Ricardo Alonso and Odilon Câmara

Bayesian persuasion with heterogeneous priors

Article (Published version)
(Refereed)

Bayesian Persuasion with Heterogeneous Priors*

RICARDO ALONSO† ODILON CÂMARA‡

London School of Economics and CEPR *University of Southern California*

July 10, 2016

Abstract

In a world in which rational individuals may hold different prior beliefs, a sender can influence the behavior of a receiver by controlling the informativeness of an experiment (public signal). We characterize the set of distributions of posterior beliefs that can be induced by an experiment, and provide necessary and sufficient conditions for a sender to benefit from persuasion. We then provide sufficient conditions for the sender to benefit from persuasion for almost every pair of prior beliefs, even when there is no value of persuasion under a common prior. Our main condition is that the receiver's action depends on his beliefs only through his expectation of some random variable.

JEL classification: D72, D83, M31.

Keywords: Persuasion, strategic experimentation, heterogeneous priors.

*An earlier version of this paper circulated under the title "Persuading Skeptics and Reaffirming Believers." We thank Dan Bernhardt, Emir Kamenica, and Anton Kolotilin for detailed comments on earlier drafts of this paper, and Clement Minaudier for excellent research assistance. We also thank Isabelle Brocas, Juan Carrillo, Maxim Ivanov, Navin Kartik, Jin Li, Tony Marino, Niko Matouschek, John Matsusaka, Tymofiy Mylovanov, Michael Powell, Luis Rayo, Joel Sobel, Eric Van den Steen, Tim Van Zandt and Yanhui Wu for their suggestions, as well as the following audiences: 2014 Conference on Media and Communication (University of Chicago), 2013 SWET, Carey Business School, Claremont Graduate University, Kellogg School of Management, London School of Economics, McMaster University, Queen's University, University of Bonn, University of British Columbia, University of Southern California, and University of Western Ontario. The paper also benefited from the helpful comments of the editor and three anonymous referees.

†LSE, Houghton Street, London WC2A 2AE, United Kingdom. R.Alonso@lse.ac.uk

‡USC FBE Dept, 3670 Trousdale Parkway Ste. 308, BRI-308 MC-0804, Los Angeles, CA 90089-0804. ocamara@marshall.usc.edu.

1 Introduction

A notable feature of organizations is that those with decision-making power are lobbied. In many cases, individuals influence decision makers by changing the information available to them. For instance, individuals can acquire and communicate hard evidence, or signal soft information. Another way of influencing decision makers' learning is through strategic experimentation — i.e., by establishing what they can learn from the outcome of a public experiment (as in, for example, Brocas and Carrillo (2007) and Kamenica and Gentzkow (2011)).

Persuasion through strategic experimentation is pervasive in economics and politics. A pharmaceutical company chooses which initial animal tests to perform, and the results influence the Food and Drug Administration's decision to approve human testing. A central bank shapes the informativeness of a market index observed by households (such as inflation) by determining which information is collected and how to compute the index. A news channel selects the questions that the host asks during an electoral debate, and the answers affect voters' opinions about the candidates. In all of these cases, modifying the characteristics of the experiment (e.g., changing the test, the rules to generate the index, or the questions asked) changes what decision makers can learn. In many relevant cases, persuasion takes place within environments in which individuals hold heterogeneous prior beliefs.[1] In this paper, we ask: how does open disagreement affect an individual's benefit from persuading others, and her choice of an optimal experiment?

The next example, in which a politician (sender) seeks to maximize the effort of a bureaucrat (receiver), illustrates our main insights. The politician has proposed a new project that must be implemented by an existing government agency. She wants to maximize the probability that the project will be successfully implemented because this increases her reelection probability. However, the probability of successful implementation depends on the effort exerted by the bureaucrat who controls the agency. Since the bureaucrat wants to maximize his career perspectives, he will exert more effort only if he thinks that the new project is

[1] Many papers study the role of heterogeneous priors in economics and politics. Giat et al. (2010) use data on pharmaceutical projects to study R&D under heterogeneous priors; Patton and Timmermann (2010) find empirical evidence that heterogeneity in prior beliefs is an important factor explaining the cross-sectional dispersion in forecasts of GDP growth and inflation; Gentzkow and Shapiro (2006) study the effects of prior beliefs on media bias.

more beneficial than other existing projects to his agency's own goals. In addition to having different goals, in many cases, the politician and the bureaucrat have heterogeneous prior beliefs about the likely effects of the policy — see Hirsch (forthcoming) for a review of the literature on principal-agent models of policymaking in political organizations, and on the empirical evidence of belief disagreement between politicians and bureaucrats. To motivate the bureaucrat to exert more effort, suppose that, prior to fully implementing the policy, the politician can design a policy experiment — a pilot test that generates a public signal about how the new policy will benefit the agency. The bureaucrat can then use the information uncovered by this experiment to update his beliefs and adjust his effort choice. How does the politician optimally design such a policy experiment?

This problem has gained increasing attention from governments around the world. For instance, in 2010, David Cameron created the Behavioural Insights Team (BIT), a unit under the Cabinet Office. The BIT would conduct small-scale tests of certain policy interventions before they were broadly implemented by the UK government. The launch of the program "was greeted with a great scepticism" (Rutter, 2015). However, it eventually had an important impact on the relationship between politicians and bureaucrats. Before the program, new governments would try to impose new ideas and projects on bureaucrats without much empirical information to persuade them about the new policy's value. With the program, the government has more flexibility and control to uncover hard information to persuade bureaucrats.[2] After the initial success of the program, the BIT now "permeates almost every area of government policy," and are setting up similar programs in Australia, Singapore, Germany and the US (Rutter, 2015).

Therefore, consider a politician who can design a policy experiment to influence a bureaucrat. For simplicity, let the politician's payoff be a, which is the bureaucrat's effort to

[2] In a recent interview, David Halpern (chief executive of BIT) said, "If you're a permanent secretary or head of department[,] you have seen lots of ideas come and go. New governments come in on a wave of new shiny ideas. But permanent secretaries can read a graph pretty well" (Rutter, 2015). This is an old concern for bureaucrats around the globe. In 1996, Richard Darman (director of the US Office of Management and Budget and a member of President Bush's Cabinet from 1989 to 1993) argued: "As a society, we have been guilty of what can fairly be termed policy corruption. In pursuit of bold visions, we have launched one risky scheme after another without anything like responsible evidence. [...] Instead of [...] new Federal programs launched at full scale, [the President] could initiate a set of bold research trials" (Darman, 1996).

implement the new project. The bureaucrat's payoff is $u_{\text{Bur}}(a, \theta) = \theta a - \frac{a^\rho}{\rho}$, where $\rho \geq 2$ is a known preference parameter, and $\theta > 0$ captures the project's uncertain benefit to the agency's goals. The bureaucrat's effort choice is, then, a concave function of his expectation, $a^* = (E_{\text{Bur}}[\theta])^{\frac{1}{\rho-1}}$. Prior to fully implementing the policy, the politician can design a policy experiment that generates a public signal about θ.[3] Can the politician benefit from persuasion? That is, can she design an experiment that, on average, leads the bureaucrat to exert more effort?

First, suppose that players have a common prior belief over θ. The linearity of the politician's payoff and the concavity of the bureaucrat's effort choice imply that the politician's expected payoff is a concave function of beliefs. Therefore, there is no experiment that benefits the politician — see Kamenica and Gentzkow (2011) (KG henceforth). Now, suppose that players have heterogeneous prior beliefs. Let $E_{\text{Pol}}[\theta]$ and $E_{\text{Bur}}[\theta]$ be the expected value of θ from the point of view of the politician and the bureaucrat. Trivially, if effort is linear in expectation ($\rho = 2$) and the bureaucrat is a "skeptic" ($E_{\text{Bur}}[\theta] < E_{\text{Pol}}[\theta]$), then the politician benefits from persuading the bureaucrat. In particular, from the politician's point of view, a fully informative experiment that reveals θ is better than no experiment.[4] One could then conjecture that if effort is too concave (high ρ) or if the bureaucrat is already a "believer" ($E_{\text{Bur}}[\theta] > E_{\text{Pol}}[\theta]$), then the politician cannot benefit from designing an experiment. Perhaps surprisingly, this conjecture is wrong. Given any finite ρ, if there are at least three possible values of θ, then the politician *generically* benefits from persuasion, where genericity is interpreted over the space of pairs of prior beliefs.

To provide some intuition for this result, suppose that $\rho = 2$ so that $a^* = E_{\text{Bur}}[\theta]$ in the previous example. Consider possible states $\theta \in \{1, 1.5, 2\}$: the politician's prior belief over states is $p_{\text{Pol}} = (0.85, 0.10, 0.05)$, while the bureaucrat's prior is $p_{\text{Bur}} = (0.10, 0.40, 0.50)$.

[3] For example, the UK government proposed a change in the way that Job Centre advisors conducted interviews with job seekers. The BIT conducted a small-scale test of the new policy (the Loughton Job Centre experiment) before the policy was scaled up to other Job Centres. According to the BIT, the pilot program showed very promising results and even increased staff happiness (see Figure 1.1 in The Behavioural Insights Team Update Report 2013-2015, available at http://www.behaviouralinsights.co.uk/publications/the-behavioural-insights-team-update-report-2013-2015/).

[4] Nevertheless, even if the bureaucrat is a skeptic, a fully informative experiment is often suboptimal. See Section 4.

The bureaucrat is then a believer of the policy, $E_{\text{Pol}}[\theta] = 1.1 < E_{\text{Bur}}[\theta] = 1.7$. Clearly, a fully revealing experiment does not benefit the politician, as she expects the bureaucrat's expectation of θ to decrease, on average. Nevertheless, the politician can still benefit from strategic experimentation. The optimal experiment determines only whether or not $\theta = 1.5$. The bureaucrat's expectation decreases to 1.5 when the experiment reveals $\theta = 1.5$, and it increases to $\frac{0.1 \times 1 + 0.5 \times 2}{0.1 + 0.5} = 1.8\overline{3}$ when the experiment shows that $\theta \neq 1.5$. With this experiment, the politician expects the average effort to increase to $0.90 \times 1.8\overline{3} + 0.10 \times 1.5 = 1.8$. To understand the result, first notice that players disagree on the likelihood of observing the different experimental outcomes, although they fully understand how the experiment is generated. The sender can then exploit this disagreement: In our example, the politician assigns more probability (0.90) than the bureaucrat (0.60) to the "beneficial" experiment result $\{\theta \neq 1.5\}$, and relatively less to the "detrimental" result $\{\theta = 1.5\}$. In fact, we show that, for this case, optimal experiments are always designed so that the sender is relatively more optimistic than the receiver regarding the likelihood of observing "better" experiment results (results that induce actions yielding a higher payoff to the sender). We also show that such experiments are (generically) available to the sender, irrespective of the receiver's beliefs.

Motivated by this example, we consider a general persuasion model in which a sender can influence a receiver's behavior by designing his informational environment. After observing the realization of a public experiment, the receiver applies Bayes' rule to update his belief, and chooses an action accordingly. The sender has no private information and can influence this action by determining what the receiver can learn from the experiment - i.e., by specifying the statistical relation of the experimental outcomes to the underlying state. We make three assumptions regarding how Bayesian players process information. First, it is common knowledge that players hold different prior beliefs about the state - i.e., they "agree to disagree." Second, this disagreement is non-dogmatic: each player initially assigns a positive probability to each possible state of the world.[5] Third, the experiment chosen by the sender is "commonly understood," in the sense that if players knew the actual realization of the state, then they would agree on the likelihood of observing each possible experimental outcome.

We start our analysis by asking: from the sender's perspective, what is the set of distri-

[5]See Galperti (2015) for the case of prior beliefs with different supports.

butions of posterior beliefs that can be induced by an experiment? We first show that, given priors p^S and p^R, posteriors q^S and q^R form a bijection — q^R is derived from q^S through a perspective transformation. Moreover, this transformation is *independent* of the actual experiment. Consequently, given prior beliefs, the probability distribution of posterior beliefs of *only one* player suffices to derive the *joint* probability distribution of posteriors generated by an arbitrary experiment. This result allows us to characterize the set of distributions of posteriors that can be induced by an experiment (Proposition 1). An important implication of our results is that belief disagreement does not expand this set - that is, it does not allow the sender to generate "more ways" to persuade the receiver. We then use the tools in KG to solve for the sender's optimal experiment (Proposition 2) and provide a necessary and sufficient condition for a sender to benefit from experimentation (Corollary 1), and for the optimal experiment to be fully revealing (Corollary 2).

In Section 4, we focus on models in which (i) the receiver's action equals his expectation of the state, $a^* = E_R[\theta]$; and (ii) the sender's payoff $u_S(a, \theta)$ is a smooth function of the receiver's action. We show that if there are three or more distinct states and $\partial u_S(a, \theta)/\partial a \neq 0$, then a sender *generically* benefits from persuasion. This result holds regardless of the relationship between the sender's payoff and the unknown state; regardless of the curvature of the sender's payoff with respect to the receiver's action; and in spite of the fact that the sender cannot induce "more" distributions over posterior beliefs than in the common-prior case.[6] To gain some intuition, consider the case $u_S(a, \theta) = a$, and note that every experiment induces a lottery over the receiver's actions. Belief disagreement over states translates to disagreement over the likelihood of different experimental outcomes and, hence, over the likelihood of different receiver's actions. We first show that persuasion is valuable whenever the sender can design a lottery in which she is relatively more optimistic than the receiver about higher, thus, more beneficial, actions. We then show that such lotteries exist for a generic pair of players' prior beliefs. In fact, any optimal experiment satisfies this property in a strong sense: the sender's relative optimism increases in the actions induced by the lottery.[7]

[6]Remarkably, the sender generically benefits from persuasion even in the most extreme case of conflict of preferences $u_S(a, \theta) = -u_R(a, \theta)$, so that the sender wants to minimize the receiver's payoff.

[7]Formally, if $\Pr_S[a]/\Pr_R[a]$ is the likelihood ratio of the probability that sender and receiver assign to the action a being induced through an experiment, then $\Pr_S[a]/\Pr_R[a]$ increases in a under an optimal

Our results show that persuasion should be widespread in situations of open disagreement. Yildiz (2004), Che and Kartik (2009), Van den Steen (2004, 2009, 2010a, 2011) and Hirsch (forthcoming) study models with heterogeneous priors in which a sender would prefer to face a like-minded receiver. In these cases, a sender believes the receiver's view to be wrong, and by providing a signal, she is likely to move the receiver's decision towards what she considers the right decision. That is, persuasion is valuable if belief disagreement is harmful to the sender. In other situations, however, the sender may benefit from belief disagreement. In our previous example, a politician interested in implementing a policy would prefer a bureaucrat that is overly optimistic about the policy's benefits. Providing a fully informative experiment to such a receiver would then be detrimental to the sender. Nevertheless, we find that persuasion is valuable even in these cases, in which belief disagreement is beneficial to the sender.

Our paper is primarily related to two strands in the literature.

Persuasion through Strategic Experimentation: Some recent papers study the gains to players from controlling the information that reaches decision makers. In Brocas and Carrillo (2007), a leader without private information sways a follower's decision in her favor by deciding the time at which a decision must be made. As information arrives sequentially, choosing the timing of the decision is equivalent to shaping (in a particular way) the information available to the follower. Duggan and Martinelli (2011) consider one media outlet that can affect electoral outcomes by choosing the "slant" of its news reports. Gill and Sgroi (2008, 2012) consider a privately informed principal who can subject herself to a test designed to provide public information about her type, and can optimally choose the test's difficulty. Rayo and Segal (2010) study optimal advertising when a company can design how to reveal its product's attributes, but it cannot distort this information. Kolotilin (2014, 2015) studies optimal persuasion mechanisms to a privately informed receiver. In a somewhat different setting, Ivanov (2010) studies the benefit to a principal of limiting the information available to a privately informed agent when they both engage in strategic communication (i.e., cheap talk). The paper most closely related to ours is KG. The authors analyze the problem of a sender who wants to persuade a receiver to change his action for arbitrary state-dependent preferences for both the sender and the receiver, and for arbitrary, *but common*, prior be-

experiment.

liefs. We contribute to this literature by introducing and analyzing a new motive for strategic experimentation: belief disagreement over an unknown state of the world.

Heterogeneous Priors and Persuasion: Several papers in economics, finance and politics have explored the implications of heterogeneous priors for equilibrium behavior and the performance of different economic institutions. In particular, Yildiz (2004), Van den Steen (2004, 2009, 2010a, 2011), Che and Kartik (2009) and Hirsch (forthcoming) show that heterogeneous priors increase agents' incentives to acquire information, as each agent believes that new evidence will back his "point of view" and, thus, "persuade" others. Our work complements this view by showing that persuasion may be valuable even when others hold "beneficial" beliefs from the sender's perspective. We also differ from this work in that we consider situations in which the sender has more leeway in shaping the information that reaches decision makers.

We present the model's general setup in Section 2. Section 3 characterizes the value of persuasion. In Section 4, we examine a class of persuasion models. Section 5 presents an extension of the model. Section 6 concludes. All proofs are in the Appendices.

2 The Model

Preferences and Prior Beliefs: Players are expected utility maximizers. The receiver selects an action a from a compact set A. The sender and the receiver have preferences over actions characterized by continuous von Neumann-Morgenstern utility functions $u_S(a, \theta)$ and $u_R(a, \theta)$, with $\theta \in \Theta$ and Θ a finite state space, common to both players.

Both players are initially uncertain about the realization of the state θ. A key aspect of our model is that players openly disagree about the likelihood of θ. Following Aumann (1976), this implies that rational players must then hold different prior beliefs.[8] Thus, let the receiver's prior be $p^R = \left(p^R_\theta\right)_{\theta \in \Theta}$ and the sender's prior be $p^S = \left(p^S_\theta\right)_{\theta \in \Theta}$. We assume that p^R and p^S belong to the interior of the simplex $\Delta(\Theta)$ - that is, players have prior beliefs that are "totally mixed," as they have full support.[9] This assumption will avoid known issues of

[8]See Morris (1994, 1995) and Van den Steen (2010b, 2011) for an analysis of the sources of heterogeneous priors and extended discussions of their role in economic theory.

[9]Actually, our results require only that players' prior beliefs have a common support, which may be a strict subset of Θ. Assuming a full support eases the exposition without any loss of generality.

non-convergence of posterior beliefs when belief distributions fail to be absolutely continuous with respect to each other (see Blackwell and Dubins, 1962, and Kalai and Lehrer, 1994).

In our base model, prior beliefs are common knowledge. We extend the base model in Section 5 to the case in which players' prior beliefs are drawn from some distribution $H(p^R, p^S)$. Depending on the support of this distribution, belief disagreement may not be common knowledge among players.

It is natural to inquire whether the sources of heterogeneous priors affect the way in which players process new information. For instance, mistakes in information processing will eventually lead players to different posterior beliefs, but will also call Bayesian updating into question. We take the view that players are Bayes rational, but may initially openly disagree on the likelihood of the state. This disagreement can come, for example, from a lack of experimental evidence or historical records that would otherwise allow players to reach a consensus on their prior views.[10] Disagreement can also come from Bayesian players that misperceive the extent to which others are differentially informed (Camerer, Lowenstein and Weber, 1989). For instance, the receiver may fail to realize that the sender had private information when selecting an experiment. A privately informed sender who is aware of this perception bias will then select an experiment as if players openly disagreed about the state.

Strategic Experimentation: All players process information according to Bayes' rule. The receiver observes the realization of an experiment π, updates his belief, and chooses an action. The sender can affect this action through the design of π. To be specific, an experiment π consists of a finite realization space Z and a family of likelihood functions over Z, $\{\pi(\cdot|\theta)\}_{\theta\in\Theta}$, with $\pi(\cdot|\theta) \in \Delta(Z)$. Note that whether or not the realization is observed by the sender does not affect the receiver's actions.

Key to our analysis is that π is a "commonly understood experiment": the receiver observes the sender's choice of π, and all players agree on the likelihoods $\pi(\cdot|\theta)$.[11] Common

[10] In fact, as argued by Van den Steen (2011), the Bayesian model specifies how new information is to be processed, but, is largely silent on how priors should be (or actually are) formed. Lacking a rational basis for selecting a prior, the assumption that individuals should, nevertheless, all agree on one may seem unfounded.

[11] Our assumption of a commonly understood experiment is similar to the notion of "concordant beliefs" in Morris (1994). Morris (1994) indicates that "beliefs are concordant if they agree about everything except the prior probability of payoff-relevant states." Technically, his definition requires both agreement over the

agreement over π generates substantial congruence: if all players knew the actual realization of the state, then they would all agree on the likelihood of observing each $z \in Z$ for any experiment π.[12]

We make two important assumptions regarding the set of available experiments. First, the sender can choose *any experiment* that is correlated with the state. Thus, our setup provides an upper bound on the sender's benefit from persuasion in a setting with a more restricted space of experiments. Second, experiments are costless to the sender. This is not a serious limitation if all experiments impose the same cost, and would not affect the sender's choice if she decides to experiment. However, the optimal experiment may change if different experiments impose different costs.[13]

Our setup is related to models that study agents' incentives to affect others' learning - e.g., through "signal jamming," as in Holmström's model of career concerns (Holmström, 1999), or through obfuscation, as in Ellison and Ellison (2009). In contrast to this literature, the sender in our model shapes the receiver's learning through the statistical specification of a public experiment. For instance, rating systems and product certification fit this framework, with consumers observing an aggregate measure of the underlying quality of firms/products. Quality tests provide another example, as a firm may not know the quality of each single product, but can control the likelihood that a test detects a defective product.

In our model of strategic experimentation, the sender has no private information when selecting an experiment. As KG show, this model is isomorphic to a model in which a sender can commit to a disclosure rule before becoming privately informed - i.e., commit to how her knowledge will map to her advice. It is also equivalent to models in which a sender is required to certifiably disclosed her knowledge while being free to choose what she actually learns (Gentzkow and Kamenica, 2014b).

Our focus is on understanding when and how the sender benefits from experimentation.

---

conditional distribution of an experiment's realizations, given the state, and that each player assigns positive probability to each realization. Our assumptions of a commonly understood experiment and totally mixed priors imply that players' beliefs are concordant in our setup.

[12]See Van den Steen (2011) and Acemoglu et al. (2006) for models in which players also disagree on the informativeness of experiments.

[13]Gentzkow and Kamenica (2014a) offer an initial exploration of persuasion with costly experiments, where the cost of an experiment is given by the expected Shannon entropy of the beliefs that it induces.

Given π, for a realization z that induces the profile of posterior beliefs $(q^S(z), q^R(z))$, the receiver's choice in any Perfect Bayesian equilibrium must satisfy

$$a(q^R(z)) \in \arg\max_{a \in A} \sum_{\theta \in \Theta} q^R_\theta(z) u_R(a, \theta),$$

while the corresponding (subjective) expected utility of the sender after z is realized is

$$\sum_{\theta \in \Theta} q^S_\theta(z) u_S(a(q^R(z)), \theta).$$

We restrict attention to equilibria in which the receiver's choice depends only on his posterior belief induced by the observed realization. To this end, we define a *language-invariant Perfect Bayesian equilibrium* as a Perfect Bayesian equilibrium in which for all experiments π and π', and realizations z and z' for which $q^R(z) = q^R(z')$, the receiver selects the same action (or the same probability distribution over actions). Our focus on language-invariant equilibria allows us to abstract from the particular realization. Given an equilibrium $a(\cdot)$, we define the sender's expected payoff v when players hold beliefs (q^S, q^R) as

$$v(q^S, q^R) \equiv \sum_{\theta \in \Theta} q^S_\theta u_S(a(q^R), \theta), \text{ with } a(q^R) \in \arg\max_{a \in A} \sum_{\theta \in \Theta} q^R_\theta u_R(a, \theta). \tag{1}$$

We concentrate on equilibria for which the function v is upper-semicontinuous. This class of equilibria is non-empty: an equilibrium in which the receiver selects an action that maximizes the sender's expected utility whenever he is indifferent between actions is a (sender-preferred) language-invariant equilibrium for which v is upper-semicontinous.[14] Given a language-invariant equilibrium that induces v, let V_π be the sender's expected payoff from experiment π, given prior beliefs. The sender's equilibrium expected utility is simply

$$V(p^S, p^R) = \max_\pi V_\pi(p^S, p^R) = \max_\pi \mathrm{E}^\pi_S \left[v(q^S(z), q^R(z)) \right], \tag{2}$$

where the maximum is computed over all possible experiments π. An optimal experiment π^* is such that $V_{\pi^*}(p^S, p^R) = V(p^S, p^R)$. We can then define the *value of persuasion* as the sender's equilibrium expected gain when, in the absence of experimentation, the receiver would remain uninformed; it is given by $V(p^S, p^R) - v(p^S, p^R)$.

[14] As noted in KG, this follows from Berge's maximum theorem. Upper-semicontinuity will prove convenient when establishing the existence of an optimal experiment.

Timing: The sender selects π after which the receiver observes a realization $z \in Z$, updates his beliefs according to Bayes' rule, selects an action, payoffs are realized and the game ends. We focus on language-invariant perfect equilibria for which v is upper-semicontinuous.

We have been silent regarding the true distribution governing the realization of θ. As our analysis is primarily positive and considers only the sender's choice of an experiment, we remain agnostic as to the true distribution of the state.

Notational Conventions: Let $card(A)$ denote the cardinality of the set A. For vectors $s, t \in \mathbb{R}^N$, let st be the component-wise product of s and t; that is, $(st)_i = s_i t_i$, and let $\langle s, t \rangle$ represent the standard inner product in $\mathbb{R}^N$, $\langle s, t \rangle = \sum_{i=1}^N s_i t_i$. As ours is a setup with heterogeneous priors, this notation proves convenient when computing expectations for which we need to specify both the information set and the individual whose perspective we are adopting. We will often refer to the subspace W of "marginal beliefs," defined as

$$W = \left\{ w \in \mathbb{R}^N : \langle 1, w \rangle = 0 \right\}. \tag{3}$$

This terminology follows from the fact that the difference between any two beliefs must lie in W. Also, we will denote by $s_{||W}$ the orthogonal projection of s onto W.

Let $r^S_\theta = \frac{p^S_\theta}{p^R_\theta}$ and $r^R_\theta = \frac{p^R_\theta}{p^S_\theta}$ be the state-θ likelihood ratios of prior beliefs. We then define

$$r^S = (r^S_\theta)_{\theta \in \Theta} = \left(\frac{p^S_\theta}{p^R_\theta} \right)_{\theta \in \Theta} \text{ and } r^R = (r^R_\theta)_{\theta \in \Theta} = \left(\frac{p^R_\theta}{p^S_\theta} \right)_{\theta \in \Theta}. \tag{4}$$

Given π, we denote by $\Pr_S[z]$ and $\Pr_S[z]$ the probabilities of realization z obtained from the sender's and the receiver's beliefs, and define the likelihood-ratios over realizations

$$\lambda^S_z \equiv \frac{\Pr_S[z]}{\Pr_R[z]} \quad \text{and} \quad \lambda^R_z \equiv \frac{\Pr_R[z]}{\Pr_S[z]}. \tag{5}$$

3 The Value of Persuasion under Open Disagreement

When does the sender benefit from experimentation? We show that, when the experiment is commonly understood, the posterior belief of one player can be obtained from that of another player without explicit knowledge of the actual experiment. This allows us to characterize the (subjective) distributions of posterior beliefs that can be induced by any experiment (Proposition 1). It also enables us to translate the search for an optimal experiment to an auxiliary problem - where the belief of each player is expressed in terms of the belief of a reference player- and then apply the techniques developed in KG to solve it (Proposition 2). We obtain necessary and sufficient conditions for a sender to benefit from experimentation (Corollary 1), and for a sender to select a fully informative experiment (Corollary 2).

3.1 Induced Distributions of Posterior Beliefs

From the sender's perspective, each experiment π induces a (subjective) distribution over profiles of posterior beliefs. In any language-invariant equilibrium, the receiver's posterior belief uniquely determines his action. Thus, knowledge of the distribution of posterior beliefs suffices to compute the sender's expected utility from π.

If players share a common prior p, KG show that the martingale property of posterior beliefs $E^{\pi}[q] = p$ is both necessary and sufficient to characterize the set of distributions of beliefs that can be induced in Bayesian rational players by some experiment. This leads us to ask: when players hold heterogeneous priors, what is the set of joint distributions of posterior beliefs that are consistent with Bayesian rationality? While the martingale property still holds when a player evaluates the induced distribution of his own posterior beliefs, it is no longer true that the sender's expectation over the receiver's posterior belief always equals the receiver's prior. Nevertheless, we next show that posteriors q^S and q^R form a bijection — q^R is derived from q^S through a perspective transformation. Moreover, this transformation is *independent* of the experiment π and realization z.

Proposition 1 *Let the prior beliefs of the sender and the receiver be the totally mixed beliefs p^S and p^R, and let $r^R = \left(r^R_\theta\right)_{\theta\in\Theta}$ be the likelihood-ratio defined by (4). From the sender's perspective, a distribution over profiles of posterior beliefs $\tau \in \Delta\left(\Delta\left(\Theta\right)\times\Delta\left(\Theta\right)\right)$ is induced by some experiment if and only if*
(i) if $(q^S, q^R) \in Supp(\tau)$, then

$$q^R_\theta = q^S_\theta \frac{r^R_\theta}{\sum_{\theta'\in\Theta} q^S_{\theta'} r^R_{\theta'}} = \frac{q^S_\theta r^R_\theta}{\langle q^S, r^R\rangle}. \tag{6}$$

(ii) $E_\tau[q^S] = p^S$.

Proposition 1 establishes that the martingale property of the sender's beliefs and the perspective transformation (6), together, characterize the set of distributions of posterior beliefs that are consistent with Bayesian rationality. It also shows that, in spite of the degrees of freedom afforded by heterogeneous priors, not all distributions are consistent with Bayesian rationality. Indeed, any two experiments that induce the same marginal distribution over the sender's posterior must necessarily induce the same marginal distribution over the receiver's posteriors.[15] In fact, (6) implies that the set of joint distributions of players posterior beliefs under common priors and heterogeneous priors form a bijection. That is, belief disagreement does not generate "more ways" to persuade the receiver. Equation (6) relies on both the assumptions of common support of priors and a commonly understood experiment. One implication of a common support of priors is that any realization that leads the receiver to revise his belief must also induce a belief update by the sender — a realization is uninformative to the receiver if and only if it is uninformative to the sender.

Expression (6) affords a simple interpretation. Heterogeneous priors over θ imply that, for given π, with realization space Z, players also disagree on how likely they are to observe each $z \in Z$. Just as the prior disagreement between the receiver and the sender is encoded in the likelihood ratio $r_\theta^R = p_\theta^R / p_\theta^S$, we can encode the disagreement over z in the likelihood ratio $\lambda_z^R = \Pr_R(z)/\Pr_S(z)$, defined by (5). The proof of Proposition 1 shows that this likelihood ratio can be obtained from r^R by

$$\lambda_z^R = \left\langle q^S(z), r^R \right\rangle . \tag{7}$$

From (6) and (7), we can relate the updated likelihood ratio $q_\theta^R(z)/q_\theta^S(z)$ to r^R and λ_z^R,

$$\frac{q_\theta^R(z)}{q_\theta^S(z)} = \frac{r_\theta^R}{\lambda_z^R}. \tag{8}$$

In words, the state-θ likelihood ratio after updating based on z is the ratio of the likelihood ratio over states to the likelihood ratio over realizations. This implies that a realization z that comes more as a "surprise" to the receiver than to the sender (so $\lambda_z^R < 1$) would lead to a larger revision of the receiver's beliefs and, thus, a component-wise increase in the updated likelihood ratio. Moreover, both likelihood ratios (r_θ^R and λ_z^R) are positively related, in the

[15]When players disagree on the likelihood functions that describe π (as is the case in Acemoglu et al., 2006 and Van den Steen, 2011), then, even for Bayesian players, knowledge of the marginal distribution of posterior beliefs of one player may not be enough to infer the entire joint distribution, and, thus, it may not be enough to compute the sender's expected utility from π.

sense that realizations that come more as a surprise to the receiver than to the sender are associated with states that the receiver perceives as less likely.[16]

As a final remark, note that the likelihood ratio r^R is the Radon-Nikodym derivative of p^R with respect to p^S. Therefore, (6) states that Bayesian updating under a commonly understood experiment simply induces a linear scaling of the Radon-Nikodym derivative, where the proportionality factor does not depend on the experiment π.

3.2 Value of Persuasion

If $\tau \in \Delta\left(\Delta\left(\Theta\right) \times \Delta\left(\Theta\right)\right)$ is a distribution over (q^S, q^R), then the sender's problem is

$$\begin{aligned} V(p^S, p^R) &= \sup_{\pi} \mathrm{E}_S^{\pi}\left[v(q^S(z), q^R(z))\right] \\ &\quad \text{s.t. } \tau \text{ is induced by } \pi, \end{aligned} \tag{9}$$

where τ obtains from π and the sender's prior p^S, and the receiver's posterior q^R follows from applying Bayes' rule to the prior p^R. Proposition 1 allows us to translate (9) to the following equivalent, but lower dimensional, optimization problem,

$$\begin{aligned} V(p^S, p^R) &= \sup_{\sigma} \mathrm{E}_{\sigma}\left[v(q^S, q^R)\right] \\ &\quad s.t.\ \sigma \in \Delta\left(\Delta\left(\Theta\right)\right), \mathrm{E}_{\sigma}\left[q^S\right] = p^S, q^R = \frac{q^S r^R}{\langle q^S, r^R\rangle}, \end{aligned} \tag{10}$$

By writing all posterior beliefs as a function of the beliefs of a reference player (in (10), the reference player is the sender), then (10) becomes amenable to the tools developed in KG.

[16]Formally, given experiment π, consider the probability distribution $\zeta^j(\theta, z)$ in $\Theta \times Z$ defined by $\zeta^j(\theta, z) = \pi(z|\theta)p^j_\theta$. Define the random variables $r^i(\theta, z) = r^i_\theta$ and $\lambda^i(\theta, z) = \lambda^i_z$. Then, r^i and λ^i are positively (linearly) correlated under $\zeta^j(\theta, z)$. To see this, note that

$$\begin{aligned} \mathrm{E}_{\zeta^i}\left[\lambda^i r^i\right] &= \sum_{z\in Z}\sum_{\theta\in\Theta} \frac{\langle \pi(z), p^i\rangle}{\langle \pi(z), p^j\rangle} \frac{p^i_\theta}{p^j_\theta} \pi(z|\theta) p^j_\theta = \sum_{z\in Z} \left(\frac{\langle \pi(z), p^i\rangle}{\langle \pi(z), p^j\rangle}\right)^2 \langle \pi(z), p^j\rangle, \\ &\geq \left(\sum_{z\in Z} \frac{\langle \pi(z), p^i\rangle}{\langle \pi(z), p^j\rangle} \langle \pi(z), p^j\rangle\right)^2 = 1, \\ \mathrm{E}_{\zeta^i}\left[r^i\right] &= \sum_{z\in Z}\sum_{\theta\in\Theta} \frac{p^i_\theta}{p^j_\theta} \pi(z|\theta) p^j_\theta = 1, \\ \mathrm{E}_{\zeta^i}\left[\lambda^i\right] &= \sum_{z\in Z}\sum_{\theta\in\Theta} \frac{\langle \pi(z), p^i\rangle}{\langle \pi(z), p^j\rangle} \pi(z|\theta) p^j_\theta = \sum_{z\in Z} \langle \pi(z), p^i\rangle = 1. \end{aligned}$$

The next proposition provides properties of optimal experiments. For this purpose, and following KG, for an arbitrary real-valued function f, define $\widetilde{f}$ as the concave closure of f,

$$\widetilde{f}(q) = \sup \{w | (q, w) \in co(f)\},$$

where $co(f)$ is the convex hull of the graph of f. In other words, $\widetilde{f}$ is the smallest upper semicontinuous and concave function that (weakly) majorizes the function f.

Proposition 2 *(i) An optimal experiment exists. Furthermore, there exists an optimal experiment with realization space Z such that $card(Z) \leq \min\{card(A), card(\Theta)\}$.*
(ii) Define the function V_S by

$$V_S\left(q^S\right) = v\left(q^S, \frac{q^S r^R}{\langle q^S, r^R \rangle}\right). \tag{11}$$

The sender's expected utility under an optimal experiment is

$$V(p^S, p^R) = \widetilde{V}_S\left(p^S\right). \tag{12}$$

Expression (12) implies that the value of persuasion is $\widetilde{V}_S\left(p^S\right) - V_S\left(p^S\right)$. Direct application of Proposition 2 to establish whether this value is positive would require the derivation of the concave closure of an upper-semicontinous function. Nevertheless, the following corollary provides conditions that make it easier to verify whether experimentation is valuable.

Corollary 1 *There is no value of persuasion if and only if there exists a vector $\gamma \in \mathbb{R}^{card(\Theta)}$ such that*

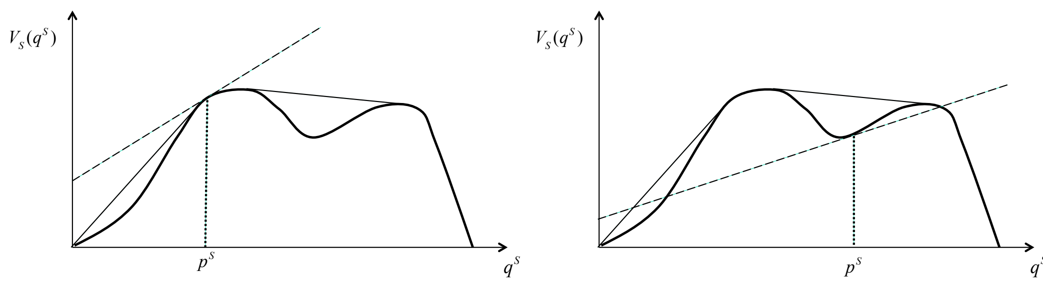
$$\left\langle \gamma, q^S - p^S \right\rangle \geq V_S\left(q^S\right) - V_S\left(p^S\right), \ q^S \in \Delta(\Theta). \tag{13}$$

In particular, if V_S is differentiable at p^S, then there is no value of persuasion if and only if

$$\left\langle \nabla V_S\left(p^S\right), q^S - p^S \right\rangle \geq V_S\left(q^S\right) - V_S\left(p^S\right), \ q^S \in \Delta(\Theta). \tag{14}$$

This corollary provides a geometric condition for the value of persuasion to be zero: a sender does not benefit from experimentation if and only if V_S admits a supporting hyperplane at p^S. This observation is based on the characterization of concave functions as the infimum of affine functions, and Figure 1 depicts this insight graphically.

If (13) is violated, then the sender will choose to experiment. Corollary 2 shows when the sender will choose an experiment that perfectly reveals the state. For this purpose, let $\mathbf{1}_\theta$ be the posterior belief that puts probability 1 on state θ.



(a) No Value of Persuasion (b) Positive Value of Persuasion

Figure 1: Illustration of Corollary 1

Corollary 2 *A perfectly informative experiment is optimal if and only if*

$$\sum_{\theta\in\Theta} q^S_\theta u_S(a(\mathbf{1}_\theta),\theta) \geq V_S\left(q^S\right),\ q^S \in \Delta\left(\Theta\right). \tag{15}$$

Condition (15) admits a simple interpretation. Suppose that players observe a realization that induces q^S in the sender. The right-hand side of (15) is the sender's expected utility if she discloses no more information, while the left-hand side of (15) is the sender's expected utility if she allows the receiver to perfectly learn the state. Then, a sender does not benefit from garbling a perfectly informative experiment if and only if for *every* possible experiment π and realization z, she is not worse off by fully revealing the state.

In some applications, it will be convenient to rewrite the sender's problem as follows. Define $\check{u}_S(a,\theta) = u_S(a,\theta) r^S_\theta$, where the likelihood ratio r^S_θ is defined by (4). For any experiment $\pi = (Z, \{\pi\left(\cdot|\theta\right)\}_{\theta\in\Theta})$ and receiver's decision rule $a(z)$, $z \in Z$, we have

$$\mathrm{E}_S\left[u_S(a(z),\theta)\right] = \sum_{\theta\in\Theta}\sum_{z\in Z}\pi(z|\theta)p^S_\theta u_S(a(z),\theta) = \sum_{\theta\in\Theta}\sum_{z\in Z}\pi(z|\theta)p^R_\theta u_S(a(z),\theta)r^S_\theta = \mathrm{E}_R\left[\check{u}_S(a(z),\theta)\right].$$

That is, the expected utility of a sender with prior p^S and utility u_S is the same as the expected utility of a sender who shares the receiver's prior p^R, but has utility $\check{u}_S$. Thus, under a commonly understood experiment, one can convert the original problem to one with common priors as follows. Rewrite (1) as $\check{v}\left(q^S, q^R\right) \equiv \sum_{\theta\in\Theta} q^S_\theta \check{u}_S(a(q^R),\theta)$, and define

$$V_R\left(q^R\right) = \check{v}\left(q^R, q^R\right). \tag{16}$$

Remark: *The claims of Proposition 2 remain valid if one substitutes $V_R\left(q^R\right)$ for $V_S\left(q^S\right)$.*

Note, however, that in many cases, the transformed utility $\check{u}_S$ is hard to interpret and defend on economic grounds. Moreover, by maintaining the original formulation, one is able to gather a better economic understanding of the implications of heterogeneous priors. For example, an important result in Section 4 is that on the space of pairs of prior beliefs, the sender generically benefits from persuasion. Such a result would be hard to postulate and interpret if one examined only the transformed problem.

4 Skeptics and Believers

How might a sender gain from designing a receiver's access to information? The literature has explored two broad sources of value under the assumption of a common prior. One source is based on the value of information: a sender who benefits from decisions that are adapted to the underlying state would certainly benefit from providing an informative experiment to a decision maker that shares her preferences. The other source is based on conflicting interests. For instance, if the sender's utility is independent of the state — "pure persuasion" —, then she would draw no value from learning the state if she could make decisions herself. However, KG and Brocas and Carrillo (2007) show that she can still benefit from experimentation if, instead, it is a receiver who makes decisions — when players share a common prior, the sender can exploit non-concavities in the receiver's action or in her own utility.

Van den Steen (2004, 2010a) and Che and Kartik (2009) show that the presence of heterogeneous priors can increase the incentives of influencers to persuade a decision maker who holds unfavorable beliefs. In this paper, we explore the extent to which open disagreement provides a third, distinct rationale for a sender to benefit from experimentation. To be sure, there are situations in which belief disagreement does not lead to experimentation. Proposition 3 provides necessary and sufficient conditions for the sender not to benefit from persuasion for *every* pair of mixed prior beliefs (p^R, p^S). We then provide sufficient conditions for the sender to benefit from persuasion for almost every pair of prior beliefs. Our main condition is that the receiver's action depends on his beliefs only through his expectation of some random variable. In this case, belief disagreement *generically* induces the sender to experiment, even when there is no value of persuasion under a common prior. Moreover, the optimal experiment is often not fully revealing of the state.

4.1 No Positive Value of Persuasion

We can express the sender's payoff $V_R\left(q^R\right)$ in (16) as

$$V_R\left(q^R\right) = \sum\nolimits_{\theta\in\Theta} \frac{p_\theta^S}{p_\theta^R} q_\theta^R u_S\left(a(q^R),\theta\right). \tag{17}$$

With common prior beliefs, KG show that there is no value of persuasion for every pair of common priors if and only if the expectation $\sum_{\theta\in\Theta} q_\theta^R u_S\left(a(q^R),\theta\right)$ is everywhere concave in q^R. With heterogeneous priors, this condition must be satisfied for each possible state.

Proposition 3 *The value of persuasion is zero for every pair of mixed prior beliefs if and only if for each state θ, the function $q_\theta^R u_S\left(a(q^R),\theta\right)$ is everywhere concave in q^R .*

The following example illustrates Proposition 3.

Example 1: Let $\Theta = \{\theta_L, \theta_H\}$, with $\theta_L < \theta_H$. Consider quadratic payoffs $u_R = -(a-\theta)^2$ and $u_S = -(a-f(\theta))^2$, where f captures the possible misalignment in preferences. The receiver's optimal action is, then, $a(q^R) = E_R[\theta]$. Using the condition from Proposition 3, the value of persuasion is zero for every pair of prior beliefs if and only if $f(\theta_H) \leq \theta_L < \theta_H \leq f(\theta_L)$. ■

The example shows that heterogeneous priors may not be enough for senders to engage in experimentation. In the example, this result follows from two forces. First, an application of Proposition 1 to a binary state shows that any realization that makes the receiver more optimistic about the state being θ_H also leads the sender to raise the likelihood of θ_H. Second, when $f(\theta_H) \leq \theta_L < \theta_H \leq f(\theta_L)$, the misalignment in preferences is extreme: the receiver would choose a higher action if he is more confident that $\theta = \theta_H$, while the sender would prefer a lower action if $\theta = \theta_H$ becomes more likely. Overall, the receiver would adversely adjust his action after any realization of any experiment, regardless of the prior disagreement.

4.2 Generic Positive Value of Persuasion

Consider the following model of persuasion. Let $A, \Theta \subset \mathbb{R}$. Our main assumption is that the receiver's action depends on his beliefs only through his expectation of some random variable, which we take to be the state θ. Formally, $a(q^R) = F\left(\left\langle q^R, \theta\right\rangle\right)$, with F twice continuously differentiable. We normalize the receiver's action by incorporating F into the sender's payoff:

(A1): *The receiver's action is* $a(q^R) = \langle q^R, \theta \rangle$.

(A2): *The sender's payoff* $u_S(a, \theta)$ *is a twice continuously differentiable function of* a.[17]

In Section 4.5, we provide a series of economic applications in which both assumptions hold.

Our first result is a sufficient condition for the sender to benefit from experimentation. We start by listing some definitions. For each state θ, let

$$u'_{S,\theta} \equiv \left. \frac{\partial u_S(a, \theta)}{\partial a} \right|_{a = \langle p^R, \theta \rangle}$$

be the sender's state-contingent marginal utility at the receiver's action chosen at his prior belief. Define the vector $u'_S \equiv \left(u'_{S,\theta} \right)_{\theta \in \Theta}$. Finally, we recall the following definition.

Definition: Vectors v and w are *negatively collinear with respect to the subspace* W, defined by (3), if there exist $\lambda < 0$ such that the projections[18] $v_{||W}$ and $w_{||W}$ satisfy

$$v_{||W} = \lambda w_{||W}. \tag{18}$$

We now state our first proposition in this section.

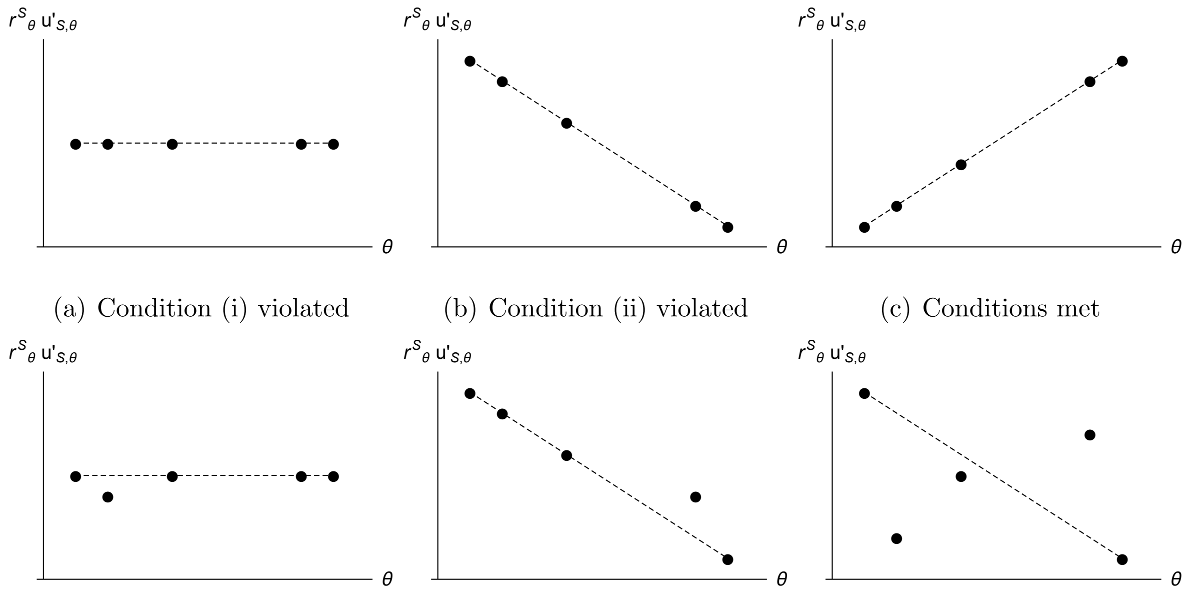
Proposition 4 *Suppose that **(A1)** and **(A2)** hold. If (i)* $\left(r^S \cdot u'_S \right)_{||W} \neq 0$, *and (ii)* $r^S \cdot u'_S$ *and* θ *are not negatively collinear with respect to* W*, then the sender benefits from persuasion.*

Conditions (i) and (ii) are easy to illustrate. For each state θ, we plot the point $(\theta, r^S_\theta u'_{S,\theta})$ on a two-dimensional graph. Condition (i) is violated if and only if *all* points fall on a single horizontal line (see Figure 2(a))— that is, if the term $r^S_\theta u'_{S,\theta}$ is constant across all states. Condition (ii) is violated if and only if *all* points fall on a single line with a strictly negative slope[19] (see Figure 2(b)). Figures 2(c) to (f) provide examples in which both conditions are satisfied; hence, the sender benefits from persuasion.

[17] It is immediate to rewrite our results for the case $a(q^R) = F\left(\langle q^R, x(\theta) \rangle\right)$, so that $x(\theta)$ is the random variable relevant to defining the receiver's action, and θ is the random variable relevant to the sender's payoff.

[18] Given vector $v = (v_1, \ldots, v_N)$, the projection $v_{||W}$ captures the deviation of each element of v from the mean of the elements of v: $v_{||W} = (v_1 - \sum_{n=1}^N v_n/N, \ldots, v_N - \sum_{n=1}^N v_n/N)$.

[19] For example, recall Example 1 from Section 4.1. Condition (ii) is violated whenever $(r^S_{\theta_H} u'_{S,\theta_H} - r^S_{\theta_L} u'_{S,\theta_L}) < 0$. If $f(\theta_H) \leq \theta_L < \theta_H \leq f(\theta_L)$, then $u'_{S,\theta_H} < 0$ and $u'_{S,\theta_L} > 0$ for every prior belief of the receiver. Hence, $(r^S_{\theta_H} u'_{S,\theta_H} - r^S_{\theta_L} u'_{S,\theta_L}) < 0$ for all r^S (for every pair of prior beliefs).



(a) Condition (i) violated (b) Condition (ii) violated (c) Conditions met

(d) Conditions met (e) Conditions met (f) Conditions met

Figure 2: Illustration of Conditions (i) and (ii) from Proposition 4

In the proof of Proposition 4, we exploit (16), which is the sender's payoff as a function of the receiver's belief, $V_R(q^R)$. The vector $r^S \cdot u'_S$ then represents the sender's expected marginal utility evaluated according to the receiver's prior belief:

$$E_S\left[u'_S|p^S\right] = \langle p^S, u'_s \rangle = \langle p^R \cdot r^S, u'_s \rangle = \langle p^R, r^S \cdot u'_s \rangle = E_R\left[r^S \cdot u'_S|p^R\right]. \tag{19}$$

Thus, $\left(r^S \cdot u'_S\right)_{||W}$ is the direction in the space of the receiver's beliefs along of highest rate of increase of the sender's expected utility. Likewise, $\theta_{||W}$ provides the direction in the space of the receiver's beliefs along which his expectation of θ, and, hence, his action, increases at the highest rate. Proposition 4 then states that the sender benefits from strategic experimentation whenever these two directions are not opposite to each other.[20] In this case, the proof of Proposition 4 shows that there exists a direction such that the sender's payoff V_R is locally strictly convex at p^R.

We now provide further intuition for Proposition 4. To do so, we construct a binary experiment that improves upon non-experimentation whenever $r^S \cdot u'_S$ and θ are not negatively collinear with respect to W. Intuitively, this binary experiment increases the receiver's action

[20]Note that Proposition 4 also applies to the case of common prior beliefs, so that $r^S = \mathbf{1}$. In this case, the sender benefits from experimentation if $u'_{S||W}$ and $\theta_{||W}$ are not negatively collinear.

only for beliefs where the sender's expected marginal utility is higher than under her prior belief.

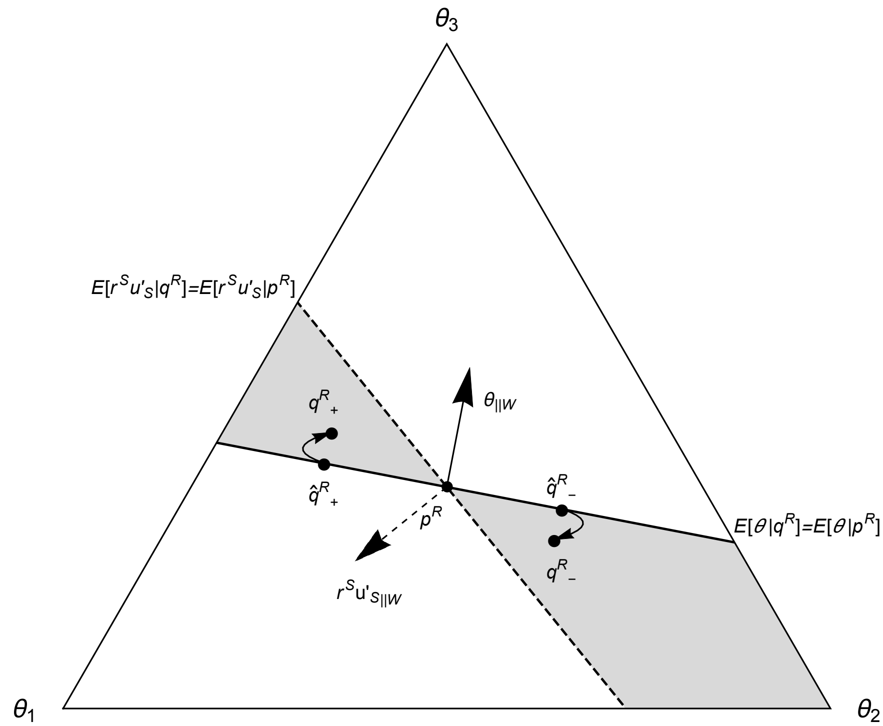


Figure 3: Finding a Beneficial Experiment

Figure 3 provides a graphical illustration of this beneficial experiment, which we construct in two steps. Consider, first, a binary experiment $\hat{\pi}$ with two equally likely outcomes that do not change the receiver's prior action. That is, under $\hat{\pi}$, the receiver can have one of two posterior beliefs, $\hat{q}^R_+ = p^R + w$ and $\hat{q}^R_- = p^R - w$, where $\left\langle \hat{q}^R_+ - p^R, \theta \right\rangle = \langle w, \theta \rangle = 0$. Clearly, the sender does not benefit from this experiment and $V_{\hat{\pi}} = 0$. Starting with $\hat{\pi}$, consider, now, a binary experiment π that induces one of two equally likely beliefs in the receiver, $q^R_+ = \hat{q}^R_+ + \varepsilon\theta_{\|W}$ and $q^R_- = \hat{q}^R_- - \varepsilon\theta_{\|W}$, with $\varepsilon > 0$. Under π, the receiver changes his action by $\Delta a = a(q^R_+) - a(p^R) = \varepsilon \left\|\theta_{\|W}\right\|^2$ if the realization induces q^R_+ and by $-\Delta a$ if it induces q^R_-. To understand the sender's gain from π, we compare the sender's expected gain from the realizations q_+^R under π and realization $\hat{q}_+^R$ under $\hat{\pi}$

$$\begin{aligned} V_\pi^+ - V_{\hat{\pi}}^+ &= \Pr{}_S\left[q_+^R\right] E_S[u_S(a\left(q_+^R\right),\theta)] - \Pr{}_S\left[\hat{q}_+^R\right] E_S[u_S(a\left(\hat{q}_+^R\right),\theta)] \\ &= \Pr{}_R\left[q_+^R\right] E_R[r^S u_S(a\left(q_+^R\right),\theta)|q_+^R] - \Pr{}_R\left[\hat{q}_+^R\right] E_S[r^S u_S(a\left(\hat{q}_+^R\right),\theta)|\hat{q}_+^R] \\ &\approx \frac{1}{2}\left(\left\langle \hat{q}_+^R, r^S\frac{\partial u_S}{\partial a}(a\left(p^R\right),\theta)\right\rangle \Delta a + \varepsilon\left\langle \theta_{\|W}, r^S u_S(a\left(p^R\right),\theta)\right\rangle\right). \end{aligned}$$

The first term is the change in the sender's expected utility from increasing the receiver's action by Δa at belief $\hat{q}_+^R$, while the second term gives the change in the sender's utility from the difference (from the sender's perspective) in the likelihood of q_+^R relative to $\hat{q}_+^R$. A similar analysis can be performed to compare the sender's expected gain under realization q_-^R under π relative to realization $\hat{q}_-^R$ under $\hat{\pi}$. Combining these two calculations, we have, after eliminating second-order terms[21]

$$\begin{aligned} V_\pi &= V_\pi - V_{\hat{\pi}} = V_\pi^+ - V_{\hat{\pi}}^+ + V_\pi^- - V_{\hat{\pi}}^- \\ &= \frac{1}{2}\left(\left\langle \hat{q}_+^R - \hat{q}_-^R, r^S\frac{\partial u_S}{\partial a}(a\left(p^R\right),\theta)\right\rangle \Delta a + \varepsilon\left\langle \theta_{\|W}, r^S\left(u_S(a\left(q_+^R\right),\theta) - u_S(a\left(q_-^R\right),\theta)\right)\right\rangle\right) \\ &\approx \left\langle w, r^S u_S'\right\rangle \Delta a \end{aligned} \tag{20}$$

Recall that the vector $w \in W$ is orthogonal to θ and $(r^S \cdot u_S')_{||W} \neq 0$. Therefore, (20) is identically zero if and only if $\left(r^S \cdot u_S'\right)_{||W}$ and $\theta_{||W}$ are collinear. If $\left(r^S \cdot u_S'\right)_{||W}$ and $\theta_{||W}$ are not collinear, however, one can find a vector w that makes (20) positive. Intuitively, under experiment π, it is more valuable for the sender to raise the receiver's action at $\hat{q}_+^R$ and less valuable at $\hat{q}_-^R$, relative to the prior belief p^R. Then, experiment π raises the sender's utility, as it induces the receiver to increase his action only for the realization for which the sender benefits relatively more from a higher action.

How often does the sender benefit from persuading the receiver? Our next result establishes sufficient conditions for the sender to generically benefit from persuasion, where genericity is interpreted over the space of pairs of prior beliefs. First, the state space must be sufficiently rich, $card\,(\Theta) > 2$. Moreover, we assume

[21] The second-order term that we eliminate is $\varepsilon\left\langle \theta_{\|W}, r^S\frac{\partial u_S}{\partial a}(a\left(p^R\right),\theta)\right\rangle \Delta a$, which captures the change in the sender's utility owing to the relative difference in the probability of q_+^R and $\hat{q}_+^R$ versus q_-^R and $\hat{q}_-^R$. The first-order term in (20) is zero if $(r^S \cdot u_S')_{||W}$ and $\theta_{||W}$ are collinear. In this case, this second-order term is positive, and, thus, the sender benefits from experiment π if $\theta_{\|W}$ and $r^S\frac{\partial u_S}{\partial a}(a\left(p^R\right),\theta)_{||W}$ are positively collinear.

(A3): *For almost every belief* p^R*, we have* $\left.\frac{\partial u_S(a,\theta)}{\partial a}\right|_{a=\langle p^R,\theta\rangle} \neq 0$ *for at least one* θ*.*

Assumption **(A3)** implies that for a generic prior belief of the receiver, changing the receiver's action marginally changes the sender's state-contingent payoff for at least one state. Condition **(A3)** holds in all applications of Section 4.5. Together, assumptions $card\,(\Theta) > 2$ and **(A3)** guarantee that both conditions (i) and (ii) from Proposition 4 hold generically.

Corollary 3 *Suppose that **(A1)** and **(A2)** hold. If* $card\,(\Theta) > 2$ *and **(A3)** hold, then the sender generically benefits from persuasion.*

A remarkable feature of Corollary 3 is that it does not impose conditions on the alignment of preferences between sender and receiver. Given a rich state space and conditions **(A1)** to **(A3)**, the sender can generically find a beneficial experiment to provide to the receiver even under extreme conflict of preferences — e.g., even if $u_S(a,\theta) = -u_R(a,\theta)$.

4.3 Pure Persuasion and Skeptics and Believers

In a world of common prior beliefs, KG describe how the value of persuasion fundamentally depends on the curvature of a sender's payoff as a function of the receiver's beliefs. In a world of heterogeneous prior beliefs, our Corollary 3 shows that if the state space is sufficiently rich and conditions **(A1)** to **(A3)** hold, then the sender generically benefits from persuasion. Furthermore, our conditions do not impose significant restrictions on the curvature of the sender's payoff other than smoothness.

Why is experimentation pervasive under open disagreement? To isolate the role of belief disagreement in strategic experimentation, we focus on the case of pure persuasion, in which the sender's utility is independent of the state:

(A2′): *The sender's payoff is* $u_S(a,\theta) = G(a)$*, with* G *twice continuously differentiable and* $G' > 0$*.*

In this case, the sender benefits from the receiver choosing a higher action, which occurs whenever he has a higher expectation of θ. We can then categorize as follows the type of receiver that the sender may face. A sender views a receiver as a *skeptic* if the sender would be made better off by a receiver who shares her point of view; that is, if $\langle q^R, \theta \rangle <$

$\langle q^S, \theta \rangle$ .Conversely, a sender views a receiver as a *believer* if the sender would not be made better off by a like-minded receiver; that is, if $\langle q^R, \theta \rangle \geq \langle q^S, \theta \rangle$.

From the sender's point of view, a fully revealing experiment increases the average action of an skeptic receiver, and (weakly) decreases that of a believer. Whether such experiments raise or decrease the sender's expected utility depends on her risk preferences, as captured by the curvature of G. Nevertheless, together, conditions **(A1)**, **(A2′)** and $card(\Theta) > 2$ imply that all conditions of Corollary 3 hold. Thus, persuasion is generically valuable, regardless of whether the sender is facing a skeptic or a believer, and regardless of her risk attitude.

We now derive a more intuitive interpretation of our collinearity condition in Proposition 4 when applied to the case of pure persuasion. We start by defining some relevant sets of beliefs. Let the *set of beneficial beliefs* A^+ be the set of the receiver's beliefs that would result in his choosing a (weakly) higher action than under the prior belief p^R, and A^- be the *set of detrimental beliefs*. That is,

$$\begin{aligned} A^+ &= \left\{ q^R \in \Delta(\theta) | \left\langle q^R, \theta \right\rangle \geq \left\langle p^R, \theta \right\rangle \right\}, \\ A^- &= \left\{ q^R \in \Delta(\theta) | \left\langle q^R, \theta \right\rangle < \left\langle p^R, \theta \right\rangle \right\}. \end{aligned} \tag{21}$$

Thus, the sender faces a skeptic if and only if $p^S \in A^+$. Figure 4(a) depicts the sets of beneficial beliefs (gray area) and detrimental beliefs (white area).

Recall that players disagree on the likelihood of reaching certain posterior beliefs. It follows from (7) that for every $q^R \in \Delta(\Theta)$, we have $\Pr_S[q^R] = \Pr_R[q^R]\langle q^R, r^S \rangle$. We say that the receiver underestimates q^R if $\Pr_S[q^R] > \Pr_R[q^R]$, and he overestimates q^R if $\Pr_S[q^R] < \Pr_R[q^R]$. We then define the sets of beliefs:

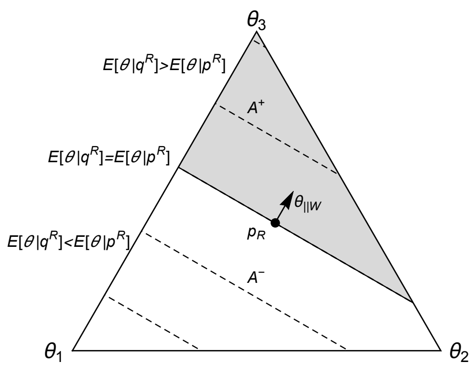
$$\begin{aligned} S^+ &= \{ q^R \in \Delta(\theta) | \langle q^R, r^S \rangle > 1 \}, \\ S^- &= \{ q^R \in \Delta(\theta) | \langle q^R, r^S \rangle < 1 \}. \end{aligned}$$

For every q^R in the support of π, the receiver underestimates q^R if and only if $q^R \in S^+$, and he overestimates q^R if and only if $q^R \in S^-$. Hence, we refer to S^+ as the *set of beliefs that the receiver underestimates*. Figure 4(b) depicts a series of hyperplanes along which $\langle q^R, r^S \rangle$ is constant. The gray area depicts S^+ and the white area depicts S^-.

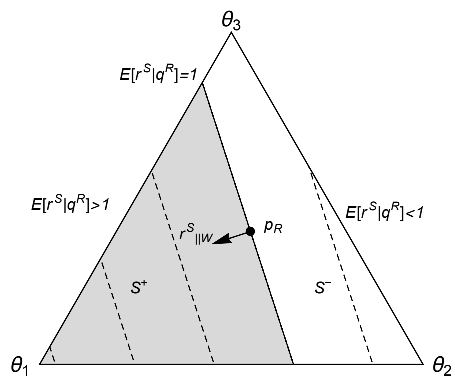
Given **(A1)** and **(A2′)**, note that the derivative $\frac{\partial u_S(a,\theta)}{\partial a} = G'(a) > 0$ is independent of the state; hence, all elements of u'_S are the same. In this case, conditions (i) and (ii) of

Proposition 4 afford a simple interpretation.

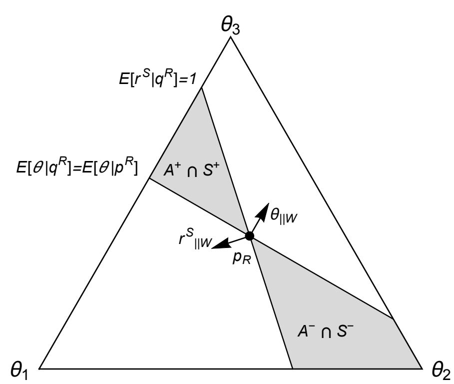
Lemma 1 *Suppose that* ***(A1)*** *and* ***(A2′)*** *hold. Then, the set of beneficial beliefs that the receiver underestimates is non-empty,* $A^+ \cap S^+ \neq \varnothing$*, if and only if (i) prior beliefs are not common, and (ii)* r^S *and* θ *are not negatively collinear with respect to* W*.*



(a) Beneficial beliefs



(b) Beliefs underestimated by the receiver



(c) Non-empty set $A^+ \cap S^+$

Figure 4: Finding a Beneficial Experiment.

Figure 4(c) describes the intersection of the sets A^+ and S^+ graphically. As the projections of θ and r^S are not negatively collinear, $A^+ \cap S^+$ is non-empty, and one can readily find posterior beliefs that are beneficial and that the sender perceives to be more likely.[22]

[22] To further highlight the importance of the sets A^+ and S^+, suppose that G is linear. Take *any* experiment π that is supported only by beliefs in the areas $A^+ \cap S^+$ and $A^- \cap S^-$. Then, the sender strictly prefers to

We can now extend Proposition 4 by providing both necessary and sufficient conditions for a positive value of persuasion.

Proposition 5 *Suppose that* ***(A1)*** *and* ***(A2′)*** *hold.*
(i) If $A^+ \cap S^+ \neq \varnothing$, then the sender benefits from persuasion.
(ii) If the sender's payoff G is concave, then she benefits from persuasion if and only if $A^+ \cap S^+ \neq \varnothing$.

Proposition 5(i) shows that the sender will experiment as long as there are beneficial beliefs underestimated by the receiver. Proposition 5(ii) then shows that if the sender's utility is a concave function of the receiver's expectation, so that experimentation is never valuable under a common prior, then the only reason for experimentation is that the sender is more optimistic about some beneficial realization. Such realizations generically exist in the space of prior beliefs, even if the receiver is a believer.

Corollary 4 *Suppose that* ***(A1)*** *and* ***(A2′)*** *hold. If $card\,(\Theta) > 2$, then $A^+ \cap S^+ \neq \varnothing$ for a generic pair of prior beliefs.*

We end this section by studying when the optimal experiment would fully reveal the state. That is, when would a sender not gain from garbling the realizations of a fully informative experiment? To answer this question, we apply Corollary 2 to the function V_S in (11) when **(A1)** and **(A2′)** hold, so that

$$V_S(q^S) = G\left(E_R[\theta]\right) = G\left(\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}\right). \tag{22}$$

Expression (22) suggests that the sender's gain from a fully informative experiment depends both on her "risk attitudes" (i.e., on the curvature of G) and the type of receiver she is facing. The next proposition formalizes this intuition. To present this proposition, recall that p^S dominates p^R in the likelihood-ratio sense, $p^S \succeq_{LR} p^R$, if $r^S_\theta = p^S_\theta / p^R_\theta$ (weakly) increases in θ — see Shaked and Shanthikumar (2007, pg 42).

provide experiment π over no experimentation. Conversely, the sender prefers no experimentation over any experiment that is supported only in the areas $A^+ \cap S^-$ and $A^- \cap S^+$.

Proposition 6 *Suppose that **(A1)** and **(A2′)** hold.*
(i) If G is convex and $p^S \succeq_{LR} p^R$, then a fully-revealing experiment is optimal.
(ii) If there exist states θ and θ' such that

$$(\theta' - \theta)\left(\left(r^S_{\theta'}\right)^2 G'(\theta') - \left(r^S_{\theta}\right)^2 G'(\theta)\right) < 0, \tag{23}$$

then a fully revealing experiment is not optimal.

Note that likelihood ratio orders are preserved under Bayesian updating. In particular, if $p^S \succeq_{LR} p^R$, then the receiver remains a skeptic after any realization that does not fully reveal the state, meaning that by fully revealing the state, the sender can increase the receiver's average action. As any garbling reduces the variance of the receiver's posterior beliefs, if u_S is convex and the receiver remains a skeptic after every partially informative realization, then the sender cannot do better than letting him fully learn the state. Nevertheless, Proposition 6(ii) argues that if at least one of these conditions is relaxed, then the sender would prefer to garble a fully informative experiment as long as (23) holds. In particular, if G is linear, then a fully-revealing experiment is optimal if and only if $p^S \succeq_{LR} p^R$. Therefore, a fully informative experiment is often suboptimal, even when the sender faces a skeptic.

4.4 Persuading Skeptics and Believers

When experimentation is valuable, what is the optimal experiment? To provide some intuition, we now restrict attention to the case where G in condition **(A2′)** is concave, so that according to Proposition 5(ii), experimentation is valuable if and only if $A^+ \cap S^+ \neq \varnothing$.

An important property of optimal experiments is time-consistent disclosure: there is no value in further releasing any information after each realization of an optimal experiment. In our case, this implies that $A^+ \cap S^+ = \varnothing$ after each realization of an optimal experiment — ex-post, the sender is never more optimistic about any beneficial belief. This leads to the following property of optimal experiments.

Proposition 7 *Suppose that **(A1)** and **(A2′)** hold, and consider a concave G. Let Z^* be the set of realizations of an optimal experiment, and define $\lambda^S_z = \Pr_S[z] / \Pr_R[z]$ and $a_z = E_R[\theta|z]$. Then,*

$$\lambda^S_{z'} \geq \lambda^S_z \iff a_{z'} \geq a_z.$$

The proposition states that if one considers the distribution of actions induced by an optimal experiment, the sender always assigns more probability to higher actions by the receiver than the receiver does. Actually, the sender's belief (as given by $\Pr_S[a_z]$) dominates the receiver's belief (as given by $\Pr_R[a_z]$) in the likelihood ratio sense. In a nutshell, regardless of whether she is facing a skeptic or a believer, the sender always selects an experiment about whose beneficial realizations she is always more optimistic. In the online Appendix B we show how the sender can construct optimal experiments for particular cases, most notably for the case when G is linear.

4.5 Applications

Attempts to persuade others are pervasive in economics and politics. Politicians and managers try to persuade bureaucrats and workers to exert more effort. Bureaucrats and workers try to influence the policy and managerial choices of politicians and executives. Interest groups and firms try to influence governments' and consumers' expenditure decisions.

An example of persuasion that has gained increasing attention from governments around the world is the use of small-scale policy experiments. The information uncovered by these experiments can influence the actions of legislators, bureaucrats and voters. For example, "the Perry Preschool Project, the Manhattan Bail Bond Experiment, the Work-Welfare Experiments, and the National Job Training Partnership Act (JTPA) Study have all had clear, direct impacts on the adoption or continuation of specific policies or (in the case of JTPA) major funding changes for an ongoing program" (Orr, 1999, pg. 234). It is important to note that the experiments' results do not always meet the designer expectations. According to David Halpern (chief executive of BIT), "one or two in every 10 trials [conducted by the BIT] fail" (Rutter, 2015). Therefore, the sender might benefit from strategically designing the experiment to better influence the receiver.[23]

[23] A recent example illustrates how the designer might strategically garble the experiment. Some local police departments in the US conducted experiments to evaluate how body-worn video technology impacts police-citizen behavior and crime. The test designers wanted legislators to approve a set of proposed rules for the use of this new technology. The experiment designers chose not to test one important aspect of the new policy: all police officers in the trial were allowed to watch the recorded video before writing their reports. Many critics argued that watching the video would greatly influence the reports; therefore, the

One of the main contributions of our paper is to show how the presence of belief disagreement will fundamentally alter how much information is released. In this section, we apply our results to show that persuasion should be widespread in all these cases. Throughout this section, we implicitly assume that there are at least three states.

Application 1 (Motivating Effort): Consider an incumbent politician (or manager) who wants to persuade a bureaucrat (or worker) to exert more effort. Although politicians usually hold the power to define policies, bureaucrats' actions affect the actual implementation and enforcement of policies — see Bertelli (2012) for an overview of the related literature. Moreover, empirical evidence suggests that there is often open disagreement between politicians and bureaucrats — see references in Hirsch (forthcoming). [24]

For concreteness, suppose that a politician wishes to implement a new policy — e.g., she wants to change the flat-wage payment of public school teachers to a pay-for-performance scheme. In order for the policy to be successful, a bureaucrat (e.g., the school district superintendent) must exert effort to implement it. State $\theta > 0$ captures the uncertainty regarding how this new policy will affect voters' and the bureaucrat's payoff. Let $u_R(a, \theta) = \theta a - \frac{a^\rho}{\rho}$ be the payoff of the bureaucrat, where $\rho \geq 2$ is a known preference parameter. Let $u_S(a, \theta) = f(\theta)a$ be the payoff of voters (hence, the payoff of the politician who seeks reelection), where the function $f > 0$ captures voters' preferences. Before implementing the new policy, the politician can run a policy experiment that will provide information to influence the bureaucrat's effort — e.g., design a pilot test in selected schools. Assumptions **(A1)** to **(A3)** hold; therefore, persuasion is generically valuable, independent of the shape of the politician's preference f and the alignment of interests between players.

Application 2 (Influencing Policies): In the previous application, the politician (or manager) had the authority to design and implement the experiment. However, in some

officers should be required to write the report first. While most people agree that watching the video before writing the report has some influence on the report, we do not know (and might have different priors over) how big this influence is. To measure the actual impact of this aspect of the policy, the experimenter could have easily (at no additional monetary cost) randomly assigned some of the officers already participating in the trial to write the report before watching the video. But the designer strategically chose not to do that.

[24]For related models of a manager motivating the effort of a worker under heterogeneous prior beliefs, see Van den Steen (2004, 2009, 2010a, 2011).

situations, the bureaucrat (or worker) is the one who controls the generation of information that the politician uses in choosing policies (or that the manager uses to make decisions).

Suppose that the school superintendent (sender) is an independent elected official who has the authority to run pilot policy tests in the school district. The information uncovered influences the policy a chosen by the incumbent politician (receiver). The politician maximizes the payoff of voters, $u_R(a, \theta) = -(a - \theta)^2$, $0 \leq \theta \leq 1$, so that $a^* = E_R[\theta] \in [0, 1]$. For example, the politician selects the compensation of schoolteachers, where $a = 0$ represents a flat wage and $a = 1$ represents a very steep pay-for-performance scheme. State θ then represents the optimal policy from the politician's point of view. The superintendent's payoff is $u_S(a, \theta) = -(a - f(\theta))^2$, where function f captures the possible misalignment in preferences. Assumptions **(A1)** to **(A3)** also hold in this case; therefore, persuasion is generically valuable, independent of the shape of bureaucrat's preference f and the alignment of interests between the players.[25] In summary, even under extreme conflicts of interest, hard information still flows in the government — communication does not shut down.

This application is closely related to the "Lobbying" example proposed by KG. In the example, the authors consider "a setting where a lobbying group commissions a study with the goal of influencing a benevolent politician. [...] The tobacco lobby has spent large sums funding studies about the health effects of smoking [...]. Would it make sense for the lobbyist to commission such studies even if the politician is rational and knows the lobbyist is providing information with the goal of influencing her decision? Would the optimal study in this case be biased toward supporting the lobbyist's position or fully revealing of the true state?" (KG, pg. 2605)

KG's conclusion, assuming common priors, is that "the lobbyist either commissions a fully revealing study or no study at all. This contrasts with the observation that industry-funded studies often seem to produce results more favorable to the industry than independent studies. The model suggests that commissioning such biased studies when policymakers are rational may not be optimal from the industry's perspective." (KG, pg. 2606)

Our results might help explain this apparent puzzle. If the lobbyist and the politician

[25] Note that Application 2 is equivalent to Example 1 in Section 4.1. If there are only two states, then Example 1 defines the preference misalignment that eliminates the value of persuasion for all prior beliefs. However, if there are three or more states, then persuasion is generically valuable.

have heterogeneous priors, then the lobbyist generically benefits from persuasion, the optimal experiment is often partial information disclosure, and the optimal experiment is such that the sender is more optimistic than the receiver about the expected results.

Application 3 (Seeking Resources): In certain cases, the public signal is better interpreted as the sender's ability to commit to a certain information disclosure rule, such as, the ability of a government agency (or a private firm) to commit to a certain disclosure rule about its activities, services and products. This information, in turn, affects the amount of resources it receives from the government (or the demand from consumers).

For concreteness, consider a government agency or independent institution that produces a public good g (e.g., an environmental agency in charge of protecting the rain forest). The bureaucrat who is the head of the institution (sender) wants to maximize the amount of resources she receives from the government. The incumbent politician (receiver) chooses the proportional income tax rate $a \in [0, 1]$ that is used to finance the institution. The politician is office-motivated and wants to maximize the payoff of a representative voter. The voter cares about her consumption of a private good c and the public good g according to $c^\rho + \theta g$, where $\rho \in (0, 1)$ is a known preference parameter and θ is the unknown marginal benefit of the public good. Let $c = (1 - a)y_m$ and $g = aY$, where y_m is the pre-tax income of the representative (median) voter; Y is the total income of the population; and aY is the total tax revenue used to finance the institution. Hence, the bureaucrat's payoff is $u_S(a, \theta) = aY$. Assuming that $\theta > \frac{\rho y_m^\rho}{Y}$, it follows that the politician's optimal choice is $a(q^R) = 1 - \left(\frac{\rho y_m^\rho}{E_R[\theta]Y}\right)^{\frac{1}{1-\rho}}$. Because the receiver's action depends only on his beliefs through his expectation of θ, without loss of generality, we can normalize his action so that assumption **(A1)** holds.

The bureaucrat can commit to disclose information about the marginal value of the public good (e.g., to a disclosure rule about the information it gathers about the dynamics of the fauna and flora of the different regions). Since the politician's action is a strictly increasing, strictly concave function of her expectation $E_R[\theta]$, under common priors, it is optimal not to disclose any information. However, conditions **(A1)** to **(A3)** apply, and the bureaucrat generically benefits from persuasion. That is, persuasion is valuable even if the incumbent politician strongly believes in the value of protecting the forests and in spite of the fact that the politician's financial decision is a strictly concave function of her expectation.

We can rewrite the model as a firm committing to disclose certain information about the quality of its products and services to a consumer. Persuasion is then generically valuable, even when the consumer is overly optimistic about the quality of the firm's products.

Application 4 (Extreme Conflict): Consider a situation of direct conflict between sender and receiver — e.g., two politicians competing for the same office or two firms competing for market share. To highlight the importance of belief disagreement to persuasion, consider the extreme case $u_S(a,\theta) = -u_R(a,\theta)$. If the receiver chooses a, when would the sender benefit from providing information about θ?

For concreteness, consider an incumbent politician whose political platform is known by voters, against a challenger who needs to choose a campaign platform (or a known incumbent firm against a potential entrant who must choose how to enter the market). The challenger (entrant) wants to choose the action that maximizes his probability of election (or market share): $u_R(a,\theta) = 1 - (a-\theta)^2$, where $0 \leq \theta \leq 1$, so that $a(q^R) = E_R[\theta] \in [0,1]$. From the challenger's point of view, his expected payoff from an optimal action decreases in the variance, $E_R[u_R(a(q^R),\theta)] = -\mathrm{VAR}_R[\theta]$. The incumbent's objective is to minimize the challenger's probability of election, $u_S(a,\theta) = -u_R(a,\theta)$. Remarkably, persuasion is generically valuable even in this extreme case, since assumptions **(A1)** to **(A3)** hold. Note that from the sender's point of view, her expected payoff can be written as $(E_S(\theta) - E_R[\theta])^2 + \mathrm{VAR}_S[\theta]$. That is, the sender benefits from the size of the receiver's "mistake," captured by the term $(E_S(\theta) - E_R[\theta])^2$, and from the degree of uncertainty, captured by $\mathrm{VAR}_S[\theta]$. Any informative experiment decreases $\mathrm{VAR}_S[\theta]$, which hurts the sender. However, the sender can generically design an experiment that sufficiently increases the expected mistake, so that persuasion is valuable. ■

5 Private Priors

We can extend the analysis to a case in which the sender is uncertain about the receiver's prior beliefs when designing π. Suppose that prior beliefs are drawn from a distribution $H(p^R, p^S)$ with conditional distribution $h(p^R|p^S)$.[26] Proposition 1 still applies for each (p^R, p^S). Con-

[26] Note that the receiver's preferences are unaffected by his beliefs about the sender's prior. Therefore, the sender's choice of experiment conveys no additional information to the receiver. This would not be true if

sequently, given p^S and $h(p^R|p^S)$, knowledge of the sender's posterior q^S suffices to compute the joint distribution of posterior beliefs. Moreover, the restriction to language-invariant equilibria implies that, given (p^R, p^S), the receiver's choice depends only on his posterior belief q^R. Therefore, we can compute the sender's expected payoff V_S using the implied distribution of q^R. More specifically, (11) translates to

$$V_S\left(q^S\right) = E_S[v(q^S, q^R)|p^S] = \int v\left(q^S, \frac{q^S \frac{p^R}{p^S}}{\left\langle q^S, \frac{p^R}{p^S}\right\rangle}\right) dh(p^R|p^S). \tag{24}$$

With this modification, the expected utility of a sender under an optimal experiment is $\widetilde{V}_S\left(p^S\right)$, and the sender would benefit from persuasion under the conditions of Corollary 1. Moreover, the expected value to the sender of a perfectly informative experiment is independent of the receiver's prior belief. Therefore, the value of garbling is positive whenever (24) satisfies the conditions in Corollary 2.

As an application of (24), consider the pure persuasion model from Section 4.3. When the sender knows the receiver's prior, Proposition 5(i) provides conditions on the likelihood ratio of priors for persuasion to be valuable. Suppose that these conditions are met, and the sender strictly benefits from providing experiment π to a particular receiver. By a continuity argument, the same π strictly benefits the sender when she faces another receiver whose prior belief is not too different. Consequently, even if the sender does not know the receiver's prior, persuasion remains beneficial when the receiver's possible priors are not too dispersed. Proposition B.1 in Online Appendix B shows that this is, indeed, the case and provides an upper bound on how dispersed these beliefs can be.

6 Conclusion

In this paper, we study the gain to an individual (sender) from controlling the information available to a decision maker (receiver) when they openly disagree about their views of the world. We first characterize the set of distributions over posterior beliefs that can be induced through an experiment, under our assumption of a "commonly understood experiment" (i.e., when players agree on the statistical relation of the experiment to the payoff-relevant state).

the sender privately observed a signal about the state, see Sethi and Yildiz (2012).

This allows us to compute the gains from persuasion.

In Section 4, we provide necessary and sufficient conditions for *some* belief disagreement to render experimentation valuable to the sender. We then define a large class of models in which the sender gains from experimentation for *almost every* pair of prior beliefs, even when there is no value of persuasion under a common prior. Our main conditions are: (i) the receiver's action depends on his beliefs only through his expectation of some random variable; and (ii) there are more than two states. The fact that these conditions hold in many important applications emphasizes our main finding that persuasion should be widespread in situations of open disagreement.

For a case in which experimentation is not valuable under a common prior, we show that optimal experiments under heterogeneous priors have an intuitive property: the sender is relatively more optimistic than the receiver in inducing beneficial outcomes. Indeed, we show that the sender's relative optimism is quite strong — her prior belief over realizations of an optimal experiment dominates the receiver's prior in the likelihood-ratio sense. This allows us to clarify why even a sender facing a "believer" can design an experiment about whose outcomes she is more optimistic.

One important example of persuasion that has gained increasing attention from governments around the world is the use of small-scale policy experiments. Many policy experiments have had real impacts on policies later adopted (see examples in Section 4.5). There are many econometric books explaining how to conduct the most informative experiment. However, many of these experiments are paid for and controlled by a politician or a bureaucrat. Given the preferences and beliefs of the parts involved, the experiment might be strategically designed (garbled) to influence others. We hope that our results might guide future empirical investigations that aim to identify which experiments conducted around the world were, indeed, strategically modified.

To focus on the role of heterogeneous priors on strategic experimentation, we restrict our analysis in several ways. First, the sender has no private information. Second, we consider a single receiver. In many situations, however, the sender may want to affect the beliefs of a collective, where she is typically constrained to use a public signal. Third, we consider a fixed decision-making process. However, sometimes the sender can both offer a contract and provide some information to a receiver — i.e., the sender designs a grand mechanism specifying the information to be released and several contractible variables. Similarly, one can examine how the optimal experiment varies across different mechanisms of preference aggregation (e.g., Alonso and Câmara (2015, forthcoming) examine persuasion in a voting model). We leave all of these promising extensions for future work.

A Proofs

Proof of Proposition 1: *Necessity*: Consider an experiment $\pi = \left(Z, \{\pi(\cdot|\theta)\}_{\theta\in\Theta}\right)$ that induces, from the sender's perspective, the distribution τ, and let $\pi(z) = (\pi(z|\theta))_{\theta\in\Theta}$ and $q^R(z)$ and $q^S(z)$ be the posterior beliefs of the receiver and the sender if $z \in Z$ is realized. The marginal distribution over the sender's posterior beliefs satisfies the martingale property — i.e., $E_\tau[q^S] = p^S$. Furthermore, as priors are totally mixed, the receiver assigns positive probability to z if and only if the sender also assigns positive probability to z.[27] Suppose, then, that $\pi(z) \neq \mathbf{0}$. Bayesian updating implies that after observing z,

$$q^S_\theta(z) = \frac{\pi(z|\theta)p^S_\theta}{\langle \pi(z), p^S\rangle},$$

so we can write

$$q^S_\theta(z)\left\langle \pi(z), p^S\right\rangle \frac{p^R_\theta}{p^S_\theta} = \pi(z|\theta)p^R_\theta,$$

and summing over $\theta \in \Theta$, we obtain

$$\left\langle \pi(z), p^S\right\rangle \left\langle q^S(z), r^R\right\rangle = \left\langle \pi(z), p^R\right\rangle.$$

Then, we can relate the two posterior beliefs by

$$q^R_\theta(z) = \frac{\pi(z|\theta)p^R_\theta}{\langle \pi(z), p^R\rangle} = \frac{\pi(z|\theta)p^S_\theta}{\langle \pi(z), p^S\rangle \langle q^S(z), r^R\rangle}\frac{p^R_\theta}{p^S_\theta} = q^S_\theta(z)\frac{r^R_\theta}{\langle q^S(z), r^R\rangle}.$$

Sufficiency: Given a distribution τ satisfying (i) and (ii), let $\tau_S(q^S)$ be the marginal distribution of the sender's posterior beliefs and define the realization space $Z = \left\{q^S : q^S \in Supp(\tau_S)\right\}$ and the likelihood functions $\pi(q^S|\theta) = \frac{q^S_\theta \Pr_{\tau_S} q^S}{p^S_\theta}$. Then, simple calculations reveal that the experiment $\pi = \left(Z, \left\{\pi(q^S|\theta)\right\}_{\theta\in\Theta}\right)$ induces τ. ■

Proof of Proposition 2: Part (i) See KG. Part (ii) As (10) can be seen as a persuasion model with a common prior, the claim then follows from KG (Corollary 2: pg. 2597).

[27] Indeed, we have $\Pr_R[z] = \left\langle \pi(z), p^R\right\rangle = 0 \Leftrightarrow \pi(z|\theta) = 0, \theta \in \Theta \Leftrightarrow \Pr_S[z] = \left\langle \pi(z), p^S\right\rangle = 0$.

Proof of Corollary 1: Condition (13) can be rephrased in terms of the subdifferential $\partial V(p)$ of a function V evaluated at p, and simply states that the sender does not benefit from persuasion if and only if $\partial\left(-V_S(p^S)\right) \neq \varnothing$. Condition (14) then follows immediately as, if V_S is differentiable at p^S, then $\partial\left(-V_S(p^S)\right)$ can have at most one element.

Sufficiency: As the concave closure $\widetilde{V}_S$ is the lower envelope of all affine functions that majorize V_S and, by assumption, the majorizing affine function $f\left(q^S\right) = V_S\left(p^S\right) + \left\langle \gamma, q^S - p^S \right\rangle$ satisfies $V_S\left(p^S\right) = f\left(p^S\right)$, then

$$V_S\left(p^S\right) = f\left(p^S\right) \geq \widetilde{V}_S\left(p^S\right) \geq V_S\left(p^S\right),$$

implying that $\widetilde{V}_S\left(p^S\right) = V_S\left(p^S\right)$ and, by Proposition 2, there is no value of persuasion.

Necessity: Suppose that there is no value of persuasion. From Proposition 2 this implies that $\widetilde{V}_S\left(p^S\right) = V_S\left(p^S\right)$. As $\widetilde{V}_S$ is the concave closure of an upper-semicontinuous function in a compact set, the differential of $-\widetilde{V}_S\left(q^S\right)$ is non-empty for all $q^S \in int(\Delta(\Theta))$. Any element of $\partial\left(-\widetilde{V}_S(p^S)\right)$ would then satisfy (13). ■

Proof of Corollary 2: *Sufficiency*: Suppose that (15) is satisfied. Then, any π that induces the distribution over posterior beliefs σ must satisfy $\mathrm{E}_\sigma\left[q^S\right] = p^S$, implying that

$$\sum_{\theta\in\Theta} p^S_\theta u_S(a(\mathbf{1}_\theta),\theta) = \mathrm{E}_\sigma\left[\sum_{\theta\in\Theta} q^S_\theta u_S(a(\mathbf{1}_\theta),\theta)\right] \geq \mathrm{E}_\sigma\left[V_S\left(q^S\right)\right].$$

Thus, a fully informative experiment weakly dominates any π and is, thus, optimal.

Necessity: Fix any belief $q^S \in \Delta(\Theta)$ and let $\bar{\delta}$ be defined as

$$\bar{\delta} = \max\left\{\delta : p^S_\theta - \frac{\delta}{1-\delta}(q^S_\theta - p^S_\theta) \geq 0, \delta \in [0,1]\right\}.$$

As the prior belief $p^S \in int(\Delta(\Theta))$ we have $1 > \bar{\delta} > 0$. Letting $\mathbf{1}_\theta$ be the belief that assigns probability 1 to state θ, consider, now, an experiment that induces belief q^S with probability $\bar{\delta}$ and belief $\mathbf{1}_\theta$ with probability $(1-\bar{\delta})\left(p^S_\theta - \frac{\bar{\delta}}{1-\bar{\delta}}(q^S_\theta - p^S_\theta)\right) = p^S_\theta - \bar{\delta} q^S_\theta \geq 0$ for each $\theta \in \Theta$. The expected utility of the sender under this experiment is

$$\delta V_S\left(q^S\right) + \sum_{\theta\in\Theta}\left(p^S_\theta - \bar{\delta} q^S_\theta\right) u_S(a(\mathbf{1}_\theta),\theta) = \bar{\delta}\left(V_S\left(q^S\right) - \sum_{\theta\in\Theta} q^S_\theta u_S(a(\mathbf{1}_\theta),\theta)\right) + \sum_{\theta\in\Theta} p^S_\theta u_S(a(\mathbf{1}_\theta),\theta).$$

Full disclosure is optimal by assumption; therefore, we must have

$$\bar{\delta}\left(V_S\left(q^S\right) - \sum_{\theta\in\Theta} q^S_\theta u_S(a(\mathbf{1}_\theta),\theta)\right) + \sum_{\theta\in\Theta} p^S_\theta u_S(a(\mathbf{1}_\theta),\theta) \leq \sum_{\theta\in\Theta} p^S_\theta u_S(a(\mathbf{1}_\theta),\theta),$$

from which, given that $\bar{\delta} > 0$, we must then necessarily have (15). ■

Proof of Proposition 3: *Necessity*: We prove the contrapositive: if for some θ', $q^R_{\theta'} u_S\left(a(q^R), \theta'\right)$ is not concave, then there exists a pair of mixed prior beliefs p^R and p^S such that the sender benefits from persuasion. Let $n = card\,(\Theta)$, and suppose that for θ', the function $q^R_{\theta'} u_S\left(a(q^R), \theta'\right)$ is not concave. Then, there exist $q^+, q^- \in int\,(\Delta\,(\Theta))$, and ν, $0 < \nu < 1$, such that

$$\nu q^+_{\theta'} u_S\left(a(q^+), \theta'\right) + (1-\nu) q^-_{\theta'} u_S\left(a(q^-), \theta'\right) - p^R_{\theta'} u_S\left(a(p^R), \theta'\right) = \Psi > 0,$$

where $p^R \in int\,(\Delta\,(\Theta))$ is given by $p^R = \nu q^+ + (1-\nu) q^-$. Since $u_S\left(a(q^R), \theta\right)$ is bounded, let

$$\bar{\Psi} = \min_{\theta \in \Theta} \left(\frac{\nu q^+_\theta u_S\left(a(q^+), \theta\right) + (1-\nu) q^-_\theta u_S\left(a(q^-), \theta'\right) - p_\theta u_S\left(a(p), \theta'\right)}{p^R_\theta} \right).$$

Define the belief p^S such that $p^S_\theta = \psi$ if $\theta \neq \theta'$ and $p^S_{\theta'} = 1 - (n-1)\psi$, where ψ is defined by

$$\psi = \min \left(\frac{1}{n\,(n-1)\left(\Psi + p^R_{\theta'}\left|\bar{\Psi}\right|\right)}, \frac{1}{n} \right) > 0.$$

Consider an experiment $\hat{\pi}$ with $Z = \{q^+, q^-\}$, which induces posterior beliefs q^+ and q^- in a receiver with prior p^R. The value of experiment $\hat{\pi}$ to a sender with prior p^S, is

$$\begin{aligned}
V_{\hat{\pi}} - v(p^S, p^R) &= \nu V_R\left(q^+\right) + (1-\nu) V_R\left(q^+\right) - V_R\left(p^R\right) = \\
&= \sum\nolimits_{\theta \in \Theta} \frac{p^S_\theta}{p^R_\theta} \left(\nu q^+_\theta u_S\left(a(q^+), \theta\right) + (1-\nu) q^-_\theta u_S\left(a(q^-), \theta'\right) - p_\theta u_S\left(a(p), \theta'\right)\right) \\
&\geq \frac{1 - (n-1)\,\psi}{p^R_{\theta'}} \Psi - (n-1)\,\psi \left|\bar{\Psi}\right| \geq \frac{1 - \frac{1}{n}}{p^R_{\theta'}} > 0.
\end{aligned}$$

Therefore, a sender with prior p^S benefits from persuading a receiver with prior p^R.

Sufficiency: Suppose that $q^R_\theta u_S\left(a(q^R), \theta\right)$ is everywhere concave in q^R for every $\theta \in \Theta$. Then, for any pair of totally mixed priors, $V_R\left(q^R\right) = \sum_{\theta \in \Theta} \frac{p^S_\theta}{p^R_\theta} q^R_\theta u_S\left(a(q^R), \theta\right)$ is concave as a positive linear combination of concave functions. Thus, $\tilde{V}_R\left(q^R\right) = V_R\left(q^R\right)$ for all q^R and Proposition **2** implies that the value of persuasion is zero. ■

The following two lemmas are used in the proof of our next propositions.

Lemma A.1 *Let $x, y \in \mathbb{R}^N$, and W defined by (3). Then,*

$$\frac{1}{2}\left(\left\|x_{\|W}\right\| \left\|y_{\|W}\right\| + \left\langle x_{\|W}, y_{\|W}\right\rangle\right) = \max \langle x, v\rangle \langle y, v\rangle\,, s.t., v \in W, \|v\| = 1. \qquad (25)$$

Proof of Lemma A.1: Let $\rho(x,y)$ be the angle formed by the vectors x and y. If $v \in W$, then $\langle v,x\rangle = \left\langle v, x_{\|W}\right\rangle$ and $\langle v,y\rangle = \left\langle v, y_{\|W}\right\rangle$. Therefore, for every $v \in W$, $\|v\| = 1$, we have

$$\begin{aligned}
\langle x,v\rangle \langle y,v\rangle &= \left\langle v, x_{\|W}\right\rangle \left\langle v, y_{\|W}\right\rangle = \left\|x_{\|W}\right\| \left\|y_{\|W}\right\| \|v\|^2 \cos\rho\left(v, x_{\|W}\right)\cos\rho\left(v, y_{\|W}\right) \\
&= \left\|x_{\|W}\right\| \left\|y_{\|W}\right\| \frac{\cos\left(\rho\left(v, x_{\|W}\right) + \rho\left(v, y_{\|W}\right)\right) + \cos\left(\rho\left(v, x_{\|W}\right) - \rho\left(v, y_{\|W}\right)\right)}{2} \\
&= \left\|x_{\|W}\right\| \left\|y_{\|W}\right\| \frac{\cos\left(2\rho\left(v, x_{\|W}\right) + \rho\left(x_{\|W}, y_{\|W}\right)\right) + \cos\left(\rho\left(x_{\|W}, y_{\|W}\right)\right)}{2},
\end{aligned}$$

which implies that

$$\begin{aligned}
&\max_{v\in W,\|v\|=1} \langle x,v\rangle \langle y,v\rangle \\
&= \left\|x_{\|W}\right\| \left\|y_{\|W}\right\| \left[\frac{\cos\left(\rho\left(x_{\|W}, y_{\|W}\right)\right)}{2} + \max_{v\in W,\|v\|=1} \frac{\cos\left(2\rho\left(v, x_{\|W}\right) + \rho\left(x_{\|W}, y_{\|W}\right)\right)}{2}\right] \\
&= \left\|x_{\|W}\right\| \left\|y_{\|W}\right\| \left[\frac{\cos\left(\rho\left(x_{\|W}, y_{\|W}\right)\right)}{2} + \frac{1}{2}\right],
\end{aligned}$$

where the maximum is achieved by selecting a vector v such that $\rho\left(v, x_{\|W}\right) = -\frac{1}{2}\rho\left(x_{\|W}, y_{\|W}\right)$. Rewriting this last expression, one obtains (25). ■

Lemma A.2 *Suppose that $N = card(\Theta) \geq 3$, and consider the subspace $W = \left\{w \in \mathbb{R}^N : \langle w, 1\rangle = 0\right\}$ with the derived topology. Then, for $x \notin W$, the rational function $\langle w,x\rangle / \langle w,y\rangle$, $w \in W$, is bounded in a neighborhood of $\mathbf{0}$ if and only if $x_{\|W}$ and $y_{\|W}$ are collinear.*

Proof of Lemma A.2: Consider the linear subspace $W_{x,1} = \left\{w \in \mathbb{R}^N : \langle w,x\rangle = 0, \langle w,1\rangle = 0\right\}$. As, by assumption, $x \notin W$, then $W_{x,1}$ is a linear subspace of dimension $N-2 \geq 1$. Consider, now, the subspace $W_y = \left\{w \in \mathbb{R}^N : \langle w,y\rangle = 0\right\}$. The ratio $\langle w,x\rangle / \langle w,y\rangle$ is locally unbounded in W iff $W_{x,1} \cap W_y^c \neq \varnothing$. First, if the projections $x_{\|W}$ and $y_{\|W}$ are not collinear, then the orthogonal projection $y_{\|W_{x,1}}$ is non-zero, implying that $\left\langle y_{\|W_{x,1}}, x\right\rangle = 0$ but $\left\langle y_{\|W_{v,1}}, y\right\rangle > 0$. This establishes that $W_{x,1} \cap W_y^c \neq \varnothing$. Now suppose that $x_{\|W} = \lambda\, y_{\|W}$ for some $\lambda \neq 0$. Then, $\left\langle w, x_{\|W}\right\rangle = 0$ iff $\left\langle w, y_{\|W}\right\rangle = 0$, implying $W_{x,1} \cap W_y^c = \varnothing$. ■

Proof of Proposition 4: Define the vectors $u_S(a) = (u_S(a,\theta))_{\theta\in\Theta}$ and $\partial u_S(a) = \left(\frac{\partial u_S(a,\theta)}{\partial a}\right)_{\theta\in\Theta}$, so that at the prior belief, we have $u_S' = \partial u_S\left(\left\langle p^R, \theta\right\rangle\right)$ The representation (17) can be concisely written as $V_R\left(q^R\right) = \left\langle q^R, r^S u_S\left(\left\langle q^R, \theta\right\rangle\right)\right\rangle$, and has gradient at p^R

$$\nabla V_R(p^R) = \left\langle p^R, r^S u_S'\right\rangle \theta + r^S u_S\left(\left\langle p^R, \theta\right\rangle\right).$$

Corollary 1 implies that the value of persuasion is zero if and only if

$$\left\langle \nabla V_R(p^R), q^R - p^R \right\rangle \geq V^R(q^R) - V^R(p^R),\ q^R \in \Delta(\Theta),$$

which, in our case, leads to

$$\left\langle p^R, r^S u_S' \right\rangle \left\langle \theta, q^R - p^R \right\rangle - \left\langle q^R, r^S \left(u_S \left(\left\langle q^R, \theta \right\rangle \right) - u_S \left(\left\langle p^R, \theta \right\rangle \right) \right) \right\rangle \geq 0,\ q^R \in \Delta(\Theta). \tag{26}$$

To ease notation, let $\varepsilon = q^R - p^R \in W$ and define $\triangle$ as the left-hand side of (26)

$$\triangle = \left\langle p^R, r^S u_S' \right\rangle \langle \theta, \varepsilon \rangle - \left\langle q^R, r^S \left(u_S \left(\left\langle q^R, \theta \right\rangle \right) - u_S \left(\left\langle p^R, \theta \right\rangle \right) \right) \right\rangle. \tag{27}$$

We now show that if $r^S u'_{S\|W} \neq 0$ and if θ and $r^S u_S'$ are not negatively collinear with respect to W, we can find a feasible q^R such that $\triangle < 0$. First, with the help of the identities

$$r^S \left(u_S \left(\left\langle q^R, \theta \right\rangle \right) - u_S \left(\left\langle p^R, \theta \right\rangle \right) \right) = \left(\int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} r^S_\theta \frac{\partial u_S(t,\theta)}{\partial a} dt \right)_{\theta \in \Theta}$$

and

$$\begin{aligned}
& \left\langle p^R, r^S \left(u_S \left(\left\langle q^R, \theta \right\rangle \right) - u_S \left(\left\langle p^R, \theta \right\rangle \right) \right) \right\rangle - \langle \theta, \varepsilon \rangle \left\langle p^R, r^S u_S' \right\rangle \\
= & \int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \left\langle p^R, r^S \partial u_S(t) \right\rangle dt - \int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \left\langle p^R, r^S u_S' \right\rangle dt \\
= & \int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \left\langle p^R, r^S \left(\frac{\partial u_S(t,\theta)}{\partial a} - \frac{\partial u_S(\left\langle p^R, \theta \right\rangle, \theta)}{\partial a} \right) \right\rangle dt \\
= & \int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \int_{\langle p^R, \theta \rangle}^{t} \left\langle p^R, r^S \frac{\partial^2 u_S(\tau,\theta)}{\partial^2 a} \right\rangle d\tau dt,
\end{aligned}$$

we can rewrite $\triangle$ in (27) as

$$\triangle = - \int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \int_{\langle p^R, \theta \rangle}^{t} \left\langle p^R, r^S \frac{\partial^2 u_S(\tau,\theta)}{\partial^2 a} \right\rangle d\tau dt - \int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \left\langle \varepsilon, r^S \partial u_S(t) \right\rangle dt. \tag{28}$$

The smoothness condition **(A2)** implies that $\frac{\partial u_S(a,\theta)}{\partial a}$ and $\frac{\partial^2 u_S(a,\theta)}{\partial^2 a}$ are bounded in the compact set $A = \left\{ a : a = \left\langle q^R, z \right\rangle, q^R \in \Delta(\Theta) \right\}$. Let $M_S = \max_{a \in A, \theta \in \Theta} \left| \frac{\partial^2 u_S(a,\theta)}{\partial^2 a} \right|$, which, for some $\phi \in \left[\left\langle p^R, \theta \right\rangle, \left\langle q^R, \theta \right\rangle \right]$, allow us to write the following second-order expansion

$$\begin{aligned}
\int_{\langle p^R, \theta \rangle}^{\langle q^R, \theta \rangle} \left\langle \varepsilon, r^S \partial u_S(t) \right\rangle dt &= \left\langle \varepsilon, r^S u_S' \right\rangle \langle \varepsilon, \theta \rangle + \frac{1}{2} \left\langle \varepsilon, r^S \frac{\partial^2 u_S(\phi,\theta)}{\partial^2 a} \right\rangle (\langle \varepsilon, \theta \rangle)^2 \\
&\geq \left\langle \varepsilon, r^S u_S' \right\rangle \langle \varepsilon, \theta \rangle - \frac{1}{2} M_S \left\langle |\varepsilon|, r^S \right\rangle (\langle \varepsilon, \theta \rangle)^2.
\end{aligned}$$

Then,

$$\begin{aligned}
\triangle &\leq M_S \int_{\langle p^R,\theta\rangle}^{\langle q^R,\theta\rangle} \int_{\langle p^R,\theta\rangle}^{t} d\tau dt - \left\langle \varepsilon, r^S u'_S \right\rangle \langle \varepsilon, \theta\rangle + \frac{1}{2} M_S \left\langle |\varepsilon|, r^S \right\rangle (\langle \varepsilon, \theta\rangle)^2 \\
&= \frac{1}{2} \langle \varepsilon, \theta\rangle^2 \left(1 + \left\langle |\varepsilon|, r^S \right\rangle\right) M_S - \left\langle \varepsilon, r^S u'_S \right\rangle \langle \varepsilon, \theta\rangle \\
&= \langle \varepsilon, \theta\rangle^2 \left(\frac{1 + \left\langle |\varepsilon|, r^S \right\rangle}{2} M_S - \frac{\left\langle \varepsilon, r^S u'_S \right\rangle}{\langle \varepsilon, \theta\rangle} \right).
\end{aligned}$$

From Lemma A.1, if $r^S u'_{S||W} \neq \mathbf{0}$, and θ and $r^S u'_S$ are not negatively collinear wrt W, then there exists a neighborhood $N(\mathbf{0})$ of $\mathbf{0}$ in W such that $\left\langle \varepsilon, r^S u'_S \right\rangle / \langle \varepsilon, \theta\rangle$ admits no upper bound. This establishes the existence of $\varepsilon \in N(\mathbf{0})$, and, thus, a feasible $q^R = p^R + \varepsilon$, such that

$$\frac{1 + \left\langle |\varepsilon|, r^S \right\rangle}{2} M_S - \frac{\left\langle \varepsilon, r^S u'_S \right\rangle}{\langle \varepsilon, \theta\rangle} < 0,$$

implying that $\triangle < 0$. ■

Proof of Corollary 3: Fix a mixed prior p^R, and define the sets

$$\begin{aligned}
O &= \left\{ p \in int\left(\Delta\left(\Theta\right)\right) : p_\theta \frac{u'_{S,\theta}}{p^R_\theta} = k, k \in \mathbb{R}, \theta \in \Theta \right\}, \text{ and} \\
P &= \left\{ p \in int\left(\Delta\left(\Theta\right)\right) : \left(p_\theta \frac{u'_{S,\theta}}{p^R_\theta} - p_{\theta'} \frac{u'_{S,\theta'}}{p^R_{\theta'}} \right) = -\lambda_1 \left(\theta - \theta'\right), \lambda_1 > 0, \theta, \theta' \in \Theta \right\}.
\end{aligned}$$

The sets O and P capture the conditions in Proposition **4** since (i) $\left(r^S \cdot u'_S\right)_{||W} = 0$ iff $p^S \in O$, and (ii) $r^S \cdot u'_S$ and θ are negatively collinear with respect to W iff $p^S \in P$. We first show that each set is contained in a one-dimensional subspace of W

We start by studying the set O. If $u'_{S,\theta} = 0$ for all θ, then $O = \Delta\left(\Theta\right)$. However, this condition would violate assumption (A3). If $u'_{S,\theta} \neq 0$ and $u'_{S,\theta'} = 0$ for some $\theta' \neq \theta$, then the set $O = \varnothing$ as O does not contain a mixed prior. Finally, if $u'_{S,\theta} \neq 0$ for all θ, then O is contained in the one-dimensional subspace $\left\{ p \in \mathbb{R}^{card(\Theta)} : p_\theta = k \frac{p^R_\theta}{u'_{S,\theta}}, k \in \mathbb{R} \right\}$.

Now consider the set P. If $u'_{S,\theta} = u'_{S,\theta'} = 0$ for two distinct states $\theta \neq \theta'$, then $P = \varnothing$. Suppose, now, that $u'_{S,\theta} \neq 0$ for all θ. Then, P is contained in the one-dimensional subspace

$$\left\{ p \in \mathbb{R}^{card(\Theta)} : p_\theta = \left(\lambda_0 \frac{1}{u'_{S,\theta}} - \lambda_1 \frac{\theta}{u'_{S,\theta}} \right) p^R_\theta, \sum \left(\lambda_0 \frac{1}{u'_{S,\theta}} - \lambda_1 \frac{\theta}{u'_{S,\theta}} \right) p^R_\theta = 1, \lambda_0, \lambda_1 \in \mathbb{R} \right\}.$$

Overall, for every sender's prior, the set in which the conditions in Proposition 4 are violated, given by the union of O and P, is contained in the union of two one-dimensional subspaces. If $card\,(\Theta) > 2$, then $\dim(\Delta\,(\Theta)) > 1$, and this set is a non-generic set of $\Delta\,(\Theta)$. Since this is true for every mixed prior $p^R \in int\,(\Delta\,(\Theta))$, the conditions in Proposition 4 are violated in a non-generic set of pairs of mixed prior beliefs. ■

Proof of Lemma 1: Let $\varepsilon = q^R - p^R \in W$ with $q^R \in \Delta\,(\Theta)$. Posterior belief $q^R \in A^+$ if and only if $\langle \varepsilon, \theta \rangle \geq 0$, while (7) implies $q^R \in S^+$ if and only if $\left\langle \varepsilon, r^S \right\rangle > 0$. We now show that $A^+ \cap S^+ = \varnothing$ iff $p^R = p^S$ or r^S and θ are negatively collinear with respect to W.

First, if $p^R = p^S$, then $r^S_\theta = 1$ and $\left\langle \varepsilon, r^S \right\rangle = \left\langle q^R - p^R, 1 \right\rangle = 0$, so $S^+ = \varnothing$. Second, suppose that $p^R \neq p^S$. Then, since $-\varepsilon \in W$ if $\varepsilon \in W$,then $A^+ \cap S^+ = \varnothing$ iff

$$\langle \varepsilon, \theta \rangle \left\langle \varepsilon, r^S \right\rangle \leq 0 \ , \varepsilon = q^R - p^R, q^R \in \Delta\,(\Theta)\,.$$

Since the set $\left\{\varepsilon : \varepsilon = q^R - p^R, q^R \in \Delta\,(\Theta)\right\} \subset W$ contains a neighborhood of 0 in W, then the previous condition is satisfied if and only if the following global condition is true:

$$\langle \varepsilon, \theta \rangle \left\langle \varepsilon, r^S \right\rangle \leq 0 \text{ for } \varepsilon \in W,$$

or, in other words, iff the quadratic form $\langle \varepsilon, \theta \rangle \left\langle \varepsilon, r^S \right\rangle$ is negative semidefinite in W.

Consider the orthogonal decompositions $\theta = \theta_{\| W} + \alpha_\theta 1$ and $r^S = r^S_{\| W} + \alpha_r 1$. Whenever $\varepsilon \in W$, we have $\langle \varepsilon, \theta \rangle = \left\langle \varepsilon, \theta_{\| W} \right\rangle$ and $\left\langle \varepsilon, r^S \right\rangle = \left\langle \varepsilon, r^S_{\| W} \right\rangle$, implying that negative semidefiniteness of $\langle \varepsilon, \theta \rangle \left\langle \varepsilon, r^S \right\rangle$ in W is equivalent to negative semidefiniteness of $\left\langle \varepsilon, \theta_{\| W} \right\rangle \left\langle \varepsilon, r^S_{\| W} \right\rangle$ in W. From Lemma A.1, we have

$$0 = \max_{\varepsilon \in W, \|\varepsilon\| = 1} \left\langle \varepsilon, \theta_{\| W} \right\rangle \left\langle \varepsilon, r^S_{\| W} \right\rangle \Leftrightarrow \left\langle \theta_{\| W}, r^S_{\| W} \right\rangle = -||\theta_{\| W}||||r^S_{\| W}||,$$

Since $\theta_{\| W} \neq 0$ and $r^S_{\| W} \neq 0$, then $\left\langle \theta_{\| W}, r^S_{\| W} \right\rangle = -||\theta_{\| W}||||r^S_{\| W}||$ iff $\cos\left(\theta_{\| W}, r^S_{\| W}\right) = -1$, which is equivalent to the existence of $\alpha > 0$ such that $\theta_{\| W} = -\alpha r^S_{\| W}$.■

Proof of Proposition 5: The representation (16) in our setup gives $V_R(q^R) = G(\left\langle q^R, \theta \right\rangle) \left\langle q^R, r^S \right\rangle$. Let $\triangle$ be defined in (27), which translates in our case to

$$\triangle = G'(\left\langle p^R, \theta \right\rangle) \langle \theta, \varepsilon \rangle - \left\langle q^R, r^S \right\rangle \left(G(\left\langle q^R, \theta \right\rangle) - G(\left\langle p^R, \theta \right\rangle) \right). \tag{29}$$

The proof of Proposition **4** shows that the value of persuasion is zero if and only if $\triangle \geq 0$. *Part (i)*- Follows from applying Proposition **4** to **(A1)** and **(A2')**.

Part (ii)- We show that if G is concave, then the condition on θ and r^S is also necessary for the sender to benefit from persuasion. We prove the contrapositive: if θ and r^S are negatively collinear wrt W, then the value of persuasion is zero.

Concavity of G yields the following bound

$$G(\langle q^R, \theta \rangle) - G(\langle p^R, \theta \rangle) \leq G'(\langle p^R, \theta \rangle) \langle \varepsilon, \theta \rangle ,$$

which, applied to (29) and noting that $1 - \langle q^R, r^S \rangle = \langle \varepsilon, r^S \rangle$, implies that

$$\triangle \geq -G'(\langle p^R, \theta \rangle) \langle \varepsilon, \theta \rangle \langle \varepsilon, r^S \rangle . \tag{30}$$

As θ and r^S are negatively collinear wrt W, Lemma 1 implies that

$$\langle \varepsilon, \theta \rangle \langle \varepsilon, r^S \rangle \leq 0 \ for \ \varepsilon \in W,$$

which applied to (30) leads to

$$\triangle \geq -u_S'(\langle p^R, \theta \rangle) \langle \varepsilon, \theta \rangle \langle \varepsilon, r^S \rangle \geq 0 \ for \ \varepsilon \in W, .$$

As $\triangle \geq 0$ for all beliefs, Corollary 1 establishes that the value of persuasion is zero. ■

Proof of Corollary 4: Assumption **(A2')** implies that $\frac{\partial u_S(a,\theta)}{\partial a} = G'(a) > 0$, so that Assumption (**A3**) is satisfied. The claim then follows from applying Corollary 3 to this particular case. ■

Proof of Proposition 6: *Part (i)* - First, likelihood ratio orders are preserved by Bayesian updating with commonly understood experiments (Whitt, 1979; Milgrom, 1981). Thus, induced posteriors satisfy $q^S(z) \succeq_{LR} q^R(z)$ if $p^S \succeq_{LR} p^R$ for any π and realization z, so we must then have $\langle q^S(z), \theta \rangle \geq \langle q^R(z), \theta \rangle$. Therefore,

$$q_\theta^S G(\langle \mathbf{1}_\theta, \theta \rangle) \geq G(\langle q^S, \theta \rangle) \geq G(\langle q^R, \theta \rangle) = V_S\left(q^S\right), \ q^S \in \Delta\left(\Theta\right),$$

where the first inequality follows from convexity of G. Corollary 2 then implies that a fully-revealing experiment is optimal.

Part (ii) - Consider two states θ and θ' and the indexed family of receiver and sender's posterior beliefs $q^R(\delta)$ and $q^S(\delta)$ given by

$$\begin{aligned} q^R(\delta) &= \delta \mathbf{1}_{\theta'} + (1-\delta)\mathbf{1}_\theta, \delta \in [0,1], \\ q^S(\delta) &= \lambda(\delta)\mathbf{1}_{\theta'} + (1-\lambda(\delta))\mathbf{1}_\theta, \text{with } \lambda(\delta) = \delta r_{\theta'}^S / (\delta r_{\theta'}^S + (1-\delta) r_\theta^S). \end{aligned}$$

Define $W(\delta,\theta,\theta')$ as

$$W(\delta,\theta,\theta') = \lambda(\delta)G(\theta') + (1-\lambda(\delta))G(\theta') - G(\delta\theta' + (1-\delta)\theta').$$

From Corollary 2, if for some (δ,θ,θ'), we have $W(\delta,\theta,\theta') < 0$, then the value of garbling is positive. After some algebraic manipulations, we can express $W(\delta,\theta,\theta')$ as

$$W(\delta,\theta,\theta') = \frac{\delta(1-\delta)}{(\delta r^S_{\theta'} + (1-\delta) r^S_{\theta})} S(\delta,\theta,\theta'),$$

with

$$S(\delta,\theta,\theta') = r^S_{\theta'} \frac{1}{(1-\delta)} \int_{\delta\theta'+(1-\delta)\theta}^{\theta'} G'(t)\,dt - r^S_{\theta} \frac{1}{\delta} \int_{\theta}^{\delta\theta'+(1-\delta)\theta} G'(t)\,dt.$$

Evaluating $S(\delta,\theta,\theta')$ at the extremes, we obtain

$$S(0,\theta,\theta') = (\theta'-\theta)\left(r^S_{\theta'}\bar{G}' - r^S_{\theta}G'(\theta)\right), \tag{31}$$

$$S(1,\theta,\theta') = (\theta'-\theta)\left(r^S_{\theta'}G'(\theta') - r^S_{\theta}\bar{G}'\right), \tag{32}$$

with

$$\bar{G}' = \frac{1}{(\theta'-\theta)} \int_{\theta}^{\theta'} G'(t)\,dt.$$

By assumption, there exist θ' and θ, $\theta' > \theta$, such that $\left(r^S_{\theta'}\right)^2 G'(\theta') < \left(r^S_{\theta}\right)^2 G'(\theta)$. This implies that $\frac{r^S_{\theta'}}{r^S_{\theta}} G'(\theta') < \frac{r^S_{\theta}}{r^S_{\theta'}} G'(\theta)$, which means that either $S(0,\theta,\theta')$ or $S(1,\theta,\theta')$ is strictly negative. To see this, suppose, for example, that $S(0,\theta,\theta') \geq 0$. Then,

$$\frac{r^S_{\theta'}}{r^S_{\theta}} G'(\theta') - \bar{G}' < \frac{r^S_{\theta}}{r^S_{\theta'}} G'(\theta) - \bar{G}' = -\frac{S(0,\theta,\theta')}{(\theta'-\theta)\, r^S_{\theta'}} \leq 0 \Rightarrow S(1,\theta,\theta') < 0. \blacksquare$$

Proof of Proposition 7: Consider a pair of realizations z and z' of an optimal experiment π. Consider a new experiment $\hat{\pi}$, which is identical to π except that realizations z and z' are merged into a single realization. The difference in the sender's expected utility from these two experiments is

$$\begin{aligned}
V_{\hat{\pi}} - V_{\pi} &= (\Pr_S[z] + \Pr_S[z'])\, G\left(\frac{\Pr_R[z]}{\Pr_R[z] + \Pr_R[z']} a_z + \frac{\Pr_R[z]}{\Pr_R[z] + \Pr_R[z']} a_{z'}\right) \\
&\quad - (\Pr_S[z] G(a_z) + \Pr_S[z'] G(a_{z'})) \\
&\geq (\Pr_S[z] + \Pr_S[z']) \frac{\Pr_R[z]}{\Pr_R[z] + \Pr_R[z']} G(a_z) + \frac{\Pr_R[z]}{\Pr_R[z] + \Pr_R[z']} G(a_{z'}) \\
&\quad - (\Pr_S[z] G(a_z) + \Pr_S[z'] G(a_{z'})) \\
&= \frac{\Pr_R[z] \Pr_R[z']}{\Pr_R[z] + \Pr_R[z']} \left(\lambda^S_{z'} - \lambda^S_z\right) (G(a_z) - G(a_{z'})).
\end{aligned}$$

Optimality of π requires that $0 \geq V_{\hat{\pi}} - V_{\pi}$ so that $0 \geq \left(\lambda^S_{z'} - \lambda^S_z\right)(G(a_z) - G(a_{z'}))$. Since G is increasing, if $\lambda^S_{z'} > \lambda^S_z$, then we must have $a_{z'} \geq a_z$. $\blacksquare$

References

[1] Acemoglu, D., V. Chernozhukov, and M. Yildiz (2006): "Learning and Disagreement in an Uncertain World," *NBER Working Paper No. 12648.*

[2] Alonso, R. and O. Câmara. (forthcoming): "Persuading Voters," *American Economic Review.*

[3] Alonso, R. and O. Câmara. (2015): "Political Disagreement and Information in Elections," *mimeo.*

[4] Aumann, R. J. (1976): "Agreeing to Disagree," *The Annals of Statistics*, 4(6), 1236-1239.

[5] Bertelli, A. M. (2012): *The political economy of public sector governance*, Cambridge University Press.

[6] Blackwell, D., and L. Dubins. (1962): "Merging of Opinions with Increasing Information," *The Annals of Mathematical Statistics*, 33(3), 882-886.

[7] Brocas, I., and J. Carrillo (2007): "Influence through Ignorance," *Rand Journal of Economics*, 38(4), 931-947.

[8] Camerer, C., Loewenstein, G. and M. Weber (1989): "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, 97(5), 1234-1254.

[9] Che, Y.-K., and N. Kartik (2009): "Opinions as Incentives," *Journal of Political Economy*, 117(5), 815-860.

[10] Darman, R. (1996): "Riverboat Gambling With Government," *The New York Times*, 12/01/1996, retrieved on 03/28/2016 from http://www.nytimes.com/1996/12/01/magazine/riverboat-gambling-with-government.html

[11] Duggan, J., and C. Martinelli (2011): "A Spatial Theory of Media Slant and Voter Choice," *Review of Economic Studies*, 78(2), 640-666.

[12] Ellison, G., and S. F. Ellison (2009): "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77, 427-452.

[13] Galperti, S. (2015): "Hide or Surprise: Persuasion without Common-Support Priors" *mimeo.*

[14] Gentzkow, M. and E. Kamenica (2014a): "Costly Persuasion," *American Economic Review*, 104(5), pp. 457-462.

[15] Gentzkow, M. and E. Kamenica (2014b): "Disclosure of Endogenous Information," *mimeo*, University of Chicago.

[16] Gentzkow, M. and J. M. Shapiro (2006): "Media Bias and Reputation," *Journal of Political Economy*, 114(2), 280-316.

[17] Giat, Y., Hackman, S. and A. Subramanian (2010): "Investment under Uncertainty, Heterogeneous Beliefs, and Agency Conflicts," *Review of Financial Studies*, 23(4), 1360-1404.

[18] Gill, D., and D. Sgroi (2008): "Sequential Decisions with Tests," *Games and Economic Behavior*, 63(2), 663-678.

[19] Gill, D., and D. Sgroi (2012): "The Optimal Choice of Pre-Launch Reviewer," *Journal of Economic Theory*, 147(3), 1247-1260.

[20] Hirsch, A. V. (forthcoming): "Experimentation and Persuasion in Political Organizations," *American Political Science Review*.

[21] Holmström, B. (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169-182.

[22] Ivanov, M. (2010): "Informational Control and Organizational Design," *Journal of Economic Theory*, 145(2), 721-751.

[23] Kalai, E., and E. Lehrer (1994): "Weak and Strong Merging of Opinions." *Journal of Mathematical Economics*, 23(1), 73-86.

[24] Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590-2615.

[25] Kolotilin, A. (2014): "Optimal Information Disclosure: Quantity vs. Quality," *mimeo*.

[26] Kolotilin, A. (2015): "Experimental Design to Persuade," *Games and Economic Behavior*, 90, 215-226.

[27] Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2), 380-91.

[28] Morris, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62(6), 1327-1347.

[29] Morris, S. (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11, 227-227.

[30] Orr, L. L. (1999): *Social Experiments: Evaluating Public Programs With Experimental Methods*, Sage Publications.

[31] Patton, A. J. and A. Timmermann (2010): "Why do Forecasters Disagree? Lessons from the Term Structure of Cross-sectional Dispersion," *Journal of Monetary Economics*, 57(7), 803-820.

[32] Rayo, L., and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118(5), 949-987.

[33] Rutter, T. (2015): "The rise of nudge — the unit helping politicians to fathom human behaviour," *The Guardian*, 07/23/2015, retrieved on 03/23/2016 from http://www.theguardian.com/public-leaders-network/2015/jul/23/rise-nudge-unit-politicians-human-behaviour

[34] Sethi, R., and M. Yildiz (2012): "Public Disagreement," *American Economic Journal: Microeconomics*, 4(3), 57-95.

[35] Shaked, M., and J. G. Shanthikumar (2007): *Stochastic Orders*, Springer.

[36] Van den Steen , E. (2004): "Rational Overoptimism (and Other Biases)," *American Economic Review*, 94(4), 1141-1151.

[37] Van den Steen , E. (2009): "Authority versus Persuasion," *American Economic Review: P&P*, 99(2), 448-453.

[38] Van den Steen , E. (2010a): "Interpersonal Authority in a Theory of the Firm," *American Economic Review*, 100(1), 466-490.

[39] Van den Steen , E. (2010b): "On the Origin of Shared Beliefs (and Corporate Culture)," *Rand Journal of Economics*, 41(4), 617-648.

[40] Van den Steen , E. (2011): "Overconfidence by Bayesian-Rational Agents," *Management Science*, 57(5), 884-896.

[41] Whitt, W. (1979): "A Note on the Influence of the Sample on the Posterior Distribution," *Journal of the America Statistical Association*, 74(366a), 424-426.

[42] Yildiz, M. (2004): "Waiting to Persuade," *The Quarterly Journal of Economics*, 119(1), 223-248.

B Online Appendix

In this Online Appendix we provide formal statements and proofs of the claims made in Section 4.4 and 5 of "Bayesian Persuasion with Heterogeneous Priors," by Alonso and Câmara.

B.1 Optimal Experiments to Persuade Skeptics and Believers

We complete Section 4.4 by characterizing properties of optimal experiments and describe a procedure to derive an optimal experiment. To describe this procedure, we now restrict attention to the case in which the sender is risk-neutral over the receiver's beliefs.

Proposition 8 *Suppose that **(A1)** and **(A2′)** hold, with G linear, card$(\Theta) > 2$, and that for each triplet $\theta_i, \theta_j, \theta_k \in \Theta$ of states, $(\theta_i, \theta_j, \theta_k)$ and (r_i^S, r_j^S, r_k^S) are not negatively collinear with respect to W. For each pair of states (θ_i, θ_j), define*

$$\Delta_{(i,j)} = -\left(r_j^S - r_i^S\right)\left(\theta_j - \theta_i\right). \tag{33}$$

If π^* is an optimal experiment, then, after each realization of π^*, the receiver puts positive probability in at most two states. Furthermore, for each state θ_i, there is a threshold $\xi_i \geq 0$ such that there is a realization of π^* induced by both states θ_i and θ_j if and only if $\Delta_{(i,j)} \geq \xi_i$.

Consequently, for every subset of states $\{\theta_i, \theta_j, \theta_k\}$, if either $\Delta_{(i,j)} \leq \min\{\Delta_{(i,k)}, \Delta_{(k,j)}\}$ or $\Delta_{(i,j)} < 0$, then there is no realization supported on both θ_i and θ_j.

Consider any pair $\theta_j > \theta_i$. The term $\Delta_{(i,j)}$ captures the value to the sender of "bundling" states θ_i and θ_j — the value of pooling these states into the same realization of the experiment. Pooling the states has positive value if and only if the receiver is a believer ($r_j^S < r_i^S$), conditional on the partition $\{\theta_i, \theta_j\}$. A positive-value bundle becomes more valuable when the differences $r_i^S - r_j^S$ and $\theta_j - \theta_i$ are larger. If state θ_i has more than one positive-value bundle, then the sender optimally allocates probability mass from θ_i across these bundles according to their value. Bundles with low positive value may be broken so that more probability mass can be assigned to higher-value bundles.

We now apply Proposition 8 to construct an algorithm to solve for the optimal experiment when there are three states, $\theta_1 < \theta_2 < \theta_3$ (see the proof of Proposition 8 for details):

Step 1: Compute the ratios $\frac{r_2^S - r_1^S}{\theta_2 - \theta_1}$ and $\frac{r_3^S - r_2^S}{\theta_3 - \theta_2}$. If the ratios are equal to each other and (weakly) negative, then no experimentation is optimal. Otherwise, proceed to Step 2.

Step 2: Compute the pooling values $\Delta_{(1,2)}$, $\Delta_{(2,3)}$ and $\Delta_{(3,1)}$. If all values are (weakly) negative, then a fully informative experiment is optimal. Otherwise, proceed to Step 3.

Step 3: Let θ_i and θ_j be the states with the lowest pooling value $\Delta_{(i,j)}$, and θ_k the remaining state. Construct experiment π_α as follows. There is a binary realization space $Z = \{z_i, z_j\}$. Likelihood functions are: state θ_i induces realization z_i with probability one; state θ_j induces z_j with probability one; state θ_k induces realization z_i with probability α and induces z_j with probability $1-\alpha$. The optimal experiment π_{α^*} is the one with α^* that maximizes the sender's expected payoff

$$\max_{\alpha\in[0,1]} \Pr{}_S[z_i|\pi_\alpha]E_R[\theta|z_i,\pi_\alpha] + \Pr{}_S[z_j|\pi_\alpha]E_R[\theta|z_j,\pi_\alpha]. \tag{34}$$

We can use this algorithm to solve the example from the introduction: $\Theta = \{1, 1.5, 2\}$, $p^S = (0.85, 0.10, 0.05)$ and $p^R = (0.10, 0.40, 0.55)$. The condition in Step 1 is not met, so we proceed to Step 2 and compute $\Delta_{1,1.5} = 4.125$, $\Delta_{1.5,2} = 0.075$ and $\Delta_{1,2} = 8.4$. Since they are positive, we proceed to Step 3. The lowest pooling value is $\Delta_{1.5,2}$; hence, we construct the binary realization space $Z = \{z_{1.5}, z_2\}$. State $\{1.5\}$ induces $z_{1.5}$ with probability one; state $\{2\}$ induces z_2 with probability one; and state $\{1\}$ induces $z_{1.5}$ with probability α. Given this experiment, (34) becomes

$$\begin{aligned}\max_{\alpha\in[0,1]} \quad & (\alpha 0.85 + 0.1)\left(1\frac{\alpha 0.85}{0.1+\alpha 0.85} + 1.5\frac{0.1}{0.1+\alpha 0.85}\right) \\ & + ((1-\alpha)0.85 + 0.05)\left(1\frac{(1-\alpha)0.85}{(1-\alpha)0.85+0.05} + 2\frac{0.05}{(1-\alpha)0.85+0.05}\right),\end{aligned}$$

and the sender's optimal choice is $\alpha^* = 1$.

In summary, the sender's primary concern is which bundles should be broken and which should be kept. When there are more than three states, the logic above can be used to eliminate all bundles with negative value and, for each triplet of states, eliminate the bundle with the lowest value. After all the "weak" bundles are eliminated, each group of states no longer "connected" with other groups of states can then be treated independently in the design of an optimal experiment.

Proof of Proposition 8: Proposition 5.i shows that the condition on each triplet $\theta_i, \theta_j, \theta_k \in \Theta$ implies that any realization of an optimal experiment leads to posterior beliefs supported on at most two states. For each pair (θ_i, θ_j), we now investigate under what conditions the optimal experiment has a realization induced by states θ_i and θ_j.

Denote by z_{ij} a realization induced by both states θ_i and θ_j. In particular, we allow z_{ii} to be a realization induced only by θ_i (and, thus, that fully reveals the state). For any experiment π, we have that the sender's expectation over its posterior expectations must equal the prior expectation — i.e., $\mathrm{E}_S^\pi[\mathrm{E}_S[\theta|z]] = \mathrm{E}_S[\theta]$. Therefore, if an experiment π^* maximizes the sender's expectation of the receiver's posterior expectation, it also maximizes the sender's expectation of the difference between the receiver's and the sender's expectation. That is, for an arbitrary π,

$$\mathrm{E}_S^{\pi^*}[\mathrm{E}_R[\theta|z]] \geq \mathrm{E}_S^{\pi}[\mathrm{E}_R[\theta|z]] \Leftrightarrow \mathrm{E}_S^{\pi^*}[\mathrm{E}_R[\theta|z] - \mathrm{E}_S[\theta|z]] \geq \mathrm{E}_S^{\pi}[\mathrm{E}_R[\theta|z] - \mathrm{E}_S[\theta|z]].$$

If a sender seeks to maximize the difference between the receiver's and her expectation of the state, her expected utility from an experiment π can be written as

$$\begin{aligned}
\mathrm{E}_S^{\pi}[\mathrm{E}_R[\theta|z] - \mathrm{E}_S[\theta|z]] &= \sum \Pr_S[z]\left(\left\langle q^R(z), \theta\right\rangle - \left\langle \frac{q^R(z) r^S}{\left\langle q^R(z), r^S\right\rangle}, \theta\right\rangle\right) \\
&= \sum \Pr_R[z]\left(\left\langle q^R(z), \theta\right\rangle \left\langle q^R(z), r^S\right\rangle - \left\langle q^R(z) r^S, \theta\right\rangle\right).
\end{aligned}$$

If an experiment induces realizations z_{ij} that are only supported on at most two states, then

$$\begin{aligned}
\left\langle q^R(z_{ij}), \theta\right\rangle \left\langle q^R(z_{ij}), r^S\right\rangle - \left\langle q^R(z_{ij}) r^S, \theta\right\rangle &= -q_i^R(z_{ij}) q_j^R(z_{ij}) \left(r_j^S - r_i^S\right)(\theta_j - \theta_i) \\
&= q_i^R(z_{ij}) q_j^R(z_{ij}) \Delta_{(i,j)},
\end{aligned}$$

so that we can write

$$\mathrm{E}_S^{\pi}[\mathrm{E}_R[\theta|z] - \mathrm{E}_S[\theta|z]] = \sum \Pr_R[z_{ij}]\, q_i^R(z_{ij}) q_j^R(z_{ij}) \Delta_{(i,j)}. \tag{35}$$

Letting $\alpha_{ij}^i = \Pr[z_{ij}|\theta_i] \Pr_S[\theta_i]$, and denoting by $H(p,q)$ the harmonic mean of p and q, so that $H(p,q) = \frac{2pq}{p+q}$, we can write (35) as

$$\mathrm{E}_S^{\pi}[\mathrm{E}_R[\theta|z] - \mathrm{E}_S[\theta|z]] = \frac{1}{2} \sum H(\alpha_{ij}^i, \alpha_{ij}^j) \Delta_{(i,j)}. \tag{36}$$

As previously noted, an experiment that maximizes (35) also maximizes $\mathrm{E}_S^\pi[\mathrm{E}_R[\theta|z]]$. Therefore, an optimal experiment under **(A1)** and **(A2′)** also solves the following program:

$$\max \sum H(\alpha_{ij}^i, \alpha_{ij}^j) \Delta_{(i,j)}, s.t. \alpha_{ij}^i, \alpha_{ij}^j \geq 0, \sum_{\theta_k \in \Theta} \alpha_{ik}^i = p_{\theta_i}^R. \tag{37}$$

Consider a fixed state θ_i. We now investigate which realizations will be induced by θ_i. First, if $\alpha^i_{ij}, \alpha^j_{ij} > 0$, we must have $\Delta_{(i,j)} > 0$, as the sender could otherwise improve by having the experiment fully reveal θ_i and θ_j if z_{ij} is realized. Second, as

$$\frac{\partial H(\alpha^i_{ij}, \alpha^j_{ij})}{\partial \alpha^i_{ij}} = \left(\frac{\alpha^j_{ij}}{\alpha^i_{ij} + \alpha^j_{ij}}\right)^2 \leq 1,$$

the marginal return to increasing α^i_{ij} in $H(\alpha^i_{ij}, \alpha^j_{ij})$ is largest when $\alpha^i_{ij} = 0$, in which case it equals 1. Now suppose that under an optimal experiment, we have that $\alpha^i_{ij} > 0$ and $\alpha^i_{ik} = 0$. Then, we must have that $\Delta_{(i,j)} \geq \Delta_{(i,k)}$. Otherwise, if $\Delta_{(i,j)} < \Delta_{(i,k)}$, marginally increasing α^i_{ik} while reducing α^i_{ij} would generate a gain to the sender

$$\frac{\partial H(\alpha^i_{ij}, \alpha^k_{ik})}{\partial \alpha^i_{ik}}\Delta_{(i,k)} - \frac{\partial H(\alpha^i_{ij}, \alpha^j_{ij})}{\partial \alpha^i_{ij}}\Delta_{(i,j)} = \Delta_{(i,k)} - \frac{\partial H(\alpha^i_{ij}, \alpha^j_{ij})}{\partial \alpha^i_{ij}}\Delta_{(i,j)} > \Delta_{(i,k)} - \Delta_{(i,j)} \geq 0.$$

To prove the last claim, suppose by way of contradiction that $\Delta_{(i,j)} \leq \min\{\Delta_{(i,k)}, \Delta_{(k,j)}\}$ and yet $\Pr_S[z_{i,j}] > 0$. First, this requires $\Delta_{(i,j)} \geq 0$. Second, applying the first part of Proposition 8 implies that $\Delta_{(i,j)} \geq \xi_i$, and since $\Delta_{(i,k)} \geq \Delta_{(i,j)}$, we must have $\Pr_S[z_{i,k}] > 0$. Similarly, $\Delta_{(k,j)} \geq \Delta_{(i,j)} \geq \xi_j$ implies that $\Pr_S[z_{k,j}] > 0$. Finally, the fact that all elements $\Delta_{(i,j)}, \Delta_{(i,k)}$,and $\Delta_{(k,j)}$ are positive implies that r^S decreases for a higher state — i.e., for $\theta_j > \theta_i$, we must have $r^S_j < r^S_i$.

Suppose, wlog, that the three states are ordered $\theta_i < \theta_j < \theta_k$. Since (7) can be rewritten as $\lambda^S_z = \langle q^R(z), r^S \rangle$, $\Pr_S[z_{i,j}], \Pr_S[z_{j,k}] > 0$ implies $r^S_i > \lambda^S_{z_{ij}} > r^S_j > \lambda^S_{z_{jk}} > r^S_k$. Therefore, $a_{z_{ij}} < a_{z_{jk}}$, but $\lambda^S_{z_{ij}} > \lambda^S_{z_{jk}}$, which violates the conclusion of Proposition 7, and, thus, this experiment cannot be optimal. ■

B.2 Private Priors

Consider the extended model with private priors described in Section 5. As an application of (24), consider the pure persuasion model from Section 4.3. When the sender knows the receiver's prior, Proposition 5(i) provides conditions on the likelihood ratio of priors such that persuasion is valuable. Suppose that these conditions are met and the sender strictly benefits from providing experiment π to a particular receiver. By a continuity argument, the same π strictly benefits the sender when she faces another receiver whose beliefs are not too different. Consequently, even if the sender does not know the receiver's prior, persuasion remains beneficial when the receiver's possible priors are not too dispersed. Proposition B.1 provides an upper bound on how dispersed these beliefs can be. To this end, let R be the set of likelihood ratios induced by the priors in the support of $h(p^R|p^S)$,

$$R = \left\{ r^R : \{ r^R_\theta = p^R_\theta / p^S_\theta \}_{\theta \in \Theta},\ p^R \in Supp(h(p^R|p^S)) \right\}. \tag{38}$$

Proposition B.1 *Suppose that r^R and $r^R\theta$ are not collinear w.r.t. W for all $r^R \in R$, and let $m = \frac{1}{2}\frac{\max\left|u''_S(a)\right|}{\min u'_S(a)} > 0$. If for all $r^R, r^{R'} \in R$*

$$\left\| r^R - r^{R'} \right\| \leq \beta, \tag{39}$$

with β given by (47), then the sender benefits from persuasion.

The condition on r^R and $r^R\theta$ implies that if the sender knew the receiver's prior, then she could find an experiment with a positive value (cf. Proposition 5). The bound β is defined below by (47), as a function of the curvature of u_S. From (39), β represents a lower bound on the cosine of the angle between any two likelihood ratios in the support of $h(p^R|p^S)$. Therefore, (39) describes how different the receiver's possible prior beliefs can be for the sender still to benefit from persuasion, by imposing an upper bound on the angle between any two likelihood ratios in R.

Proof: The proof of this Proposition will make use of the following lemma:

Lemma B.1 *Let R be defined by (38) and $m = \frac{1}{2}\frac{\max\left|u''_S(a)\right|}{\min u'_S(a)} > 0$, and for each $r^R \in R$, define $\Delta_S = \frac{\left\langle q^S, r^R\theta \right\rangle}{\left\langle q^S, r^R \right\rangle} - \left\langle p^R, \theta \right\rangle$, and define $l_{r^R}(\varepsilon)$ as*

$$l_{r^R}(\varepsilon) = \frac{\left\langle \varepsilon, r^R \right\rangle}{\Delta_S}. \tag{40}$$

For any ε and $r^R \in R$ such that

$$l_{r^R}(\varepsilon) < -m \text{ and } \Delta_S > 0, \text{ with } p^S + \varepsilon \in \Delta(\Theta), \tag{41}$$

there exists an experiment π with the following properties: (i) Some realization of π induces in the sender the belief $p^S + \varepsilon$; and (ii) π increases the expected utility of the sender when the receiver's associated likelihood ratio is r^R.

Proof: The function $l_{r^R}(\varepsilon)$ has an immediate interpretation as a measure of disagreement: the numerator $\langle \varepsilon, r^R \rangle$ is the difference in the probability that the receiver and sender attach to a realization inducing a posterior $q_S = p_S + \varepsilon$ on the sender, divided by the probability that the sender ascribes to such realization, while the denominator is the change in the receiver's action when the sender changes her belief to q_S. We first show that if some ε satisfies (41), then the value of information control is positive. Consider V_S defined in (11), which in this case can be written as

$$V_S(q^S) = u_S\left(\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}\right),$$

with gradient at p^S

$$\nabla V_S(p^S) = u_S'(\langle p^R, \theta\rangle)\left(r^R\theta - \langle p^R, \theta\rangle r^R\right).$$

By Corollary 1, the value of information control is positive if and only if there exists ε, with $p^S + \varepsilon \in \Delta(\Theta)$, such that

$$\langle \nabla V_S(p^S), \varepsilon\rangle < V_S(p^S + \varepsilon) - V_S(p^S). \tag{42}$$

We now show that an ε satisfying (41) also satisfies (42). Since

$$u_S\left(\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}\right) - u_S(\langle p^R, \theta\rangle) - u_S'(\langle p^R, \theta\rangle)\left(\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle} - \langle p^R, \theta\rangle\right) = \int_{\langle p^R,\theta\rangle}^{\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}} \int_{\langle p^R,\theta\rangle}^{t} u_S''(\tau)d\tau dt,$$

we can rewrite (42) as

$$u_S'(\langle p^R, \theta\rangle)\langle \varepsilon, r^R\rangle \Delta_S < \int_{\langle p^R,\theta\rangle}^{\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}} \int_{\langle p^R,\theta\rangle}^{t} u_S''(\tau)d\tau dt.$$

By the mean value theorem, we have

$$\int_{\langle p^R,\theta\rangle}^{\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}} \int_{\langle p^R,\theta\rangle}^{t} u_S''(\tau)d\tau dt \geq -\max|u_S''(a)| \int_{\langle p^R,\theta\rangle}^{\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}} \int_{\langle p^R,\theta\rangle}^{t} d\tau dt = -\frac{1}{2}\max|u_S''(a)|\Delta_S^2.$$

Moreover, if ε satisfies (41), then it also satisfies

$$\langle \varepsilon, r^R\rangle \min u_S'(a) < -\frac{1}{2}\max|u_S''(a)|\Delta_S,$$

implying that ε also satisfies (42) since

$$u_S'(\langle p^R, \theta\rangle)\langle \varepsilon, r^R\rangle \Delta_S < \langle \varepsilon, r^R\rangle \Delta_S \min u_S'(a) < -\frac{1}{2}\max|u_S''(a)|\Delta_S^2 \leq \int_{\langle p^R,\theta\rangle}^{\frac{\langle q^S, r^R\theta\rangle}{\langle q^S, r^R\rangle}} \int_{\langle p^R,\theta\rangle}^{t} u_S''(\tau)d\tau dt.$$

For each ε satisfying (41), we now construct an experiment that improves the sender's expected utility and that has a realization that induces belief $p^S+\varepsilon$ in the sender. Let υ be the excess of the right-hand side over the left-hand side in (42),

$$\upsilon = V_S(p^S+\varepsilon) - V_S(p^S) - \left\langle \nabla V_S(p^S), \varepsilon \right\rangle > 0. \tag{43}$$

Consider the experiment $\pi(\varepsilon,\delta)$ with $Z=\{\varepsilon^+,\varepsilon^-\}$, such that $\Pr_S[z=\varepsilon^+]=\delta$, and if $z=\varepsilon^+$, then the sender's posterior is $p^S+\varepsilon$. A taylor series expansion of $V_S(q^S)$ yields

$$V_S(q^S) = V_S(p^S) + \left\langle \nabla V_S(p^S), q^S-p^S \right\rangle + L\left(q^S-p^S\right), \text{ with } \lim_{t\to 0}\frac{L\left(t\left(q^S-p^S\right)\right)}{t} = 0. \tag{44}$$

Then, the sender's gain from $\pi(\varepsilon,\delta)$ is

$$\begin{aligned}
\Delta_{\pi(\varepsilon,\delta)} &= \delta\left(V_S(p^S+\varepsilon)-V_S(p^S)\right) + (1-\delta)\left(V_S(p^S-\frac{\delta}{1-\delta}\varepsilon)-V_S(p^S)\right)\\
&= \delta\left(\upsilon+\left\langle \nabla V_S(p^S),\varepsilon\right\rangle\right) - \delta\left\langle \nabla V_S(p^S),\varepsilon\right\rangle + L\left(-\frac{\delta}{1-\delta}\varepsilon\right)\\
&= \delta\left(\upsilon-(1-\delta)\frac{L\left(-\delta\varepsilon/(1-\delta)\right)}{\left(-\delta/(1-\delta)\right)}\right).
\end{aligned}$$

The convergence to zero of the second term in the parentheses when δ tends to zero and $\upsilon>0$ guarantees the existence of $\delta>0$ such that $\Delta_{\pi(\varepsilon,\delta)}>0$. ■

Proof of Proposition B.1: First, we introduce additional notation. With $l_{r^R}(\varepsilon)$ defined as in (40), define the sets $M(r^R)$ by

$$M(r^R) = \left\{\varepsilon : l_{r^R}(\varepsilon) < -m,\ \Delta_S > 0, p^S+\varepsilon \in \Delta(\Theta)\right\}.$$

Note that r^S and θ are negatively collinear if and only if r^R and $r^R\theta$ are positively collinear. That is, the condition on Proposition 5 could instead be stated in terms of collinearity of r^R and $r^R\theta$. Moreover, if r^R and $r^R\theta$ are not collinear, then the restriction of $l_{r^R}(\varepsilon)$ to $\{\varepsilon : \langle\varepsilon, 1\rangle = 0\}$ is surjective, and, thus, the set $M(r^R)$ is non-empty.

Define the function

$$\Psi\left(\varepsilon, r^R\right) = \left\langle \varepsilon, r^R - m f^R \right\rangle + \left(\left\langle \varepsilon, r^R\right\rangle\right)^2, \text{ with } f^R = r^R\theta - \left\langle p^S, r^R\theta\right\rangle,$$

which characterizes $M(r^R)$ since for ε such that $p^S+\varepsilon\in\Delta(\Theta)$, $\Psi\left(\varepsilon, r^R\right)\le 0$ and $\left\langle\varepsilon, f^R\right\rangle \ge 0$ if and only if $\varepsilon\in M(r^R)$. Finally, let

$$\begin{aligned}
\gamma &= 2\left(1+m\left(\max|\theta|+\|\theta\|\right)+(4+m\|\theta\|)\sup_{r^R\in R}\left\|r^R\right\|\right), &(45)\\
Z &= \min_{\varepsilon\in\{\varepsilon:p^S+\varepsilon\in\Delta(\Theta)\}, r^R\in R} \Psi\left(\varepsilon, r^R\right) \text{ s.t. } \left\langle \varepsilon, r^R\left(\theta-\left\langle p^S, r^R\theta\right\rangle\right)\right\rangle \le 0, r^R\in R. &(46)
\end{aligned}$$

Under the conditions of Proposition B.1, $Z < 0$. Finally, define β in (39) as

$$\beta = \frac{|Z|}{\gamma}. \tag{47}$$

Our proof is structured in two steps that show (i) if $\cap_{r^R \in R} M(r^R)$ is non-empty, then following Lemma B.1 allows us to design an experiment π that increases the sender's expected utility for every receiver's belief in the support of $h(p^R|p^S)$; and (ii) under the conditions of Proposition B.1, $\cap_{r^R \in R} M(r^R) \neq \varnothing$.

Step (i) - Suppose that $\varepsilon \in \cap_{r^R \in R} M(r^R)$. Consider υ as defined by (43). As υ is a continuous function of r^R in the compact set R, it achieves a minimum $\underline{\upsilon}= \min_{r^R \in R} \upsilon > 0$. Then, define $\underline{\delta}$ as

$$\underline{\delta} = \min \left\{ \delta : \underline{\upsilon} + \frac{L\left(-\frac{\delta}{1-\delta}\varepsilon\right)}{\delta} \geq 0 \right\},$$

with the function L given by (44). Now, define the experiment $\pi(\varepsilon, \delta')$ as in the proof of Lemma B.1— i.e., $Z = \{\varepsilon^+, \varepsilon^-\}$, $q^S(\varepsilon^+) = p^S + \varepsilon$ and $\Pr_S[z = \varepsilon^+] = \delta'$, and set $\delta' = \underline{\delta}$. Then, the sender's gain from $\pi(\varepsilon, \delta')$ is positive for any receiver's prior in $Supp(h(p^R|p^S))$.

Step (ii) - Fix $p^{R'}$ with associated likelihood ratio $r^{R'} \in R$. For any $r^R \in R$ with $\eta = r^R - r^{R'}$, we have

$$\Psi\left(\varepsilon, r^R\right) - \Psi\left(\varepsilon, r^{R'}\right) = \left(1 + m\left\langle p^S, r^{R'}\theta\right\rangle + \left\langle \varepsilon, r^R + r^{R'}\right\rangle\right)\langle \varepsilon, \eta\rangle - m\langle \varepsilon, \eta\theta\rangle + m\left\langle p^S, \eta\theta\right\rangle \langle \varepsilon, r\rangle .$$

The following bounds make use of the Cauchy-Schwartz inequality (in particular, the implication that $|\langle \varepsilon, \eta\theta\rangle| \leq \|\varepsilon\| \|\eta\| \|\theta\|$—see Steele, 2004)[28] and the fact that $\left\|p^S\right\| \leq 1$ and $\|\varepsilon\| = \left\|q^S - p^S\right\| \leq 2$,

$$\begin{aligned}
\left|1 + m\left\langle p^S, r^{R'}\theta\right\rangle + \left\langle \varepsilon, r^R + r^{R'}\right\rangle\right| &\leq 1 + m \max\theta + 4 \sup_{r^R \in R} \left\|r^R\right\|, \\
|m\langle \varepsilon, \eta\theta\rangle| &\leq m\|\varepsilon\| \|\eta\| \|\theta\| \leq 2m\|\eta\| \|\theta\|, \\
\left|m\left\langle p^S, \eta\theta\right\rangle \langle \varepsilon, r\rangle\right| &\leq 2m\|\eta\| \|\theta\| \sup_{r^R \in R} \left\|r^R\right\|.
\end{aligned}$$

[28]Steele, J. M. (2004) "The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities," *Mathematical Association of America.*

From these bounds, we then obtain the following estimate

$$
\begin{aligned}
\left|\Psi\left(\varepsilon, r^R\right)-\Psi\left(\varepsilon, r^{R'}\right)\right| &\leq \left|1+m\left\langle p^S, r^{R'}\theta\right\rangle+\left\langle\varepsilon, r^R+r^{R'}\right\rangle\right|\|\varepsilon\|\|\eta\| \\
&\quad +|m\langle\varepsilon, \eta\theta\rangle|+\left|m\left\langle p^S, \eta\theta\right\rangle\langle\varepsilon, r\rangle\right| \\
&\leq 2\left(1+m\max\theta+4\sup_{r^R\in R}\left\|r^R\right\|\right)\|\eta\|+2m\|\theta\|\|\eta\| \\
&\quad +2m\|\theta\|\sup_{r^R\in R}\left\|r^R\right\|\|\eta\| \\
&= \gamma\|\eta\|,
\end{aligned}
$$

where γ is defined by (45). Selecting ε' an $r^{R'}$ that solve the program (46) and noting that $Z<0$, we have that for any $r^R \in R$,

$$
\Psi\left(\varepsilon', r^R\right)=\Psi\left(\varepsilon', r^{R'}\right)+\Psi\left(\varepsilon', r^R\right)-\Psi\left(\varepsilon', r^{R'}\right) \leq Z+\gamma\|\eta\| \leq Z+|Z|=0.
$$

This implies that $\varepsilon' \in M(r^R)$ for all $r^R \in R$. ■