemma 9 *Consider any equilibrium of a commitment-free mechanism* $(\sigma^*, \rho^*, \mu^*)$. *For any two messages* m *and* m' *sent with positive probability:*

$$v_m - v_{m'} \leq \max_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m) - \min_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m)$$

Proof. Supposing, w.l.o.g. that the equilibrium is truthful, for any two messages m and m' sent with positive probability it must be the case that

$$v_{m'} \geq \sum_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m)\, \mu^*\left(\omega|m'\right)$$

This implies

$$\begin{aligned} v_m - v_{m'} &\leq v_m - \sum_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m)\, \mu^*\left(\omega|m'\right) \\ &= \sum_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m)\, \mu^*(\omega|m) - \sum_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m)\, \mu^*\left(\omega|m'\right) \\ &= \sum_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m) \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right]. \end{aligned}$$

Now, note that

$$\sum_{\omega} \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] = 0$$

Let $\Omega^+ = \{\omega \in \Omega | \mu^*(\omega|m) > \mu^*(\omega|m')\}$ and $\Omega^- = \{\omega \in \Omega | \mu^*(\omega|m) < \mu^*(\omega|m')\}$. Then

$$0 < \sum_{\omega \in \Omega^+} \mu^*(\omega|m) - \mu^*\left(\omega|m'\right) = \sum_{\omega \in \Omega^-} \mu^*\left(\omega|m'\right) - \mu^*(\omega|m) \leq 1.$$

Hence,

$$\begin{aligned}
v_m - v_{m'} &\leq \sum_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m) \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] \\
&= \sum_{\omega\in\Omega^+} \int_A v(a,\omega)\, d\rho^*(a|m) \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] \\
&\quad + \sum_{\omega\in\Omega^-} \int_A v(a,\omega)\, d\rho^*(a|m) \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] \\
&= \sum_{\omega\in\Omega^+} \int_A v(a,\omega)\, d\rho^*(a|m) \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] \\
&\quad - \sum_{\omega\in\Omega^-} \int_A v(a,\omega)\, d\rho^*(a|m) \left[\mu^*\left(\omega|m'\right) - \mu^*(\omega|m)\right] \\
&\leq \sum_{\omega\in\Omega^+} \left(\max_{\omega'} \int_A v\left(a,\omega'\right) d\rho^*(a|m)\right) \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] \\
&\quad - \sum_{\omega\in\Omega^-} \left(\min_{\omega'} \int_A v\left(a,\omega'\right) d\rho^*(a|m)\right) \left[\mu^*\left(\omega|m'\right) - \mu^*(\omega|m)\right] \\
&= \left[\max_{\omega'} \int_A v\left(a,\omega'\right) d\rho^*(a|m) - \min_{\omega'} \int_A v\left(a,\omega'\right) d\rho^*(a|m)\right] \\
&\quad \times \sum_{\omega\in\Omega^+} \left[\mu^*(\omega|m) - \mu^*\left(\omega|m'\right)\right] \\
&\leq \max_{\omega'} \int_A v\left(a,\omega'\right) d\rho^*(a|m) - \min_{\omega'} \int_A v\left(a,\omega'\right) d\rho^*(a|m)
\end{aligned}$$

■

The key implication of Lemma 9 is the following. Given an equilibrium of a commitment-free mechanism, let $\mathbf{M}$ be the set of all messages sent with a positive probability in that equilibrium. Then,

Lemma 10 *If in equilibrium of a mechanism $\exists m \in \mathbf{M}$ s.t. $v_m \leq \hat{v}(\mu_0)$, then the gain from that mechanism is at most D.*

Proof. Consider an equilibrium of a commitment-free mechanism $(\sigma^*, \rho^*, \mu^*)$. Suppose $\exists m \in$

M s.t. $v_m \leq \hat{v}(\mu_0)$. Then, by Lemma 9

$$\begin{aligned} v_{m'} - \hat{v}(\mu_0) &\leq v_{m'} - v_m \\ &\leq \max_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m) - \min_{\omega} \int_A v(a,\omega)\, d\rho^*(a|m) \\ &\leq \int_A \max_{\omega} v(a,\omega)\, d\rho^*(a|m) - \int_A \min_{\omega} v(a,\omega)\, d\rho^*(a|m) \\ &= \int_A \left(\max_{\omega} v(a,\omega) - \min_{\omega} v(a,\omega)\right) d\rho^*(a|m) \\ &\leq \max_{a \in A} \left(\max_{\omega} v(a,\omega) - \min_{\omega} v(a,\omega)\right) \\ &= D. \end{aligned}$$

■

Now, say $\hat{v}$ is *without troughs* if there do not exist μ_a, μ_b, μ_c and $\gamma \in (0,1)$ s.t. $\mu_b = \gamma\mu_a + (1-\gamma)\mu_c$ and $\min\{\hat{v}(\mu_a), \hat{v}(\mu_c)\} > \hat{v}(\mu_b)$.

Lemma 11 *Consider any belief μ' and any distribution of beliefs τ with finite support s.t. $E_\tau[\mu] = \mu'$. If $\hat{v}$ is without troughs, then $\exists \mu \in Supp(\tau)$ s.t. $\hat{v}(\mu) \leq \hat{v}(\mu')$.*

Proof. Suppose $\hat{v}$ is without troughs. Given any belief μ' and any distribution of beliefs τ with finite support s.t. $E_\tau[\mu] = \mu'$, let $k = |Supp(\tau)|$. We prove the Lemma by induction on k. If $k = 2$, the desired conclusion is immediate. Now, suppose the Lemma holds for $k = l$. We need to show it holds for $k = l + 1$. Pick any $\widetilde{\mu}$ in $Supp(\tau)$. Then,

$$\mu' = \sum_{\mu \in Supp(\tau)} \tau(\mu)\,\mu = \tau(\widetilde{\mu})\,\widetilde{\mu} + (1 - \tau(\widetilde{\mu})) \sum_{\mu \in Supp(\tau)\setminus\{\widetilde{\mu}\}} \frac{\tau(\mu)}{1-\tau(\widetilde{\mu})}\mu.$$

Now, suppose $\hat{v}(\widetilde{\mu}) > \hat{v}(\mu')$. Then, since $\hat{v}$ is without troughs, we know $\hat{v}\left(\sum_{\mu \in Supp(\tau)\setminus\{\widetilde{\mu}\}} \frac{\tau(\mu)}{1-\tau(\widetilde{\mu})}\mu\right) \leq \hat{v}(\mu')$. But, since the Lemma holds for $k = l = |Supp(\tau)\setminus\{\widetilde{\mu}\}|$, we know there exists $\widetilde{\mu}' \in Supp(\tau)\setminus\{\widetilde{\mu}\}$ s.t. $\hat{v}(\widetilde{\mu}') \leq \hat{v}\left(\sum_{\mu \in Supp(\tau)\setminus\{\widetilde{\mu}\}} \frac{\tau(\mu)}{1-\tau(\widetilde{\mu})}\mu\right) \leq \hat{v}(\mu')$. ■

Lemma 11 leads to the following result:

Proposition 16 *If $\hat{v}$ is without troughs, the gain from any commitment-free mechanism is at most D.*

Proof. Suppose $\hat{v}$ is without troughs and consider an equilibrium $(\sigma^*, \rho^*, \mu^*)$ of any commitment-free mechanism. Let τ_m denote the equilibrium probability of message m:

$$\tau_m = \sum_{\omega} \sum_{s} \sigma^* (m|s) \pi (s|\omega) \mu_0 (\omega) .$$

We know that $\mu_0 (\cdot) = \sum_m \tau_m \mu^* (\cdot|m)$. Therefore, by Lemma 11, $\exists m \in \mathbf{M}$ s.t. $v_m \leq \hat{v} (\mu_0)$. Hence, by Lemma 10, the gain from the mechanism is at most D. ■

Since $\hat{v}$ is necessarily monotonic if it is without troughs, Proposition 14 is a corollary of Proposition 16.

10 Appendix B: Extension to infinite state spaces

In the main body of the paper, we assumed that Ω is finite. We also claimed this assumption was made primarily for expositional convenience. In this appendix, we show that the approach used in the paper extends to the case when Ω is a compact metric space.[33]

As before, Receiver has a continuous utility function $u (a, \omega)$ that depends on her action $a \in A$ and the state of the world $\omega \in \Omega$. Sender has a continuous utility function $v (a, \omega)$ that depends on Receiver's action and the state of the world. The action space A is assumed to be compact and the state space Ω is assumed to be a compact metric space. Let $\Delta (\Omega)$ denote the set of Borel probabilities on Ω, a compact metric space in the weak* topology. Sender and Receiver share a prior $\mu_0 \in \Delta (\Omega)$.

A *persuasion mechanism* is a combination of a signal and a message technology. A *signal* (π, S) consists of a compact metric realization space S and a measurable function $\pi : [0, 1] \to \Omega \times S$, $x \mapsto (\pi_1 (x) , \pi_2 (x))$. Assume that x is uniformly distributed on $[0, 1]$ and that Sender observes $\pi_2 (x)$. We denote a realization of $\pi_2 (x)$ by s. Note that since S is a compact metric space (hence, complete and separable), there exists a regular conditional probability (i.e., a posterior probability) obtained by conditioning on $\pi_2 (x) = s$ (Shiryaev 1996, p.230). A *message technology* c consists of a message space M and a family of functions $\left\{ c (\cdot|s) : M \to \overline{\mathbb{R}}_+ \right\}_{s \in S}$. As before, a mechanism is honest if $M = S$ and $c (m|s) = \left\{ \begin{smallmatrix} k \text{ if } s=m \\ \infty \text{ if } s \neq m \end{smallmatrix} \right.$ for some $k \in \mathbb{R}_+$. A persuasion mechanism

[33] We are very grateful to Max Stinchcombe for help with this extension.

defines a game just as before. Perfect Bayesian equilibrium is still the solution concept and we still select Sender-preferred equilibria. Definitions of value and gain are same as before.

Let $a^*(\mu)$ denote the set of actions optimal for Receiver given her beliefs are $\mu \in \Delta(\Omega)$:

$$a^*(\mu) \equiv \arg\max_a \int u(a,\omega)\, d\mu(\omega).$$

Note that $a^*(\cdot)$ is an upper hemicontinuous, non-empty valued, compact valued, correspondence from $\Delta(\Omega)$ to A.

Let $\hat{v}(\mu)$ denote the maximum expected value of v if Receiver takes an action in $a^*(\mu)$:

$$\hat{v}(\mu) \equiv \max_{a \in a^*(\mu)} \int v(a,\omega)\, d\mu(\omega).$$

Since $a^*(\mu)$ is non-empty and compact and $\int v(a,\omega)\, d\mu(\omega)$ is continuous in a, $\hat{v}$ is well defined. We first show that the main ingredient for the existence of an optimal mechanism, namely the upper semicontinuity of $\hat{v}$, remains true in this setting.

Lemma 12 *$\hat{v}$ is upper semicontinuous.*

Proof. Given any a, the random variable $v(a,\omega)$ is dominated by the constant random variable $\max_\omega v(a,\omega)$ (since v is continuous in ω and Ω is compact, the maximum is attained). Hence, by the Lebesgue's Dominated Convergence Theorem, $\int v(a,\omega)\, d\mu(\omega)$ is continuous in μ for any given a. Now, suppose that $\hat{v}$ is discontinuous at some μ. Since u is continuous, by Berge's Maximum Theorem this means that Receiver must be indifferent between a set of actions at μ, i.e., $a^*(\mu)$ is not a singleton. By definition, however, $\hat{v}(\mu) \equiv \max_{a \in a^*(\mu)} \int v(a,\omega)\, d\mu(\omega)$. Hence, $\hat{v}$ is upper semicontinuous. ■

Now, *a distribution of posteriors*, denoted by τ, is an element of the set $\Delta(\Delta(\Omega))$, the set of Borel probabilities on the compact metric space $\Delta(\Omega)$. We say a distribution of posteriors τ is *Bayes-plausible* if $\int_{\Delta(\Omega)} \mu d\tau(\mu) = \mu_0$. We say that π *induces* τ if conditioning on $\pi_2(x) = s$ gives posterior κ_s and the distribution of $\kappa_{\pi_2(x)}$ is τ given that x is uniformly distributed. Since Ω is a compact metric space, for any Bayes-plausible τ there exists a π that induces it.[34] Hence, the

[34] Personal communication with Max Stinchcombe. Detailed proof available upon request.

problem of finding an optimal mechanism is equivalent to solving

$$\max_{\tau \in \Delta(\Delta(\Omega))} \int_{\Delta(\Omega)} \hat{v}(\mu)\, d\tau(\mu)$$
$$\text{s.t.} \int_{\Delta(\Omega)} \mu d\tau(\mu) = \mu_0.$$

Now, let

$$V \equiv \sup\{z | (\mu, z) \in co(hyp(\hat{v}))\},$$

where $co(\cdot)$ denotes the convex hull and $hyp(\cdot)$ denotes the hypograph. Recall that given a subset K of an arbitrary vector space, $co(K)$ is defined as $\cap\{C | K \subset C, C \text{ convex}\}$. Let $g(\mu_0)$ denote the subset of $\Delta(\Delta(\Omega))$ that generate the point $(\mu_0, V(\mu_0))$, i.e.,

$$g(\mu_0) \equiv \left\{ \tau \in \Delta(\Delta(\Omega)) \,|\, \int_{\Delta(\Omega)} \mu d\tau(\mu) = \mu_0, \int_{\Delta(\Omega)} \hat{v}(\mu) d\tau(\mu) = V(\mu_0) \right\}.$$

Note that we still have not established that $g(\mu_0)$ is non-empty. That is the primary task of the proof of our main proposition.

Proposition 17 *Optimal mechanism exists. Value of an optimal mechanism is $V(\mu_0)$. Sender benefits from persuasion iff $V(\mu_0) > \hat{v}(\mu_0)$. An honest mechanism with a signal that induces an element of $g(\mu_0)$ is optimal.*

Proof. By construction of V, there can be no mechanism with value strictly greater than $V(\mu_0)$. We need to show there exists a mechanism with value equal to $V(\mu_0)$, or equivalently, that $g(\mu_0)$ is not empty. Without loss of generality, suppose the range of v is $[0, 1]$. Consider the set $H = \{(\mu, z) \in hyp(V) | z \geq 0\}$. Since $\hat{v}$ is upper semicontinuous, H is compact. By construction of V, H is convex. Therefore, by Choquet's Theorem (e.g., Phelps 2001), for any $(\mu', z') \in H$, there exists a probability measure η s.t. $(\mu', z') = \int_H (\mu, z)\, d\eta(\mu, z)$ with η supported by extreme points of H. In particular, there exists η s.t. $(\mu_0, V(\mu_0)) = \int_H (\mu, z)\, d\eta(\mu, z)$ with η supported by extreme points of H. Now, note that if (μ, z) is an extreme point of H, then $V(\mu) = \hat{v}(\mu)$; moreover, if $z > 0$, $z = V(\mu) = \hat{v}(\mu)$. Hence, we can find an η s.t. $(\mu_0, V(\mu_0)) = \int_H (\mu, z)\, d\eta(\mu, z)$ with support of η entirely within $\{(\mu, \hat{v}(\mu)) | \mu \in \Delta(\Omega)\}$. Therefore, there exists a $\tau \in g(\mu_0)$. ■

References

Aumann, Robert J, & Maschler, Michael B. 1995. *Repeated Games with Incomplete Information.* MIT Press.

Benabou, Roland, & Tirole, Jean. 2002. Self-confidence and personal motivation. *Quarterly Journal of Economics*, **117**(3), 871–915.

Benabou, Roland, & Tirole, Jean. 2003. Intrinsic and extrinsic motivation. *Review of Economic Studies*, **70**, 489–520.

Benabou, Roland, & Tirole, Jean. 2004. Willpower and personal rules. *Journal of Political Economy*, **112**(4), 848–885.

Bodner, Ronit, & Prelec, Drazen. 2003. Self-signaling and diagnostic utility in everyday decision making. *Pages 105–123 of:* Brocas, Isabelle, & Carrillo, Juan D. (eds), *The Psychology of Economic Decisions.* Oxford: Oxford University Press.

Brocas, Isabelle, & Carrillo, Juan D. 2007. Influence through ignorance. *RAND Journal of Economics*, **38**, 931–947.

Caillaud, Bernard, & Tirole, Jean. 2007. Consensus building: How to persuade a group. *American Economic Review*, **97**(5), 1877–1900.

Cain, Daylian M., Loewenstein, George, & Moore, Don A. 2005. The dirt on coming clean: Perverse effects of disclosing conflicts. *Journal of Legal Studies*, **34**(1), 1–25.

Carillo, Juan D., & Mariotti, Thomas. 2000. Strategic ignorance as a self-disciplining device. *Review of Economic Studies*, **67**(3), 529–544.

Crawford, Vincent, & Sobel, Joel. 1982. Strategic information transmission. *Econometrica*, **50**(6), 1431–1451.

Ettinger, David, & Jehiel, Philippe. forthcoming. A theory of deception. *American Economic Journal: Microeconomics.*

Glazer, Jacob, & Rubinstein, Ariel. 2004. On optimal rules of persuasion. *Econometrica*, **72**, 1715–1736.

Glazer, Jacob, & Rubinstein, Ariel. 2006. A study in the pragmatics of persuasion. *Theoretical Economics*, **1**, 395–410.

Green, Jerry R., & Stokey, Nancy L. M. 2007. A two-person game of information transmission. *Journal of Economic Theory*, **135**(1), 90–104.

Grossman, Sanford J. 1981. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, **24**(3), 461–483.

Grossman, Sanford J, & Hart, Oliver D. 1983. An analysis of the principal-agent problem. *Econometrica*, **51**(1), 7–45.

Grossman, Sanford J, & Hart, Oliver D. 1986. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, **94**(4), 691–719.

Hart, Oliver, & Moore, John. 1990. Property rights and the nature of the firm. *Journal of Political Economy*, **98**(6), 1119–58.

Hiriart-Urruty, Jean-Baptiste, & Lemarechal, Claude. 2004. *Fundamentals of convex analysis*. Springer.

Holmstrom, Bengt. 1979. Moral hazard and observability. *Bell Journal of Economics*, **10**(1), 74–91.

Ivanov, Maxim. 2008. Informational control and organizational design. *Working Paper*.

Johnson, Justin P., & Myatt, David P. 2006. On the simple economics of advertising, marketing, and product design. *American Economic Review*, **96**(3), 756–784.

Jovanovic, Boyan. 1982. Truthful disclosure of information. *Bell Journal of Economics*, **13**(1), 36–44.

Kartik, Navin. forthcoming. Strategic communication with lying costs. *Review of Economic Studies*.

Lazear, Edward P. 2006. Speeding, terrorism, and teaching to the test. *Quarterly Journal of Economics*, **121**(3), 1029–1061.

Lewis, Tracy R., & Sappington, David E. M. 1994. Supplying information to facilitate price discrimination. *International Economic Review*, **35**(2), 309–327.

McCloskey, Donald, & Klamer, Arjo. 1995. One quarter of GDP is persuasion. *American Economic Review Papers and Proceedings*, **85**(2), 191–195.

Milgrom, Paul. 1981. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, **12**(2), 380–391.

Milgrom, Paul, & Roberts, John. 1986. Relying on the information of interested parties. *RAND Journal of Economics*, **17**(1), 18–32.

Milgrom, Paul R, & Weber, Robert J. 1982. A theory of auctions and competitive bidding. *Econometrica*, **50**(5), 1089–1122.

Mullainathan, Sendhil, Schwartzstein, Joshua, & Shleifer, Andrei. 2008. Coarse thinking and persuasion. *Quarterly Journal of Economics*, **123**(2), 577–619.

Myerson, Roger B. 1979. Incentive compatability and the bargaining problem. *Econometrica*, **47**(1), 61–73.

Ostrovsky, Michael, & Schwarz, Michael. 2008. Information dislosure and unraveling in matching markets. *Working Paper*.

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

Prendergast, Canice. 1992. The insurance effect of groups. *International Economic Review*, **33**(3), 567–81.

Rayo, Luis, & Segal, Ilya. 2008. Optimal information disclosure. *Working Paper*.

Shin, Hyun Song. 2003. Disclosures and asset returns. *Econometrica*, **71**(1), 105–33.

Shiryaev, A. N. 1996. *Probability*. Springer.

Shmaya, Eran, & Yariv, Leeat. 2009. Foundations for Bayesian updating. *Working Paper*.

Spence, A. Michael. 1973. Job market signaling. *Quarterly Journal of Economics*, **87**(3), 355–374.

Taub, Bart. 1997. Dynamic agency with feedback. *RAND Journal of Economics*, **28**(3), 515–543.