Hide or Surprise?
Persuasion without Common-Support Priors

Simone Galperti*

UCSD

February 27, 2015

Abstract

Often persuaders have a richer understanding of the world than their audience. This paper models such situations by letting persuader and audience have prior beliefs with different supports. This asymmetry adds unexplored aspects to the persuasion problem: Persuaders can hide their superior knowledge or surprise their audience with unexpected information; After surprises Bayes' rule cannot describe the audience's response. The paper examines persuaders' incentives to hide and surprise and their resulting communication strategies. Moreover, it derives necessary and sufficient conditions for persuaders to surprise their audience as well as to hide some information.

KEYWORDS: persuasion, information control, common support, hiding information, surprise, non-Bayesian updating, concavification.
JEL CLASSIFICATION: D82, D83, K41, M30.

*I thank S. Nageeb Ali, Eddie Dekel, Laura Doval, Mark Machina, Marciano Siniscalchi, Joel Sobel, and Joel Watson for their comments and suggestions. John N. Rehbeck provided excellent research assistance. All remaining errors are mine. First Draft: December 2014

1 Introduction

Persuasion plays an important role across economic activities.[1] In many situations, the persuader has a superior understanding of the subject about which she wants to persuade her audience. This asymmetry can be better information—a more precise assessment of the likelihood of the possible cases—but more often it is better knowledge—a richer, broader, or more accurate theory of what is possible. Examples include experts and policymakers, physicians and patients, parents and young children, sellers of new products and consumers. Due to her superior knowledge, the persuader may view as possible things that her audience ignores, deems impossible, or is unaware of.

This asymmetry between persuader and audience adds unexplored aspects to the persuasion problem, which this paper aims to investigate. First of all, the persuader (Sender, she) understands that part of the information she can convey is completely unexpected for her audience (Receiver, he). This raises new questions: How will Receiver react to unexpected information? Given this, should Sender hide such information (if possible) or disclose it, thus surprising Receiver? What are the benefits and drawbacks of hiding information and surprising Receiver? Can Sender combine hiding and surprising to her best advantage? How? The paper answers these questions in general persuasion games à la Kamenica and Gentzkow (2011), in which the asymmetry between Sender's and Receiver's 'theory of the world' is modeled by letting their subjective priors over an arbitrary state space have different supports.

To illustrate the analysis and ease the comparison with the standard case of common-support priors, it is helpful to revisit the court example in Kamenica and Gentzkow (2011).[2] A lawyer (Sender) wants to convince the judge (Receiver) that the defendant caused a damage to her client, so as to obtain a sentence to an equal refund. The damage can be of four levels (states): 0, 1, 2, or 3 (thousands/millions of) dollars. The lawyer's prior assigns them probability 0.35, 0.4, 0.15, and 0.1, respectively. By contrast, the judge thinks that the damage is either \$0 or \$2, with prior probabilities 0.7 and 0.3.[3] Assume that the lawyer's prior is correct. The judge's goal is to match sentence to actual damage: in each state, he gets payoff 1 from doing so and zero otherwise. The lawyer maximizes the expected refund.[4] To do so, she conducts an investigation whose

[1] See, e.g., McCloskey and Klamer (1995).

[2] This example is formally solved in Section 5.1.

[3] Note that conditional on the event {\$0,\$2} the lawyer and the judge have the same belief. To be clear, the case in which both *priors* are $(0.7, 0, 0.3, 0)$ corresponds to Kamenica and Gentzkow's (2011) example.

[4] The general model allows Sender's utility function to be state dependent as well.

outcome must be fully disclosed as per due-process law. Formally, an investigation is a family π of distributions—one for each state—specifying the probabilities of observing different signals (or evidence). In particular, the lawyer must produce some evidence also in states \$1 and \$3, hence she cannot hide them by simply remaining silent.

The priors' different supports immediately raise a conceptual issue: how does the judge respond to unexpected evidence? If an investigation perfectly reveals the damage, for instance, he may observe signals to which ex ante he assigned zero probability. To describe his response to such signals we cannot directly invoke Bayes' rule, the backbone of the entire literature on persuasion (see below). However, allowing the judge to have any posterior belief seems unsatisfactory, for instance simply because the observed signal may rule out some states. It seems also reasonable that his posteriors satisfy some regularity and predictability.

This paper considers different models of how Receiver responds to unexpected information. The first is an application of Ortoleva's (2012) axiomatic model of "paradigm change." The second assumes that Receiver is endowed with a lexicographic belief system (LBS), whose first element corresponds to his prior (e.g., $(0.7, 0, 0.3, 0)$ for the judge).[5] In both models, after expected signals Receiver updates his initial prior using Bayes' rule, as usual. After unexpected signals, however, he first looks for a new prior (or theory) that accounts for the observed evidence—following different procedures depending on the model. Once he finds his new prior, Receiver again updates it using Bayes' rule. In our example, let the judge's LBS have two priors with the secondary one equal to the lawyer's prior.[6]

With these assumptions, we can formally study a key dilemma Sender faces in our setting. Should the lawyer confirm the judge's theory that the defendant is either innocent or guilty of a \$2 damage, or disprove it? Always producing evidence the judge expects will confirm his theory and lead at most to a \$2 refund. By contrast, producing unexpected evidence will disprove that theory. It is intuitive that if the damage is \$3, the lawyer would like the judge to know. But what about the \$1 damage?

The solution to Sender's dilemma relies on a general characterization of the joint distributions over posterior beliefs she can achieve by controlling information (one of the paper's main results). In settings with common-support priors, Receiver's and Sender's posteriors are always in a one-to-one relationship.[7] This is no longer true without common-

[5]This second class of models is related to the lexicographic sequences of hypotheses in Kreps and Wilson (1982) (see Section 2).

[6]This case also corresponds to the simplest version of Ortoleva's (2012) model (see Section 4.1).

[7]See Kamenica and Gentzkow (2011) and Alonso and Câmara (2013).

support priors: each Sender's posterior corresponds to a unique Receiver's posterior, but not vice versa. Indeed, Sender's posterior tells us whether Receiver's theory has been confirmed or disproved, which prior in his LBS he updates, and the ensuing posterior. In Ortoleva's (2012) model, it tells us how Receiver changes paradigm and forms his posterior. Finally, the marginal distribution over Sender's posteriors has to satisfy the usual constraint: her expected posterior must equal her prior.

The feasible distributions over posteriors have specific properties, absent in the standard setting. We can illustrate them with our example. First, the judge's posterior cannot assign positive probability both to elements in {\$0,\$2} and to elements in {\$1,\$3}. If the lawyer confirms his theory, he continues to think only \$0 and \$2 are possible; to disprove it, the lawyer must show that the damage is *neither* \$0 *nor* \$2. Second, the judge's posterior varies discontinuously in the lawyer's as we go from expected to unexpected signals; this discontinuity captures formally the idea of surprise.[8] Third, after expected signals, the judge's posterior depends only on the lawyer's posterior *conditional on the states the former deemed possible at the outset*, i.e., the event {\$0,\$2}. As a result, the lawyer can have any posterior conditional on the remaining states. More generally, this says that Receiver's reluctance to abandon his theory in light of inconclusive evidence allows Sender to 'hide' states outside that theory by pooling them with signals on states inside it; moreover, she can learn about the states she hides, without affecting Receiver's behavior. Concretely, if an investigation yields signal x when the damage is either \$1 or \$2 and y when it is either \$2 or \$3, the judge will always conclude that it is \$2, but the lawyer learns about state \$1 from x and state \$3 from y.

These general properties drive how Sender will communicate. To analyze this, the paper has to extend and modify the 'concavification method' now common in the persuasion literature (see below). This method allows us to characterize Sender's expected payoff from feasible distributions over posteriors and Receiver's ensuing actions, which contains useful information on her communication strategy.

To build intuition, suppose first that Sender always confirms Receiver's theory, hiding all states outside it. In this case, her payoff coincides with the one she would get by dividing her problem as follows. She first learns only whether the true state is consistent with Receiver's theory—the event {\$0,\$2} or {\$1,\$3} in the example. If it is, her interim belief has the same support as Receiver's prior, and she adopts the optimal communication strategy as in Kamenica and Gentzkow (2011) and Alonso and Câmara (2013). Otherwise, she optimally hides the states as follows. For each state, she produces one

[8] This is different from the notion of surprise in Ely et al. (2013).

signal which leads her to assign *almost* probability 1 to it and Receiver to choose the best action for her in that state *among all actions she can induce while confirming his theory*. Concretely, the lawyer hides the \$1 and \$3 damages with an investigation that maps each to a different signal which, however, does not rule out entirely the \$2 damage—the same signal also arises with a tiny probability in this state. This sophisticated strategy defines the value of hiding states—\$2 in the example—and hence is the key to Sender's incentives to surprise Receiver.

Based on these observations, the paper provides a simple necessary and sufficient condition for Sender to surprise Receiver. There must exist *some* posterior of Sender such that, given Receiver's response to the surprise, her expected payoff exceeds that from optimally hiding *all* states to which she assigns positive probability. Our example immediately satisfies this condition for the posterior assigning probability 1 to state \$3, since the value of optimally hiding it is \$2.

To understand Sender's communication and payoff when she surprises Receiver, we can again imagine that she divides her problem into two parts. If the true state is consistent with Receiver's theory (\$0 or \$2), she communicates as before. If not, she optimally combines hiding states and surprising Receiver. This combination's payoff can be computed as the value taken at Sender's interim belief (i.e., given event {\$1,\$3}) by the concavification of the maximum between two functions: for each posterior, one gives her expected payoff from optimally hiding states, the other that from surprising Receiver. This procedure also delivers the ex-ante probability that Sender surprises Receiver, how she hides states, and a necessary and sufficient condition for her to never hide any state.

All results on Sender's payoffs and communication are in terms of primitives and do not involve geometric properties of concavifications. These results are derived even though in the present model an optimal signal device may formally not exist, due to the natural discontinuity of Receiver's posterior after surprises. This technical difficulty is overcome by considering the value function of Sender's problem.

To finish our example, although a truly optimal π does not exists, the lawyer can *approximately* get an expected refund of \$1.8—without communication it would be \$0. Her investigation has four signals: The first arises if the defendant is innocent, revealing it; The second arises in states \$0 and \$2 and makes the judge assign them equal probability;[9] The third arises in states \$1 and \$2 and makes the lawyer assign probability arbitrarily close to 1 to *\$1*, while the judge assigns probability 1 to *\$2*; The fourth arises in states \$1

[9] These first two signals are reminiscent of the optimal investigation in Kamenica and Gentzkow's (2011) environment.

and \$3 and makes the judge assign them equal probability. So, even though the lawyer can always hide the most likely \$1 damage with the better \$2 one, she reveals it with positive probability! In so doing, she makes the judge less sure when he orders a \$3 refund, but she increases its overall probability. For comparison, the expected refund can at most be \$1.6 if the investigation always hides states \$1 and \$3, and \$1.5 if it never hides states. So, by optimally combining hiding and surprising, the lawyer increases her expected payoff by 12.5% and 20% respectively.

The paper applies the main results to the classic model in which Sender's and Receiver's payoffs are quadratic loss functions and their ideal actions differ for each state. Since Crawford and Sobel (1982), this model has been used to study communication in many contexts.[10] Finally, the paper shows with an example that Sender's superior knowledge need not benefit her: she may be strictly better off if initially Receiver shared her richer theory of the world, rather than viewing some states as impossible. In this case, Sender would like to 'persuade' Receiver to switch theory before communication occurs, but no signal technology allows her to do so.

2 Related Literature

This paper contributes to the literature on games of persuasion and information control. It is the first to study games in which Sender and Receiver disagree on their subjective theory of the possible states of the world. A common question addressed in the literature has been if and when Sender benefits from revealing (expected) information. By contrast, one of this paper's main questions is if and when Sender benefits from surprising Receiver with unexpected information, rather than hiding it. This paper differs from the literature in another important aspect. With common-support priors, Bayes' rule always dictates how Receiver responds to information and represents *the* constraint on what Sender can achieve. As noted, this is no longer true without common-support priors. While moving beyond Bayesian rationality, this paper does not introduce bounded rationality or behavioral features on Receiver's side.[11]

The closest papers in the persuasion literature are the following. In Brocas and Carrillo (2007), Sender has access to a fixed device producing i.i.d. signals about a binary state. She chooses sequentially whether to produce another signal, or stop and let Re-

[10]Examples include organizational design (Dessein (2002); Alonso et al. (2008)), political economy (Grossman and Helpman (2002)), legal-dispute resolution (Goltsman et al. (2009)), lobbying (Kamenica and Gentzkow (2011)), and financial advising (Morgan and Stocken (2003)).

[11]For a study of persuasion with these features see, e.g., Mullainathan et al. (2008).

ceiver act. The paper examines Sender's optimal stopping rule and how much she benefits from controlling information. Kamenica and Gentzkow (2011) study persuasion in more general settings with common priors and no restrictions on signal devices. They show that the problem of choosing a device can be conveniently reformulated as choosing a distribution over posteriors subject to only one constraint imposed by Bayesian rationality: the expected posterior must equal the prior. They also find conditions for Sender to benefit from informative communication, characterize its properties, and examine the effects of changing the alignment of Sender's and Receiver's preferences. Within Kamenica and Gentzkow's (2011) settings, Alonso and Câmara (2013) examine how disagreement in priors (with common support) affects Sender's communication, showing that it increases the scope for benefiting from revealing information. They also show that disagreement does not expand the set of feasible distributions over posteriors. As noted, this set is qualitatively very different without common-support priors.[12]

As in the previous literature (Kamenica and Gentzkow (2011), Alonso and Câmara (2013), Ely (2014)), this paper's analysis relies on the 'concavification method' proposed in the seminal work of Aumann and Maschler (1995) on repeated games with incomplete information. However, it modifies this method in several ways to examine Sender's payoff separately when she confirms and disproves Receiver's theory. This method also turns out to be helpful to characterize Sender's behavior despite the possible non-existence of optimal information devices.

This paper models responses to unexpected information in ways related to several papers in the literature. On the one hand, it borrows Ortoleva's (2012) axiomatic model of "change of paradigm," which provides specific predictions on how Receiver 'chooses' a new prior and updates it after surprises. A key difference is that in Ortoleva (2012) the information structure is exogenous; here it is not, which will require some care in applying his model. On the other hand, the use of lexicographic belief systems (LBS's) is inspired by Kreps and Wilson (1982) and Karni and Vierø (2013).[13] In their work on sequential equilibria, Kreps and Wilson (1982) imagine that each player has a system of "hypotheses" on how the game is played, the primary one corresponding to equilibrium play. A player always applies his primary hypothesis on the equilibrium path. But if an off-path information set is reached—a zero-probability event under the primary hypothesis—the player attempts to apply other hypotheses until one predicts what happened. The player

[12]Other papers study problems of information design, both static and dynamic (see, e.g., Forges and Koessler (2008), Rayo and Segal (2010), Horner and Skrzypacz (2010), Ely et al. (2013), Ely (2014)). All these papers assume common priors and address different questions from the present paper.

[13]LBS's appear also in Blume et al. (1991a, 1991b), but in these papers they work in a fundamentally different way: the agent always takes into account, though lexicographically, all layers of his LBS.

always updates his current hypothesis using Bayes' rule. In Kreps and Wilson (1982), players have a common prior over Nature's moves and the hypotheses are on each others' strategies; here the common-prior assumption is absent, and we can interpret Receiver's hypotheses as being on player Nature.

Karni and Vierø (2013) model how an agent's beliefs evolve when he either discovers that states he considered impossible are actually possible, or becomes aware of new states. They suggest and axiomatize a notion of belief consistency similar to that used in the present paper. A key difference is that here not only Receiver may have to expand his subjective set of possible states and hence form new priors; he also has to update such priors after observing Sender's information.

Finally, this paper is related to the literature on strategic information transmission following Crawford and Sobel (1982), but differs from it in several aspects shared with all previously mentioned papers on persuasion. In that literature Sender learns the state before choosing how to communicate; here he commits to a signal device before the state occurs. This property changes deeply the incentive problems Sender faces and eliminates multiplicity of equilibria. Also, that literature has focussed on settings with common-support priors.

3 Model

The primitives of the model are as in Kamenica and Gentzkow (2011), except of course for the assumptions on priors.

There are two agents, called Sender (she) and Receiver (he). Ω is a finite set of mutually exclusive states of the world. Sender and Receiver have a common language to describe each $\omega \in \Omega$ and agree on the meaning of this description. Each state ω is an exhaustive description of reality. If ω occurs, we say that ω is 'true;' otherwise, we say that ω is 'false.'

At the beginning, neither Sender nor Receiver know the true state ω. Sender has a subjective prior belief σ with support $\mathcal{S} = \Omega$ and Receiver has a subjective prior belief ρ_0 with support $\mathcal{R} \subsetneq \Omega$.[14] Hereafter, let $\overline{\mathcal{R}} = \mathcal{S} \setminus \mathcal{R}$ and $\mathbf{supp}\,\sigma$ denote the support of σ, and adopt this notation for any other probability distribution. We shall call σ Sender's theory of the world, and ρ_0 Receiver's theory.

[14] In Kamenica and Gentzkow (2011), $\mathcal{S} = \mathcal{R}$ and $\sigma = \rho_0$. In Alonso and Câmara (2013), $\mathcal{S} = \mathcal{R}$ but σ may differ from ρ_0. In the present model, $\mathcal{R}$ may differ from $\mathcal{S}$ without $\mathcal{R} \subsetneq \mathcal{S}$. This case, however, comprises all key aspects of the different-support assumption as explained in Section 6.3.

Following the literature on persuasion, Sender and Receiver interact as follows. Sender commits to a signal device to provide Receiver with information on the true state, with the goal of steering his behavior.[15] After observing a signal realization, Receiver chooses an action from the compact set A, with $|A| > 1$, which affects both agents' payoffs. As in Kamenica and Gentzkow (2011), a signal device π is defined by a family of conditional distributions $\{\pi(\cdot|\omega)\}_{\omega\in\Omega}$ over a finite set X_π of signal realizations: for each $\omega \in \Omega$, $\pi(\cdot|\omega) \in \Delta(X_\pi)$ where $X_\pi = \cup_{\omega\in\Omega}\mathbf{supp}\,\pi(\cdot|\omega)$. The set of all signal devices is Π. As in Alonso and Câmara (2013), signal devices are "commonly understood:" Sender and Receiver agree on how π generates signals given each state ω. Hereafter, the term *signal* will refer to a particular realization x of a *device* π; the term *message* will refer to any pair (x, π) with $x \in X_\pi$.

Regarding payoffs, Sender design her device so as to maximize her subjective expected utility with cardinal utility function $u_S : A \times \Omega \to \mathbb{R}$, continuous in a for every ω. For Receiver, let his utility function be $u_R : A \times \Omega \to \mathbb{R}$, again continuous in a for every ω. Receiver's behavior after every signal will be specified below.[16]

The paper's goal is to examine how Sender communicates with Receiver in this environment.

Interpretation of Priors' Different Supports

Many reasons can explain why Sender and Receiver have different priors.[17] A natural—certainly not unique—one is that Sender is an expert with a more complete and accurate understanding of the world and consequently her theory is the 'correct' one. But this is not necessary: we can view the entire analysis as from the *subjective* ex-ante perspective of Sender. That is, the paper examines how Sender communicates if she thinks that Receiver has a different theory ρ_0 from her σ and responds to signals as described below.

The mathematical property that the supports of Sender's and Receiver's priors differ can be interpreted in several ways. For the sake of clarity, most of the paper will focus on one: Receiver is aware of all states in $\mathcal{S}$ and views them as well-specified hypotheses, but simply thinks that some states are impossible based on his theory of the world. For instance, the statement "the Earth goes around the Sun" was formally correct and

[15]For an extensive discussion and justification of the commitment assumption, see Kamenica and Gentzkow (2011). Modifying this assumption is beyond the scope of the present paper.

[16]The functions u_S and u_R should be interpreted as reduced forms in an Anscombe and Aumann (1963) setting, where subjective beliefs are well defined. Letting $\mathcal{G} = \{g_a\}_{a\in A}$ be a set of Anscombe and Aumann's acts where A is a set of 'labels,' we have $u_i(a, \omega) = \hat{u}_i(g_a(\omega))$ for $i = R, S$.

[17]See also the discussion in Morris (1995).

scientists could understand its meaning even before the Copernican Revolution; yet they deemed such a statement impossible according to their theories. In the more recent debate on climate change, deniers and believers in manmade global warming understand each other's theories; yet each side views the other's theory as impossible. Receiver may assign zero probability to some states for different reasons. Some may be exogenous and subjective, like religion in the previous examples. Sender's and Receiver's theories may be logically inconsistent, so that they cannot have the same set of possible predictions. In some cases, Receiver may frame the situation at hand in some narrow way that leads him to ignore some states.[18] Finally, Receiver may be boundedly rational or face high cognitive costs; as a result, he may be able to reason only in terms of simple theories, which again ignore some states.

Before observing any signal, Receiver has no doubt on his theory in the following sense. If he assigned a positive probability to another theory which incorporates states outside $\mathcal{R}$ as possible, then a correct probabilistic description of his overall theory should include these additional states in its support. Hence, if $\mathcal{R} \neq \mathcal{S}$, it means that ex ante Receiver is fully confident that the states outside $\mathcal{R}$ are impossible—even though he may know that Sender has a different opinion. Nonetheless, of course Receiver will have to abandon his theory if a signal unambiguously proves that the true state is outside $\mathcal{R}$.

Another possible interpretation of the model is that Receiver's prior assigns zero probability to states outside $\mathcal{R}$ because he is unaware of them. Though perhaps natural, this interpretation is more delicate and requires careful explanation in the present environment. We shall defer its discussion until Section 6.1.

4 Feasible Distributions over Posteriors

To understand how Sender communicates with Receiver, we first need to answer the following question: which Receiver's probabilistic assessments over Ω can Sender induce using signal devices in Π?

This question does not have an immediate answer in our environment. If ρ_0's support were Ω, we would usually invoke Bayesian rationality and combine ρ_0 with any π using Bayes' rule to get a unique posterior of Receiver for each signal. Here, however, Receiver may endogenously observe unexpected messages, evidence to which he assigns zero probability ex ante. In this case Bayes' rule does not apply. So one possibility is to say that

[18] Ahn and Ergin (2010), for instance, develop a model of framing in which an agent can 'overlook' some events.

after surprises Receiver can have any belief; but this approach would be unsatisfactory. We would like his posteriors to feature some regularity and predictability and to take into account properties of π even after unexpected signals. For instance, a posterior should assign zero probability to states that cannot lead to message (x, π).

To this end, we shall first consider Ortoleva's (2012) axiomatically founded model of 'change of paradigm.' Section 4.2 answers our main question within this model. Section 6.2 will consider other models of how Receiver responds to signals and shows that the paper's main conclusions continue to hold.

4.1 A Model of Receiver's Response to Information

Hereafter, we say that message (x, π) *confirms* ρ_0 if Receiver assigns positive probability to observing x from π given ρ_0; otherwise, (x, π) *disproves* ρ_0. Let C_π be the set of signals such that (x, π) confirms ρ_0 and D_π the set of signals such that (x, π) disproves ρ_0:[19]

$$C_\pi = \{x \in X_\pi : \pi(x|\omega) > 0 \text{ for some } \omega \in \mathcal{R}\}, \quad \text{and} \quad D_\pi = X_\pi \setminus C_\pi.$$

Note that C_π is always nonempty, but D_π may be empty.

According to Ortoleva's (2012) model of 'change of paradigm,' Receiver responds to Sender's messages as follows. Receiver has a prior over priors $\mu \in \Delta(\Delta(\Omega))$ with finite support. Initially, he adopts the prior with the highest likelihood under μ as his theory of the world—i.e., ρ_0. After expected messages, he updates ρ_0 using Bayes' rule. But after unexpected messages, he first updates μ obtaining $\hat{\mu}$, adopts prior $\hat{\rho}$ with the highest likelihood under $\hat{\mu}$, and then again updates $\hat{\rho}$ using Bayes' rule. Formally, let μ be such that for every $\omega \in \Omega$ there exists $\rho \in \mathbf{supp}\,\mu$ with $\rho(\omega) > 0$. Then, given message (x, π), let the probability the updated prior over priors assigns to ρ for every $\rho \in \Delta(\Omega)$ be

$$\hat{\mu}(\rho|x, \pi) = \frac{\left[\sum_{\omega \in \Omega} \pi(x|\omega)\rho(\omega)\right] \mu(\rho)}{\sum_{\tilde{\rho} \in \mathbf{supp}\,\mu} \left[\sum_{\omega \in \Omega} \pi(x|\omega)\tilde{\rho}(\omega)\right] \mu(\tilde{\rho})}. \tag{1}$$

Also, let

$$M = \arg\max_{\rho} \mu(\rho) \quad \text{and} \quad M(x, \pi) = \arg\max_{\rho} \hat{\mu}(\rho|x, \pi).$$

In general, M and $M(x, \pi)$ need not be singletons. Indeterminacies in Receiver's 'choice' of a prior, however, would make his behavior ill defined. Following Ortoleva (2012), we

[19] Both sets depend on ρ_0, but this dependence is left implicit to simplify notation. Also, recall that X_π includes only signals in the support of $\pi(\cdot|\omega)$ for some ω.

endow Receiver with a strict linear order $\succ$ over $\Delta(\Omega)$ which he uses to 'choose' a prior when the maximum-likelihood criterion is inconclusive.[20] For simplicity, assume that M is singleton.

Assumption 1 (A1: Hypothesis-Testing Model (Ortoleva (2012))). *Receiver has a prior over priors $\mu \in \Delta(\Delta(\Omega))$ with finite support and such that for every $\omega \in \Omega$ there exists $\rho \in \mathbf{supp}\,\mu$ with $\rho(\omega) > 0$ and $M = \{\rho_0\}$. Also, (c) if $x \in C_\pi$, Receiver updates ρ_0 using Bayes' rule; (d) if $x \in D_\pi$, he updates ρ using Bayes' rule where ρ is $\succ$-maximal in $M(x, \pi)$.*

A1 corresponds to Ortoleva's (2012) Hypothesis-Testing Representation with $\varepsilon = 0$, which is consistent with our definition of C_π. As a simple example of A1, suppose that $\mathbf{supp}\,\mu = (\rho_0, \rho_1)$ with $\mu(\rho_0) > \frac{1}{2}$ and $\mathbf{supp}\,\rho_1 = \Omega$. Then, for $x \in C_\pi$ Receiver uses his primary prior ρ_0, but for $x \in D_\pi$ he switches to his secondary prior ρ_1. In both cases, Receiver updates it using Bayes' rule.

A few remarks on part (c) of A1 are in order. If message (x, π) confirms ρ_0, it gives Receiver no objective reason to doubt his theory; therefore he does not abandon it. A1(c) holds if and only if Receiver's behavior is dynamically consistent (Theorem 1 in Ortoleva (2012)). It is also in the spirit of the notion of Perfect Bayesian Equilibrium and its refinements, which require that Bayes' rule hold whenever possible.[21] It is consistent with the interpretation that, at the outset, Receiver thinks that his theory is correct. It is also consistent with the property that signal devices, per se, contain no information; so Receiver should respond only to the evidence given by signal *realizations*. Finally, it is consistent with a phenomenon known in the psychology literature as 'confirmatory bias:' when presented with *inconclusive* evidence, people tend to interpret it in favor of their initial hypothesis (see, e.g., Rabin and Schrag (1999) and references therein).[22]

4.2 Characterization of Distributions over Posteriors

As usual, Bayesian rationality implies a tight link between each signal device and a distribution over *Sender's* posteriors. Studying distributions over posteriors is helpful, for

[20]Since in our setting Receiver's prior over priors μ is predetermined while the evidence he observes is endogenous, it is logically impossible to construct μ so that the maximum-likelihood criterion always gives a unique answer as in Ortoleva (2012).

[21]We can always think of a third player, Nature, who chooses ω at the beginning of the game and has flat preferences.

[22]Ortoleva's (2012) representation allows for the possibility that Receiver views (x, π) as disproving ρ_0 if the probability of observing x from π given ρ_0 is less than some $\varepsilon > 0$. This violates the usual dynamic-consistency condition which characterizes Bayesian updating. Studying such violations in persuasion games is clearly orthogonal to and beyond this paper's scope: they may arise even if $\rho_0 = \sigma$.

it focuses attention on the heart of information revelation and sets aside possibly fragile aspects that may depend on irrelevant details of signal devices (like language or frames). Previous papers have shown that in settings with common-support priors Receiver's and Sender's posteriors induced by any π are always in a one-to-one relationship.[23] As we will see, this property does not hold without common-support priors, opening the door to new, specific properties of the distributions over posteriors.

Any π induces a distribution over Sender's posteriors as follows. By Bayes' rule, given (x, π) the probability she assigns to ω satisfies

$$q(\omega|x, \pi) = \frac{\pi(x|\omega)\sigma(\omega)}{\sum_{\omega' \in \mathcal{S}} \pi(x|\omega')\sigma(\omega')}. \tag{2}$$

Applying (2) across x's delivers a distribution τ with finite support over Sender's posteriors, where the probability that τ assigns to any $q \in \Delta(\mathcal{S})$ is

$$\tau(q) = \sum_{\{x: q(\cdot|x,\pi)=q\}} \sum_{\omega \in \mathcal{S}} \pi(x|\omega)\sigma(\omega).$$

As usual, we must have $\sum_{q \in \mathbf{supp}\, \tau} q\tau(q) = \sigma$ for any $\tau \in \Delta(\Delta(\mathcal{S}))$ induced by some $\pi \in \Pi$.

Definition 1. Fix Sender's prior σ. A distribution $\tau \in \Delta(\Delta(\mathcal{S}))$ is *feasible* if and only if $\mathbf{supp}\, \tau$ is finite and $\sum_{q \in \mathbf{supp}\, \tau} q\tau(q) = \sigma$. The set of feasible distributions is $\mathcal{F}_\sigma$.

Definition 1 is equivalent to Kamenica and Gentzkow's (2011) notion of Bayes plausibility. As they showed, for any $\tau \in \mathcal{F}_\sigma$ there exists $\pi \in \Pi$ that induces τ through Bayes' rule.

This is a useful result. But if we know only Sender's posterior—not the message that led to it—can we infer whether Receiver's prior has been confirmed or disproved? Can we always recover his posterior? We will show that the answer to both questions is yes.

Supports of Sender's and Receiver's Posteriors

It is useful to first consider the support of Sender's posterior induced by message (x, π)—and hence in the support of some $\tau \in \mathcal{F}_\sigma$. Given $q(\cdot|x, \pi)$ in (2), its support reveals whether x belongs to C_π or D_π. Indeed, $\mathbf{supp}\, q(\cdot|x, \pi)$ coincides with the set of states that are consistent with (x, π), defined by

$$\Omega_\pi(x) = \{\omega \in \Omega : \pi(x|\omega) > 0\}. \tag{3}$$

[23] See Kamenica and Gentzkow (2011) and Alonso and Câmara (2013).

Therefore,

$$x \in C_\pi \Leftrightarrow \mathbf{supp}\, q(\cdot|x,\pi) \cap \mathcal{R} \neq \varnothing.$$

Of course, then $x \in D_\pi$ if and only if $\mathbf{supp}\, q(\cdot|x,\pi) \subset \overline{\mathcal{R}}$. So, for any $\tau \in \mathcal{F}_\sigma$, the support of any Sender's posterior q arising under τ allows us to tell whether ρ_0 has been confirmed or disproved, without even knowing the message that led to q.

We can then classify Sender's posteriors as follows. Given the whole set $\Delta(\mathcal{S})$, let

$$\Delta^d = \Delta(\overline{\mathcal{R}}) \quad \text{and} \quad \Delta^c = \Delta(\mathcal{S}) \setminus \Delta^d.$$

Abusing terminology, we say that q *confirms* ρ_0 if $q \in \Delta^c$, otherwise q *disproves* ρ_0.[24] For any $\tau \in \mathcal{F}_\sigma$, then define

$$D_\tau = \{q \in \mathbf{supp}\, \tau : q \in \Delta^d\} \quad \text{and} \quad C_\tau = \mathbf{supp}\, \tau \setminus D_\tau.$$

The next preliminary result links the supports of Sender's and Receiver's posteriors.[25]

Proposition 1. *Given any (x,π), let $p(\cdot|x,\pi)$ be Receiver's posterior. Under A1, if $q(\cdot|x,\pi) \in \Delta^c$, then $\mathbf{supp}\, p(\cdot|x,\pi) \subset \mathcal{R}$; if $q(\cdot|x,\pi) \in \Delta^d$, then $\mathbf{supp}\, p(\cdot|x,\pi) \subset \overline{\mathcal{R}}$.*

Proposition 1 highlights a constraint on the posteriors of Receiver that Sender can induce in settings with different-support priors, which does not arise with common-support priors.[26] Here Sender cannot make Receiver assign positive probability to both elements in $\mathcal{R}$ and in $\overline{\mathcal{R}}$. She faces a strict decision: either confirm ρ_0, in which case Receiver will assign positive probability only to states in $\mathcal{R}$, or disprove ρ_0, which requires showing that *all* states in $\mathcal{R}$ are false.

Proposition 1 also restricts the probability Sender assigns ex ante to confirming and disproving ρ_0. For any τ, these are $\tau^c = \tau(C_\tau)$ and $\tau^d = \tau(D_\tau)$.

Corollary 1. *For any $\tau \in \mathcal{F}_\sigma$, the probability τ^c of confirming ρ_0 is at least $\sigma(\mathcal{R}) > 0$.*

However Sender decides to disclose information, there is always a strictly positive probability that Receiver will not abandon his theory of the world. From Sender's viewpoint, the largest probability of disproving ρ_0 is $\sigma(\overline{\mathcal{R}}) < 1$; this happens if π is fully revealing, for instance. By contrast, Sender can always design a π which confirms ρ_0 with

[24]Since $\overline{\mathcal{R}} \subsetneq \mathcal{S}$, Δ^d is a strictly lower dimensional subset of $\Delta(\mathcal{S})$ and $\Delta^d \cap int(\Delta(\mathcal{S})) = \varnothing$. Δ^d is also a face of the convex set $\Delta(\mathcal{S})$ which lies on its (relative) boundary.

[25]All proofs are in Appendix B.

[26]See Kamenica and Gentzkow (2011) and Alonso and Câmara (2013).

probability 1—by ensuring that, for every $\omega \in \overline{\mathcal{R}}$, $\textbf{supp}\,\pi(\cdot|\omega) \subset \textbf{supp}\,\pi(\cdot|\omega')$ for some $\omega' \in \mathcal{R}$. This feature highlights an asymmetry between confirming an 'incorrect' theory and disproving it by giving information on its predictions (as opposed to logical arguments uncovering its contradictory predictions, for instance).

Relationship Between Sender's and Receiver's Posteriors

Given any message (x, π), we need to know only Sender's posterior $q = q(\cdot|x, \pi)$ to know which prior Receiver updates. Of course, if q confirms ρ_0 (i.e., $q \in \Delta^c$), then Receiver updates ρ_0. If q disproves ρ_0 (i.e., $q \in \Delta^d$), even without knowing (x, π), we can recover how Receiver 'changes his paradigm' after observing x and selects a new prior ρ'. To see this, note that how he updates his prior over priors μ depends on the total probability each $\rho' \in \textbf{supp}\,\mu$ assigns to x under π (see (1)). These probabilities can be uniquely inferred from Sender's posterior $q(\cdot|x, \pi)$: using (2), we can write

$$\pi(x|\omega)\rho'(\omega) = q(\omega|x, \pi)\frac{\rho'(\omega)}{\sigma(\omega)}\left[\sum_{\tilde{\omega}\in\Omega}\pi(x|\tilde{\omega})\sigma(\tilde{\omega})\right].$$

Therefore, for every $\rho' \in \textbf{supp}\,\mu$, expression (1) becomes

$$\hat{\mu}(\rho'|x, \pi) = \frac{\left[\sum_{\omega\in\Omega} q(\omega|x, \pi)\frac{\rho'(\omega)}{\sigma(\omega)}\right]\mu(\rho')}{\sum_{\tilde{\rho}\in\textbf{supp}\,\mu}\left[\sum_{\omega\in\Omega} q(\omega|x, \pi)\frac{\tilde{\rho}(\omega)}{\sigma(\omega)}\right]\mu(\tilde{\rho})}. \tag{4}$$

Given $q(\cdot|x, \pi)$, this expression depends neither on x nor on π. Hence, let $\hat{\mu}(\cdot; q)$ be Receiver's updated prior over priors given Sender's posterior q and

$$M(q) = \arg\max_{\rho'}\ \hat{\mu}(\rho'; q).$$

By A1, Receiver picks the unique prior in $M(q)$ which is maximal according to the strict order $\succ$. Thus, define the following function over the set of Sender's posteriors $\Delta(\mathcal{S})$:

$$\rho(q) = \begin{cases} \rho_0 & \text{if } q \in \Delta^c \\ \rho' \text{ s.t. } \rho' \succ \tilde{\rho} \text{ for all } \tilde{\rho} \in M(q),\ \tilde{\rho} \neq \rho' & \text{if } q \in \Delta^d \end{cases}. \tag{5}$$

Once we know which prior Receiver updates, we can conclude that Sender's posterior also tells us everything about Receiver's posterior. The argument generalizes that in

Alonso and Câmara (2013).[27] Note that, since by assumption for every $\omega \in \Omega$ there is a $\rho' \in \mathbf{supp}\, \mu$ with $\rho'(\omega) > 0$, for every $q \in \Delta(\mathcal{S})$ the prior $\rho(q)$ in (5) must assign positive probability to observing the (x, π) that induced q; that is, $\sum_{\omega \in \Omega} \pi(x|\omega)\rho(\omega; q) > 0$.

Lemma 1. *Given any message (x, π), let $q(\cdot|x, \pi)$ be Sender's posterior, $\rho \in \Delta(\Omega)$ be such that $\sum_{\omega \in \Omega} \pi(x|\omega)\rho(\omega) > 0$, and $p(\cdot|x, \pi)$ be Receiver's posterior after updating ρ. Then, for all $\omega \in \Omega$,*

$$p(\omega|x, \pi) = \frac{q(\omega|x, \pi)\frac{\rho(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega'|x, \pi)\frac{\rho(\omega')}{\sigma(\omega')}}. \tag{6}$$

If $\rho(\omega) = 0$, expression (6) is well-defined and $p(\omega|x, \pi) = 0$, even though $q(\omega|x, \pi)$ may be positive. Moreover, if (x, π) and (y, π) are such that $q(\cdot|x, \pi) = q(\cdot|y, \pi)$, then $p(\cdot|x, \pi) = p(\cdot|y, \pi)$.

These observations lead to the first main result of the paper: a full characterization of the set of feasible joint distributions over posteriors under A1. Even though here Receiver's responses to information is more complicated than in models with common-support priors, we can describe his posteriors only in terms of Sender's posteriors and without reference to signal devices.

Proposition 2. *Consider any $\tau \in \mathcal{F}_\sigma$. Under A1, for every $q \in \mathbf{supp}\, \tau$ and $\omega \in \Omega$, Receiver's posterior satisfies*

$$p(\omega; q) = \frac{q(\omega)\frac{\rho(\omega;q)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega')\frac{\rho(\omega';q)}{\sigma(\omega')}}, \tag{7}$$

where $\rho(q)$ is given in (5).

Proposition 2 implies several key properties for the rest of the analysis:

- When Sender's posterior q confirms ρ_0, Receiver's posterior does not depend on the probability q assigns to states in $\overline{\mathcal{R}}$. Every $q \in \Delta^c$ assigns positive probability to the event $\mathcal{R}$, as $\mathbf{supp}\, q \cap \mathcal{R} \neq \varnothing$; moreover, $\rho_0(\omega) = 0$ for $\omega \in \overline{\mathcal{R}}$. So

$$p(\omega; q) = \frac{\frac{q(\omega)}{q(\mathcal{R})}\frac{\rho_0(\omega)}{\sigma(\omega)/\sigma(\mathcal{R})}}{\sum_{\omega' \in \mathcal{R}} \frac{q(\omega')}{q(\mathcal{R})}\frac{\rho_0(\omega')}{\sigma(\omega')/\sigma(\mathcal{R})}} = \frac{q(\omega|\mathcal{R})\frac{\rho_0(\omega)}{\sigma(\omega|\mathcal{R})}}{\sum_{\omega' \in \mathcal{R}} q(\omega'|\mathcal{R})\frac{\rho_0(\omega')}{\sigma(\omega'|\mathcal{R})}}. \tag{8}$$

  Given $q \in \Delta^c$, Receiver has the same posterior he would have in a world in which priors were ρ_0 and $\sigma(\cdot|\mathcal{R})$ and Sender had posterior $q(\cdot|\mathcal{R})$.

[27] Though immediate, the steps of this generalization are reproduced in Appendix B to make the paper self contained.

- Conditional on confirming ρ_0 and inducing posterior p, Sender can have any posterior over $\overline{\mathcal{R}}$. Hence we lose the one-to-one relationship between Sender's and Receiver's posteriors which characterizes models with common-support priors. To see this, take any $q' \in \Delta^c$ and $q'' \in \Delta(\overline{\mathcal{R}})$. For every $\omega \in \mathcal{S}$, let $\hat{q}(\omega) = \alpha q'(\omega|\mathcal{R}) + (1-\alpha)q''(\omega)$ with $\alpha \in (0,1)$. Clearly, $\hat{q} \in \Delta(\mathcal{S})$. It is easy to see that $\hat{q}(\cdot|\mathcal{R}) = q'(\cdot|\mathcal{R})$ and $\hat{q}(\cdot|\overline{\mathcal{R}}) = q''$.

- Since there are no restrictions on π, Sender can induce any posterior $q(\cdot|\mathcal{R})$ and Receiver's corresponding posterior under (8). Indeed, any $q \in \Delta^c$ must arise from a signal $x \in C_\pi$ for some device π, i.e., $q = q(\cdot|x,\pi)$. Then, for any $\omega \in \mathcal{R}$,

$$q(\omega|\mathcal{R}) = q(\omega|(x,\pi),\mathcal{R}) = \frac{\pi(x|\omega)\sigma(\omega|\mathcal{R})}{\sum_{\omega'\in\mathcal{R}}\pi(x|\omega')\sigma(\omega'|\mathcal{R})},$$

  which depends only on $\pi(\cdot|\omega)$ for $\omega \in \mathcal{R}$.

Analogous properties hold when Sender disproves ρ_0. For instance, let $\mathbf{supp}\,\mu = (\rho_0, \rho_1)$ with $\mu(\rho_0) > \frac{1}{2}$. Then, for any $q \in \Delta^d$ Receiver updates ρ_1 and expression (7) can be written as

$$p(\omega; q) = \frac{q(\omega)\frac{\rho_1(\omega|\overline{\mathcal{R}})}{\sigma(\omega|\overline{\mathcal{R}})}}{\sum_{\omega'\in\overline{\mathcal{R}}} q(\omega')\frac{\rho_1(\omega'|\overline{\mathcal{R}})}{\sigma(\omega'|\overline{\mathcal{R}})}}. \tag{9}$$

So Receiver has the same posterior he would have in a world in which priors were $\rho_1(\cdot|\overline{\mathcal{R}})$ and $\sigma(\cdot|\overline{\mathcal{R}})$ and Sender had posterior q. Moreover, Sender can again induce any posterior $q \in \Delta^d$ and Receiver's corresponding posterior under (9). Indeed, any $q \in \Delta^d$ must arise from a signal $x \in D_\pi$ for some π. That is, $q = q(\cdot|x,\pi)$ so that

$$q(\omega|x,\pi) = \frac{\pi(x|\omega)\sigma(\omega)}{\sum_{\omega'\in\overline{\mathcal{R}}}\pi(x|\omega')\sigma(\omega')}.$$

The claim follows because Sender can construct any π with $\pi(x|\omega) = 0$ if $\omega \in \mathcal{R}$.

Further distinctive features of settings without common-support priors come to light by examining how Receiver's posterior varies as a function of Sender's. By Proposition 2, $p(q)$ varies continuously in q over Δ^c. However, it always changes discontinuously as q moves from Δ^c to Δ^d; this discontinuity captures formally the idea of surprise.

Corollary 2. *For any $q \in \Delta^c$ and $q' \in \Delta^d$, we have $||p(q) - p(q')|| > ||p(q')|| > 0$.*

By contrast, in Kamenica and Gentzkow (2011) and Alonso and Câmara (2013) Receiver's posterior is always a continuous function over $\Delta(\mathcal{S})$.

When Sender disproves ρ_0 (i.e., over Δ^d), the continuity of Receiver's posterior in Sender's depends on his prior over priors μ. As can be easily checked, Receiver's posterior is continuous over Δ^d if $\mathbf{supp}\,\mu = (\rho_0, \rho_1)$ with $\mu(\rho_0) > \frac{1}{2}$, for example. Appendix A provides an example with discontinuities over Δ^d. Intuitively, *for a given prior* Receiver's updating through Bayes' rule varies continuously as q varies in Δ^d, but his choice of a prior may vary discontinuously when the maximum-likelihood criterion is inconclusive.[28] This kind of discontinuity, however, is different from that highlighted in Corollary 2. Since after unexpected evidence Receiver must abandon ρ_0, a discrete difference in posteriors with and without surprise is necessarily implied by the very nature of being surprised.

5 Optimal Communication

Building on the previous results, this section examines the signal devices that Sender chooses.

We first need to specify how Receiver chooses actions. Assume that given posterior p Receiver chooses an action that maximizes the resulting expected utility:[29]

$$a(p) \in \mathcal{A}(p) = \arg\max_{a \in A} \sum_{\omega \in \Omega} u_R(a, \omega) p(\omega).$$

In general, given p, Receiver can be indifferent among multiple actions.

Assumption 2. *If $\mathcal{A}(p)$ contains multiple actions, Sender can recommend Receiver to choose any $a \in \mathcal{A}(p)$ and Receiver follows the recommendation.*

Assumption 2 corresponds to Kamenica and Gentzkow's (2011) 'Sender-preferred' subgame-perfect equilibrium and also appears in Alonso and Câmara (2013). In their settings with common-support priors, it ensures that an optimal signal device always exists.

Despite this standard assumption, due to the natural discontinuity in Receiver's posteriors after surprises, here optimal signal devices may not exist. To see why, given posteriors q and p let Sender's expected payoff be

$$v(q, p) = \max_{a \in \mathcal{A}(p)} \sum_{\omega \in \Omega} u_S(a, \omega) q(\omega).$$

[28] The function $\hat{\mu}(\cdot)$ in (4) is continuous in q and hence $M(\cdot)$ is upper hemicontinuous. This is insufficient to conclude that the function $\rho(\cdot)$ defined in Proposition 2 is continuous on Δ^d.

[29] In Ortoleva (2012), SEU maximization is part of the Hypothesis-Testing Representation of the decision-maker's behavior.

Given any mapping from Sender's to Receiver's posterior $\hat{p} : \Delta(\mathcal{S}) \to \Delta(\Omega)$, define

$$w(q) = v(q, \hat{p}(q)) \quad \text{for all } q \in \Delta(\mathcal{S}). \tag{10}$$

As in Kamenica and Gentzkow (2011) and Alonso and Câmara (2013), $v(\cdot, \cdot)$ is upper semicontinuous.[30] Moreover, since in those papers $\hat{p}(\cdot)$ is continuous, $w(\cdot)$ is upper semicontinuous as well. By contrast, here $w(\cdot)$ need not be upper semicontinuous, because $p(\cdot)$ in Proposition 2 (and 6 below) is discontinuous.

Example 1 (Non-existence of Optimal Communication). Let $\mathcal{S} = \{\omega_1, \omega_2\}$ with $\sigma = (\frac{1}{2}, \frac{1}{2})$, $\mathcal{R} = \{\omega_1\}$, and $A = \{a, b, c\}$. Sender's and Receiver's utility functions are as follows:

$u_S(\cdot,\cdot)$	ω_1	ω_2
a	1	0
b	0	1
c	-1	-1

$u_R(\cdot,\cdot)$	ω_1	ω_2
a	1	0
b	1	0
c	1	1

For Sender, a is optimal if $q(\omega_1) \geq \frac{1}{2}$ and b is optimal if $q(\omega_1) \leq \frac{1}{2}$. For any $q \neq (0, 1)$, $p(q) = (1, 0)$ and $\mathcal{A}((1, 0)) = A$. Therefore, Sender can make Receiver choose a or b depending on her posterior q. For $q = (0, 1)$, $p(q) = (0, 1)$ and $\mathcal{A}((0, 1)) = \{c\}$.

Consider device $\pi \in \Pi$ with $X_\pi = \{x_1, x_2\}$ and

$$\pi(\hat{x}|\omega_1) = \begin{cases} 1-\varepsilon & \text{if } \hat{x} = x_1 \\ \varepsilon & \text{if } \hat{x} = x_2 \end{cases} ; \quad \pi(\hat{x}|\omega_2) = \begin{cases} 1-\varepsilon & \text{if } \hat{x} = x_2 \\ \varepsilon & \text{if } \hat{x} = x_1 \end{cases} .$$

So $q(\cdot|x_1, \pi) = (1-\varepsilon, \varepsilon)$ and $q(\cdot|x_2, \pi) = (\varepsilon, 1-\varepsilon)$, each arising with probability $\frac{1}{2}$. For any $\varepsilon > 0$, Sender's expected payoff is then

$$\frac{1}{2}w(q(\cdot|x_1, \pi)) + \frac{1}{2}w(q(\cdot|x_2, \pi)) = \frac{1-\varepsilon}{2} + \frac{1-\varepsilon}{2} = 1 - \varepsilon, \tag{11}$$

However, for $\varepsilon = 0$, we have

$$\frac{1}{2}w((1, 0)) + \frac{1}{2}w((0, 1)) = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0.$$

Clearly, the supremum of Sender's expected payoff over Π is 1, which is also the maximum she can hope for. But no π can achieve 1. This would require that Sender learn the true

[30] This is because $\mathcal{A}(p)$ is a nonempty, compact-valued, upper-hemicontinuous correspondence by Berge's Maximum Theorem.

state to make Receiver choose her preferred action accordingly; but then Receiver must also learn the true state and hence will choose c in ω_2.

Importantly, these technical difficulties do not preclude an informative analysis of how Sender communicates. Under our assumptions, we can view Receiver as a 'machine' computing posteriors and associated optimal actions. Sender's problem is then essentially a single-agent decision problem. Although this problem may not have an exact solution, we can examine its *value function* and infer from it properties of signal devices that are *virtually* optimal. For instance, in example 1 expression (11) says that it is virtually optimal for Sender to adopt a π that allows her to almost perfectly learn ω but never disproves ρ_0. To formalize this approach, denote Sender's expected payoff from any τ by

$$V(\tau) = \sum_q w(q)\tau(q),$$

where $w(\cdot)$ in (10) uses the function $p(\cdot)$ in Proposition 2 (or 6 below). Sender would like to reach the highest $V(\tau)$ by choosing $\tau \in \mathcal{F}_\sigma$. This problem's value function is[31]

$$W_\sigma = \sup_{\tau \in \mathcal{F}_\sigma} V(\tau). \tag{12}$$

The key observation here is that the usual 'concavification' procedure[32] is well defined even if the function w in (10) is not upper semicontinuous.

Definition 2 (Concavification)**.** Let $g : E \subset \mathbb{R}^n \to \mathbb{R}$. The concavification $\hat{g}$ of g is the lowest concave function with $\hat{g}(e) \geq g(e)$ for all $e \in E$; it satisfies

$$\hat{g}(e) = \sup\{\xi : (e, \xi) \in \mathrm{co}(\mathrm{hyp}\, g)\} \quad \text{for all } e \in E,$$

where $\mathrm{co}(\mathrm{hyp}\, g)$ is the convex hull of the hypograph[33] of g (Rockafellar (1997), p. 52).[34]

Using the concavification of w, we obtain the following.

Lemma 2. $W_\sigma = \hat{w}(\sigma)$. Moreover, in (12), it is without loss of generality to restrict attention to distributions τ with $|\mathbf{supp}\, \tau| \leq |\mathcal{S}|$.

[31]By characterizing W_σ, we will also obtain useful information on ε-optimal τ's for any $\varepsilon > 0$, since there always exists a feasible τ with $V(\tau) \geq W_\sigma - \varepsilon$.

[32]See Section 2 for references.

[33]Given a function $g : \mathbb{R}^n \to \mathbb{R}$, its hypograph is defined as $\mathrm{hyp}\, g = \{(e, r) \in \mathbb{R}^{n+1} : r \leq g(e)\}$.

[34]In example 1, $\hat{w}(q) = 1$ for all $q \neq (0, 1)$ and $\hat{w}((0, 1)) = 0$.

By definition $\hat{w}(\sigma) \geq w(\sigma)$. If $\hat{w}(\sigma) = w(\sigma)$, Sender does not benefit from any informative π and she can clearly achieve W_σ. If $\hat{w}(\sigma) > w(\sigma)$, there exist feasible τ's that correspond to informative π's and satisfy $V(\tau) > w(\sigma)$. We therefore adopt the following terminology.

Definition 3. If $\hat{w}(\sigma) = w(\sigma)$, it is optimal for Sender not to reveal information. If $\hat{w}(\sigma) > w(\sigma)$, it is (virtually) optimal for Sender to reveal information.

Definition 3 refers to Sender's viewpoint, as some π may reveal information according to her theory of the world, σ, but no information according to Receiver's theory, ρ_0.[35]

In this paper the main question is not whether Sender reveals any information at all, but if and when she disproves Receiver's prior and how she optimally communicates overall. Clearly, if $\hat{w}(\sigma) = w(\sigma)$, for Sender it is optimal to never disprove ρ_0 and to communicate nothing. So hereafter we focus on the case with $\hat{w}(\sigma) > w(\sigma)$.[36]

To address our questions, we need a different concavification argument, which considers the function w over Δ^c and Δ^d *separately.* Let w^c and w^d be the restrictions of w to Δ^c and Δ^d and $\hat{w}^c$ and $\hat{w}^d$ their concavifications (Definition 2). For $i = c, d$, $\hat{w}^i(q) \leq \hat{w}(q)$ for all $q \in \Delta^i$ since $\text{co}(\text{hyp}\, w^i) \subset \text{co}(\text{hyp}\, w)$; moreover, $\lim_{q' \to q} \hat{w}^i(q') \geq \hat{w}^i(q)$ for all $q \in \Delta^i$.[37] As shown later, $\hat{w}^d$ is continuous if w^d is continuous, which holds when $p(\cdot)$ is continuous over Δ^d.

Lemma 3. *If $\hat{w}^c(\sigma) < \hat{w}(\sigma)$, there exists $\tau \in \mathcal{F}_\sigma$ such that $V(\tau) > \hat{w}^c(\sigma)$. Moreover, if $V(\tau) > \hat{w}^c(\sigma)$, then the probability τ^d of disproving ρ_0 is strictly positive.*

This result suggests the following terminology.

Definition 4. If $\hat{w}^c(\sigma) = \hat{w}(\sigma)$, it is (virtually) optimal for Sender to always confirm ρ_0. If $\hat{w}^c(\sigma) < \hat{w}(\sigma)$, it is (virtually) optimal for Sender to disprove ρ_0.

---

[35] In example 1, $w(\sigma) = \frac{1}{2}$, $W_\sigma = 1$, and Sender can approximate W_σ using a feasible τ with $\textbf{supp}\, \tau = \{(\varepsilon, 1-\varepsilon), (1, 0)\}$ for ε arbitrarily small. With this τ, however, Receiver's posteriors are $p((1,0)) = p((\varepsilon, 1-\varepsilon)) = (1,0) = \rho_0$. So, from Receiver's viewpoint, no information is revealed.

[36] As shown in Corollary 3 and Proposition 5 below, conditional on $\mathcal{R}$ Sender communicates as in the standard settings of Kamenica and Gentzkow (2011) and Alonso and Câmara (2013). So the sufficient conditions for Sender to optimally reveal some information in those settings can be adapted to ensure that $\hat{w}(\sigma) > w(\sigma)$.

[37] Being concave, $\hat{w}^c$ is continuous at every $q \in int\Delta^c$ by Theorem 10.1 in Rockafellar (1997). By Theorem 10.3 in Rockafellar (1997), there exists only one way to extend $\hat{w}^c$ from $int\Delta^c$ to a continuous finite convex function on $\Delta(\mathcal{S})$. In fact, this extension equals $-\text{cl}(-\hat{w}^c)$ on $\Delta(\mathcal{S})$ where $\text{cl}(-\hat{w}^c)$ is the closure of the convex function $-\hat{w}^c$ (see the proof of Theorem 10.3 and p. 52 of Rockafellar (1997)). Therefore, for any $q \in \Delta^c \setminus int\Delta^c$, we have $\lim_{q' \to q} \hat{w}^c(q') \geq \hat{w}^c(q)$ since $\text{cl}(-\hat{w}^c) \leq -\hat{w}^c$. A similar argument applies for $\hat{w}^d$.

The next result characterizes how Sender communicates with Receiver when she always confirms ρ_0. This is also a key step in understanding her incentives to disprove ρ_0. Recall that

$$\hat{w}^c(\sigma) \quad = \sup_{\{\tau\in\mathcal{F}_\sigma:\mathbf{supp}\,\tau\subset\Delta^c\}} V(\tau) = \sup_{\mathcal{F}^c_\sigma}\sum_m \tau_m w(q_m), \tag{13}$$

where

$$\mathcal{F}^c_\sigma = \left\{\{(q_n,\tau_n)\}_{n=1}^N : \sum_{m=1}^N \tau_m q_m = \sigma, \sum_{m=1}^N \tau_m = 1, \tau_m \geq 0, q_m \in \Delta^c, \forall m\right\}.$$

Proposition 3. *To compute $\hat{w}^c(\sigma)$ in (13), it is without loss of generality to consider distributions τ with the following properties:*
(1) for every $q \in \mathbf{supp}\,\tau$, there exists at most one $\omega \in \overline{\mathcal{R}}$ such that $q(\omega) \in (0,1)$ and[38]

$$q = (1-q(\omega))q(\cdot|\mathcal{R}) + q(\omega)\delta_\omega;$$

(2) for every $\omega \in \overline{\mathcal{R}}$, there exists a unique $q \in \mathbf{supp}\,\tau$ such that $q(\omega) > 0$ and, given Receiver's action $a(p(q))$,[39]

$$u_S(a(p(q)),\omega) = u^*_S(\omega) \equiv \max_{q'\in\Delta(\mathcal{R})}\{\max_{a\in\mathcal{A}(p(q'))} u_S(a,\omega)\}.$$

When Sender always confirms ρ_0—either as an assumption or because it is optimal—she communicates as follows. First, each message leads to a posterior q' such that she rules out all states that are impossible under ρ_0 except at most one, say $\omega' \in \overline{\mathcal{R}}$. Since Receiver's posterior $p(q')$ continues to assign probability zero to ω', we will refer to this strategy as *hiding ω' in posterior q'*. Second, Sender hides each $\omega' \in \overline{\mathcal{R}}$ in only one posterior and, among all those confirming ρ_0, she selects q' so that, having posterior $p(q')$, Receiver chooses the best action for Sender in state ω'. We will refer to this strategy as *optimal hiding of ω'* and to $u^*_S(\omega')$ as its payoff.

The intuition is this. Property (1) arises because Sender can replace a posterior assigning positive probability to, say, two states in $\overline{\mathcal{R}}$ with two posteriors such that both are identical to the original one conditional on $\mathcal{R}$—and hence lead to the same posterior for Receiver—but each assigns positive probability to only one of the two states. This is possible because the priors' different supports allow Sender to become more informed

[38]For every ω, δ_ω represents the distribution that assigns probability 1 to ω.
[39]Note that $\Delta(\mathcal{R})$ is compact and $\max_{a\in\mathcal{A}(p(q'))} u_S(a,\omega)$ is upper semicontinuous in q, so u^*_S is well defined.

without necessarily making Receiver more informed as well. Property (2) arises because Sender can ensure that she assigns positive probability to $\omega' \in \overline{\mathcal{R}}$ at a posterior q' for which Receiver's action is most beneficial for her if ω' is true; moreover, she can make q' assign almost probability 1 to ω'. This also means that if the true state is *not* ω', almost certainly her signals will *not* induce q'. In this way, for states in $\mathcal{R}$, her device can yield with almost certainty signals that make Receiver behave in her best interest for those states.

Proposition 3 allows us to obtain a simple expression for $\hat{w}^c(\sigma)$.

Corollary 3. *Sender's expected payoff from always confirming* ρ_0 *is*

$$\hat{w}^c(\sigma) = \sigma(\mathcal{R})\hat{w}^c(\sigma(\cdot|\mathcal{R})) + \sum_{\omega \in \overline{\mathcal{R}}} \sigma(\omega) u_S^*(\omega).$$

The value of always confirming ρ_0 is the value Sender would achieve if she could divide her problem into two steps as follows. First, she learns whether $\mathcal{R}$ or $\overline{\mathcal{R}}$ occurred. Conditional on $\mathcal{R}$, her belief has the same support as Receiver's prior and she chooses the optimal signal device as in settings with common-support priors—this is the case studied in Kamenica and Gentzkow (2011) and Alonso and Câmara (2013). Conditional on $\overline{\mathcal{R}}$, she chooses a fully informative device, where her payoff in each state ω is replaced by that from optimally hiding ω.

We can now obtain a simple necessary and sufficient condition for Sender to optimally disprove ρ_0 (i.e., for $\hat{w}(\sigma) > \hat{w}^c(\sigma)$). To this end, define $h : \Delta^d \to \mathbb{R}$ as

$$h(q) = \sum_{\omega \in \overline{\mathcal{R}}} u_S^*(\omega) q(\omega).$$

This is the payoff Sender expects, at posterior q, from optimally hiding all states for which she is providing supporting evidence (i.e., all $\omega \in \mathbf{supp}\, q$). In words, Sender disproves ρ_0 with positive probability if and only if there exists *some* posterior q at which she disproves ρ_0 and, given Receiver's response, her expected payoff strictly exceeds $h(q)$.

Proposition 4. *It is optimal for Sender to disprove* ρ_0 *if and only if* $w^d(q) > h(q)$ *for some* $q \in \Delta^d$.

Note that this condition can be directly checked without using any concavification.

The next result characterizes Sender's payoff and communication strategy when she disproves ρ_0. Some preliminary observations are in order. If w^d is continuous, so is $\hat{w}^d$

and hence $\hat{w}^d(q)$ is achieved for every $q \in \Delta^d$.[40] To deal with the general case, we slightly modify w^d by considering the lowest upper-semicontinuous function that is pointwise larger than w^d, denoted w^d_*, and let $\hat{w}^d_*$ be its concavification (Definition 2). Then $\hat{w}^d_*$ is continuous, it agrees with $\hat{w}^d$ on the interior of Δ^d, and $\hat{w}^d_*(q)$ is achieved for every $q \in \Delta^d$ (see Lemma 4 in Appendix B). Now define the function $m : \Delta^d \to \mathbb{R}$ by

$$m = \max\{h, w^d_*\}.$$

For every q disproving ρ_0, $m(q)$ captures whether with belief q Sender expects to do better by actually surprising Receiver or by optimally hiding states. Since h and w^d_* are upper semicontinuous, so is m. Then its concavification $\hat{m}$ is continuous and $\hat{m}(q)$ is achieved for all $q \in \Delta^d$ by the argument establishing these properties for $\hat{w}^d_*$.[41]

Proposition 5. *If Sender optimally disproves ρ_0, then*

$$\begin{aligned} \hat{w}(\sigma) &= \tau^c \hat{w}^c(q^c) + (1-\tau^c)\hat{w}^d_*(q^d) && (14) \\ &= \sigma(\mathcal{R})\hat{w}^c(\sigma(\cdot|\mathcal{R})) + \sigma(\overline{\mathcal{R}})\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) && (15) \end{aligned}$$

where $q^c \in \Delta^c$, $q^d \in \Delta^d$, $\tau^c \in [\sigma(\mathcal{R}), 1)$, $\tau^c q^c + (1-\tau^c)q^d = \sigma$, and

$$\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = \frac{\tau^c q^c(\overline{\mathcal{R}})}{\sigma(\overline{\mathcal{R}})} h(q^c(\cdot|\overline{\mathcal{R}})) + \frac{1-\tau^c}{\sigma(\overline{\mathcal{R}})}\hat{w}^d_*(q^d).$$

Moreover, any (conditional) distribution on Δ^d achieving $\hat{w}^d_*(q^d)$ assigns positive probability only to $q \in \Delta^d$ such that $w^d_*(q) \geq h(q)$. Finally, $\tau^c = \sigma(\mathcal{R})$ if and only if $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = \hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}}))$.

Proposition 5 can be interpreted as follows. Expression (14) highlights the dichotomy in Sender's problem between confirming and disproving ρ_0. To solve her problem, she can find the best among all convex combinations averaging to σ and involving only *two* posteriors: one, q^c, confirms ρ_0 and one, q^d, disproves it. To q^c, she assigns the expected payoff of the fictitious scenario in which q^c is her *prior* and she always confirms ρ_0 (Corollary 3). To q^d, she assigns the expected payoff of the fictitious scenario in which q^d is again her prior, but Receiver responds to messages in the possibly non-Bayesian way captured by the function $p(\cdot)$ in Proposition 2 after surprises. The weight τ^c she assigns

[40] The claimed property follows from Part (i) of Lemma 4 in Appendix B.

[41] To see that m is upper semicontinuous (u.s.c.), note that $\text{hyp}\, m = \text{hyp}\, h \cup \text{hyp}\, w^d_*$ where both h and w^d_* are u.s.c.. By Theorem 7.1 in Rockafellar (1997), a function is u.s.c. if and only if its hypograph is closed.

to q^c pins down the ex-ante probabilities of confirming and disproving ρ_0.

Expression (15) highlights the difference in Sender's communication between states inside and outside Receiver's theory of the world and allows us to compute her overall payoff from primitives. Conditional on $\mathcal{R}$, Sender gets the expected payoff from optimally communicating with Receiver in the fictitious (standard) world in which their priors are $\sigma(\cdot|\mathcal{R})$ and ρ_0. Conditional on $\overline{\mathcal{R}}$, Sender gets the expected payoff from optimally combining disproving ρ_0 and hiding states *given 'prior'* $\sigma(\cdot|\overline{\mathcal{R}})$. This is clear from the expression of $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$, which can be computed by finding $\hat{w}^d_*$ and the best convex combination $\gamma h(q_1)+(1-\gamma)\hat{w}^d_*(q_2)$ with $q_1, q_2 \in \Delta(\overline{\mathcal{R}})$ and $\gamma q_1+(1-\gamma)q_2 = \sigma(\cdot|\overline{\mathcal{R}})$. Given the optimal γ, we get the probability of disproving ρ_0, $\tau^d = (1-\gamma)\sigma(\overline{\mathcal{R}})$, the probability of hiding states, $\tau^c q^c(\overline{\mathcal{R}}) = \gamma\sigma(\overline{\mathcal{R}})$, and $q^c = q_2 + \frac{1}{\tau^c}(\sigma - q_2)$ since $q_2 = q^d$. Proposition 5 also says—as should be expected—that when Sender ends up with q disproving ρ_0, she never expects that she could do strictly better by hiding the states for which she is providing supporting evidence.

Finally, Proposition 5 gives a necessary and sufficient condition for Sender to optimally disprove ρ_0 whenever the true state would call for it. The key is that, *at the specific 'prior'* $\sigma(\cdot|\overline{\mathcal{R}})$, Sender has *one* way to maximize her expected payoff which never involves optimally hiding states. Note, however, that τ^c in Proposition 5 need not be unique. Of course, a stronger sufficient condition for $\tau^c = \sigma(\mathcal{R})$ is $w^d \geq h$ with strict inequality for some $q \in \Delta^d$.

The next result gives a simpler sufficient condition for Sender *not* to disprove ρ_0 whenever required by the state. Let $T(\sigma(\cdot|\overline{\mathcal{R}}))$ be the set of distributions $\tau \in \Delta(\Delta^d)$ such that $\sum_q q\tau(q) = \sigma(\cdot|\overline{\mathcal{R}})$ and $\sum_q w^d_*(q)\tau(q) = \hat{w}^d_*(\sigma(\cdot|\overline{\mathcal{R}}))$. Recall that $T(\sigma(\cdot|\overline{\mathcal{R}})) \neq \varnothing$.

Corollary 4. *Suppose $\hat{w}(\sigma) > \hat{w}^c(\sigma)$. If there exists $\hat{\tau} \in T(\sigma(\cdot|\overline{\mathcal{R}}))$ and $q \in \mathbf{supp}\,\hat{\tau}$ such that $w^d_*(q) < h(q)$, then $\tau^d < \sigma(\overline{\mathcal{R}})$.*

In words, Sender will not disprove ρ_0 whenever required by ω if there is *an* optimal way of doing so that involves a posterior at which she could do strictly better, in expectation, by optimally hiding all states for which she provides supporting evidence. All elements in this condition can be directly computed from primitives of the model. The condition is weaker than requiring that $w^d(q) < h(q)$ for all $q \in \Delta^d$, which of course implies $\tau^d = 0$. Note that we can have $\hat{w}^d_*(\sigma(\cdot|\overline{\mathcal{R}})) > h(\sigma(\cdot|\overline{\mathcal{R}}))$ and yet $\tau^d < \sigma(\overline{\mathcal{R}})$. Also, conditional on $\overline{\mathcal{R}}$ Sender and Receiver could have sufficiently aligned preferences, so that $\hat{w}^d_*(\sigma(\cdot|\overline{\mathcal{R}}))$ is achieved by fully revealing all states in $\overline{\mathcal{R}}$. Yet Sender may prefer to optimally hide some $\omega \in \overline{\mathcal{R}}$, thus confirming ρ_0 in that sate.

Proposition 5 also implies that a (virtually) optimal τ has to satisfy the following "no regret" property. When Sender ends up disproving ρ_0, she cannot strictly prefer any action that Receiver takes when she confirms ρ_0.

Corollary 5 ('No Regret')**.** *Take any $\tau \in \mathcal{F}_\sigma$ with $\tau^d > 0$. Suppose that for some $\hat{q} \in D_\tau$ and $q' \in C_\tau$*

$$w^d(\hat{q}) < \sum_{\omega \in \Omega} u_S(a(p(q')), \omega)\hat{q}(\omega).$$

Then, $V(\tau) < \hat{w}(\sigma)$.

This is again because of the flexibility Sender enjoys in managing her posteriors, while not changing Receiver's. She can have posterior $\hat{q}$ conditional on $\overline{\mathcal{R}}$ and q' conditional on $\mathcal{R}$, and at the same time assign very low probability to the states in $\mathcal{R}$. In this way, she makes Receiver choose $a(p(q'))$ while having a posterior arbitrarily close to $\hat{q}$ and hence an expected payoff close to $\sum_{\omega \in \Omega} u_S(a(p(q')), \omega)\hat{q}(\omega)$.

5.1 Illustrative Example: Court

We can now formally solve the court example from the introduction. Recall the data of the problem: $\mathcal{S} = \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $\mathcal{R} = \{\omega_0, \omega_2\}$, $\sigma = (0.35, 0.4, 0.15, 0.1)$, $\rho_0 = (0.7, 0, 0.3, 0)$, and $\rho_1 = \sigma$.[42] The judge (Receiver) only cares about matching defendant's refund with the state. Letting $f(\omega_i) = i$ (where i stands for i thousands/millions of dollars) and $A = \{0, 1, 2, 3\}$, we have

$$u_R(a, \omega_i) = \begin{cases} 1 & \text{if } a = f(\omega_i) \\ 0 & \text{otherwise} \end{cases}.$$

Concerning the lawyer (Sender), we have $u_S(a, \omega_i) = a$, the refund amount.

By Proposition 4 Sender must disprove ρ_0 with strictly positive probability. For Sender the best way to hide both ω_1 and ω_3 is to ensure that Receiver chooses $a = 2$. This occurs at posteriors q assigning at least probability 0.5 to ω_2 *conditional on* $\mathcal{R}$. Therefore, $u_S^*(\omega_1) = u_S^*(\omega_3) = 2$ and hence $h(q) = 2$ for all $q \in \Delta^d = \Delta(\{\omega_1, \omega_3\})$. Since $w^d(\delta_{\omega_3}) = 3$, Proposition 4 implies the claimed property.

To obtain Sender's payoff $\hat{w}(\sigma)$ in (15), we first compute $\hat{w}^c(\sigma(\cdot|\mathcal{R}))$ with the help of Figure 1. As in Kamenica and Gentzkow (2011), when both parties have prior $\sigma(\cdot|\mathcal{R}) =$

[42]Note that, with these priors, Sender's and Receiver's posteriors coincide if either $q \in \Delta(\mathcal{R})$ or $q \in \Delta(\overline{\mathcal{R}})$.

$(0.7, 0.3)$, it is optimal for Sender to generate two posteriors: At the first, both parties assign probability 1 to ω_0 (i.e., $q = (1, 0, 0, 0)$) and Receiver chooses $a = 0$; At the second, both assign equal probabilities to ω_0 and ω_2 (i.e., $q' = (0.5, 0, 0.5, 0)$) and Receiver chooses $a = 2$. The full red dots in Figure 1 represent the corresponding expected payoffs for Sender. Hence, $\hat{w}^c(\sigma(\cdot|\mathcal{R})) = \frac{0.3}{0.5}2 = 1.2$.

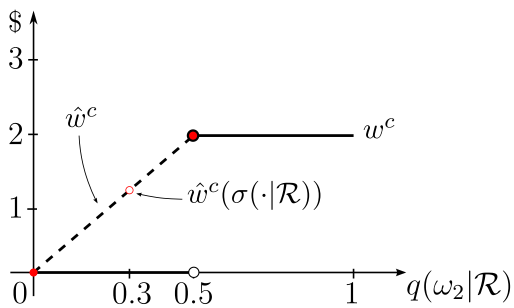

Figure 1: Confirming ρ_0

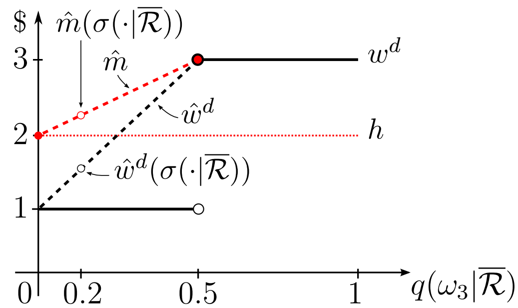

Figure 2: Disproving ρ_0

Second, we compute $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$ in (15). Since for all $q \in \Delta^d$ Receiver's posterior equals Sender's, $w^d(q) = 1$ if $q(\omega_3) < 0.5$ and $w^d(q) = 3$ otherwise. Therefore, $m = \max\{2, w^d\}$ equals 2 if $q(\omega_3) < 0.5$ and 3 otherwise. As illustrated in Figure 2, again when both parties have prior $\sigma(\cdot|\overline{\mathcal{R}}) = (0.8, 0.2)$ *and* Sender has payoff function m, she optimally induces two posteriors: At the first, both parties assign equal probability to ω_1 and ω_3 (i.e., $q^d = (0, 0.5, 0, 0.5)$) and Receiver chooses $a = 3$; At the second, both parties assign probability 1 to ω_1 (i.e., $\hat{q} = (0, 1, 0, 0)$) and Sender get's payoff $u_S^*(\omega_1)$. So,

$$\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = \frac{0.2}{0.5} w^d(q^d) + \frac{0.3}{0.5} h(\delta_{\omega_1}) = 2.4.$$

Given this, Sender's overall expected payoff (the expected refund for her client) is

$$W_\sigma = 0.5(1.2) + 0.5(2.4) = 1.8.$$

Consider now the probability with which Sender disproves ρ_0. By Proposition 5, this probability is strictly less than $0.5 = \sigma(\overline{\mathcal{R}})$, because

$$\hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}})) = \frac{0.2}{0.5} w^d(q^d) + \frac{0.3}{0.5} w^d(\delta_{\omega_1}) = \frac{0.2}{0.5}3 + \frac{0.3}{0.5}1 = 1.8$$

which is lower than $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$.[43] In fact,

$$\tau^d = \sigma(\overline{\mathcal{R}})\frac{0.2}{0.5} = 0.2.$$

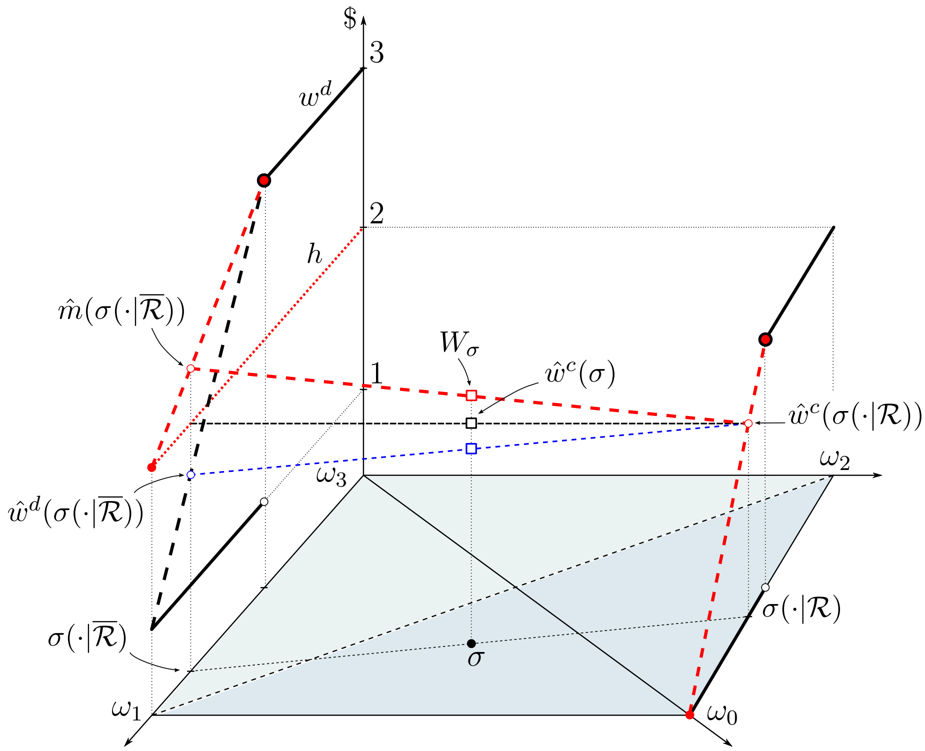

Figure 3: Optimal Signal Device

Figure 3 helps us describe Sender's (virtually) optimal signal device. It involves four signals: three confirm ρ_0 and one disproves it. Signal x_0 arises only in state ω_0, revealing it. Signal x_2 arises in ω_0 and ω_2 and leads both parties to assign them equal probability. Signal x_1 arises in ω_1 and ω_2 and leads Sender to assign ω_1 probability *arbitrarily close* to 1, while Receiver assigns probability 1 to ω_2. Finally, signal x_3 arises in ω_1 and ω_3 and leads both parties to assign them equal probability. Graphically, the solid red dots in Figure 3 represent Sender's expected payoff for each signal.

For comparison, consider Sender's payoff from the strategies of never and always disproving ρ_0. In the first case, her maximal expected payoff is given in Corollary 3:

$$\hat{w}^c(\sigma) = 0.5(1.2) + 0.5(2) = 1.6.$$

This corresponds to the black square in Figure 3. In the second case, her maximal

[43] In this example, we could have immediately used Corollary 4 to reach the same conclusion since $w^d(\delta_{\omega_1}) < h(\delta_{\omega_1})$.

expected payoff is

$$0.5\hat{w}^c(\sigma(\cdot|\mathcal{R})) + 0.5\hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}})) = 0.5(1.2) + 0.5(1.8) = 1.5.$$

This corresponds to the blue square in Figure 3. So, by optimally combining hiding and surprising, Sender improves her payoff by 12.5% relative to always hiding and by 20% relative to always surprising.

5.2 Application with Quadratic Payoffs

This section applies the previous results to a setting in which states and actions take values on the real line and Sender's and Receiver's payoffs are quadratic loss functions of the gap between an ideal and the implemented action. Such settings have been applied to study many concrete problems.[44]

The setting is defined as follows. Let $\Omega = \{\omega_1, \ldots, \omega_n\} \subset \mathbb{R}_{++}$ where states are ordered so that $\omega_i < \omega_{i+1}$ for all i, and let $A = \mathbb{R}$. Payoffs functions are $u_S(a, \omega) = -(a-\omega)^2$ and $u_R(a, \omega) = -(a - \beta(\omega))^2$ where $\beta(\omega)$ is Receiver's ideal action in state ω. Assume that $\beta(\omega) > \omega$ for all ω and that $\beta(\cdot)$ involves a fixed and a linear component in ω, i.e., $\beta(\omega) = \kappa\omega + b$ with $\kappa > 0$ and $b \geq 0$. Receiver initially views low and high enough states as impossible: $\mathcal{R} = \{\omega_i\}_{i=\underline{m}}^{\overline{m}}$ with $1 < \underline{m} < \overline{m} < n$. Moreover, $\mathbf{supp}\,\mu = (\rho_0, \rho_1)$ with $\mu(\rho_0) > \frac{1}{2}$, $\rho_0 = \sigma(\cdot|\mathcal{R})$, and $\rho_1 = \sigma$. As usual, in this setting given posterior p Receiver chooses his expected ideal action: $a(p) = \mathbb{E}_p[\beta]$.

First of all, we need to compute Sender's expected payoffs for any posterior and from optimally hiding. For any $q \in \Delta(\mathcal{S})$,

$$\begin{aligned} w(q) &= \mathbb{E}_q[u_S(\mathbb{E}_{p(q)}[\beta], \omega)] \qquad (16) \\ &= -\kappa^2(\mathbb{E}_{p(q)}[\omega])^2 + 2\kappa\mathbb{E}_{p(q)}[\omega]\mathbb{E}_q[\omega] - 2b(\kappa\mathbb{E}_{p(q)}[\omega] - \mathbb{E}_q[\omega]) - b^2 - \mathbb{E}_q[\omega^2]. \end{aligned}$$

Recall that $p(q) = q(\cdot|\mathcal{R})$ for $q \in \Delta^c$ by (8); also, $p(q) = q$ for $q \in \Delta^d$. So $\mathbb{E}_{p(q)}[\omega]$ is always linear in q and w^d is continuous. To compute $u_S^*(\omega)$ we have to consider several cases. For $\omega < \omega_{\underline{m}}$ it is clearly optimal to hide ω so that Receiver assigns probability 1 to $\omega_{\underline{m}}$, the closest state to ω within $\mathcal{R}$; hence $u_S^*(\omega) = -(\beta(\omega_{\underline{m}}) - \omega)^2$. For $\omega > \omega_{\overline{m}}$ the optimal hiding strategy depends on β. If $\omega_{\overline{m}} < \omega < \beta(\omega_{\underline{m}})$ the situation is equivalent to the case $\omega < \omega_{\underline{m}}$. If $\beta(\omega_{\overline{m}}) \leq \omega$, it is optimal to hide ω so that Receiver assigns probability 1 to $\omega_{\overline{m}}$; hence $u_S^*(\omega) = -(\beta(\omega_{\overline{m}}) - \omega)^2$. If instead $\beta(\omega_{\overline{m}}) > \omega \geq \beta(\omega_{\underline{m}})$, there

[44] See Footnote 10.

always exists $q \in \Delta^c$ such that $\mathbb{E}_{p(q)}[\beta] = \omega$; hence $u^*_S(\omega) = 0$.

Applying Proposition 4, we have that Sender will disprove Receiver's theory with positive probability ($\tau^d > 0$) for all specifications of κ and b. Indeed, for any $\omega < \omega_{\underline{m}}$, $\delta_\omega \in \Delta^d$ and $w^d(\delta_\omega) = -(\beta(\omega) - \omega)^2$ which is strictly larger than $h(\delta_\omega) = -(\beta(\omega_{\underline{m}}) - \omega)^2$ because $\omega < \beta(\omega) < \beta(\omega_{\underline{m}})$.

Using Proposition 5, we can fully characterize Sender's communication strategy as a function of the bias parameters κ and b. To compute $\hat{w}(\sigma)$ in expression (15), consider each component separately. First, if $\tau \in \Delta(\Delta(\mathcal{S}))$ is such that $\sum_q q\tau(q) = \sigma(\cdot|\mathcal{R})$, then $\mathbf{supp}\, q \subset \mathcal{R}$ for all $q \in \mathbf{supp}\, \tau$ and hence $q = q(\cdot|\mathcal{R})$. It follows that (16) simplifies to

$$w^c(q) = \kappa(2-\kappa)(\mathbb{E}_q[\omega])^2 + 2b(\kappa - 1)\mathbb{E}_q[\omega] - \mathbb{E}_q[\omega^2] - b^2. \tag{17}$$

Note that $w^c(q)$ is strictly convex in q if and only if $\kappa < 2$. Therefore, if $\kappa \leq 2$, $\hat{w}^c(\sigma(\cdot|\mathcal{R}))$ is achieved by $\tau = \{(\delta_\omega, \sigma(\omega|\mathcal{R}))\}_{\omega\in\mathcal{R}}$ (uniquely if $\kappa < 2$), that is, by fully revealing the state; by contrast, if $\kappa > 2$, revealing no information is strictly optimal and $\hat{w}^c(\sigma(\cdot|\mathcal{R})) = w^c(\sigma(\cdot|\mathcal{R}))$. Second, for any $q \in \Delta^d$, expression (17) again holds. Hence, if $\kappa \leq 2$, $\hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}}))$ is again achieved by $\tau' = \{(\delta_\omega, \sigma(\omega|\overline{\mathcal{R}}))\}_{\omega\in\overline{\mathcal{R}}}$ (uniquely if $\kappa < 2$), whereas $\hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}})) = w^d(\sigma(\cdot|\overline{\mathcal{R}}))$ if $\kappa > 2$.

Based on these observations, consider first the case with $\kappa \leq 2$. In short, Sender always fully reveals all states below $\omega_{\overline{m}}$. For higher states, she divides them into two groups depending on Receiver's bias: those close enough to $\omega_{\overline{m}}$ are always hidden, the others are fully revealed. Finally, the stronger the Receiver's bias, the larger the set of hidden states.

Corollary 6. *If $\kappa \leq 2$, Sender's communication strategy has the following properties.*
- *For each $i \leq \overline{m}$, she fully reveals ω_i.*
- *For all $i > j > \overline{m}$, there exist thresholds $b_i(\kappa)$ and $b_j(\kappa)$ decreasing in κ (each strictly when positive) and such that $b_i(\kappa) \leq b_j(\kappa)$ (with $<$ if either is positive). If $b \leq b_{\overline{m}+1}(\kappa)$, we have $\hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}})) = \hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$ and hence $\tau^d = \sigma(\overline{\mathcal{R}})$. If $b > b_{\overline{m}+1}(\kappa)$, there exists $i^*(b,\kappa)$, non-decreasing in both κ and b, such that Sender hides with probability 1 each ω_i with $\overline{m} < i < i^*(b,\kappa)$ and fully reveals each ω_i with $i \geq i^*(b,\kappa)$.*

So, when $\kappa \leq 2$, the fixed component b of Receiver's bias plays no role in how Sender communicates *when she is not hiding states.* This point generalizes a similar observation in Kamenica and Gentzkow (2011). By contrast, however, here b affects which states Sender hides and how she optimally hides them.

The case with $\kappa > 2$ is less straightforward. Since w^d is strictly concave in q, we always have $\hat{w}^d(q^d) = w^d(q^d)$ in Proposition 5. Therefore, whenever Sender disproves ρ_0, she reveals no information and makes Receiver take the same action: $\mathbb{E}_{q^d}[\beta]$. It turns out that this action always caters to either the low states or the high states, rather than trying to achieve a compromise.

Corollary 7. *If $\kappa > 2$, then either $\mathbb{E}_{q^d}[\beta] < \beta(\omega_{\underline{m}})$ or $\mathbb{E}_{q^d}[\beta] > \beta(\omega_{\overline{m}})$, and Sender's expected payoff is*

$$-\left\{\sum_{\omega\in\mathcal{R}}(\mathbb{E}_{\sigma(\cdot|\mathcal{R})}[\beta]-\omega)^2\sigma(\omega)+\tau^d\mathbb{E}_{q^d}(\mathbb{E}_{q^d}[\beta]-\omega)^2+\tau^c\sum_{\omega\in\overline{\mathcal{R}}}\{-u^*_S(\omega)\}q^c(\omega)\right\} \tag{18}$$

where $\tau^d = 1-\tau^c > 0$ and $\tau^c q^c + \tau^d q^d = \sigma$. If $\omega_{\underline{m}} \leq \mathbb{E}_{\sigma(\cdot|\overline{\mathcal{R}})}[\omega] \leq \omega_{\overline{m}}$, then $\tau^d < \sigma(\overline{\mathcal{R}})$.

To gain intuition, suppose $\mathbb{E}_{q^d}[\beta]$ were between $\beta(\omega_{\underline{m}})$ and $\beta(\omega_{\overline{m}})$, the highest and lowest action Receiver would take if his theory is confirmed. Then, for all states $\omega < \omega_{\underline{m}}$, Sender would strictly prefer to hide them. But if she does so, $\mathbb{E}_{q^d}[\beta]$ jumps above $\beta(\omega_{\overline{m}})$. She can then combine hiding and surprising for states above $\omega_{\overline{m}}$ to achieve the best $\mathbb{E}_{q^d}[\beta]$, thus improving her payoff. A similar improvement may start from hiding states $\omega > \omega_{\overline{m}}$ which leads to $\mathbb{E}_{q^d}[\beta] < \beta(\omega_{\underline{m}})$. Overall, given κ and b, whether it is best for Sender to cater to high or to low states with $\mathbb{E}_{q^d}[\beta]$ ultimately depends on which states she thinks are more likely.

Sender's expected payoff in (18) contains information on the other actions that she persuades Receiver to choose. For states in $\mathcal{R}$, he chooses $\mathbb{E}_{\sigma(\cdot|\mathcal{R})}[\beta]$ with probability arbitrarily close to 1 (which becomes exactly 1 if $\tau^d = \sigma(\overline{\mathcal{R}})$). Sender 'uses' the remaining probability to hide states in $\overline{\mathcal{R}}$, in which case Receiver's action is in the interval $[\beta(\omega_{\underline{m}}), \beta(\omega_{\overline{m}})]$ but differs in general from $\mathbb{E}_{\sigma(\cdot|\mathcal{R})}[\beta]$. If σ is such that $\mathbb{E}_{\sigma(\cdot|\overline{\mathcal{R}})}[\omega] \in [\omega_{\underline{m}}, \omega_{\overline{m}}]$, we can say more about which states Sender hides. If at the optimum $\mathbb{E}_{q^d}[\beta] < \beta(\omega_{\underline{m}})$, there exists $\omega > \omega_{\overline{m}}$ which is optimally hidden with strictly positive probability (otherwise $\mathbb{E}_{q^d}[\beta] \geq \beta(\omega_{\underline{m}})$). Similarly, if $\mathbb{E}_{q^d}[\beta] > \beta(\omega_{\overline{m}})$ there exists $\omega < \omega_{\underline{m}}$ which is optimally hidden with strictly positive probability. In other words, Sender makes Receiver's action cater to low (high) states by optimally hiding high (low) states. Note that these qualitative properties as well as those in Corollary 7 are independent of κ and b.

5.3 Is Having a Richer Theory Always Good for Persuaders?

With the help of an example, this section aims to illustrate two simple points: (1) Sender may be strictly better off if Receiver shared her theory of the world rather than viewing some states as impossible; (2) Sender may reveal some information in the former scenario but not in the latter.[45] The key to these points is Proposition 1: by treating some states as impossible, Receiver's theory can severely limit how Sender can influence his beliefs and hence actions.

Consider a setting with two states: $\mathcal{S} = \{\omega_1, \omega_2\}$. Sender has prior $\sigma = (\frac{1}{2}, \frac{1}{2})$. Receiver has four actions, $A = \{a, b, c, d\}$, and payoffs are as follows:

$u_S(\cdot,\cdot)$	ω_1	ω_2
a	0	5
b	3	1
c	1	4
d	2	2

$u_R(\cdot,\cdot)$	ω_1	ω_2
a	2	-3
b	1	0
c	0	1
d	-3	2

The next table summarizes Receiver's best actions depending on his posterior p:

Values of $p(\omega_2)$	$[0, \frac{1}{4})$	$\{\frac{1}{4}\}$	$(\frac{1}{4}, \frac{1}{2})$	$\{\frac{1}{2}\}$	$(\frac{1}{2}, \frac{3}{4})$	$\{\frac{3}{4}\}$	$(\frac{3}{4}, 1]$
$\mathcal{A}(p)$	$\{a\}$	$\{a, b\}$	$\{b\}$	$\{b, c\}$	$\{c\}$	$\{c, d\}$	$\{d\}$

When Receiver's prior is $\rho_0 = (0, 1)$, Sender has only two options: always confirm ρ_0 and always disprove it. The expected payoff from these strategies are 2 and 1 respectively, so it is optimal for Sender not to communicate at all.

Now suppose that $\rho_0 = \sigma$ as in Kamenica and Gentzkow (2011). In this case, Sender's and Receiver's posteriors coincide for any message: $p(q) = q$ for all $q \in \Delta(\mathcal{S})$. Using the previous table, we can easily compute Sender's expected payoff as a function of q:

$$w^{KG}(q) = \begin{cases} 5q_2 & \text{if } q_2 \in [0, \frac{1}{4}) \\ 3 - 2q_2 & \text{if } q_2 \in [\frac{1}{4}, \frac{1}{2}) \\ 1 + 3q_2 & \text{if } q_2 \in [\frac{1}{2}, \frac{3}{4}] \\ 2 & \text{if } q_2 \in (\frac{3}{4}, 1] \end{cases},$$

where $q_2 = q(\omega_2)$. By Corollary 2 in Kamenica and Gentzkow (2011), Sender's expected payoff from an optimal π is $W^{KG}_\sigma = \hat{w}^{KG}(\sigma)$ and she benefits from revealing information

[45] When Sender's and Receiver's have common-support but different priors, Alonso and Câmara (2013) show that generically reveling some information is optimal for Sender.

if and only if $\hat{w}^{KG}(\sigma) > w^{KG}(\sigma)$. It is easy to see that in this example

$$\hat{w}^{KG}(q) = \begin{cases} 10q_2 & \text{if } q_2 \in [0, \frac{1}{4}) \\ \frac{17}{8} + \frac{3}{2}q_2 & \text{if } q_2 \in [\frac{1}{4}, \frac{3}{4}) \\ 7 - q_2 & \text{if } q_2 \in (\frac{3}{4}, 1] \end{cases} .$$

Therefore, $\hat{w}^{KG}(\frac{1}{2}, \frac{1}{2}) = 2.875$ whereas $w^{KG}(\frac{1}{2}, \frac{1}{2}) = 2.5$. Moreover, W_σ^{KG} exceeds Sender's expected payoff when Receiver views ω_1 as impossible ($W_\sigma = 2$).

This example raises an intriguing issue for the literature on persuasion. Starting from a situation in which Receiver's theory is $\rho_0 = (0, 1)$, Sender would like to first 'persuade' him to adopt her theory σ and only then reveal some information on the states. We have seen, however, that no signal device in Π—the class commonly studied in the literature—can make Receiver switch from ρ_0 to σ. So, if there is any technique that would make him switch theories (perhaps arguments on their internal logical consistency), it must belong to a different class of persuasion activities than those usually studied in the literature. This class may deserve further investigation.

6 Extensions

6.1 Unawareness Interpretation of the Model

It is possible, perhaps natural, to interpret this paper's model as a way to describe situations in which Sender wants to persuade a Receiver who is initially unaware of some states of the world. Receiver's unawareness of the states in $\overline{\mathcal{R}}$ would be the underlying reason why his prior assigns them zero probability.

This interpretation requires further explanation. It is reasonable to assume that, though initially unaware of ω, Receiver can fully understand its description once given to him. This description can also exhaustively specify Receiver's payoff in ω; for example, it says how actions map to monetary prizes, for which he knows his utility function. Hence, we can continue to define u_R as a function of Ω, not just $\mathcal{R}$. We can also continue to assume that Sender commits to signal devices. When π is disclosed to Receiver before any signal realizes, he understands that he may have been unaware of some conceivable states. Nonetheless, we can assume that Receiver does not abandon his prior after observing *only* π. This may happen for several reasons: (1) Per se π contains no information; (2) Receiver may be firmly skeptic and think that Sender simply invented states that are incompatible with his theory, and hence he should not abandon it unless conclusive evidence disproves it; (3) Receiver may always try to interpret any observation in favor of his initial theory, as shown in studies on "confirmatory bias."[46] Of course, another possibility is that as soon as Receiver hears the description of an $\omega \notin \mathcal{R}$, he abandons his initial theory of the world and adopts one that includes ω. In this case, however, the model essentially collapses to one with common-support priors: by construction all $\pi \in \Pi$ describe all states in Ω and therefore Sender will design π as if Receiver already assigns positive probability to all states.

A natural concern—especially under the unawareness interpretation—is that after observing a signal Sender may want to conceal some part of her device. In this way, Sender can *literally* hide states Receiver is unaware of. Formally, after observing x from π, Sender may want to communicate a message (x, π') where π' is obtained by deleting $\pi(\cdot|\omega)$ for some ω. If (x, π) does not make Receiver abandon ρ_0 and $\omega \in \overline{\mathcal{R}}$, such a change does not benefit Sender: Receiver ignores $\pi(x|\omega)$ anyways. But in other cases the change may benefit Sender: she can prevent Receiver from becoming aware of some states, after gaining better information for herself. Concealing parts of π is conceptually different from concealing some signal realization (e.g., by lumping signals together in a coarser one) and can have different consequences. Their analysis is left for future research.

6.2 Alternative Models of Receiver's Responses to Information

This section examines a set of assumptions describing other tractable ways to model how Receiver responds to unexpected evidence. These models are less sophisticated than Ortoleva's (2012). They are, however, simpler to the extent that they do not involve a prior over priors—which should be part of what Sender knows about Receiver—and assume a direct procedure by which Receiver picks new theories after surprises. These models rely on *lexicographic belief systems* (LBS's) and are similar in spirit to the 'sequences of hypotheses' in Kreps and Wilson's (1982) work on sequential equilibria. They may also fit better an unawareness interpretation of different-support priors.[47]

The first assumption describes how Receiver responds to messages confirming ρ_0 and has the same rationale as A1(c).

Assumption 3 (A3)**.** *If $x \in C_\pi$, Receiver updates ρ_0 using Bayes' rule.*

[46] See, e.g., Rabin and Schrag (1999) and references therein.

[47] Such interpretations may be compatible with Ortoleva's (2012) axioms. However, if at the outset Receiver is unaware of some states in Ω, it may be hard to imagine that he conceives a prior ρ assigning positive probability to them and, a fortiori, a prior over priors with ρ in its support.

To describe how Receiver responds to messages disproving ρ_0, we consider two possibilities. In both cases, we shall continue to assume that Receiver forms a unique posterior belief.[48] To do so, after a surprising x, he updates some prior other than ρ_0 which is 'triggered' by x. We formalize this idea using lexicographic belief systems (LBS's).

In the first assumption, Receiver has only one alternative theory of the world which contains all states in Ω.

Assumption 4 (A4: Binary LBS)**.** *Receiver has an LBS* (ρ_0, ρ_1) *with* **supp** $\rho_1 = \Omega$*. If* $x \in D_\pi$*, he updates* ρ_1 *using Bayes' rule.*

To illustrate, suppose $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{R} = \{\omega_1\}$, and that x can arise only in ω_2 under π, i.e., $\Omega_\pi(x) = \{\omega_2\}$. Given (x, π), first of all Receiver switches to viewing ω_1, ω_2, *and* ω_3 as possible; he then updates his new theory using Bayes' rule. A4 captures a Receiver who is willing to easily abandon his theory of the world and, given evidence contradicting it, admits that all states in Ω are actually possible. Together, A3 and A4 are similar to Ortoleva's (2012) model with $\mu(\rho_0) > \frac{1}{2}$ and $\mu(\rho_1) = 1 - \mu(\rho_0)$. His model, however, suggests a richer story for why Receiver first adopts ρ_0 and then turns to ρ_1 after unexpected signals.

The next assumption allows for richer LBS's. Recall that $\Omega_\pi(x)$ is the set of states that are consistent with (x, π) (see (3)).

Assumption 5 (A5: Gradual LBS)**.** *Receiver has LBS* $(\rho_0, \dots, \rho_N)$ *such that, for each* $\Omega_i \subset \Omega$ *with* $\Omega_i \supsetneq \mathcal{R}$*, there is exactly one* ρ_i *with* **supp** $\rho_i = \Omega_i$*.*[49] *If* $x \in D_\pi$*, he updates prior* ρ_i *with* **supp** $\rho_i = \mathcal{R} \cup \Omega_\pi(x)$ *using Bayes' rule.*

A5 captures a Receiver who is reluctant to abandon his theory. Given evidence contradicting it, he is willing to expand it to include only those states which are either possible under ρ_0 or consistent with (x, π). Suppose again that $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{R} = \{\omega_1\}$, and $\Omega_\pi(x) = \{\omega_2\}$. Given (x, π), now Receiver's new theory views *only* ω_1 and ω_2 as possible.

In general, we may require some consistency between layers in Receiver's LBSs.[50]

Assumption 6 (A6: Consistency)**.** *Given LBS* $(\rho_0, \dots, \rho_N)$*, for any* ρ_i *and* ρ_j *and*

[48]Alternatively, one could imagine that surprising signals may leave Receiver with some ambiguity, captured by a set of posteriors.

[49]To clarify, N in Receiver's LBS equals the number of subsets of Ω containing $\mathcal{R}$ as a strict subset.

[50]Karni and Vierø (2013) study a decision-theoretic model of growing awareness and provide an axiomatic foundation for this consistency property. This property should not be confused with the notion of consistency in Kreps and Wilson (1982).

corresponding supports Ω_i and Ω_j,

$$\frac{\rho_i(\omega|\Omega_i \cap \Omega_j)}{\rho_i(\omega'|\Omega_i \cap \Omega_j)} = \frac{\rho_j(\omega|\Omega_i \cap \Omega_j)}{\rho_j(\omega'|\Omega_i \cap \Omega_j)} \quad \textit{for all } \omega, \omega' \in \Omega_i \cap \Omega_j.$$

That is, for any ρ_i and ρ_j, Receiver assigns the same *relative* likelihood to all states that are possible under both theories. For example, let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\rho_0 = (\frac{1}{2}, \frac{1}{2}, 0, 0)$, and ρ_1 have support $\Omega_1 = \{\omega_1, \omega_2, \omega_3\}$. Then, conditional on $\{\omega_1, \omega_2\}$, ρ_1 assigns equal probability to the two states. A6 seems reasonable since each state is an exhaustive description of reality. If ω_1 and ω_2 can occur only when neither ω_3 nor ω_4 occurs, then the assessment of ω_1 and ω_2's relative likelihood knowing that $\{\omega_3, \omega_4\}$ is impossible should be the same as when knowing that $\{\omega_3, \omega_4\}$ was possible but did not occur.

Under A3-A6, we can characterize the joint distributions over posteriors Sender can achieve with signal devices, along the lines of Section 4.2. The analysis in Section 5 then applies unchanged.

The key step is to realize that Sender's posterior after any (x, π) pins down which prior Receiver will update among those in his LBS. Under A4 this is immediate: for every $x \in D_\pi$, $q(\cdot|x, \pi) \in \Delta^d$ and Receiver updates ρ_1. Consider now A5. For every $x \in D_\pi$, $\mathbf{supp}\, q(\cdot|x, \pi)$ equals the set $\Omega_\pi(x)$ of states consistent with (x, π). Given (x, π), Receiver updates ρ_i with support $\mathcal{R} \cup \Omega_\pi(x)$. So, for every $x \in D_\pi$, let

$$\Omega(q(\cdot|x, \pi)) = \mathcal{R} \cup \mathbf{supp}\, q(\cdot|x, \pi).$$

This mapping from Sender's posteriors to subsets of Ω is well defined: $\mathbf{supp}\, q(\cdot|x, \pi) = \mathbf{supp}\, q(\cdot|y, \pi)$ if $q(\cdot|x, \pi) = q(\cdot|y, \pi)$ for $x, y \in D_\pi$. Then, for every $\tau \in \mathcal{F}_\sigma$ and ρ_i in Receiver's LBS, let $D_\tau(\rho_i) \subset D_\tau$ be the set of Sender's posteriors at which Receiver updates ρ_i:

$$D_\tau(\rho_i) = \{q \in \mathbf{supp}\, \tau : \Omega(q) = \mathbf{supp}\, \rho_i\}.$$

Note that $\{D_\tau(\rho_i)\}_{i=1}^N$ forms a partition of D_τ. Relying on Lemma 1, we may then draw the following conclusion.

Proposition 6. *Consider any $\tau \in \mathcal{F}_\sigma$ and let Receiver's LBS be $(\rho_0, \ldots, \rho_N)$. Then, for every $q \in \mathbf{supp}\, \tau$ and $\omega \in \Omega$, we have*

$$p(\omega; q) = \frac{q(\omega)\frac{\rho(\omega;q)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega')\frac{\rho(\omega';q)}{\sigma(\omega')}}, \tag{19}$$

where under A4 (i.e., $N = 1$)

$$\rho(q) = \begin{cases} \rho_0 & \text{if } q \in C_\tau \\ \rho_1 & \text{if } q \in D_\tau \end{cases},$$

and under A5

$$\rho(q) = \begin{cases} \rho_0 & \text{if } q \in C_\tau \\ \rho_i & \text{if } q \in D_\tau(\rho_i) \end{cases}.$$

By Proposition 6, if A6 holds, Sender can achieve the same joint distributions over posteriors under A4 and A5. Therefore, considering the simpler case of A4 involves no loss of generality for the purpose of this paper.

Corollary 8. *Consider LBS (ρ_0^A, ρ_1^A) under A4 and LBS $(\rho_i^B)_{i=0}^N$ under A5. Suppose that $\rho_0^A = \rho_0^B$, $\rho_1^A = \rho_N^B$ with $\mathbf{supp}\, \rho_N^B = \Omega$, and $(\rho_i^B)_{i=0}^N$ satisfies A6. Then, for any $\tau \in \mathcal{F}_\sigma$,*

$$p^A(\omega; q) = p^B(\omega; q) \quad \text{for all } \omega \in \Omega \text{ and } q \in \mathbf{supp}\, \tau.$$

Proposition 6 also implies that under A4 $p(q)$ varies continuously in q over the sets Δ^c and Δ^d *separately*. By Corollary 8, the same is true under A5 if A6 holds.

Without A6, A4 and A5 can lead to different sets of joint distributions over posteriors and Receiver's posterior may be discontinuous over Δ^d. To see this, suppose that $\mathcal{S} = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\mathcal{R} = \{\omega_1\}$, $\sigma = \rho_1^A = \rho_7^B = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, and $\rho_6^B = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0)$. Let $q = (0, \frac{1}{2}, \frac{1}{2}, 0) \in \Delta^d$. Then, $\rho(q) = \rho_1^A$ under A4, but $\rho(q) = \rho_6^B$ under A5. Straightforward calculations yield

$$p^A(\omega_3; q) = \frac{1}{2} \quad \text{and} \quad p^B(\omega_3; q) = \frac{2}{3}.$$

For $\varepsilon \in (0, 1)$, consider $q_\varepsilon = (0, \frac{1-\varepsilon}{2}, \frac{1-\varepsilon}{2}, \varepsilon) \in \Delta^d$. Then, under both A4 and A5, we have

$$p^A(\omega_3; q_\varepsilon) = p^B(\omega_3; q_\varepsilon) = \frac{1-\varepsilon}{2},$$

which converges to $\frac{1}{2}$ as $\varepsilon \to 0$.

6.3 Alternative Differences in Priors' Supports

The support Receiver's prior need not be a subset of that of Sender's prior ($\mathcal{R} \subsetneq \mathcal{S}$). This case, however, comprises all key aspects of the different-support assumption. For $\mathcal{S}, \mathcal{R} \subset \Omega$, consider the alternatives (i) $\mathcal{S} \subsetneq \mathcal{R}$, (ii) $\mathcal{S} \cap \mathcal{R} \neq \varnothing$ but $\mathcal{S} \not\subset \mathcal{R}$ and $\mathcal{R} \not\subset \mathcal{S}$, and (iii) $\mathcal{S} \cap \mathcal{R} = \varnothing$. Case (i) is trivial and uninteresting for our purposes, as Receiver can never be surprised and Bayes' rule always applies. For the other cases, suppose first that Sender cannot provide information on states outside $\mathcal{S}$. In this setting, we can add a default signal—'no evidence'—which arises whenever $\omega \notin \mathcal{S}$. Then, case (ii) is equivalent to $\mathcal{R} \subsetneq \mathcal{S}$ because Sender cannot affect Receiver's beliefs for states outside $\mathcal{S}$ and thinks that such states are impossible. Similar considerations apply for case (iii), except that Receiver will always be surprised. If Sender can provide information on states outside $\mathcal{S}$, in both case (ii) and (iii) we can always rely on Bayes' rule whenever Receiver is not surprised; when he is, the problem is the same as with $\mathcal{R} \subsetneq \mathcal{S}$.

A Appendix: Discontinuity of Receiver's Posterior over Δ^d under Assumption 1

Given $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $\sigma = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, let $\mathbf{supp}\, \mu = (\rho_0, \rho_1, \rho_2)$ with $\mu(\rho_0) = \frac{1}{2}$, $\mu(\rho_1) = \mu(\rho_2) = \frac{1}{4}$, $\mathcal{R} = \{\omega_1\}$, $\rho_1 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0)$, and $\rho_2 = \sigma$. Consider Sender's posterior $q_z = (0, \frac{1-z}{2}, \frac{1-z}{2}, z) \in \Delta^d$ for $z \in (0, 1)$. Then,

$$\mu(\rho_1; q_z) = \frac{\sum_{\omega \in \Omega} q_z(\omega)\rho_1(\omega)}{\sum_{\omega \in \Omega} q_z(\omega)\rho_1(\omega) + \sum_{\omega \in \Omega} q_z(\omega)\rho_2(\omega)} = \frac{1}{1 + \frac{2}{3(1-z)}},$$

$$\mu(\rho_2; q_z) = 1 - \mu(\rho_1; q_z).$$

Hence, $\mu(\rho_1; q_z) \geq \mu(\rho_2; q_z)$ if and only if $z \leq \frac{1}{3}$. For $z = \frac{1}{3}$ Receiver will choose either ρ_1 or ρ_2 depending on how he ranks them under $\succ$.

Using Lemma 1, we can compute Receiver's posteriors starting with ρ_1 and ρ_2 when Sender has posterior q_z. Focusing on ω_3, we have

$$p_1(\omega_3; q_z) = \frac{(1-z)\frac{1}{2}}{(1-z)\frac{1}{2} + (1-z)\frac{1}{4}} = \frac{2}{3},$$

$$p_2(\omega_3; q_z) = \frac{(1-z)\frac{1}{4}}{(1-z)\frac{1}{4} + (1-z)\frac{1}{4} + 2z\frac{1}{4}} = \frac{1-z}{2}.$$

So Receiver's posterior must vary discontinuously in q_z at $z = \frac{1}{3}$.

B Appendix: Proofs

B.1 Proof of Proposition 1

If $q(\cdot|x, \pi) \in \Delta^c$, then $p(\cdot|x, \pi)$ results from updating $\rho_0 \in \Delta(\mathcal{R})$ using Bayes' rule. Hence, $\mathbf{supp}\, p(\cdot|x, \pi) \subset \mathcal{R}$. If $q(\cdot|x, \pi) \in \Delta^d$, then under A1, we have that $p(\cdot|x, \pi)$ results from updating some prior $\rho \in \Delta(\Omega)$ using Bayes' rule and hence $\mathbf{supp}\, p(\cdot|x, \pi) \subset \Omega_\pi(x)$. Finally, by definition of Δ^d, we have $\Omega_\pi(x) = \mathbf{supp}\, q(\cdot|x, \pi) \subset \overline{\mathcal{R}}$.

B.2 Proof of Lemma 1

This steps generalize the proof of Proposition 1 in Alonso and Câmara (2013). Take any π, $x \in X_\pi$, and $\rho \in \Delta(\Omega)$ such that $\sum_{\omega' \in \Omega} \pi(x|\omega')\rho(\omega') > 0$. Applying Bayes' rule, we get

$$p(\omega|x,\pi) = \frac{\pi(x|\omega)\rho(\omega)}{\sum_{\omega'\in\Omega}\pi(x|\omega')\rho(\omega')} \quad \text{for all } \omega \in \Omega.$$

For any $\omega \in \Omega$, since $\sigma(\omega) > 0$, using (2) we can write

$$\pi(x|\omega)\rho(\omega) = q(\omega|x,\pi)\frac{\rho(\omega)}{\sigma(\omega)}\left[\sum_{\omega'\in\Omega}\pi(x|\omega')\sigma(\omega')\right].$$

Hence,

$$\sum_{\omega\in\Omega}\pi(x|\omega)\rho(\omega) = \left[\sum_{\omega'\in\Omega}\pi(x|\omega')\sigma(\omega')\right]\left[\sum_{\omega\in\Omega}q(\omega|x,\pi)\frac{\rho(\omega)}{\sigma(\omega)}\right].$$

Substituting and simplifying, we obtain that

$$p(\omega|x,\pi) = \frac{q(\omega|x,\pi)\frac{\rho(\omega)}{\sigma(\omega)}}{\sum_{\omega'\in\Omega}q(\omega'|x,\pi)\frac{\rho(\omega')}{\sigma(\omega')}} \quad \text{for all } \omega \in \Omega.$$

B.3 Proof of Corollary 2

By Proposition 1, $\mathbf{supp}\, p(q) \cap \mathbf{supp}\, p(q') = \varnothing$. Therefore,

$$\left|\left|p(q) - p(q')\right|\right|^2 = \sum_{\omega\in\mathcal{R}}|p(\omega;q)|^2 + \sum_{\omega\in\overline{\mathcal{R}}}\left|p(\omega;q')\right|^2 > \sum_{\omega\in\overline{\mathcal{R}}}\left|p(\omega;q')\right|^2 > 0.$$

B.4 Proof of Lemma 2

For $q \in \Delta(\mathcal{S})$, we have $w(q) > -\infty$ by continuity of u_S and compactness of A. For all $q \in \mathbb{R}^{|\mathcal{S}|-1}\setminus\Delta(\mathcal{S})$ define $w(q) = -\infty$. By Carathéodory's Theorem (see Rockafellar (1997), Corollary 17.1.5),

$$\hat{w}(\sigma) = \sup_{T_\sigma}\sum_m \tau_m w(q_m),$$

where

$$T_\sigma = \{\{(q_m,\tau_m)\}_{m=1}^{|\mathcal{S}|} : \sum_{m=1}^{|\mathcal{S}|}\tau_m q_m = \sigma, \sum_{m=1}^{|\mathcal{S}|}\tau_m = 1, \tau_m \geq 0, q_m \in \Delta(\mathcal{S}), \forall m\}.$$

Since $T_\sigma \subset \mathcal{F}_\sigma$, it follows that $\hat{w}(\sigma) \leq W_\sigma$. By definition of W_σ, for every $\varepsilon > 0$, there exists $\tau_\varepsilon \in \mathcal{F}_\sigma$ such that $V(\tau_\varepsilon) \geq W_\sigma - \varepsilon$. However, $V(\tau_\varepsilon) \in \{\xi : (\sigma,\xi) \in \mathrm{co}(\mathrm{hyp}\, w)\}$ and hence

$V(\tau_\varepsilon) \leq \hat{w}(\sigma)$. So for every $\varepsilon > 0$, $\hat{w}(\sigma) \geq W_\sigma - \varepsilon$ which implies that $\hat{w}(\sigma) \geq W_\sigma$.

B.5 Proof of Lemma 3

The first part follows from Lemma 2. For the second part, note that by the same argument as in the proof of Lemma 2,

$$\hat{w}^c(\sigma) = \sup_{T_\sigma^c} \sum_m \tau_m w(q_m),$$

where

$$T_\sigma^c = \{\{(q_m, \tau_m)\}_{m=1}^N : N \geq 1, \sum_{m=1}^N \tau_m q_m = \sigma, \sum_{m=1}^N \tau_m = 1, \tau_m \geq 0, q_m \in \Delta^c, \forall m\}.$$

Suppose $V(\tau) > \hat{w}^c(\sigma)$ but $\tau(D_\tau) = 0$. Since $\tau \in \mathcal{F}_\sigma$, $|\mathbf{supp}\,\tau| = N$ for some finite N and hence $D_\tau = \varnothing$. Therefore, $\tau \in T_\sigma^c$ and hence $V(\tau) \leq \hat{w}^c(\sigma)$. A contradiction.

B.6 Proof of Proposition 3

Part (1): Take any $\tau \in \mathcal{F}_\sigma^c$ with $q \in \mathbf{supp}\,\tau$ such that $q(\omega) > 0$ and $q(\omega') > 0$ for some $\omega, \omega' \in \overline{\mathcal{R}}$ with $\omega \neq \omega'$. We will show that there exists another $\tau_1 \in \mathcal{F}^c$ which satisfies the first part of (1) and such that $V(\tau_1) \geq V(\tau)$. Since $\tau \in \mathcal{F}_\sigma^c$, there exists $\pi \in \Pi$ that induces τ, i.e., for every $q \in \mathbf{supp}\,\tau$ there exists x such that $q = q(\cdot|x, \pi)$ induced by π through Bayes' rule. For every $q \in \mathbf{supp}\,\tau$, let $\overline{\mathcal{R}}(q) = \{\omega \in \overline{\mathcal{R}} : q(\omega) > 0\}$. By assumption, $|\overline{\mathcal{R}}(\hat{q})| > 1$ for some $\hat{q} \in \mathbf{supp}\,\tau$.

For any such $\hat{q}$ do the following. Let $\hat{x}$ be the signal inducing it under π, i.e., $\hat{q} = q(\cdot|\hat{x}, \pi)$. Clearly, $\pi(\hat{x}|\omega) > 0$ if and only if $\omega \in \mathbf{supp}\, q(\cdot|\hat{x}, \pi)$ with includes as a strict subset $\overline{\mathcal{R}}(q(\cdot|\hat{x}, \pi))$ because $q(\cdot|\hat{x}, \pi) \in \Delta^c$. Modify π to π' as follows. For each $\omega \in \overline{\mathcal{R}}(q(\cdot|\hat{x}, \pi))$, create a signal x_ω with the following properties: (i) $\pi'(x_\omega|\omega) = \pi(\hat{x}|\omega)$, (ii) $\pi'(x_\omega|\omega') = 0$ for all $\omega' \in \overline{\mathcal{R}}(q(\cdot|\hat{x}, \pi)) \setminus \{\omega\}$, (iii) for each $\tilde{\omega} \in \mathcal{R}$

$$\pi'(x_\omega|\tilde{\omega}) = \pi(\hat{x}|\tilde{\omega}) \frac{\pi(\hat{x}|\omega)}{\sum_{\omega' \in \overline{\mathcal{R}}(q(\cdot|\hat{x}, \pi))} \pi(\hat{x}|\omega')}.$$

Note that π' is a well-defined signal device, since $\pi(\cdot|\omega)$ is a probability distribution over finitely many signals for every ω. By construction, for every $\omega \in \overline{\mathcal{R}}(q(\cdot|\hat{x}, \pi))$, $q(\omega'|x_\omega, \pi') = 0$ if $\omega' \in \overline{\mathcal{R}}(q(\cdot|\hat{x}, \pi)) \setminus \{\omega\}$ and for every $\tilde{\omega} \in \mathcal{R}$

$$\begin{aligned} q(\tilde{\omega}|x_\omega, \pi', \mathcal{R}) &= \frac{\pi'(x_\omega|\tilde{\omega})\sigma(\tilde{\omega}|\mathcal{R})}{\sum_{\omega' \in \mathcal{R}} \pi'(x_\omega|\omega')\sigma(\omega'|\mathcal{R})} \\ &= \frac{\pi(\hat{x}|\tilde{\omega})\sigma(\tilde{\omega}|\mathcal{R})}{\sum_{\omega' \in \mathcal{R}} \pi(\hat{x}|\omega')\sigma(\omega'|\mathcal{R})} = q(\tilde{\omega}|\hat{x}, \pi, \mathcal{R}). \end{aligned}$$

This implies that $\mathcal{A}(p(q(\cdot|x_\omega,\pi'))) = \mathcal{A}(p(q(\cdot|\hat{x},\pi)))$ for every $\omega \in \overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))$. Let the total probability that $q(\cdot|x_\omega,\pi')$ arises under π' be $\beta'(x_\omega) = \sum_{\omega' \in \mathcal{S}} \pi'(x_\omega|\omega')\sigma(\omega')$ and note that

$$\begin{aligned}\sum_{\omega\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} \beta'(x_\omega) &= \sum_{\omega\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} \pi(\hat{x}|\omega)\sigma(\omega) + \sum_{\omega\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} \left[\sum_{\tilde{\omega}\in\mathcal{R}} \pi'(x_\omega|\tilde{\omega})\sigma(\tilde{\omega})\right] \\ &= \sum_{\omega\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} \pi(\hat{x}|\omega)\sigma(\omega) + \sum_{\tilde{\omega}\in\mathcal{R}} \pi(\hat{x}|\tilde{\omega})\sigma(\tilde{\omega}) = \beta(\hat{x}),\end{aligned}$$

i.e., the probability that $q(\cdot|\hat{x},\pi)$ arises under π. Note also that $q(\cdot|\hat{x},\pi) = \sum_{\omega\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} q(\cdot|x_\omega,\pi')\frac{\beta'(x_\omega)}{\beta(\hat{x})}$, so $\hat{q}$ is the conditional expectation of posteriors $\{q(\cdot|x_\omega,\pi')\}_{\omega\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))}$. Indeed, if $\omega \in \overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))$, then $q(\omega|x_{\omega'}) = 0$ for all $\omega' \neq \omega$ and

$$q(\omega|x_\omega,\pi')\frac{\beta'(x_\omega)}{\beta(\hat{x})} = \frac{\pi(\hat{x}|\omega)\sigma(\omega)}{\beta'(x_\omega)}\frac{\beta'(x_\omega)}{\beta(\hat{x})} = q(\omega|\hat{x},\pi);$$

if $\omega \in \mathcal{R}$, then

$$\sum_{\omega'\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} \frac{\pi'(x_{\omega'}|\omega)\sigma(\omega)}{\beta(\hat{x})} = \sum_{\omega'\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))} \frac{\pi(\hat{x}|\omega)\sigma(\omega)\frac{\pi(\hat{x}|\omega')}{\sum_{\omega''\in\overline{\mathcal{R}}(q(\cdot|\hat{x},\pi))}\pi(\hat{x}|\omega'')}}{\beta(\hat{x})} = q(\omega|\hat{x},\pi).$$

In summary, the distribution π' replaces the posterior $\hat{q}$ allocating its probability $\tau(\hat{q})$ across a collection of posteriors $q_\omega = q(\cdot|x_\omega,\pi')$, each with probability $\tau'(q_\omega)$, such that $\hat{q} = \sum_{\omega\in\overline{\mathcal{R}}(\hat{q})} q_\omega \frac{\tau'(q_\omega)}{\tau(\hat{q})}$. Note that $\tau'(q) = \tau(q)$ for all other $q \in \mathbf{supp}\,\tau$, hence Sender's payoff changes only when posterior $\hat{q}$ arises. We want to show that this change can only be a (weak) improvement. Given posterior $\hat{q}$, let $a(\hat{q}) \in \mathcal{A}(p(\hat{q}))$ be Receiver's action. Then, Sender's conditional expected payoff from the distribution over q_ω's is

$$\begin{aligned}\sum_{\omega\in\overline{\mathcal{R}}(\hat{q})} \left\{\max_{a\in\mathcal{A}(p(\hat{q}))} \sum_{\tilde{\omega}\in\mathcal{S}} u_S(a,\tilde{\omega})q_\omega(\tilde{\omega})\right\}\frac{\tau'(q_\omega)}{\tau(\hat{q})} &\geq \sum_{\omega\in\overline{\mathcal{R}}(\hat{q})} \left\{\sum_{\tilde{\omega}\in\mathcal{S}} u_S(a(\hat{q}),\tilde{\omega})q_\omega(\tilde{\omega})\right\}\frac{\tau'(q_\omega)}{\tau(\hat{q})} \\ &= \sum_{\tilde{\omega}\in\mathcal{S}} u_S(a(\hat{q}),\tilde{\omega})\left\{\sum_{\omega\in\overline{\mathcal{R}}(\hat{q})} q_\omega(\tilde{\omega})\frac{\tau'(q_\omega)}{\tau(\hat{q})}\right\} \\ &= \sum_{\tilde{\omega}\in\mathcal{S}} u_S(a(\hat{q}),\tilde{\omega})\hat{q}(\tilde{\omega}),\end{aligned}$$

namely Sender's expected payoff at posterior $\hat{q}$.

This construction can be replicated for all $\hat{q} \in \mathbf{supp}\,\tau$ with $\left|\overline{\mathcal{R}}(\hat{q})\right| > 1$, leading to a new distribution τ_1 over Δ^c such that $V(\tau_1) \geq V(\tau)$ and, by construction, for every $q \in \mathbf{supp}\,\tau_1$ there exists at most one $\omega \in \overline{\mathcal{R}}$ such that $q(\omega) \in (0,1)$. For any such q, $q(\mathcal{R}) = \sum_{\omega'\in\mathcal{R}} q(\omega') = 1 - q(\omega)$. Hence, $q(\omega') = (1-q(\omega))q(\omega'|\mathcal{R})$ for every $\omega' \in \mathcal{R}$ and $q(\omega) = q(\omega)\delta_\omega$, so that $q = (1-q(\omega))q(\cdot|\mathcal{R}) + q(\omega)\delta_\omega$.

Part (2): Take any $\tau \in \mathcal{F}_\sigma$ with $\mathbf{supp}\,\tau \subset \Delta^c$ and satisfying property (1) in the proposition. For every $\omega \in \overline{\mathcal{R}}$ there must exist at least one $q \in \mathbf{supp}\,\tau$ with $q(\omega) > 0$: in the device π leading to τ, $\pi(\cdot|\omega)$ must assign positive probability to some signal $x \in \mathbf{supp}\,\pi(\cdot|\omega')$ for some $\omega' \in \mathcal{R}$. Let $Q(\omega) = \{q \in \mathbf{supp}\,\tau : q(\omega) > 0\}$ and suppose that for some $\omega^* \in \overline{\mathcal{R}}$ we have $|Q(\omega^*)| > 1$. Then let $T^* = \sum_{q\in Q(\omega^*)} \tau(q)$ and

$$\begin{aligned} q^* = \sum_{q\in Q(\omega^*)} q\tau(q|Q(\omega^*)) \quad &= \sum_{q\in Q(\omega^*)} [(1-q(\omega^*))q(\cdot|\mathcal{R}) + q(\omega^*)\delta_{\omega^*}]\,\tau(q|Q(\omega^*)) \\ &= \sum_{q\in Q(\omega^*)} q(\cdot|\mathcal{R})(1-q(\omega^*))\tau(q|Q(\omega^*)) \\ &\quad +\delta_{\omega^*} \sum_{q\in Q(\omega^*)} q(\omega^*)\tau(q|Q(\omega^*)), \end{aligned}$$

where the second equality follows from Part (1). So, q^* arises with probability T^* and is the convex combination of the posteriors δ_{ω^*} and $\{q(\cdot|\mathcal{R})\}_{q\in Q(\omega^*)}$.

Now consider Sender's expected payoff conditional on $Q(\omega^*)$, letting $a(q)$ be Receiver's choice at each $q \in Q(\omega^*)$:

$$\begin{aligned} &\sum_{q\in Q(\omega^*)} \{\sum_{\omega\in\mathcal{S}} u_S(a(q),\omega)q(\omega)\}\tau(q|Q(\omega^*)) \\ =&\sum_{q\in Q(\omega^*)} \{(1-q(\omega^*))\sum_{\omega\in\mathcal{S}} u_S(a(q),\omega)q(\omega|\mathcal{R}) + q(\omega^*)u_S(a(q),\omega^*)\}\tau(q|Q(\omega^*)) \\ =&\sum_{q\in Q(\omega^*)} \{\sum_{\omega\in\mathcal{S}} u_S(a(q),\omega)q(\omega|\mathcal{R})\}(1-q(\omega^*))\tau(q|Q(\omega^*)) \\ &+\sum_{q\in Q(\omega^*)} u_S(a(q),\omega^*)q(\omega^*)\tau(q|Q(\omega^*)) \\ =&(1-\xi(\omega^*))\left\{\sum_{q\in Q(\omega^*)} \{\sum_{\omega\in\mathcal{S}} u_S(a(q),\omega)q(\omega|\mathcal{R})\}\frac{(1-q(\omega^*))\tau(q|Q(\omega^*))}{1-\xi(\omega^*)}\right\} \\ &+\xi(\omega^*)\left\{\sum_{q\in Q(\omega^*)} u_S(a(q),\omega^*)\frac{q(\omega^*)\tau(q|Q(\omega^*))}{\xi(\omega^*)}\right\}, \end{aligned} \tag{20}$$

where $\xi(\omega^*) = \sum_{q\in Q(\omega^*)} q(\omega^*)\tau(q|Q(\omega^*))$. Now recall that $\mathcal{A}(p(q))$ depends only on $q(\cdot|\mathcal{R})$, so any change in q which leaves $q(\cdot|\mathcal{R})$ unaffected does not change the actions Sender can make Receiver choose. Expression (20) can only increase if, for every $q \in Q(\omega^*)$, we replace $u_S(a(q),\omega^*)$ with $u_S(a(\tilde{q}),\omega^*) \equiv \max_{q'\in Q(\omega^*)} u_S(a(q'),\omega^*)$—that is, we shift the entire weight $\xi(\omega^*)$ to the largest $u_S(a(q),\omega^*)$. This change modifies τ to τ' as follows. For every $q \notin Q(\omega^*)$, $\tau'(q) = \tau(q)$. Each $q \in Q(\omega^*)$ with $q \neq \tilde{q}$ is replaced by $q' = q(\cdot|\mathcal{R})$ and $\tilde{q}$ is replaced by

$$\tilde{q}' = \tilde{q}(\cdot|\mathcal{R})\frac{(1-\tilde{q}(\omega^*))\tau(\tilde{q}|Q(\omega^*))}{(1-\tilde{q}(\omega^*))\tau(\tilde{q}|Q(\omega^*))+\xi(\omega^*)} + \delta_{\omega^*}\frac{\xi(\omega^*)}{(1-\tilde{q}(\omega^*))\tau(\tilde{q}|Q(\omega^*))+\xi(\omega^*)}.$$

Moreover, letting $Q' = \mathbf{supp}\,\tau' \setminus \overline{Q(\omega^*)}$ (where $\overline{Q(\omega^*)}$ is the complement of $Q(\omega^*)$), we have $\tau'(q'|Q') = (1-q(\omega^*))\tau(q|Q(\omega^*))$ and $\tau'(\tilde{q}'|Q') = (1-\tilde{q}(\omega^*))\tau(\tilde{q}|Q(\omega^*))+\xi(\omega^*)$. By construction, $\sum_{q'\in Q'} q'\tau'(q'|Q') = q^*$ and $\sum_{q'\in Q'} \tau'(q') = T^*$, so that $\tau' \in \mathcal{F}^c_\sigma$. Using this notation, by the previous argument, (20) is less than or equal to

$$\sum_{q\in Q(\omega^*)} \{\sum_{\omega\in\mathcal{S}} u_S(a(q),\omega)q(\omega|\mathcal{R})\}(1-q(\omega^*))\tau(q|Q(\omega^*)) + \xi(\omega^*)u_S(a(\tilde{q}),\omega^*)$$

$$= \sum_{q'\in Q'} \{\sum_{\omega\in\mathcal{S}} u_S(a(q'),\omega)q'(\omega)\}\tau(q'|Q') \leq \sum_{q'\in Q'} \{\max_{a\in\mathcal{A}(p(q'))} \sum_{\omega\in\mathcal{S}} u_S(a,\omega)q'(\omega)\}\tau(q'|Q'),$$

where the inequality follows because, for each q' and associated original q in the construction, $a(q) \in \mathcal{A}(p(q'))$.

This shows that, for every $\omega \in \overline{\mathcal{R}}$ such that $|Q(\omega)| > 1$ under the original $\tau \in \mathcal{F}^c_\sigma$, it is possible to modify τ to obtain $\tau' \in \mathcal{F}^c_\sigma$ such that $|Q'(\omega)| = 1$ and $V(\tau') \geq V(\tau)$.

To prove the last claim, take any $\omega \in \overline{\mathcal{R}}$ and associated $q_\omega \in \mathbf{supp}\,\tau$ such that $q_\omega(\omega) > 0$. Let $a(q_\omega)$ be Receiver's choice at q_ω. Suppose that there exists $q' \in \Delta(\mathcal{R})$ such that $u_S(a(q_\omega),\omega) < \max_{a\in\mathcal{A}(p(q'))} u_S(a,\omega) = u_S(a(q'),\omega)$. By part (1), for any $\varepsilon > 0$ small enough, we can write

$$q_\omega = \lambda_\varepsilon q_\omega(\cdot|\mathcal{R}) + (1-\lambda_\varepsilon)\left[\varepsilon q_\omega(\cdot|\mathcal{R}) + (1-\varepsilon)\delta_\omega\right],$$

where $\lambda_\varepsilon \in (0,1)$ is chosen so that $q_\omega(\omega) = (1-\lambda_\varepsilon)(1-\varepsilon)$ and hence $\lambda_\varepsilon \to 1-q_\omega(\omega)$ as $\varepsilon \to 0$. For any $z > 0$, define $q_z = \hat{q} + \frac{1}{z}(\sigma - \hat{q})$ with $\hat{q} = \sum_{q\neq q_\omega} q\frac{\tau(q)}{1-\tau(q_\omega)}$. Recall that $q_\omega = q_{\tau(q_\omega)}$. Now take posterior q' and consider $\varepsilon q' + (1-\varepsilon)\delta_\omega$. Note that, for $\varepsilon > 0$ small enough, $\varepsilon q' + (1-\varepsilon)\delta_\omega$ is arbitrarily close to $\varepsilon q_\omega(\cdot|\mathcal{R}) + (1-\varepsilon)\delta_\omega$ and hence

$$q'_\omega = \lambda_\varepsilon q_\omega(\cdot|\mathcal{R}) + (1-\lambda_\varepsilon)[\varepsilon q' + (1-\varepsilon)\delta_\omega]$$

is arbitrarily close to q_ω.

Given q_ω, there exists $z_\varepsilon \geq \tau(q_\omega)$ and $\alpha_\varepsilon \in (0,1)$ such that $q_{z_\varepsilon} \in int\Delta^c$, $q_{\alpha_\varepsilon} \in \Delta^c$, and $q_{z_\varepsilon} = \alpha_\varepsilon q_{\alpha_\varepsilon} + (1-\alpha_\varepsilon)q'_\omega$. Moreover, as $\varepsilon \to 0$, we can choose z_ε and α_ε so that $z_\varepsilon \downarrow \tau(q_\omega)$ and $\alpha_\varepsilon \downarrow 0$. Hence, we can modify τ to obtain τ_ε with

$$\tau_\varepsilon(q) = \begin{cases} \frac{1-z_\varepsilon}{1-\tau(q_\omega)}\tau(q) & \text{if } q \in \mathbf{supp}\,\tau, q \neq q_\omega \\ z_\varepsilon\alpha_\varepsilon & \text{if } q = q_{\alpha_\varepsilon} \\ z_\varepsilon(1-\alpha_\varepsilon)\lambda_\varepsilon & \text{if } q = q_\omega(\cdot|\mathcal{R}) \\ z_\varepsilon(1-\alpha_\varepsilon)(1-\lambda_\varepsilon) & \text{if } q = \varepsilon q' + (1-\varepsilon)\delta_\omega \end{cases}.$$

It can be easily checked that $\tau_\varepsilon \in \mathcal{F}_\sigma$ for every $\varepsilon > 0$.

Now consider $V(\tau)$ and $V(\tau_\varepsilon)$. Letting $k = \sum_{q \neq q_\omega} w(q)\frac{\tau(q)}{1-\tau(q_\omega)}$, we have

$$V(\tau) = \tau(q_\omega)\{(1-q_\omega(\omega))\sum_{\omega' \in \mathcal{S}} u_S(a(q_\omega),\omega')q_\omega(\omega'|\mathcal{R}) + q_\omega(\omega)u_S(a(q_\omega),\omega)\} + (1-\tau(q_\omega))k,$$

$$\begin{aligned} V(\tau_\varepsilon) &= (1-z_\varepsilon)k + z_\varepsilon\alpha_\varepsilon \sum_{\omega' \in \mathcal{S}} u_S(a(q_{\alpha_\varepsilon}),\omega')q_{\alpha_\varepsilon}(\omega') \\ &+ z_\varepsilon(1-\alpha_\varepsilon)\lambda_\varepsilon \sum_{\omega' \in \mathcal{S}} u_S(a(q_\omega),\omega')q_\omega(\omega'|\mathcal{R}) \\ &+ z_\varepsilon(1-\alpha_\varepsilon)(1-\lambda_\varepsilon)\{\varepsilon \sum_{\omega' \in \mathcal{S}} u_S(a_\omega,\omega')q'(\omega') + (1-\varepsilon)u_S(a_\omega,\omega)\}, \end{aligned}$$

where $a_\omega = \arg\max_{a \in \mathcal{A}(p(q'))} u_S(a,\omega)$. Recall that Sender's expected payoff from any action and posterior q is finite. So,

$$\begin{aligned} \lim_{\varepsilon \to 0} V(\tau_\varepsilon) &= \tau(q_\omega)(1-q_\omega(\omega))\sum_{\omega' \in \mathcal{S}} u_S(a(q_\omega),\omega')q_\omega(\omega'|\mathcal{R}) \\ &+ \tau(q_\omega)q_\omega(\omega)u_S(a_\omega,\omega) + (1-\tau(q_\omega))k > V(\tau). \end{aligned}$$

Therefore, there exists $\varepsilon > 0$ small enough such that $V(\tau_\varepsilon) > V(\tau)$. This shows that we can focus without loss on distributions $\tau \in \mathcal{F}^c_\sigma$ such that $u_S(a(q_\omega),\omega) \geq \max_{q \in \Delta(\mathcal{R})}\{\max_{a \in \mathcal{A}(p(q))} u_S(a,\omega)\}$ for every $\omega \in \overline{\mathcal{R}}$.

B.7 Proof of Corollary 3

Hereafter, let $\hat{w}^c_*(\sigma)$ be the expression of $\hat{w}^c(\sigma)$ in the statement and $\mathcal{F}^{c*}_\sigma \subset \mathcal{F}^c_\sigma$ be the family of all feasible distributions satisfying the properties in Proposition 3.

Claim 1. $V(\tau) \leq \hat{w}^c_*(\sigma)$ for any $\tau \in \mathcal{F}^{c*}_\sigma$.

Proof. Given τ, for each $\omega \in \overline{\mathcal{R}}$, let q_ω be the unique posterior in $\mathbf{supp}\,\tau$ assigning positive probability to ω as in part (2) of Proposition 3. Also, let $\overline{Q} = \{q_\omega\}_{\omega \in \overline{\mathcal{R}}}$, $Q = \mathbf{supp}\,\tau \setminus \overline{Q}$, and $a(q_\omega) = a_\omega \in \arg\max_{a \in \mathcal{A}(p(q_\omega))} u_S(a,\omega)$ for every $\omega \in \overline{\mathcal{R}}$. Then,

$$\begin{aligned} V(\tau) &= \sum_{q \in Q} w(q)\tau(q) \\ &+ \sum_{q_\omega \in \overline{Q}} \left[\sum_{\tilde{\omega} \in \mathcal{S}} u_S(a_\omega,\tilde{\omega})\{(1-q_\omega(\omega))q_\omega(\tilde{\omega}|\mathcal{R}) + q_\omega(\omega)u_S(a_\omega,\omega)\}\right]\tau(q_\omega) \\ &= \sum_{q \in Q} w(q)\tau(q) + \sum_{q_\omega \in \overline{Q}} \left[\sum_{\tilde{\omega} \in \mathcal{S}} u_S(a_\omega,\tilde{\omega})q_\omega(\tilde{\omega}|\mathcal{R})\right](1-q_\omega(\omega))\tau(q_\omega) \end{aligned}$$

$$
\begin{aligned}
&+\sum_{q_\omega\in\overline{Q}} u_S(a_\omega,\omega)q_\omega(\omega)\tau(q_\omega)\\
\leq\ &\sum_{q\in Q} w(q)\tau(q)+\sum_{q_\omega\in\overline{Q}} w(q_\omega(\cdot|\mathcal{R}))(1-q_\omega(\omega))\tau(q_\omega) \qquad (21)\\
&+\sum_{q_\omega\in\overline{Q}} u_S^*(\omega)q_\omega(\omega)\tau(q_\omega).
\end{aligned}
$$

Since

$$
\sigma=\sum_{q\in Q} q\tau(q)+\sum_{q_\omega\in\overline{Q}} q_\omega(\cdot|\mathcal{R})(1-q_\omega(\omega))\tau(q_\omega)+\sum_{q_\omega\in\overline{Q}} \delta_\omega q_\omega(\omega)\tau(q_\omega),
$$

we have $\sigma(\omega)=q_\omega(\omega)\tau(q_\omega)$ for every $\omega\in\overline{\mathcal{R}}$ and $\sum_{q_\omega\in\overline{Q}} q_\omega(\omega)\tau(q_\omega)=\sigma(\overline{\mathcal{R}})$. This in turn implies that for each $\tilde{\omega}\in\mathcal{R}$,

$$
\frac{1}{\sigma(\mathcal{R})}\left[\sum_{q\in Q} q(\tilde{\omega})\tau(q)+\sum_{q_\omega\in\overline{Q}} q_\omega(\tilde{\omega}|\mathcal{R})(1-q_\omega(\omega))\tau(q_\omega)\right]=\frac{\sigma(\tilde{\omega})}{\sigma(\mathcal{R})}=\sigma(\tilde{\omega}|\mathcal{R}).
$$

The first two terms in (21) are then a convex combination of values $w^c(q)$ with $q\in\Delta(\mathcal{R})$ and with average posterior $\sigma(\cdot|\mathcal{R})$. Therefore, expression (21) is bounded above by

$$
\sigma(\mathcal{R})\hat{w}^c(\sigma(\cdot|\mathcal{R}))+\sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)\sigma(\omega)=\hat{w}_*^c(\sigma).
$$

□

Claim 2. For any $\varepsilon>0$ there exists $\tau_\varepsilon\in\mathcal{F}_\sigma^{c*}$ such that $V(\tau_\varepsilon)\geq\hat{w}_*^c(\sigma)-\varepsilon$. Hence, $\hat{w}_*^c(\sigma)$ is the least upper bound of the values of $V(\tau)$ over $\mathcal{F}_\sigma^{c*}$.

Proof. Starting from any $\tau\in\mathcal{F}_\sigma^{c*}$, first construct a sequence $\{\tau_n\}_{n=1}^\infty\subset\mathcal{F}_\sigma^{c*}$ with $\tau_0=\tau$ as follows. Define Q and $\overline{Q}$ as in the proof of Claim 1. For every $q\in Q$, let $\tau_n(q)=\tau_0(q)$. For each $q_\omega\in\overline{Q}$ and each $n\geq 1$, split $\tau_0(q_\omega)$ by replacing it with

$$
\tau_n(q')=\begin{cases}\tau_0(q_\omega)z_{\omega,n} & \text{for } q'=q_\omega(\cdot|\mathcal{R})\\ \tau_0(q_\omega)(1-z_{\omega,n}) & \text{for } q'=q_{\omega,n}\equiv\frac{1}{K_\omega^n}q_\omega(\cdot|\mathcal{R})+(1-\frac{1}{K_\omega^n})\delta_\omega\end{cases},
$$

with $z_{\omega,n}\in(0,1)$ so that $q_\omega(\omega)=(1-z_{\omega,n})(1-\frac{1}{K_\omega^n})$ for every n, where $K_\omega>1$ is chosen large enough so as to satisfy this condition for $n=1$ and hence for all $n\geq 1$. By construction, $z_{\omega,n}\uparrow(1-q_\omega(\omega))$ and $q_{\omega,n}\to\delta_\omega$ as $n\to\infty$, and for every n

$$
q_\omega=z_{\omega,n}q_\omega(\cdot|\mathcal{R})+(1-z_{\omega,n})q_{\omega,n}.
$$

So, for every n, the conditional distribution defined by $z_{\omega,n+1}$ is a mean-preserving spread around

q_ω of that defined by $z_{\omega,n}$. It follows that, for every $q_\omega \in \overline{Q}$,

$$\begin{aligned}\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_\omega(\tilde\omega) \quad \le \quad & z_{\omega,1}\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_\omega(\tilde\omega|\mathcal{R}) \\ & +(1-z_{\omega,1})\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_{\omega,1}(\tilde\omega),\end{aligned}$$

and for all $n \ge 1$

$$\begin{aligned}& z_{\omega,n}\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_{\omega,n}(\tilde\omega|\mathcal{R}) + (1-z_{\omega,n})\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_{\omega,n}(\tilde\omega) \\ \le & z_{\omega,n+1}\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_\omega(\tilde\omega|\mathcal{R}) + (1-z_{\omega,n+1})\max_{a\in\mathcal{A}(p(q_\omega))}\sum_{\tilde\omega\in\mathcal{S}} u_S(a,\tilde\omega)q_{\omega,n+1}(\tilde\omega).\end{aligned}$$

For every n, letting $Q'_n = \mathbf{supp}\,\tau_n \setminus Q$, by construction we also have $\sum_{q_\omega\in\overline{Q}} q_\omega\tau_0(q_\omega) = \sum_{q'\in Q'_n} q'\tau_n(q')$ and $\sum_{q_\omega\in\overline{Q}}\tau_0(q_\omega) = \sum_{q'\in Q'_n}\tau_n(q')$—hence each $\tau_n \in \mathcal{F}^{c*}_\sigma$. We conclude that $V(\tau_n) \le V(\tau_{n+1})$ for every n.

Now, for each n, let $Z_n = \sum_{q\in Q}\tau_n(q) + \sum_{\omega\in\overline{\mathcal{R}}}\tau_0(q_\omega)z_{\omega,n}$ and express $V(\tau_n)$ as

$$\underbrace{Z_n\sum_{q\in Q} w(q)\frac{\tau_0(q)}{Z_n} + \sum_{\omega\in\overline{\mathcal{R}}} w(q_\omega(\cdot|\mathcal{R}))\frac{\tau_0(q_\omega)z_{\omega,n}}{Z_n}}_{B_n} + \underbrace{\sum_{\omega\in\overline{\mathcal{R}}} w(q_{\omega,n})\tau_0(q_\omega)(1-z_{\omega,n})}_{B'_n}.$$

Since $z_{\omega,n}\uparrow(1-q_\omega(\omega))$ for every $\omega\in\overline{\mathcal{R}}$ as $n\to\infty$ and we must have

$$\sigma(\omega) = \tau_0(q_\omega)(1-z_{\omega,n})(1-\frac{1}{K^n_\omega}) = \tau(q_\omega)q_\omega(\omega)$$

for all n, it follows that $\lim_{n\to\infty} B'_n = \sum_{\omega\in\overline{\mathcal{R}}} u^*_S(\omega)\sigma(\omega)$. Note also that $\lim_{n\to\infty} Z_n = 1 - \lim_{n\to\infty}\sum_{\omega\in\overline{\mathcal{R}}}\tau_0(q_\omega)(1-z_{\omega,n}) = \sigma(\mathcal{R})$.

Regarding the term B_n, first observe that

$$\frac{1}{Z_n}\left[\sum_{q\in Q} q\tau_n(q) + \sum_{\omega\in\overline{\mathcal{R}}} q_\omega(\cdot|\mathcal{R})\tau_0(q_\omega)z_{\omega,n}\right] = \hat q_n \in \Delta(\mathcal{R}),$$

and hence $\lim_{n\to\infty}\hat q_n = \sigma(\cdot|\mathcal{R})$. This implies that, for every n, B_n is a convex combination of values $w^c(q)$ with $q\in\Delta(\mathcal{R})$ and with average posterior $\hat q_n$. Recall that, restricted to $\Delta(\mathcal{R})$, the function $\hat w^c(q)$ is continuous in q. Letting $\chi = \hat w^c(\sigma(\cdot|\mathcal{R}))$ and $\zeta = \sum_{\omega\in\overline{\mathcal{R}}} u^*_S(\omega)\sigma(\omega)$ for simplicity, we can write

$$\hat w^c_*(\sigma) - V(\tau_n) \le |\chi|\,|Z_n - \sigma(\mathcal{R})| + Z_n\,|B_n - \chi| + \left|B'_n - \zeta\right|.$$

Given any $\varepsilon > 0$, there exists N_1 such that $|\chi|\,|Z_n - \sigma(\mathcal{R})| + |B'_n - \zeta| \leq \frac{\varepsilon}{2}$ for all $n \geq N_1$. Also, there exists N_2 such that $|\chi - \hat{w}^c(\hat{q}_n)| \leq \frac{\varepsilon}{2}$ for all $n \geq N_2$. So, fix $n^* \geq \max\{N_1, N_2\}$ and consider the distribution $\hat{\tau} \in \Delta(\Delta(\mathcal{R}))$ that achieves $\hat{w}^c(\hat{q}_{n^*})$. Define the distribution τ_ε as follows:

$$\tau_\varepsilon(q) = \begin{cases} Z_{n^*}\hat{\tau}(q) & \text{if } q \in \mathbf{supp}\,\hat{\tau} \\ \tau_{n^*}(q_{\omega,n^*}) & \text{if } q = q_{\omega,n^*} \\ 0 & \text{otherwise.} \end{cases}$$

By construction, $V(\tau_\varepsilon) \geq \hat{w}^c_*(\sigma) - \varepsilon$ and

$$\begin{aligned} \sum_{q \in \mathbf{supp}\,\tau_\varepsilon} q\tau_\varepsilon(q) &= Z_{n^*} \sum_{q \in \mathbf{supp}\,\hat{\tau}} q\hat{\tau}(q) + \sum_{\omega \in \overline{\mathcal{R}}} q_{\omega,n^*}\tau_{n^*}(q_{\omega,n^*}) \\ &= \sum_{q \in Q} q\tau_0(q) + \sum_{\omega \in \overline{\mathcal{R}}} q_\omega(\cdot|\mathcal{R})\tau_0(q_\omega)z_{\omega,n^*} \\ &\quad + \sum_{\omega \in \overline{\mathcal{R}}} q_{\omega,n^*}\tau_0(q_\omega)(1 - z_{\omega,n^*}) = \sigma. \end{aligned}$$

□

B.8 Proof of Proposition 4

The result follows from the next two claims.

Claim 3. If $w^d(q) \leq h(q)$ for all $q \in \Delta^d$, then $\hat{w}(\sigma) \leq \hat{w}^c(\sigma)$.

Proof. For any $\tau \in \mathcal{F}_\sigma$, let $q^c = \sum_{q \in C_\tau} q\tau(q|C_\tau)$ and $q^d = \sum_{q \in D_\tau} q\tau(q|D_\tau)$ so that $q^c \in \Delta^c$, $q^d \in \Delta^d$, $\sigma = \tau^c q^c + (1 - \tau^c)q^d$. Then,

$$\begin{aligned} V(\tau) &\leq \tau^c\hat{w}^c(q^c) + (1 - \tau^c) \sum_{q \in D_\tau} w^d(q)\tau(q|D_\tau) \\ &\leq \tau^c\hat{w}^c(q^c) + (1 - \tau^c) \sum_{q \in D_\tau} h(q)\tau(q|D_\tau) \\ &= \tau^c\hat{w}^c(q^c) + (1 - \tau^c) \sum_{q \in D_\tau} \left[\sum_{\omega \in \overline{\mathcal{R}}} u^*_S(\omega)q(\omega) \right] \tau(q|D_\tau) \\ &= \tau^c\hat{w}^c(q^c) + \sum_{\omega \in \overline{\mathcal{R}}} u^*_S(\omega)\beta(\omega), \end{aligned} \tag{22}$$

$\beta(\omega) = \tau^d \sum_{q \in D_\tau} q(\omega)\tau(q|D_\tau)$. Note that $\mathbf{supp}\,q^c \supset \mathcal{R}$ and $q^c \in int\Delta(\mathbf{supp}\,q^c)$. Therefore, we can view q^c as Sender's prior in the fictitious environment with $\tilde{\Omega} = \mathbf{supp}\,q^c$ and $\tilde{\rho} = \rho_0$. By

Proposition 3, we then have

$$\hat{w}^c(q^c) = q^c(\mathcal{R})\hat{w}^c(q^c(\cdot|\mathcal{R})) + \sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)q^c(\omega).$$

Moreover, we have

$$\sigma = \tau^c q^c(\mathcal{R})q^c(\cdot|\mathcal{R}) + \sum_{\omega\in\overline{\mathcal{R}}} \delta_\omega\{\beta(\omega) + \tau^c q^c(\omega)\},$$

which implies that $\sigma(\omega) = \beta(\omega) + \tau^c q^c(\omega)$ for all $\omega \in \overline{\mathcal{R}}$ and hence $q^c(\cdot|\mathcal{R}) = \sigma(\cdot|\mathcal{R})$. Therefore, (22) is equal to

$$\sigma(\mathcal{R})\hat{w}(\sigma(\cdot|\mathcal{R})) + \sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)\sigma(\omega) = \hat{w}^c(\sigma).$$

Using the definition of W_σ and Lemma 2, we conclude that $\hat{w}(\sigma) \leq \hat{w}^c(\sigma)$.

□

Claim 4. If $w^d(q) > h(q)$ for some $q \in \Delta^d$, then there exists τ such that $V(\tau) > \hat{w}^c(\sigma)$ and hence $\hat{w}(\sigma) > \hat{w}^c(\sigma)$.

Proof. Let q^* be any element of $\Delta(\overline{\mathcal{R}})$ with $w^d(q^*) > h(q^*)$. Since $\sigma \in int\Delta(\mathcal{S})$, there exists $\lambda \in (0,1)$ and $q^c \in int\Delta(\mathcal{S})$ such that $\sigma = \lambda q^c + (1-\lambda)q^*$. By the same argument in the proof of Claim 3

$$\begin{aligned}(1-\lambda)w^d(q^*) + \lambda\hat{w}^c(q^c) &> (1-\lambda)\sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)q^*(\omega) + \lambda\hat{w}^c(q^c)\\ &= \sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)\{(1-\lambda)q^*(\omega) + \lambda q^c(\omega)\} + \lambda q^c(\mathcal{R})\hat{w}^c(q^c(\cdot|\mathcal{R}))\\ &= \hat{w}^c(\sigma),\end{aligned}$$

where the last equality follows again from observing that $\sigma(\omega) = (1-\lambda)q^*(\omega) + \lambda q^c(\omega)$ for all $\omega \in \overline{\mathcal{R}}$ and hence $q^c(\cdot|\mathcal{R}) = \sigma(\cdot|\mathcal{R})$. Therefore, there exists $\tau \in \mathcal{F}_\sigma$ such that $V(\tau) > \hat{w}^c(\sigma)$.

□

B.9 Lemma 4

Lemma 4. *The function $\hat{w}_*^d$ satisfies the following properties:*
(i) for every $q \in \Delta^d$, there exists $\tau \in \Delta(\Delta^d)$ such that $\hat{w}_*^d = \sum_{q'} w_*^d(q')\tau(q')$ with $q = \sum_{q'} q'\tau(q')$ and $|\textbf{supp}\,\tau| \leq |\overline{\mathcal{R}}|$;
(ii) $\hat{w}^d \leq \hat{w}_*^d$ with equality over $int\Delta^d$;
(iii) $\hat{w}_*^d = \mathrm{cl}\hat{w}_*^d$ and hence it is continuous.

Proof. *Part (i):* By Corollary 17.1.5 in Rockafellar (1997),

$$\hat{w}_*^d(q) = \sup_{T(q)} \sum_{m=1}^{|\overline{\mathcal{R}}|} w_*^d(q_m)\tau_m$$

where

$$T(q) = \left\{ \{(q_m, \tau_m)\}_{m=1}^{|\overline{\mathcal{R}}|} : \sum_{m=1}^{|\overline{\mathcal{R}}|} q_m\tau_m = q, \sum_{m=1}^{|\overline{\mathcal{R}}|} \tau_m = 1, \tau_m \geq 0, q_m \in \Delta(\overline{\mathcal{R}}), \forall m \right\}.$$

Since w_*^d is upper semicontinuous and $T(q)$ is compact, by standard arguments $\hat{w}_*^d(q)$ is achieved for every $q \in \Delta^d$.

Part (ii): Given a function $f : \Delta^d \to \mathbb{R}$, let hyp$f$ be the hypograph of f: hyp$f = \{(q, \xi) : q \in \Delta^d, \xi \in \mathbb{R}, \xi \leq f(q)\}$. Note that $\text{hyp}\, w_*^d = \overline{\text{hyp}\, w^d}$. Therefore, for all $q \in \Delta^d$,

$$\begin{aligned} \hat{w}^d(q) &= \sup\{\xi : (q, \xi) \in \text{co}(\text{hyp}\, w^d)\} \\ &\leq \sup\{\xi : (q, \xi) \in \text{co}(\overline{\text{hyp}\, w^d})\} = \hat{w}_*^d(q). \end{aligned}$$

Now consider the closure of $\hat{w}^d$, $\text{cl}\hat{w}^d$, which is the unique continuous extension of $\hat{w}^d$ to Δ^d by Theorem 10.3 in Rockafellar (1997), is concave, and satisfies $\text{cl}\hat{w}^d \geq \hat{w}^d \geq w^d$. So, for every $q \in \Delta^d$,

$$w_*^d(q) = \limsup_{q' \to q} w^d(q') \leq \limsup_{q' \to q} \text{cl}\hat{w}^d(q') = \text{cl}\hat{w}^d(q).$$

Hence, $\text{cl}\hat{w}^d$ is a concave function majorizing w_*^d. Since $\hat{w}_*^d$ is the smallest of such functions, $\text{cl}\hat{w}^d \geq \hat{w}_*^d$. Finally, since $\text{cl}\hat{w}^d = \hat{w}^d$ over $int\Delta^d$, property (i) follows.

Part (iii): We already know that $\hat{w}_*^d = \text{cl}\hat{w}_*^d$ over $int\Delta^d$. By definition, $\text{hyp}\,\text{cl}\hat{w}_*^d = \overline{\text{hyp}\, \hat{w}_*^d}$. If $\text{hyp}\, \hat{w}_*^d$ is closed, then $\text{hyp}\, \hat{w}_*^d = \text{hyp}\,\text{cl}\hat{w}_*^d$ and hence we are done. Indeed, by definition $\hat{w}_*^d \leq \text{cl}\hat{w}_*^d$. So, suppose there exists $q \in \partial\Delta^d$ such that $\hat{w}_*^d(q) < \text{cl}\hat{w}_*^d(q)$. Then there exists $\xi \in \mathbb{R}$ such that $\hat{w}_*^d(q) < \xi \leq \text{cl}\hat{w}_*^d$, which is a contradiction. So, we need to prove that $\text{hyp}\, \hat{w}_*^d$ is closed.

First, for every $q \in \Delta^d$, by property $\hat{w}_*^d(q) = \max\{\xi : (q, \xi) \in \text{co}(\text{hyp}\, w_*^d)\}$ and therefore $\text{hyp}\, \hat{w}_*^d = \text{co}(\text{hyp}\, w_*^d)$. Second, define $\underline{w}_*^d = \inf_{q \in \Delta^d} w_*^d(q)$ so that we can express $\text{hyp}\, w_*^d$ as $G \cup H$ where

$$G = \{(q, \xi) : q \in \Delta^d, \underline{w}_*^d - 1 \leq \xi \leq w_*^d(q)\} \quad \text{and} \quad H = \{(q, \xi) : q \in \Delta^d, \xi \leq \underline{w}_*^d - 1\}.$$

Now $\text{co}(\text{hyp}\, w_*^d) = (\text{co}G) \cup (\text{co}H) = (\text{co}G) \cup H$. One inclusion is trivial. Now consider $(q, \xi) \in \text{co}(\text{hyp}\, w_*^d)$. Then, by Theorem 2.3 in Rockafellar (1997), (q, ξ) is a convex combination of points (q_n, ξ_n) in $\text{hyp}\, w_*^d$. Therefore, $q \in \Delta^d$ as the latter is a convex set and $\xi = \sum_n \alpha_n \xi_n \leq$

$\sum_n \alpha_n w_*^d(q_n)$ as $\alpha_n \geq 0$ for all n. But $(\text{co}G) \cup H$ contains all convex combinations of points in $\text{hyp}\, w_*^d$ that satisfy this property. Finally, note that H is closed and G is bounded and closed since w_*^d is upper semicontinuous. Therefore, $\text{co}(G)$ is also closed by Theorem 17.2 in Rockafellar (1997). We conclude that $\text{co}(\text{hyp}\, w_*^d) = (\text{co}G) \cup H$ is closed, as desired.

□

B.10 Proof of Proposition 5

Claim 5. If $\hat{w}(\sigma) > \hat{w}^c(\sigma)$, then

$$\hat{w}(\sigma) = \tau^c \hat{w}^c(q^c) + (1 - \tau^c)\hat{w}_*^d(q^d)$$

with $q^c \in \Delta^c$, $q^d \in \Delta^d$, $\sigma(\mathcal{R}) \leq \tau^c < 1$, and $\sigma = \tau^c q^c + (1 - \tau^c) q^d$.

Proof. Take any $\tau \in \mathcal{F}_\sigma$ with $\mathbf{supp}\, \tau \leq |\mathcal{S}|$ and $\tau^d = \tau(D_\tau) > 0$, which is necessary by Lemma 3. Define[51] $\tau(q|C_\tau) = \frac{\tau(q)}{\tau(C_\tau)}$ and

$$\tau(q|D_\tau) = \begin{cases} 0 & \text{if } \tau(D_\tau) = 0 \text{ or } q \notin D_\tau \\ \frac{\tau(q)}{\tau(D_\tau)} & \text{if } \tau(D_\tau) > 0 \text{ and } q \in D_\tau \end{cases}.$$

We can then write

$$V(\tau) = \tau^c \sum_{q \in C_\tau} w(q)\tau(q|C_\tau) + \tau^d \sum_{q \in D_\tau} w(q)\tau(q|D_\tau).$$

If we define $q^c = \sum_{q \in C_\tau} q\tau(q|C_\tau)$ and $q^d = \sum_{q \in D_\tau} q\tau(q|D_\tau)$, we have $q^c \in \Delta^c$, $q^d \in \Delta^d$, $\sigma = \tau^c q^c + (1 - \tau^c) q^d$. Since $V(\tau) \leq \tau^c \hat{w}^c(q^c) + (1 - \tau^c)\hat{w}^d(q^d)$, we must have

$$\hat{w}(\sigma) = W_\sigma = \sup_{\tau \in \mathcal{F}_\sigma} V(\tau) \leq \sup_{\mathcal{T}} \{\tau^c \hat{w}^c(q^c) + (1 - \tau^c)\hat{w}^d(q^d)\}, \tag{23}$$

where

$$\mathcal{T} = \{(\tau^c, q^c, q^d) : \tau^c \in [\sigma(\mathcal{R}), 1], q^c \in \Delta^c, q^d \in \Delta^d, \sigma = \tau^c q^c + (1 - \tau^c) q^d\};$$

moreover, the inequality in (23) must be an equality, otherwise there would be $\tau \in \mathcal{F}_\sigma$ with $V(\tau) > W_\sigma$. Since $\tau^c \in [\sigma(\mathcal{R}), 1]$, we must have $q^c(\omega) = \frac{1}{\tau^c}\sigma(\omega) \geq \sigma(\omega)$ for all $\omega \in \mathcal{R}$. The function $\hat{w}^c$ is continuous over $\Delta(\mathcal{R})$ and therefore, by Corollary 3, $\hat{w}^c(q^c)$ is continuous over $Q^c = \{q \in \Delta^c : q(\omega) \geq \sigma(\omega), \forall \omega \in \mathcal{R}\}$, which is a compact subset of Δ^c. Construct $\mathcal{T}'$ by replacing Δ^c with Q^c in $\mathcal{T}$. So, since $\hat{w}^d \leq \text{cl}\hat{w}^d = \text{cl}\hat{w}_*^d = \hat{w}_*^d$ and $\hat{w}_*^d$ is continuous by Lemma

[51] Recall that $\tau(C_\tau) > 0$ always by Corollary 1

4, the right-hand side of (23) equals

$$\sup_{\mathcal{T}'}\{\tau^c\hat{w}^c(q^c)+(1-\tau^c)\hat{w}^d(q^d)\}\leq\max_{\mathcal{T}'}\{\tau^c\hat{w}^c(q^c)+(1-\tau^c)\hat{w}^d_*(q^d)\}.$$

First, note that the maximum on the right-hand side must be attained at $\tau^c<1$, because by assumption $\hat{w}^c(\sigma)<\hat{w}(\sigma)$. Second, note that the inequality must be an equality. This is immediate if $\left|\overline{\mathcal{R}}\right|=1$, since in this case $\hat{w}^d=w^d=w^d_*=\hat{w}^d_*$. So, suppose $\left|\overline{\mathcal{R}}\right|>1$ and define $\Delta^d_n=\{q\in\Delta(\overline{\mathcal{R}}):q(\omega)\geq\left|\overline{\mathcal{R}}\right|^{-n},\forall\omega\in\overline{\mathcal{R}}\}$ for $n\geq1$. Construct $\mathcal{T}'_n$ by replacing Δ^d with Δ^d_n in $\mathcal{T}'$ for every n. Note that, for all n, $\Delta^d_n\subset int\Delta^d$, $\Delta^d_n\subset\Delta^d_{n+1}$, and $\Delta^d_n\to\Delta^d$ as $n\to\infty$. Since $\hat{w}^d=\hat{w}^d_*$ over $int\Delta^d$ by Lemma 4, for all n we have

$$\max_{\mathcal{T}'_n}\{\tau^c\hat{w}^c(q^c)+(1-\tau^c)\hat{w}^d_*(q^d)\}\leq\sup_{\mathcal{T}'}\{\tau^c\hat{w}^c(q^c)+(1-\tau^c)\hat{w}^d(q^d)\}.$$

Since the left-hand side forms an increasing sequence in n that converges to the maximum over $\mathcal{T}'$, the desired equality follows.

□

Claim 6. If $\hat{w}(\sigma)>\hat{w}^c(\sigma)$, then (15) holds. Moreover, if $\hat{\tau}\in\Delta(\Delta^d)$ is such that $\hat{w}^d_*(q^d)=\sum_q w^d_*(q)\hat{\tau}(q)$ and $q^d=\sum_q q\hat{\tau}(q)$, then $w_*{}^d(q)\geq h(q)$ for all $q\in\mathbf{supp}\,\hat{\tau}$ with strict inequality for some q.

Proof. Again by Lemmas 3 and Claim 5, $\tau^c\in(0,1)$ and

$$\begin{aligned}\hat{w}(\sigma)&=\tau^c\hat{w}^c(q^c)+(1-\tau^c)\hat{w}^d_*(q^d)\\&=\tau^cq^c(\mathcal{R})\hat{w}^c(q^c(\cdot|\mathcal{R}))+\tau^c\sum_{\omega\in\overline{\mathcal{R}}}u^*_S(\omega)q^c(\omega)+(1-\tau^c)\hat{w}^d_*(q^d)\end{aligned}$$

with

$$\sigma=\tau^cq^c(\mathcal{R})q^c(\cdot|\mathcal{R})+\sum_{\omega\in\overline{\mathcal{R}}}\delta_\omega\{\tau^cq^c(\omega)+(1-\tau^c)q^d(\omega)\}.$$

Therefore, $\sigma(\omega)=\tau^cq^c(\omega)+(1-\tau^c)q^d(\omega)$ for all $\omega\in\overline{\mathcal{R}}$ and hence $\tau^cq^c(\mathcal{R})=\sigma(\mathcal{R})$ and $q^c(\cdot|\mathcal{R})=\sigma(\cdot|\mathcal{R})$. Note also that $(1-\tau^c)+\tau^c\sum_{\omega\in\overline{\mathcal{R}}}q^c(\omega)=\sigma(\overline{\mathcal{R}})$ and therefore

$$\sigma=\sigma(\mathcal{R})\sigma(\cdot|\mathcal{R})+\sigma(\overline{\mathcal{R}})\left[\frac{\tau^cq^c(\overline{\mathcal{R}})}{\sigma(\overline{\mathcal{R}})}\sum_{\omega\in\overline{\mathcal{R}}}\delta_\omega q^c(\omega|\overline{\mathcal{R}})+\frac{1-\tau^c}{\sigma(\overline{\mathcal{R}})}q^d\right].$$

Hence, for every $\omega\in\overline{\mathcal{R}}$

$$\delta_\omega\frac{\tau^cq^c(\omega)}{\sigma(\overline{\mathcal{R}})}+\frac{1-\tau^c}{\sigma(\overline{\mathcal{R}})}q^d(\omega)=\sigma(\omega|\overline{\mathcal{R}}).$$

So, we obtain

$$\hat{w}(\sigma) = \sigma(\mathcal{R})\hat{w}^c(\sigma(\cdot|\mathcal{R})) + \sigma(\overline{\mathcal{R}}) \underbrace{\left[\gamma h(q^c(\cdot|\overline{\mathcal{R}})) + (1-\gamma)\hat{w}_*^d(q^d)\right]}_{\xi^*}, \tag{24}$$

where $\gamma = \frac{\tau^c q^c(\overline{\mathcal{R}})}{\sigma(\overline{\mathcal{R}})}$ and $\gamma q^c(\cdot|\overline{\mathcal{R}}) + (1-\gamma)q^d = \sigma(\cdot|\overline{\mathcal{R}})$.

Now consider any $\hat{\tau} \in \Delta(\Delta^d)$ such that $\hat{w}_*^d(q^d) = \sum_q w_*^d(q)\hat{\tau}(q)$ and $q^d = \sum_q q\hat{\tau}(q)$. Suppose $w_*^d(q') < h(q')$ for any $q' \in \mathbf{supp}\,\hat{\tau}$. Then,

$$\sum_q w_*^d(q)\hat{\tau}(q) < \sum_{\{q: q\neq q'\}} w_*^d(q)\hat{\tau}(q) + \hat{\tau}(q') \sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)q'(\omega).$$

But this implies the existence of $\tau \in \mathcal{F}_\sigma$ with $V(\tau) > W_\sigma$, a contradiction. So, for all $q \in \mathbf{supp}\,\hat{\tau}$, $w_*^d(q) \geq h(q)$. Finally, suppose $w_*^d(q) = h(q)$ for all $q \in \mathbf{supp}\,\hat{\tau}$. Then ξ^* becomes

$$\gamma h(q^c(\cdot|\overline{\mathcal{R}})) + (1-\gamma)\sum_q h(q)\hat{\tau}(q) = \sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)\sigma(\omega|\overline{\mathcal{R}}),$$

and hence $\hat{w}(\sigma) = \hat{w}^c(\sigma)$ by Corollary 3, contradicting $\hat{w}(\sigma) > \hat{w}^c(\sigma)$.

Similarly, in (24) we must have $h(q^c(\cdot|\overline{\mathcal{R}})) \geq w_*^d(q^c(\cdot|\overline{\mathcal{R}}))$ because otherwise it would again be possible to improve upon W_σ. Hence, ξ^* in (24) belongs to the set $\{\xi : (\sigma(\cdot|\overline{\mathcal{R}}), \xi) \in \text{co}(\max\{h, w_*^d\})\}$ and must equal to its maximum, which exists and equals $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$.

□

Claim 7. $\tau^c = \sigma(\mathcal{R})$ if and only if $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = \hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}}))$.

Proof. If $1 - \tau^c = \sigma(\overline{\mathcal{R}})$, then $\gamma = 0$ in (24). This implies that $q^c(\overline{\mathcal{R}}) = 0$, $q^d = \sigma(\cdot|\overline{\mathcal{R}})$, and $\xi^* = \hat{w}_*^d(q^d)$; moreover, since $\sigma(\cdot|\overline{\mathcal{R}}) \in int\Delta^d$, $\hat{w}_*^d(\sigma(\cdot|\overline{\mathcal{R}})) = \hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}}))$ by Lemma 4. Conversely, suppose $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = \hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}}))$. Then, $\xi^* = \hat{w}_*^d(\sigma(\cdot|\overline{\mathcal{R}}))$ in (24) and hence $\gamma = 0$, which implies that $q^c(\overline{\mathcal{R}}) = 0$ and hence $\sigma(\overline{\mathcal{R}}) = \tau^c q^c(\overline{\mathcal{R}}) + 1 - \tau^c = 1 - \tau^c$.

□

B.11 Proof of Corollary 5

First, observe that

$$\sum_{\omega\in\Omega} u_S(a(p(q')),\omega)\hat{q}(\omega) \leq \max_{a\in\mathcal{A}(p(q'))} \sum_{\omega\in\Omega} u_S(a,\omega)\hat{q}(\omega) \leq \sum_{\omega\in\Omega} u_S^*(\omega)\hat{q}(\omega) = h(\hat{q}).$$

Hence, in the expression of $V(\tau)$

$$\sum_{q} w^d(q)\tau(q|D_\tau) < \sum_{q\neq\hat{q}} w^d(q)\tau(q|D_\tau) + h(\hat{q})\tau(\hat{q}|D_\tau),$$

which implies that there exists $\tau' \in \mathcal{F}_\sigma$ such that $V(\tau) < V(\tau') \leq \hat{w}(\sigma)$.

B.12 Proof of Corollary 6

For $\kappa \leq 2$, we already know that Sender fully reveals all states in $\mathcal{R}$. Since w^d is convex, so is $m = \max\{h, w^d\}$. Therefore, $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$ is achieved by $\tau' = \{(\delta_\omega, \sigma(\omega|\overline{\mathcal{R}}))\}_{\omega\in\overline{\mathcal{R}}}$ (uniquely if $\kappa < 2$) and hence

$$\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = \sum_{\omega\in\overline{\mathcal{R}}} \max\{w^d(\delta_\omega), h(\delta_\omega)\}\sigma(\omega|\overline{\mathcal{R}}).$$

For each $i < \underline{m}$, we already know that $w^d(\delta_{\omega_i}) > h(\delta_{\omega_i})$ and therefore it is optimal to fully reveal ω_i. Now consider $i > \overline{m}$. For each value of κ there exists a value $b_i(\kappa)$ such that $w^d(\delta_{\omega_i}) \geq h(\delta_{\omega_i})$ if and only if $b \leq b_i(\kappa)$: this threshold is given by

$$b_i(\kappa) = \max\{\omega_i - \frac{\kappa}{2}(\omega_{\overline{m}} + \omega_i), 0\} = \max\{(1 - \frac{\kappa}{2})\omega_i - \frac{\kappa}{2}\omega_{\overline{m}}, 0\}.$$

So each $b_i(\kappa)$ is decreasing in κ (strictly when positive) and $b_i(\kappa) \leq b_j(\kappa)$ if and only if $i < j$ (with $<$ if either threshold is positive). So if $b \leq b_{\overline{m}+1}(\kappa)$, we have that $\hat{w}^d(\sigma(\cdot|\overline{\mathcal{R}})) = \hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$ and hence $\tau^d = \sigma(\overline{\mathcal{R}})$ by Proposition 5. On the other hand, if $b > b_{\overline{m}+1}(\kappa)$, we have that $w^d(\delta_{\omega_{\overline{m}+1}}) < h(\delta_{\omega_{\overline{m}+1}})$ and hence $\tau^d < \sigma(\overline{\mathcal{R}})$ by Corollary 4. Moreover, in this case let $i^*(b,\kappa) = \min\{i > \overline{m} : b_i(\kappa) \geq b\}$, which is non-decreasing in both κ and b. Then, it is optimal to fully hide all states ω_i with $\overline{m} < i < i^*(b,\kappa)$ and fully reveal all others states in $\overline{\mathcal{R}}$.

B.13 Proof of Corollary 7

By Proposition 5, $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}}))$ is given by

$$\max_{\gamma\in[0.1],q_1,q_2\in\Delta^d} \gamma \sum_{\omega\in\overline{\mathcal{R}}} u_S^*(\omega)q_1(\omega) + (1-\gamma)\sum_{\omega\in\overline{\mathcal{R}}}\{-(\mathbb{E}_{q_2}[\beta] - \omega)^2\}q_2(\omega)$$

subject to $\gamma q_1 + (1-\gamma)q_2 = \sigma(\cdot|\overline{\mathcal{R}})$. By continuity of h and w^d, a solution (γ, q_1, q_2) to this problem exists. Suppose that (γ, q_1, q_2) implies $\beta(\omega_{\underline{m}}) \leq \mathbb{E}_{q_2}[\beta] \leq \beta(\omega_{\overline{m}})$. We will show that there exists a feasible (γ', q_1', q_2') which strictly dominates (γ, q_1, q_2). Since β is strictly increasing, we must have $\omega_i, \omega_j \in \mathbf{supp}\, q_2$ for some $i < \underline{m}$ and $j > \overline{m}$. Suppose first that $\omega > \beta(\omega_{\underline{m}})$ for some $\omega > \omega_{\overline{m}}$. Then, for any $\xi \in [\beta(\omega_{\underline{m}}), \beta(\omega_{\overline{m}})]$,

$$\begin{aligned} -\sum_{\omega \in \overline{\mathcal{R}}} (\xi - \omega)^2 q_2(\omega) \quad < \quad & -\sum_{\omega \in \overline{\mathcal{R}}} (\beta(\omega_{\underline{m}}) - \omega)^2 \mathbf{1}\{\omega < \beta(\omega_{\underline{m}})\} q_2(\omega) \\ & -\sum_{\omega \in \overline{\mathcal{R}}} (\beta(\omega_{\overline{m}}) - \omega)^2 \mathbf{1}\{\omega > \beta(\omega_{\overline{m}})\} q_2(\omega) = \sum_{\omega \in \overline{\mathcal{R}}} u_S^*(\omega). \end{aligned}$$

This means that $\gamma' = 1$ and $q_1' = \sigma(\cdot|\overline{\mathcal{R}})$ strictly dominates (γ, q_1, q_2). Now, suppose that $\omega \leq \beta(\omega_{\underline{m}})$ for all $\omega > \omega_{\overline{m}}$. If $\beta(\omega_{\underline{m}}) < \mathbb{E}_{q_2}[\beta] \leq \beta(\omega_{\overline{m}})$, then again $\gamma' = 1$ and $q_1' = \sigma(\cdot|\overline{\mathcal{R}})$ strictly dominates (γ, q_1, q_2). If $\mathbb{E}_{q_2}[\beta] = \beta(\omega_{\underline{m}})$, then $\hat{m}(\sigma(\cdot|\overline{\mathcal{R}})) = h(\sigma(\cdot|\overline{\mathcal{R}}))$. But we know that always hiding all states in $\overline{\mathcal{R}}$ is not optimal: $\hat{w}(\sigma) > \hat{w}^c(\sigma)$ since $w^d(\delta_{\omega_1}) > h(\delta_{\omega_1})$. Therefore, (γ, q_1, q_2) is again strictly dominated.

Finally, if σ is such that $\omega_{\underline{m}} \leq \mathbb{E}_{\sigma(\cdot|\overline{\mathcal{R}})}[\omega] \leq \omega_{\overline{m}}$, then $\tau^d = \sigma(\overline{\mathcal{R}})$ implies that $q^d = \sigma(\cdot|\overline{\mathcal{R}})$ and hence $\beta(\omega_{\underline{m}}) \leq \mathbb{E}_{\sigma(\cdot|\overline{\mathcal{R}})}[\beta] \leq \beta(\omega_{\overline{m}})$.

B.14 Proof of Corollary 8

Given any $\tau \in \mathcal{F}_\sigma$, the claim follows immediately from (7) for every $q \in C_\tau$. Now consider any $q \in D_\tau$. If $q \in D_\tau(\rho_N)$, then again the claim follows directly from (7). So suppose $q \in D_\tau(\rho_i)$ for some $i \neq N$. Then,

$$p^A(\omega; q) = \frac{q(\omega)\frac{\rho_1^A(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega')\frac{\rho_1^A(\omega')}{\sigma(\omega')}} \quad \text{and} \quad p^B(\omega; q) = \frac{q(\omega)\frac{\rho_i^B(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega')\frac{\rho_i^B(\omega')}{\sigma(\omega')}}.$$

Note that $q(\omega) = 0$ implies both $p^A(\omega; q) = 0$ and $p^B(\omega; q) = 0$. So, restrict attention to $\mathbf{supp}\, q$ which contains both $\mathbf{supp}\, p^A(q)$ and $\mathbf{supp}\, p^B(q)$. By definition, $\rho_1^A(\omega) > 0$ for all $\omega \in \Omega$; therefore, $\mathbf{supp}\, q = \mathbf{supp}\, p^A(q)$. Also, by definition, $\mathbf{supp}\, q \subset \mathbf{supp}\, \rho_i^B = \Omega(q)$; therefore, $\mathbf{supp}\, q = \mathbf{supp}\, p^B(q)$ as well. Now, restrict attention to $\Omega(q) \subsetneq \Omega = \mathbf{supp}\, \rho_1^A$. By A6, we have

$$\frac{\rho_1^A(\omega|\Omega(q))}{\rho_1^A(\omega'|\Omega(q))} = \frac{\rho_i^B(\omega)}{\rho_i^B(\omega')}$$

for all $\omega, \omega' \in \Omega(q)$, where $\rho_1^A(\omega|\Omega(q)) = \frac{\rho_1^A(\omega)}{\rho_1^A(\Omega(q))}$. So, fixing one $\hat{\omega} \in \Omega(q)$, we have

$$\rho_1^A(\omega|\Omega(q)) = \frac{\rho_1^A(\hat{\omega}|\Omega(q))}{\rho_i^B(\hat{\omega})}\rho_i^B(\omega).$$

It follows that, for all $\omega \in \mathbf{supp}\, q$,

$$\begin{aligned} p^A(\omega;q) &= \frac{\rho_1^A(\Omega(q))q(\omega)\frac{\rho_1^A(\omega)}{\sigma(\omega)}}{\rho_1^A(\Omega(q))\sum_{\omega'\in\Omega_i} q(\omega')\frac{\rho_1^A(\omega')}{\sigma(\omega')}} = \frac{q(\omega)\frac{\rho_1^A(\omega|\Omega(q))}{\sigma(\omega)}}{\sum_{\omega'\in\Omega(q)} q(\omega')\frac{\rho_1^A(\omega'|\Omega(q))}{\sigma(\omega')}} \\ &= \frac{\frac{q(\omega)}{\sigma(\omega)}\frac{\rho_1^A(\hat\omega|\Omega(q))}{\rho_i^B(\hat\omega)}\rho_i^B(\omega)}{\sum_{\omega'\in\Omega(q)}\frac{q(\omega')}{\sigma(\omega')}\frac{\rho_1^A(\hat\omega|\Omega(q))}{\rho_i^B(\hat\omega)}\rho_i^B(\omega')} = \frac{\frac{q(\omega)}{\sigma(\omega)}\rho_i^B(\omega)}{\sum_{\omega'\in\Omega(q)}\frac{q(\omega')}{\sigma(\omega')}\rho_i^B(\omega')} = p^B(\omega;q). \end{aligned}$$

References

Ahn, D. S. and H. Ergin (2010). Framing contingencies. *Econometrica 78*(2), 655–695.

Alonso, R. and O. Câmara (2013). Persuading skeptics and reaffirming believers. *Available at SSRN 2306820*.

Alonso, R., W. Dessein, and N. Matouschek (2008). When does coordination require centralization? *The American economic review 98*(1), 145–179.

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, 199–205.

Aumann, R. J. and M. Maschler (1995). *Repeated games with incomplete information.* MIT press.

Blume, L., A. Brandenburger, and E. Dekel (1991a). Lexicographic probabilities and choice under uncertainty. *Econometrica*, 61–79.

Blume, L., A. Brandenburger, and E. Dekel (1991b). Lexicographic probabilities and equilibrium refinements. *Econometrica*, 81–98.

Brocas, I. and J. D. Carrillo (2007). Influence through ignorance. *The RAND Journal of Economics 38*(4), 931–947.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica*, 1431–1451.

Dessein, W. (2002). Authority and communication in organizations. *The Review of Economic Studies 69*(4), 811–838.

Ely, J. (2014). Beeps. *mimeo, Northwestern University*.

Ely, J., A. Frankel, and E. Kamenica (2013). Suspense and surprise. *forthcoming, Journal of Political Economy*.

Forges, F. and F. Koessler (2008). Long persuasion games. *Journal of Economic Theory 143*(1), 1–35.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009). Mediation, arbitration and negotiation. *Journal of Economic Theory 144*(4), 1397–1420.

Grossman, G. M. and E. Helpman (2002). *Special interest politics*. MIT press.

Horner, J. and A. Skrzypacz (2010). Selling information. *Cowles Foundation Discussion Paper*.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6).

Karni, E. and M.-L. Vierø (2013). Reverse bayesianism: A choice-based theory of growing awareness. *American Economic Review 103*(7), 2790–2810.

Kreps, D. M. and R. Wilson (1982). Sequential equilibria. *Econometrica*, 863–894.

McCloskey, D. and A. Klamer (1995). One quarter of gdp is persuasion. *American Economic Review*, 191–195.

Morgan, J. and P. C. Stocken (2003). An analysis of stock recommendations. *RAND Journal of Economics*, 183–203.

Morris, S. (1995). The common prior assumption in economic theory. *Economics and Philosophy 11*(02), 227–253.

Mullainathan, S., J. Schwartzstein, and A. Shleifer (2008). Coarse thinking and persuasion*. *The Quarterly Journal of Economics 123*(2), 577–619.

Ortoleva, P. (2012). Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review 102*(6), 2410–2436.

Rabin, M. and J. L. Schrag (1999). First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics*, 37–82.

Rayo, L. and I. Segal (2010). Optimal information disclosure. *Journal of Political Economy 118*(5), 949–987.

Rockafellar, R. T. (1997). *Convex analysis*. Princeton University Press.