Persuasion: The Art of Changing Worldviews[*]

Simone Galperti[†]

PRELIMINARY: NOT FOR GENERAL CIRCULATION

August 28, 2017

Abstract

Persuaders often have greater understanding of a subject matter than their listener. By conveying something unexpected for the listener, they can change his worldview. These aspects of persuasion cannot be captured in the standard Bayesian framework, and call for a new framework. The paper identifies distinctive features of the expert persuader's problem, derives necessary and sufficient conditions for her to change the listener's worldview, and shows when and how she conceals her superior knowledge and takes advantage of the listener's poorer worldview.

KEYWORDS: persuasion, non-Bayesian, conceal, surprise, worldview change, different prior.

JEL CLASSIFICATION: D82, D83, K41, M30.

[*]I thank for comments and suggestions S. Nageeb Ali, Riccardo Alonso, Alessandro Bonatti, Benjamin Brooks, Rahul Deb, Eddie Dekel, Laura Doval, Maria Goltsman, Philippe Jehiel, Emir Kamenica, Navin Kartik, Jay Lu, Mark Machina, Paul Niehaus, Marciano Siniscalchi, Joel Sobel, Charles Sprenger, Joel Watson, and seminar participants at CETC 2015, ESSET Gerzensee 2015, ESWC 2015, SWET 2015, Yale (Cowles), and USC. John N. Rehbeck provided excellent research assistance. All remaining errors are mine.

[†]Department of Economics, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA, 92093. E-mail address: sgalperti@ucsd.edu.

Awareness of resistance is valuable to the creator of new vision.
(Gardner (2006), *Changing Minds*, p. 127)

Resistance hounds persuasion the way friction frustrates motion.
(Knowles and Linn (2004), *Resistance and Persuasion*, p. 3)

1 Introduction

Persuasion plays an important role across economic activities and beyond.[1] In many situations, the persuader has a greater understanding of a subject matter than does the listener. Examples include visionary leaders, scientists, experts advising policymakers, or sellers launching game-changing products. In these cases, persuasion often involves making the listener think harder about a problem, examine it through a new lens (or worldview), and expand his universe of possible worlds. This is the opposite of what happens in a standard Bayesian framework, where new information always causes the listener's universe to shrink within his existing worldview.

Despite its pervasiveness, this role of persuaders as creators of worldviews has received little to no attention. This paper tries to formalize it and investigate its consequences for persuasion. What constraints do self-interested parties face in persuading agents who have poorer understanding of the world and may be resistant to change? Do they force such agents to revise their worldviews, by providing "surprising" evidence, or would they rather keep them in the dark even if their worldview is incorrect? When and how can persuaders benefit from concealing their superior knowledge?

As in Kamenica and Gentzkow (2011) (hereafter, KG), the persuader, called Sender (she), seeks to affect an action of the listener, called Receiver (he). Their payoffs depend on this action and a payoff-relevant variable, called the state of the world. Sender controls which evidence Receiver can observe about the state: Starting from a situation where both have no evidence, she designs and commits to a public experiment π which, conditional on each state, defines a distribution over pieces of evidence.

KG's Bayesian framework has to be modified in order to introduce the possibility of changing Receiver's worldview. This modification in itself is part of the paper's contribution. A worldview is modeled as a subjective prior belief over the commonly-known state space. To help intuition, think about the state as an observable prediction of some underlying theory of the world. For example, the state can be a specific increase in the Earth's temperatures, Sender is a climate scientist who believes in global warming, while Receiver is a climate-change denier. The scientist thinks that temperature increases of several degrees can happen, while the denier deems them impossible. To capture such situations, the paper relaxes KG's common-prior assumption as follows. Sender's prior is fixed and assigns positive probability

[1] See, for example, McCloskey and Klamer (1995).

to all states. Receiver's prior instead assigns a very small probability ε (possibly, zero) to some states, which form a subset $\mathcal{I}$ of the state space. Let $\mathcal{P}$ contain all other states. Think of $\mathcal{P}$ as the foundation of Receiver's worldview. For example, the climate-change denier is convinced that temperatures will not rise ($\mathcal{P}$) and rules out any noticeable increment ($\mathcal{I}$).

The idea that people have different worldviews and agree to disagree is consistent not only with daily observations, but also with foundational work on how they form beliefs (see Section 7). Case-based reasoning which applies similarity judgments to experience and searches for regularities in raw data can lead people to have different beliefs and agree to disagree, even if they use the same experience and data. This is especially likely for new issues with few or no precedents, such as climate change or the ramifications of a nuclear war. Stark worldview differences are also common for issues affected by religious, moral, social, or political principles. Sender and Receiver may also represent different organizations whose worldview is determined by internal practices, expertise, and culture.

A second key departure from the Bayesian framework is that here some evidence can *disprove* Receiver's worldview, in the sense of making him doubt its correctness. This usually happens when one sees unexpected or surprising evidence. If $\varepsilon = 0$, this would be evidence conclusively proving that some state in $\mathcal{I}$ obtained. More generally, unexpected evidence is defined as evidence which lowers the probability that Receiver assigns to $\mathcal{P}$ below some breaking point (possibly, larger than zero) where he decides that his worldview is no longer tenable.[2]

Modeling changes of worldview raises a host of conceptual issues, reaching far beyond economics.[3] Here, the main question is how Sender predicts that Receiver will react to evidence that shatters his worldview. The answer cannot come from Bayes' rule, the backbone of the entire literature on persuasion.[4] The paper suggests *an* answer by adapting to the present setting the few models proposed in the literature of how agents react to unexpected events. The main premise is that people are reluctant to change worldview, which has significant empirical support discussed in Section 7. Thus, the stronger is Receiver's confidence in the foundation $\mathcal{P}$ of his theory—that is, the larger $1-\varepsilon$ is—the more surprising the evidence has to be for him to doubt his worldview. When this occurs, Receiver forms a new worldview (i.e., prior). In the baseline model, this prior is always the same and independent of the evidence. Section 6 generalizes this. Receiver updates his worldview—old or new—using Bayes' rule.

[2]A natural question is whether the phenomena analyzed here can be "approximated" by a standard model where the probability that Receiver's full-support prior puts to some states converges to zero. Section 7 shows that the answer is no.

[3]See, for instance, Fermé and Hansson (2011) for a review of the literature in philosophical logic on how to revise a system of propositions, representing one's beliefs, in light of a new inconsistent proposition.

[4]This is not just a technicality. Bayes' rule is a procedure to *update* an existing worldview in light of new evidence consistent with it. But if the evidence disproves that worldview, it is not clear why the same procedure should also guide how one *forms* a new worldview.

Since Receiver's posterior beliefs determine his actions, the first part of the paper characterizes which posteriors experiments can induce. Compared to persuasion without worldview changes, three main differences emerge, which are also illustrated with an example in the next section. First, the set of feasible posteriors is smaller than in the standard case, and the more so, the stronger is Receiver's reluctance to change worldview. This is a novel constraint. Intuitively, thinking that global warming is a hoax "blinds" Receiver to the signs of rising temperatures, unless these are crystal clear. Thus, his posterior that global warming is true can only be very low (zero) or very high (one). A second related property is that worldview changes cause discontinuous jumps in posteriors. This captures formally what we loosely call *surprise*. As the signs of rising temperatures accumulate, a denier of climate change keeps thinking that it is a hoax until the evidence is so compelling that he changes his mind. At this point, his posterior jumps from zero to one. The possibility of drastic belief swings calls for more attention when designing π. Finally, Receiver's "blindness" to evidence inconsistent with it grant Sender a dimension of flexibility which has no analog in standard persuasion: She can essentially *conceal* each state in $\mathcal{I}$ by pooling it with evidence on states in $\mathcal{P}$, after which Receiver continues to base his inference and decisions only on $\mathcal{P}$. The scientist does not have to worry if, in order to support some idea, she has to present evidence supporting also climate change. Its denier will just ignore this content and focus on her idea.

The second part of the paper characterizes the optimal π. Note that the equivalence between every π and a distribution over Receiver's posteriors whose average equals to his prior (see KG) no longer holds here, because the prior he updates can change. That equivalence is still valid, however, for *Sender's* posteriors. Moreover, our assumptions on Receiver's response to evidence imply that each Sender's posterior maps to a well-identified posterior of Receiver—even after worldview changes. Thanks to this result, we can still derive π using familiar "concavification arguments" which have become standard in the literature (Aumann and Maschler (1995) and KG). The characterization focuses on the case of $\varepsilon = 0$, which is shown to be at all effects the limit as $\varepsilon \to 0$ of the more general model. This robustness result justifies using the former in applications.

To understand if and when Sender will disprove Receiver's worldview, we need to assign a value to the option of concealing states. This is called the *opportunity cost of surprising.* If the scientist keeps her listener thinking that global warming is a hoax, she can induce him to have any posterior regarding other aspects of the issue at hand (i.e., $\mathcal{P}$) and choose any action he finds optimal at such posteriors. Her highest payoff *from these actions* in the state corresponding to a specific level of warming defines her opportunity cost of proving that state, which would surprise her listener. Given this, suppose that there exists *some* piece of evidence supporting only states in $\mathcal{I}$ such that, if the scientist observed it, her expected payoff from making it public would exceed her expected opportunity cost of surprising. Proposition 4 shows that this condition—called "there exists *knowledge* Sender would share with Receiver"—is necessary and sufficient for the scientist to disprove her listener's worldview.[5]

Regarding the optimal π, the paper shows how its characterization can be divided into simpler parts and solved by adapting familiar tools from the literature. Intuitively, for all states in $\mathcal{P}$, *every* experiment will produce evidence consistent with Receiver's worldview as in standard persuasion settings. Hence, the design of π follows the same principles and have the same properties as in those settings (see KG and Alonso and Câmara (2016)). Regarding the states in $\mathcal{I}$, Sender has to choose which states to conceal, which to reveal, which evidence to use, and which prior she wants Receiver to adopt—intuitively, how deeply she wants him to think to form a new worldview. The answer to all these questions is given by the concavification of two functions: One gives Sender's expected payoff from surprising evidence, the other its expected opportunity cost. This result also provides simple conditions, easily derived from primitives, to find which states π conceals without having to characterize π.

Overall, these results offer insights into how and to what extent Sender can take advantage of Receiver's poorer worldview. A general persuasion principle is that bundling "bad" with "good" states in the same piece of evidence benefits Sender because, as long as this evidence does not render Receiver too pessimistic about the good states, it increases the chances of a desirable outcome. In the present setting, this principle faces new opportunities as well as new constraints. Sender can bundle a bad state in $\mathcal{I}$ with a good state in $\mathcal{P}$, *without* decreasing Receiver's confidence in the latter (concealment). On the other hand, if she has to change Receiver's worldview to obtain a desirable outcome, she cannot bundle a bad state in $\mathcal{P}$ with a good state in $\mathcal{I}$: Doing so would actually destroy any chance of getting that outcome. This is also illustrated in the next section. Thus, different worldviews can weaken Sender's incentives to disclose information by creating opportunities for concealment (perhaps not surprisingly), but they can also strengthen those incentives by forcing her to be more clear and explicit.

To examine how the optimal π may depend on the parties' conflict of interests and compare this with Bayesian persuasion, the paper presents a lobbying application using a classic framework with quadratic payoffs as in KG.[6] The main difference is that here the lobbyist is an expert who can change the worldview of her target politician. In contrast to KG's setting, the lobbyist's π is never fully uninformative and can be only partially informative. More importantly, π tends to produce results in favor of the interests represented by the lobbyist, a phenomenon which seems to occur in reality, but is at odds with the

[5] This condition differs substantively from KG's "there exists *information* Sender would share with Receiver," as explained in Section 5.

[6] This framework has been used to study communication in other contexts. 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)).

predictions of KG's Bayesian model (see KG, p. 2606). This suggests that incompatible worldviews may be as important as conflicting interests in shaping persuader's strategies.

1.1 An Illustrative Example: Persuading Climate-change Deniers

An scientist on energy and environmental issues (Sender) faces a policymaker (Receiver) who denies man-made climate change. The policy choice at hand involves measures to promote renewable or fossil energy sources. The scientist is hired by the renewables industry to influence that choice. She is asked to design an experiment π producing evidence about the following two dimensions: Renewables generate higher economic returns (profits, jobs, etc.) than fossil sources or not (r or f); climate change is true or a hoax (t or h). The state are then ω_{hf}, ω_{hr}, ω_{tf}, and ω_{tr}. The policymaker denies climate change and thinks that with 70% probability fossil sources dominate renewables economically. The scientist thinks that with a high probability climate change is true and that either type of sources can economically dominate. These worldviews are shown in Table 1.[7]

	State	Receiver's prior (ρ_0)	Sender's prior (σ)
$\mathcal{P} =$	ω_{hf}	0.7	0.07
	ω_{hr}	0.3	0.03
$\mathcal{I} =$	ω_{tf}	0	0.5
	ω_{tr}	0	0.4

Table 1: Worldviews

Policy	Profits (in \$bn)
a_{hf}	1
a_{hr}	3
a_{tf}	2
a_{tr}	4

Table 2: Policies and Profits

The policymaker seeks to be re-elected: His payoff is 1 if he succeeds and 0 otherwise. To succeed, his policy choice has to be ideal given the realized state (which becomes publicly known ex post). Such policies are as follows, where subscripts indicate the state: a_{hf} is the business-as-usual subsidization of fossil sources; a_{hr} promotes the development of renewables; a_{tf} regulates CO_2 emissions from fossil sources while preserving their economic dominance on renewables; a_{tr} implements a drastic shift from fossil to renewable sources.

The scientist is asked to design π so as to maximize the expected profits of the renewables industry, which depend on a as shown in Table 2. The assumption here is that profits do not depend on ω. For each ω, π defines a probability distribution $\pi(\cdot|\omega)$ over pieces of evidence. As in KG, all such distributions are feasible—though this may be a stretch in some situations. An important assumption is that some π can produce unexpected evidence which disproves

[7]The assumption that ρ_0 coincides with Sender's prior conditional on $\mathcal{P}$ simplifies the comparison with KG's common-prior setting and some calculations later.

the policymaker's worldview. In this example, such evidence has to completely rule out $\mathcal{P}$; in the general model, it will have to be sufficiently incompatible with $\mathcal{P}$.

The policymaker responds to evidence as follows. After observing evidence consistent with $\mathcal{P}$, he updates ρ_0 via Bayes' rule. After unexpected evidence, he first adopts a new worldview ρ^1 and then updates it via Bayes' rule. As in the baseline model of the paper, ρ^1 is always the same and puts positive probability on all states. Here assume $\rho^1 = \sigma$.

We can now preview the differences between the posterior beliefs that experiments can induce here and in standard persuasion without worldview changes. Note that whenever inconclusive evidence leaves open the possibility that climate change may be a hoax, the policymaker filters this ambiguity through his worldview and continues to think that *it is* a hoax. Since he cannot "see" the signs of climate change through the lens of his worldview, it takes strongly opposite evidence to shatter it and change his mind. As a result, his beliefs will always be extreme: In this example, they assign probability 0 or 1 to climate change. This also means that worldview changes manifest themselves in discontinuous shifts of beliefs. No π can then induce moderate beliefs and minor tweaks to π can have big effects on them, which imposes new constraints for the scientist. On the other hand, the policymaker's "blindness" to partial evidence on climate change also opens new opportunities: Using such evidence, π can essentially conceal ω_{tf} and ω_{tr}, in the sense that he will continue to deny climate change and base his decision only on ω_{hf} and ω_{hr}.

These constraints and opportunities shape the optimal π^*. As shown in Table 3, π^* uses three pieces of evidence (or signals), where subscripts indicate the induced policy. Consider first ω_{hf} and ω_{hr}. Given these states, any signal will confirm the policymaker's worldview, thereby inducing a_{hf} or a_{hr}. For him to choose a_{hr}, his posterior belief that ω_{hr} obtained has to be at least 0.5. Since this belief increases in the ratio $\frac{\pi^*(s_{hr}|\omega_{hr})}{\pi^*(s_{hr}|\omega_{hf})}$, we must have $\pi^*(s_{hr}|\omega_{hr}) = 1$. However, $\pi^*(s_{hr}|\omega_{hf})$ is also positive: This lowers the policymaker's beliefs that ω_{hr} obtained, but raises the ex-ante chances of getting a_{hr}. To maximize these chances while keeping that belief above 0.5, $\pi^*(s_{hr}|\omega_{hf})$ has to be $\frac{3}{7}$. With the remaining probability π^* has to produce other signals in ω_{hf}, namely, s_{hf}. The reader may have noticed that so far π^* is *identical* to the optimal π in KG's example. This property is general and has a twofold intuition. First, conditional on $\mathcal{P}$ the policymaker's worldview cannot be changed *and* agrees with the scientist's (as in KG's common-prior settings). Second, even if a signal consistent with $\mathcal{P}$ arises also in ω_{tf} or ω_{tr}, the policymaker ignores this and chooses a_{hf} or a_{hr}; so the scientist can use s_{hf} or s_{hr} to conceal ω_{tf} and ω_{tr} and induce whichever of a_{hf} or a_{hr} she prefers.

Consider the rest of π^*. Signals disproving the policymaker can only lead to a_{tf} or a_{tr}. While a_{tf} is the best for the scientist, she prefers a_{hr} to a_{tf}. One might then expect that she always conceals ω_{tf}. However, after changing worldview, the policymaker chooses a_{tr} if his posterior belief that ω_{tr} obtained is at least 0.5. Thus, for the same reason as before,

$\pi^*(s\|\omega)$	s_{hf}	s_{hr}	s_{tr}
ω_{hf}	$\frac{4}{7}$	$\frac{3}{7}$	0
ω_{hr}	0	1	0
ω_{tf}	0	$\frac{1}{5}$	$\frac{4}{5}$
ω_{tr}	0	0	1

Table 3: Optimal Experiment π^*

the scientist wants s_{tr} to arise for sure in ω_{tr}, but also in other states so as to increase the chances of getting a_{tr}. She cannot use ω_{hf} nor ω_{hr}, otherwise s_{tr} would not disprove the policymaker, which defeats the purpose. She must then use ω_{tf}, thereby strategically revealing it with some probability. To keep the belief that ω_{tr} obtained above 0.5 while maximizing the chances of a_{tr}, $\pi^*(s_{tr}|\omega_{tf})$ has to be $\frac{4}{5}$ and so ω_{tf} must again lead to some other signal. The key difference is that this signal can be s_{hr}, which induces a_{hr}, rather than some fourth signal, which would fully reveal ω_{tf} and induce a_{tf}.

In short, we can interpret π^* as showing either that fossil sources are economically dominant for sure (s_{hf}), or that renewables are dominant with 50-50 chance (s_{hr}), or that this is the case and in addition climate change is true (s_{tf})—with the probabilities in Table 3.

2 Related Literature

This paper contributes to the literature that studies persuasion in the form of strategic experimentation.[8] In a common-prior setting, Brocas and Carrillo (2007) study how Sender benefits by sequentially revealing information by repeating an exogenously fixed experiment. KG introduce a fully general framework with common priors and unrestricted experiments. They characterize when Sender benefits from informative experiments, their properties, and their dependence on the conflict of interests with Receiver. In Alonso and Câmara (2016) (hereafter, AC), Sender and Receiver have common-support but different priors. AC show that this form of disagreement *expands* the scope for Sender to benefit from informative experiments.[9] In all these papers, there is no sense in which Sender has superior knowledge (not information) and can change Receiver's worldview, who may be reluctant to do so.

To model Receiver's response to evidence, this paper builds on the game- and decision-theory literature. One model uses lexicographic belief systems à la Kreps and Wilson (1982). They imagine that each player in a game has a list of "hypotheses" about how the game is

[8] Persuasion can of course take other interesting forms and rely on different tools (see, for example, Mullainathan et al. (2008), Glazer and Rubinstein (2012), and Glazer and Rubinstein (2014)).

[9] Other papers which study information-design problems include Forges and Koessler (2008), Rayo and Segal (2010), Horner and Skrzypacz (2014), Ely (2016), and Ely et al. (2015). All of them assume common priors and address questions different from the ones treated in the present paper.

played. The list ranking is interpreted as the player's (fixed) judgment of each hypothesis' likelihood to be "correct;" hence, he always seeks to apply the lowest-ranked hypothesis consistent with what he observes.[10] The present paper replaces those hypotheses with Receiver's worldviews and uses the lexicographic procedure to capture his reluctance to change worldview. The paper also studies the implications of Ortoleva's (2012) updating model. In contrast to the previous model, now a surprised Receiver uses the evidence to reassess the likelihood that each worldview is "correct" before adopting one.

Finally, this paper is related to the literature on strategic communication. Unlike the aforementioned literature, this literature studies settings where Sender observes some information before choosing how to communicate with Receiver. Also, in cheap-talk settings à la Crawford and Sobel (1982), Sender communicates by sending unverifiable messages, rather than by committing to a verifiable protocol. As a result, how Receiver interprets Sender's messages is now part of the equilibrium of the game. Importantly, without verifiable information Sender may not be able to change Receiver's worldview: After seeing unexpected messages, he can always think that Sender deviated from her communication strategy and conclude that there is no need to change worldview. A key insight from Crawford and Sobel (1982) is that communication is hampered by the conflict between parties' interests. The present paper suggests that another obstacle can be their different worldviews and reluctance to change them. The strategic-communication literature almost entirely assumes common priors. One exception is Milgrom and Roberts (1986). Using Milgrom's (1981) framework with verifiable information, they examine settings where Receiver ignores some relevant aspects of the world. However, they also assume that there are many Senders and their interests are sufficiently in conflict that, in equilibrium, they always reveal everything they know. This property allows the authors to model Receiver as an automaton, thus avoiding the updating issues in the present paper. One can view the present analysis as a benchmark against which we can evaluate the benefits of competition among expert persuaders.

3 The Model

Players. An agent, called Sender (she), aims to influence the behavior of another agent, called Receiver (he). To do so, she produces evidence about some payoff-relevant state of the world ω, contained in some finite set Ω. Ex ante neither agent knows ω, and their information is symmetric throughout the game.

After observing Sender's evidence Receiver chooses an action a from a finite set A. Both agents maximize expected utility: Sender's cardinal utility is $v : A \times \Omega \to \mathbb{R}$ and Receiver's is $u : A \times \Omega \to \mathbb{R}$. As in KG, if Receiver is indifferent between some actions, Sender can

[10]The lexicographic belief systems in Blume et al. (1991a, 1991b) work in a fundamentally different way: The agent always takes into account, though lexicographically, all the beliefs of his system.

break ties as she prefers.

Experiments. Sender provides evidence via an experiment, π, which consists of a finite realization space S (for "signals") and a family of distributions $\{\pi(\cdot|\omega)\}_{\omega\in\Omega}$ over S. A piece of evidence consists of the pair (s,π), as π determines the meaning of s.

Besides the definition of π, the following assumptions are shared with KG: (1) Sender designs and commits to π, which is the only source of evidence about ω in the game; (2) she can choose any family of distributions π; (3) she and Receiver commonly understand π, that is, both understand the rules and methods designed to conduct π before gathering data, which define its statistical properties captured by $\{\pi(\cdot|\omega)\}_{\omega\in\Omega}$. In particular, even though the agents may have different worldviews (to be introduced shortly), they agree on the inherent randomness of experiments. By (2), Sender can design a fully informative π, namely, a π that yields a distinct s for every ω and thus conclusively proves which state obtained. As KG note, this assumption is weaker than it may seem: One can always take the most informative π available—whatever that is—and interpret Ω as its possible realizations. Sender's problem is then to choose how much to garble that π.

The paper departs from the literature (in particular KG and AC) by allowing for the possibility that Receiver changes worldview after seeing unexpected evidence. Formalizing this requires a bit of motivation, explanation, and definitions.

Worldviews. Ex ante the agents have different worldviews, represented by subjective prior beliefs, and agree to disagree. Sender's worldview is σ with support $\mathbf{supp}\,\sigma=\Omega$. Receiver's *initial* worldview is ρ_ε, where $\varepsilon\geq 0$. It partitions Ω into two nonempty sets, $\mathcal{P}$ and $\mathcal{I}$, which satisfy $\rho_\varepsilon(\omega)>\varepsilon$ for all $\omega\in\mathcal{P}$, $\rho_\varepsilon(\mathcal{I})=\varepsilon$, and $\rho_\varepsilon(\omega)>0$ for all $\omega\in\mathcal{I}$ if $\varepsilon>0$. Think of ε as being very small, possibly zero. Intuitively, $\mathcal{P}$ represents the states that Receiver deems probable, plausible, or simply possible—the core foundation of his worldview. $\mathcal{I}$ instead represents states he deems improbable, implausible, or even impossible. For instance, if $\varepsilon=0$, he thinks that no state in $\mathcal{I}$ can obtain and hence no π can ever prove any of them. It is useful to interpret $1-\varepsilon$ as Receiver's degree of confidence that his worldview centered on $\mathcal{P}$ is correct. The next assumption gives content to this interpretation; it also simplifies later analysis that varies ε.

Assumption 1 (A1)**.** *For all $\omega\in\Omega$, $\rho_\varepsilon(\omega)$ can be written as*

$$\rho_\varepsilon(\omega)=(1-\varepsilon)\rho_\varepsilon(\omega|\mathcal{P})+\varepsilon\rho_\varepsilon(\omega|\mathcal{I}),$$

where $\rho_\varepsilon(\omega|\mathcal{P})$ and $\rho_\varepsilon(\omega|\mathcal{I})$ do not depend on ε.

Changing vs. Updating Worldviews. In standard persuasion models Receiver updates his fixed prior via Bayes' rule for all (s,π).[11] Here, by contrast, some evidence will lead him

[11]Note that in both KG and AC this is the case, no matter how small the (positive) probability assigned by Receiver to some states is.

to doubt his worldview and change it altogether. This entails departing from Bayes' rule, which describes how agents *update* beliefs, not how they *change* their view of the world. This point transcends the mere fact that Bayes' rule does not apply after zero-probability events (when $\varepsilon = 0$).

Worldview changes render Sender's problem of predicting how Receiver responds to evidence significantly more complex and delicate: Which evidence triggers a change? Does he form a new worldview? Following what procedure? These are challenging questions, whose answer is of course a key part of the model. However, there is no single, broadly accepted, standard answer comparable to Bayesianism. To attempt a first analysis of persuasion with worldview changes, the paper considers several models of Receiver's response to evidence (inspired by the literature). They should be interpreted as Sender's ex-ante subjective prediction, and nothing is assumed or implied here regarding its ex-post correctness.

To define which evidence is unexpected and triggers a worldview change, we cannot simply use the ex-ante probability Receiver assigns to s given π. To see this, suppose that $\pi(\hat{s}|\mathcal{P}) > 0$ but tiny and $\pi(\hat{s}|\mathcal{I}) = 0$. Receiver deems $\hat{s}$ very unlikely, yet observing it is perfectly consistent with his worldview. If instead $\pi(\hat{s}|\mathcal{P}) = 0$ and $\pi(\hat{s}|\mathcal{I}) > 0$, he would consider $\hat{s}$ surprising. Specifically, conditional on $\hat{s}$ he would assign a very low—in this case, zero—probability to $\mathcal{P}$, the foundation of his worldview. Generalizing this observation, (s, π) is going to lead Receiver to doubt his worldview if it induces him to assign—under ρ_ε—a sufficiently low probability to $\mathcal{P}$ relative to $\mathcal{I}$. Formally, for $\varepsilon > 0$ let

$$\ell(s, \pi) = \frac{\sum_{\omega \in \mathcal{P}} \pi(s|\omega)\rho_\varepsilon(\omega)}{\sum_{\omega' \in \mathcal{I}} \pi(s|\omega')\rho_\varepsilon(\omega')};$$

for $\varepsilon = 0$, let

$$\ell(s, \pi) = \begin{cases} \infty & \text{if } \{\omega \in \mathcal{P} : \pi(s|\omega) > 0\} \neq \varnothing, \\ 0 & \text{if } \{\omega \in \mathcal{P} : \pi(s|\omega) > 0\} = \varnothing. \end{cases}$$

Definition 1 (Unexpected Evidence)**.** For every $\varepsilon \geq 0$, let $\theta_\varepsilon \in [0, \frac{1-\varepsilon}{\varepsilon})$. Receiver deems (s, π) unexpected if and only if $\ell(s, \pi) \leq \theta_\varepsilon$.

We will say that unexpected (s, π) *disproves* ρ_ε; otherwise, for simplicity we will say that (s, π) *confirms* ρ_ε. Note that Definition 1 relies only on ρ_ε so that Receiver's doubt or 'embarrassment' about his worldview precedes (also logically) any consideration of alternative worldviews. The parameter θ_ε represents a tolerance threshold or breaking point for Receiver. Clearly, it has to be smaller that the prior likelihood ratio of $\mathcal{P}$ and $\mathcal{I}$.

Individuals' confidence in their worldview may vary, possibly affecting which evidence they consider unexpected. Intuitively, the more probable Receiver deems $\mathcal{P}$ relative to $\mathcal{I}$ ex ante, the less he should tolerate evidence contradicting this assessment. At the same time, the stronger his confidence is in his worldview, the harder it should be to shake it. Besides everyday experience, both psychology and the history of science document individuals' reluctance to change worldview (see Section 7.1). This is captured by the next assumption; its explanation follows.

Assumption 2 (A2). *θ_ε increases continuously as ε decreases and* $\lim_{\varepsilon\to 0}\theta_\varepsilon = \infty$. *Also, $\theta_\varepsilon \frac{\varepsilon}{1-\varepsilon}$ decreases as ε decreases and* $\lim_{\varepsilon\to 0}\theta_\varepsilon \frac{\varepsilon}{1-\varepsilon} = 0$.

For $\varepsilon > 0$, we have that

$$\ell(s,\pi) = \frac{\pi(s|\mathcal{P})}{\pi(s|\mathcal{I})} \times \frac{\rho_\varepsilon(\mathcal{P})}{\rho_\varepsilon(\mathcal{I})}.$$

Suppose that, given $\frac{\rho_\varepsilon(\mathcal{P})}{\rho_\varepsilon(\mathcal{I})} = \frac{1-\varepsilon}{\varepsilon}$, s is sufficiently more likely to arise given $\mathcal{P}$ than $\mathcal{I}$, so that Receiver does not doubt his worldview (i.e., $\frac{\pi(s|\mathcal{P})}{\pi(s|\mathcal{I})} > \theta_\varepsilon \frac{\varepsilon}{1-\varepsilon}$). If he deems $\mathcal{P}$ even more likely ex ante (i.e., ε falls), it should be easier for him to argue that his worldview is tenable given the same $\frac{\pi(s|\mathcal{P})}{\pi(s|\mathcal{I})}$.[12] Thus, stronger evidence against it is needed to shake his confidence: $\frac{\pi(s|\mathcal{P})}{\pi(s|\mathcal{I})}$ needs to be below some smaller threshold $\theta_\varepsilon \frac{\varepsilon}{1-\varepsilon}$. For $\varepsilon = 0$, to disprove Receiver s must arise only from states in $\mathcal{I}$, in which case $\frac{\pi(s|\mathcal{P})}{\pi(s|\mathcal{I})} = 0$. Thus, A2 establishes consistency between the cases of $\varepsilon > 0$ and $\varepsilon = 0$, so that one can also interpret the latter as the limit for $\varepsilon \to 0$. A2 implies that $\theta_\varepsilon > 0$ when $\varepsilon > 0$ is small enough (our maintained assumption). That is, evidence highly (as opposed to fully) incompatible with ρ_ε suffices to disprove it.

What happens after unexpected evidence? It is plausible to imagine that Receiver forms a new worldview in an attempt to rationalize the evidence, possibly via reassessing past experience, introspection, or both. That is, he forms a new prior. He then updates this prior via Bayes' rule taking into account (s,π). We shall first consider the simplest model where the new prior is always the same, denoted by ρ^1—Section 6 relaxes this. We start from this baseline model because it is easier and cleaner to study. Also, assume that $\rho^1(\omega) \geq \underline{\rho}^1$ for all $\omega \in \Omega$, where $\underline{\rho}^1 > \varepsilon$ and is independent of ε. That is, ρ^1 deems no state impossible or even highly improbable.

To recap, facing evidence (s,π), Receiver updates his initial worldview ρ_ε as long as (s,π) confirms it. Otherwise, he changes worldview to ρ^1 and then updates it. Either way, updating follows Bayes' rule and results in a well-defined posterior belief, denoted by $p \in \Delta(\Omega)$.[13]

4 Feasible Posterior Beliefs

Since Receiver's beliefs determine his behavior, this section characterizes which posteriors p Sender can induce with experiments. In settings without worldview changes, every distribution over p's whose average equals Receiver's prior is feasible. This is no longer true here. The strategy will then be to use *Sender's* posterior—which satisfies standard properties—as

[12] A1 ensures that, as we change ε, given π the ratio $\frac{\pi(s|\mathcal{P})}{\pi(s|\mathcal{I})}$ remains fixed.

[13] Given a set Y, $\Delta(Y)$ denotes the set of probability distributions over Y; δ_y denotes a distribution that assigns probability one to y.

an intermediate variable and show that it always maps to a well-identified p, even though Receiver may update different priors. This mapping will shed light on the unique aspects of persuasion when it involves changing worldviews.

To formalize this, let q denote Sender's posterior (computed via Bayes' rule) and let τ denote any distribution over q's. From Aumann and Maschler (1995) and KG, we know that τ is induced by some π if and only if $\mathbf{supp}\,\tau$ is finite and $\mathbb{E}_\tau[q] = \sigma$. Let the set of such τ's be $\mathcal{T}$. Now suppose every q maps to a unique p through a function $\mathbf{p}_\varepsilon$ that is independent of π. Then, $(\mathcal{T}, \mathbf{p}_\varepsilon)$ characterizes all feasible distributions over Receiver's posteriors.

To build intuition for deriving $\mathbf{p}_\varepsilon$, it is useful to review a result from the benchmark case without worldview changes.

Proposition 1 (Alonso and Câmara (2016)). *Suppose Receiver always updates ρ using Bayes' rule, where $\mathbf{supp}\,\rho = \mathbf{supp}\,\sigma = \Omega$. There exists a function from Sender's to Receiver's posterior, $\mathbf{p} : \Delta(\Omega) \to \Delta(\Omega)$, which satisfies*

$$\mathbf{p}(\omega; q) = \frac{q(\omega)\frac{\rho(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega')\frac{\rho(\omega')}{\sigma(\omega')}}, \quad \omega \in \Omega.$$

This follows from observing that, if p and q are the posteriors induced by (s, π), they satisfy

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

Each $\pi(s|\omega)$ can then be written as $\frac{q(\omega|s,\pi)}{\sigma(\omega)} \times \mathit{const}$ and substituted into the first expression of p. The next implication is immediate but crucial for the rest of the paper.

Corollary 1 (Homeomorphism). $\mathbf{p}$ *is continuous, one-to-one, onto.*

The same logic can be used in settings with worldview changes, *if* we know which prior Receiver updates. It turns out that this is *also* pinned down by Sender's posterior, without having to know (s, π). For the case of $\varepsilon = 0$, this is immediate: $\mathbf{supp}\,q(\cdot|s,\pi)$ reveals which states (s, π) rules out. Therefore, if we let D_0 contain all q's induced by evidence disproving ρ_0 and C_0 contain all q's induced by evidence confirming ρ_0, we get

$$D_0 = \{q : \mathbf{supp}\,q \subseteq \mathcal{I}\} = \Delta(\mathcal{I}) \quad \text{and} \quad C_0 = \Delta(\Omega) \setminus D_0.$$

The case of $\varepsilon > 0$ follows by adapting AC's logic to express $\ell(s, \pi)$ as a function of q only.[14]

Lemma 1. *For every $\varepsilon \geq 0$ there exists a partition $(D_\varepsilon, C_\varepsilon)$ of $\Delta(\Omega)$, where $q(\cdot|s,\pi) \in D_\varepsilon$ if (s,π) disproves ρ_ε and $q(\cdot|s,\pi) \in C_\varepsilon$ if (s,π) confirms ρ_ε. Moreover, $C_\varepsilon \subset C_{\varepsilon'}$ and $D_\varepsilon \supset D_{\varepsilon'}$ if $\varepsilon > \varepsilon'$, $\lim_{\varepsilon \to 0} D_\varepsilon = D_0$, and $\lim_{\varepsilon \to 0} C_\varepsilon = C_0$.*

[14]The main results are proven in Appendix A. All other proofs are in Appendix B.

The second part reflects in terms of Sender's posteriors the property that, as Receiver becomes more confident in his worldview, the space of evidence sufficiently strong to change it shrinks (monotonically).

These observations lead to the first main result of the paper. Its proof is by now immediate and therefore omitted.

Proposition 2. *For every $\varepsilon \geq 0$ there exists a function from Sender's to Receiver's posterior, $\mathbf{p}_\varepsilon : \Delta(\Omega) \to \Delta(\Omega)$, which satisfies*

$$\mathbf{p}_\varepsilon(\omega; q) = \frac{q(\omega)\frac{\mathbf{r}(\omega;q)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q(\omega')\frac{\mathbf{r}(\omega';q)}{\sigma(\omega')}}, \quad \omega \in \Omega, \tag{1}$$

where $\mathbf{r}(q) = \rho_\varepsilon$ if $q \in C_\varepsilon$ and $\mathbf{r}(q) = \rho^1$ if $q \in D_\varepsilon$.

Comparing $\mathbf{p}_\varepsilon$ with $\mathbf{p}$ reveals notable differences between persuasion with and without worldview changes, summarized by the next three corollaries. Recall that we should think of ε as being small.

The first difference formalizes what is loosely called *surprise*. Recall that $\mathbf{p}$ is continuous.

Corollary 2 (Surprises). *For $\varepsilon \geq 0$ small enough, $\mathbf{p}_\varepsilon$ is not continuous at the boundary between C_ε and D_ε.*[15]

Such a discontinuity captures the consequences of changing worldview and its fundamental difference from Bayesian updating. Imagine a smooth transition from some expected to some unexpected evidence. Along the way Sender's bayesian posterior adjusts gradually. By contrast, at some (s', π') in between Receiver reaches a breaking point where he can no longer cling to ρ_ε and switches to ρ^1. At that point, he interprets (s', π') from a totally fresh perspective and derives substantively different conclusions. Receiver's belief jumps, while Sender's varies smoothly. Drastic swings in beliefs can cause similar changes in behavior, which calls for extra care when designing experiments.

The second difference formalizes the serious constraints that Receiver's reluctance to change worldview imposes on the posteriors Sender can induce. These belong to the sets $\mathbf{p}_\varepsilon(C_\varepsilon)$ and $\mathbf{p}_\varepsilon(D_\varepsilon)$ when she confirms and disproves ρ_ε, respectively.

Corollary 3 (Constraints). *As ε falls, both $\mathbf{p}_\varepsilon(C_\varepsilon)$ and $\mathbf{p}_\varepsilon(D_\varepsilon)$ shrink, and $\sup_{p \in \mathbf{p}_\varepsilon(C_\varepsilon)} p(\mathcal{I})$ and $\sup_{p \in \mathbf{p}_\varepsilon(D_\varepsilon)} p(\mathcal{P})$ decrease monotonically to zero. Also, $\lim_{\varepsilon \to 0} \mathbf{p}_\varepsilon(C_\varepsilon) = \Delta(\mathcal{P})$ and $\lim_{\varepsilon \to 0} \mathbf{p}_\varepsilon(D_\varepsilon) = \Delta(\mathcal{I})$.*

Thus, $\mathbf{p}_\varepsilon$ is not onto in contrast to $\mathbf{p}$. Intuitively, Receiver's worldview determines the extent to which he is receptive to new pieces of evidence and filters out those incompatible with it.

[15] This property does not require ε to be zero. In fact, one can show that it holds for every $\varepsilon > 0$ in the interior of $\Delta(\Omega)$. The function $\mathbf{p}_\varepsilon$ remains continuous over C_ε and D_ε, separately.

Since only strongly incompatible evidence shakes his confidence and leads him to fully take into account states in $\mathcal{I}$, his posteriors will be extreme relative to the sets $\mathcal{P}$ and $\mathcal{I}$. As a result, Receiver's poorer understanding of the world need not help Sender: In some settings, she would strictly prefer that he shared her level of understanding.[16]

Receiver's reluctance to change worldview creates constraints, but also opportunities. As he becomes more entrenched in his opinions, he also becomes less receptive to evidence inconsistent with it. This opens a growing gap between what he and Sender learn from expected evidence: While C_ε grows as ε falls, $\mathbf{p}_\varepsilon(C_\varepsilon)$ shrinks. As a result, she can conceal states inconsistent with his worldview by pooling them in evidence about states consistent with it. Consider the case of $\varepsilon = 0$. For every $\eta \in (0,1)$, $q' \in \Delta(\mathcal{P})$, and $\omega \in \mathcal{I}$, we have

$$q_\eta = (1-\eta)q' + \eta\delta_\omega \in C_0 \quad \text{and} \quad \mathbf{p}_0(\omega; q_\eta) = 0. \tag{2}$$

Also, $\Delta(\mathcal{P}) \subseteq \mathbf{p}_\varepsilon(C_\varepsilon)$ for every $\varepsilon \geq 0$. Therefore, Sender can become almost sure that $\omega \in \mathcal{I}$ obtained (i.e., $\eta \approx 1$) and, at the same time, keep Receiver thinking that ω is impossible *and* having any posterior in $\Delta(\mathcal{P})$.[17] This essentially continues to hold for sufficiently small $\varepsilon > 0$.

Corollary 4 (Concealment). *Let q_η be defined as in (2). For every $\varepsilon > 0$, there exists $\eta_\varepsilon < 1$ such that, if $\eta < \eta_\varepsilon$, then $q_\eta \in C_\varepsilon$ for all $q' \in \Delta(\mathcal{P})$ and $\omega \in \mathcal{I}$. Also, η_ε increases as ε decreases and $\lim_{\varepsilon \to 0} \eta_\varepsilon = 1$. Finally, if $\hat{q}, \tilde{q} \in C_\varepsilon$ for some ε and satisfy $\hat{q}(\cdot|\mathcal{P}) = \tilde{q}(\cdot|\mathcal{P})$, then $\lim_{\varepsilon \to 0} ||\mathbf{p}_\varepsilon(\hat{q}) - \mathbf{p}_\varepsilon(\tilde{q})|| = 0$.*

Thus, $\mathbf{p}_0$ is not one-to-one in contrast to $\mathbf{p}$: Every $q \in C_0$ with the same conditional $q(\cdot|\mathcal{P})$ maps to the same p. The last limit shows that this property is almost true for small $\varepsilon > 0$, even though $\mathbf{p}_\varepsilon$ remains one-to-one. Thus, Sender can design experiments that are very informative about states in $\mathcal{I}$, yet affect only marginally Receiver's *overall* posteriors.

Example 1 (Feasible posteriors). Recall the introduction example: $\Omega = \{\omega_{hf}, \omega_{hr}, \omega_{tf}, \omega_{tr}\}$, $\rho_0 = (0.7, 0.3, 0, 0)$, and $\rho^1 = \sigma = (0.07, 0.03, 0.5, 0.4)$. As Figure 1(a) shows, D_0 is the ω_{tf}–ω_{tr} edge of the simplex $\Delta(\Omega)$ and C_0 is the rest. Figure 1(b) shows that $\mathbf{p}_0$ is not onto: C_0 maps to the ω_{hf}–ω_{hr} edge and D_0 to the ω_{tf}–ω_{tr} edge. Figure 1(c) shows that $\mathbf{p}_0$ is not one-to-one: For example, every q in the shaded region $Q \subseteq C_0$ maps to the same $\mathbf{p}_0(Q)$. Figure 1(d) illustrates the discontinuity due to surprises: When q moves from Q to D_0, p jumps from the ω_{hf}–ω_{hr} edge to the ω_{tf}–ω_{tr} edge.

[16] Details of an example are available upon request.

[17] Note that, for every $\omega \in \mathcal{P}$, $\mathbf{p}_0(\omega; q_\eta)$ is entirely determined by q':

$$\mathbf{p}_0(\omega; q_\eta) = \frac{q_\eta(\omega)\frac{\rho_0(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \Omega} q_\eta(\omega')\frac{\rho_0(\omega')}{\sigma(\omega')}} = \frac{q_\eta(\omega|\mathcal{P})\frac{\rho_0(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \mathcal{P}} q_\eta(\omega'|\mathcal{P})\frac{\rho_0(\omega')}{\sigma(\omega')}} = \frac{q'(\omega)\frac{\rho_0(\omega)}{\sigma(\omega)}}{\sum_{\omega' \in \mathcal{P}} q'(\omega')\frac{\rho_0(\omega')}{\sigma(\omega')}}.$$

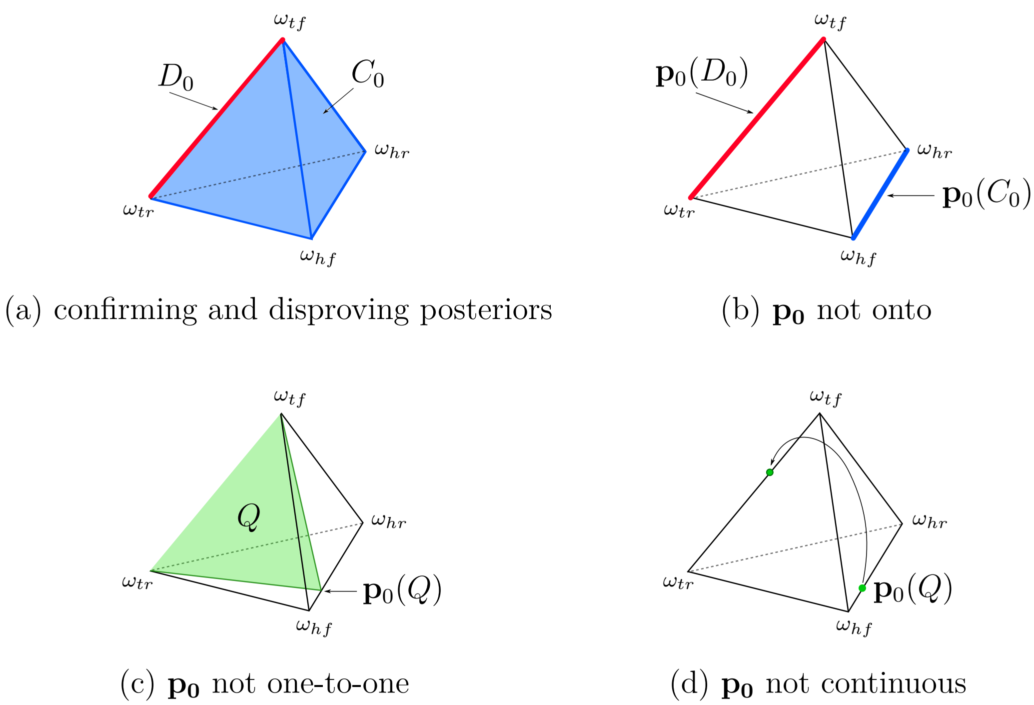

Figure 1: From Sender's to Receiver's posterior

5 Optimal Persuasion

This section examines how Sender chooses experiments in terms of choosing a feasible distribution over her and Receiver's posteriors. To this end, for every p let $a(p)$ be the action he implements, and for every q define her expected payoff by

$$\hat{v}_\varepsilon(q) = \mathbb{E}_q[v(a(\mathbf{p}_\varepsilon(q)), \omega)].$$

Sender's problem is to find $\tau \in \mathcal{T}$ that maximizes $\mathbb{E}_\tau[\hat{v}_\varepsilon(q)]$. In contrast to standard persuasion, this problem may not have an exact solution due to the lack of continuity of $\mathbf{p}_\varepsilon$ (see Online Appendix B.1 for an example). Importantly, this technicality does not preclude an instructive analysis of Sender's problem. Under our assumptions about Receiver, it boils down to a single-agent decision problem. We can then study its *value function*,

$$V_\varepsilon^* = \sup_{\tau \in \mathcal{T}} \mathbb{E}_\tau[\hat{v}_\varepsilon(q)], \tag{3}$$

and infer from it properties of τ's and associated π's that are at least *virtually* optimal.

We first need a couple of technical lemmas and some definitions. Thanks to Proposition 2, we can still rely on the "concavification" procedure of Aumann and Maschler (1995) and KG. Let V_ε be the *concavification* of $\hat{v}_\varepsilon$, namely, the smallest concave function w that satisfies $w(q) \geq \hat{v}_\varepsilon(q)$ for all q. The next lemma allows us to express whether it is optimal for Sender to reveal any information using V_ε and $\hat{v}_\varepsilon$.

Lemma 2. $V_\varepsilon^* = V_\varepsilon(\sigma)$. *In (3), it suffices to restrict attention to τ's with $|\mathbf{supp}\,\tau| \leq |\Omega|$.*

Definition 2 (Optimality of Revealing Information)**.** If $V_\varepsilon(\sigma) = \hat{v}_\varepsilon(\sigma)$, we will say that it is optimal to reveal no information ($\tau = \delta_\sigma$ is optimal). If $V_\varepsilon(\sigma) > \hat{v}_\varepsilon(\sigma)$, we will say that it is (virtually) optimal to reveal some information (some $\tau' \in \mathcal{T}$ strictly dominates $\tau = \delta_\sigma$).

In this paper the main question is not whether Sender reveals any information at all, but if and when she disproves ρ_ε and how she designs π accordingly. Since fully uninformative π's always confirm ρ_ε, hereafter we shall assume that it is optimal to reveal some information.[18]

To study the incentives to disprove ρ_ε, it is useful to consider the function $\hat{v}_\varepsilon$ over C_ε and D_ε *separately*. Let its restriction to C_ε and D_ε be $\hat{v}^c_\varepsilon$ and $\hat{v}^d_\varepsilon$, with concavifications V^c_ε and V^d_ε. Note that $V^c_\varepsilon(q) \leq V_\varepsilon(q)$ for $q \in C_\varepsilon$ because this is a proper subset of the domain of $\hat{v}_\varepsilon$. The next lemma allows us to express whether it is optimal to disprove ρ_ε using V_ε and V^c_ε.

Lemma 3. *If $V^c_\varepsilon(\sigma) < V_\varepsilon(\sigma)$, there exists $\tau \in \mathcal{T}$ such that $\mathbb{E}_\tau[\hat{v}_\varepsilon(q)] > V^c_\varepsilon(\sigma)$. Moreover, if $\mathbb{E}_\tau[\hat{v}_\varepsilon(q)] > V^c_\varepsilon(\sigma)$, then the probability of disproving ρ_ε, denoted by τ^d, is strictly positive.*

Definition 3 (Optimality of Disproving Receiver)**.** If $V^c_\varepsilon(\sigma) < V_\varepsilon(\sigma)$, we will say that it is (virtually) optimal to disprove ρ_ε. If $V^c_\varepsilon(\sigma) = V_\varepsilon(\sigma)$, we will say that it is (virtually) optimal not to disprove ρ_ε.

For the sake of exposition, hereafter the analysis focuses on the case of $\varepsilon = 0$. This will also be justified later by showing that $\lim_{\varepsilon\to 0} V^c_\varepsilon(\sigma) = V^c_0(\sigma)$ and $\lim_{\varepsilon\to 0} V_\varepsilon(\sigma) = V_0(\sigma)$, which involves a constructive and non-trivial proof due to the lack of continuity of $\mathbf{p}_\varepsilon$. For practical purposes, one can then use the model with $\varepsilon = 0$ as an approximation of the model with small $\varepsilon > 0$. While it is hard to work with the latter, the former is more amenable to analysis (see, e.g., Section 5.1). To simplify notation, we will drop the subscript 0 when there is no risk of confusion.

The decision to disprove Receiver depends on the following trade-off. The benefit is simple: When τ induces $q \in D$, Sender's payoff is $\hat{v}^d(q)$. Measuring the cost is more subtle. Intuitively, what she gives up by disproving ρ_0 is the best payoff she can get from concealing the states in $\mathcal{I}$ while confirming ρ_0 (recall Corollary 4). To compute this payoff and characterize the optimal concealing strategies, let's focus on feasible τ's that always confirm ρ_0 (i.e., $\mathbf{supp}\,\tau \subseteq C$) and derive

$$V^c(\sigma) = \sup_{\{\tau \in \mathcal{T}:\, \mathbf{supp}\,\tau \subseteq C\}} \mathbb{E}_\tau[\hat{v}(q)].$$

To do so, let $\mathcal{A}_0$ contain all actions that Receiver would choose for some posterior in $\Delta(\mathcal{P})$—that is, $\mathcal{A}_0 = \cup_{p \in \Delta(\mathcal{P})} \mathcal{A}(p)$, where $\mathcal{A}(p) = \arg\max_{a \in A} \mathbb{E}_p[u(a, \omega)]$. Recall that each q is induced by some signal s.

[18] As shown below, for states in $\mathcal{P}$ Sender designs π as in KG and AC. Thus, their sufficient conditions for her to optimally reveal some information can be adapted here to ensure that $V_\varepsilon(\sigma) > \hat{v}_\varepsilon(\sigma)$.

Proposition 3 (Optimal Concealment)**.** *To compute $V^c(\sigma)$ it suffices to consider distributions τ with the following properties:*
(1) Each signal conceals no more than one state: *For every $q \in \mathbf{supp}\,\tau$, $q(\omega) > 0$ for at most one $\omega \in \mathcal{I}$.*
(2) Each state is concealed in one signal: *For every $\omega \in \mathcal{I}$, $q(\omega) > 0$ for only one $q \in \mathbf{supp}\,\tau$.*
(3) Each concealing signal induces an optimal action for the concealed state: *If $q \in \mathbf{supp}\,\tau$ and $q(\omega) > 0$ for some $\omega \in \mathcal{I}$, then $a(\mathbf{p}(q)) \in \arg\max_{a\in\mathcal{A}_0} v(a,\omega)$.*
Letting $h(\omega) = \max_{a\in\mathcal{A}_0} v(a,\omega)$ for every $\omega \in \mathcal{I}$, we have

$$V^c(\sigma) = \sigma(\mathcal{P})V^c(\sigma(\cdot|\mathcal{P})) + \sum_{\omega\in\mathcal{I}} h(\omega)\sigma(\omega). \quad (4)$$

To see the intuition, recall that Receiver interprets every (s,π) confirming his worldview through its lens, thereby disregarding the content of s about states he deems impossible. Therefore, whether one or multiple such states give rise to s has no effect on his action, which explains property (1). Sender can then conceal each $\omega \in \mathcal{I}$ in only one of the pieces of confirming evidence, which explains (2). And if this piece arises with very high probability only in ω, the resulting action essentially will not matter for any other state; so she can freely manipulate this action to cater to her interests in ω (property (3)).

We can now define the *opportunity cost of surprising* Receiver. Every $q \in D$ is induced by evidence supporting only states in $\mathcal{I}$. Its opportunity cost for Sender is then the payoff she would expect, with belief q, if she instead optimally concealed all those states (i.e., all $\omega \in \mathbf{supp}\,q$).

Definition 4 (Opportunity Cost of Surprising)**.** The opportunity cost of surprising Receiver is measured by the function $\hat{h} : D \to \mathbb{R}$ defined as

$$\hat{h}(q) = \mathbb{E}_q[h(\omega)].$$

Given this, we will say that there is *knowledge* Sender would share if

$$\hat{v}^d(q) > \hat{h}(q) \text{ for some } q \in D. \quad (5)$$

That is, there exists *some* evidence supporting only states outside Receiver's worldview such that, if Sender had this evidence, she would rather share it with him than mask it as consistent with his worldview in the most advantageous manner for her.

Proposition 4. *It is optimal for Sender to disprove Receiver's worldview if and only if there is knowledge she would share.*

For common-prior settings without worldview changes, KG show that a key condition for Sender to benefit from persuasion is that "there is *information* she would share:" $\hat{v}(q) > \mathbb{E}_q[v(a(\rho_0),\omega)]$ for some $q \in \Delta(\Omega)$. This condition differ substantively from (5). Withholding information on something Receiver already takes into account has the only consequence of him acting on his prior. By contrast, sharing knowledge that Receiver ignores involves more intricate considerations due to the many ways one can conceal it.

Example 2 (Knowledge Sender would share)**.** For the scientist, the best way to conceal ω_{tf} and ω_{tr} is to induce the policymaker to choose a_{hr}, which occurs if $q(\omega_{hr}|\mathcal{P}) \geq 0.5$ (recall Figure 2). Since $h(\omega_{tf}) = h(\omega_{tr}) = 3$ while $\hat{v}^d(\delta_{\omega_{tr}}) = 4$, there is knowledge the scientist would share.

The next result characterizes Sender's payoff and experiment when it is optimal to disprove Receiver. To keep track of its cost and benefits, define the function $\hat{m} = \max\{\hat{v}^d, \hat{h}\}$ and denote its concavification by M.

Proposition 5. *If it is optimal for Sender to disprove Receiver, then*

$$V(\sigma) = \sigma(\mathcal{P})V^c(\sigma(\cdot|\mathcal{P})) + \sigma(\mathcal{I})M(\sigma(\cdot|\mathcal{I})), \tag{6}$$

where

$$M(\sigma(\cdot|\mathcal{I})) = \max_{\gamma, q^c, q^d}\{\gamma V^d(q^d) + (1-\gamma)\hat{h}(q^c)\}, \tag{7}$$

subject to $\gamma \in [0,1]$, $q^c, q^d \in D$, and $\gamma q^d + (1-\gamma)q^c = \sigma(\cdot|\mathcal{I})$. Any distribution on D that achieves $V^d(q^d)$ assigns positive probability only to q's such that $\hat{v}^d(q) \geq \hat{h}(q)$. Given an optimal γ^* in (7), the probability of disproving Receiver is $\tau^d = \gamma^*\sigma(\mathcal{I})$, where $\gamma^* = 1$ if and only if $M(\sigma(\cdot|\mathcal{I})) = V^d(\sigma(\cdot|\mathcal{I}))$.

Expression (6) helps us understand how Sender designs π, by breaking down this problem into parts and algorithms which are familiar from standard persuasion without worldview changes. Sender can divide her task between the states in $\mathcal{P}$ and in $\mathcal{I}$, and think about how she would design π conditional on knowing only whether $\mathcal{P}$ or $\mathcal{I}$ occurred. Given $\mathcal{P}$, trivially every π produces evidence confirming Receiver's worldview. Her problem is then equivalent to those studied by KG and AC with prior $\sigma(\cdot|\mathcal{P})$ for Sender and ρ_0 for Receiver. As they show, the properties of the optimal π can be inferred from the concavification of her payoff function at that prior, namely, $V^c(\sigma(\cdot|\mathcal{P}))$. Thus, their characterizations apply here as well.

Given $\mathcal{I}$, instead, Sender has to decide when to confirm Receiver's worldview by concealing states, when to disprove it, and with which evidence. This is a complex problem. Its solution, however, is again given by a concavification at the specific belief $\sigma(\cdot|\mathcal{I})$, yet of the new function $\hat{m}$. To obtain this solution, Proposition 5 also shows a simpler way using only V^d, $\hat{h}$, and two posteriors, q^d and q^c. The optimal q^c immediately tells us which states Sender conceals with some probability (namely, those in $\mathbf{supp}\, q^c$). Concealment follows the properties in Proposition 3: For each ω, $h(\omega)$ tells us which action Receiver implements and hence which evidence should be used to conceal ω. Finally, $V^d(q^d)$ tells us the properties of the optimal π when it disproves Receiver, again as in KG and AC.

The introduction example helps illustrate these points.

Example 3 (Optimal π)**.** We will compute $V^c(\sigma(\cdot|\mathcal{P}))$ and $M(\sigma(\cdot|\mathcal{I}))$ in (6) for the example and derive from them the optimal τ^* and associated π^*. We already know that $\hat{h}(q) = 3$ independently of q (Example 2). We need to find $\hat{v}^c$ for $q \in \Delta(\mathcal{P})$ and $\hat{v}^d$. Recall that $A = \{a_{hf}, a_{hr}, a_{tf}, a_{tr}\}$, $v(a, \omega) = v(a)$ as shown in Table 2, and $u(a, \omega)$ equals 1 if a matches ω and 0 otherwise. Note that $\mathbf{p}_0(q) = q$ for $q \in \Delta(\mathcal{P}) \cup D_0$ since $\rho_0 = \sigma(\cdot|\mathcal{R})$ and $\rho^1(\cdot|\mathcal{I}) = \sigma(\cdot|\mathcal{I})$. Figure 2 then illustrates $\hat{v}^c$ over $\Delta(\mathcal{P})$ as a solid line; Figure 3 illustrates $\hat{h}$ and $\hat{v}^d$ (solid line).

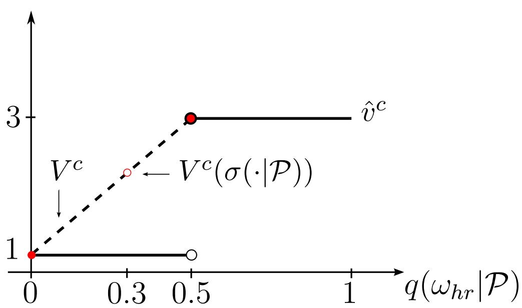

Figure 2: Confirming Receiver

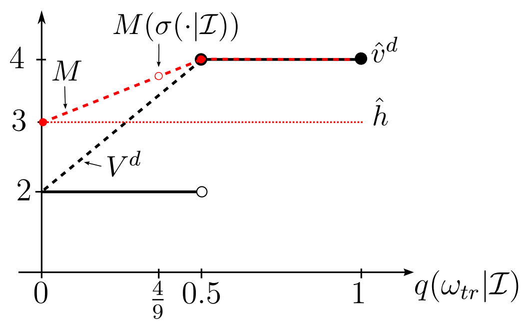

Figure 3: Disproving Receiver

Given $\sigma(\cdot|\mathcal{P}) = (0.7, 0.3)$, we obtain $V^c(\sigma(\cdot|\mathcal{P}))$ by inducing q' and q'' in Table 4, which map to the payoffs denoted by solid red circles in Figure 2. Since $\alpha q' + (1-\alpha)q'' = \sigma(\cdot|\mathcal{P})$ if and only if $\alpha = 0.4$, we have that $\tau^*(q') = \alpha\sigma(\mathcal{P}) = 0.04$ and $\tau^*(q'') = (1-\alpha)\sigma(\mathcal{P}) = 0.06$.

Now consider $M(\sigma(\cdot|\mathcal{I}))$, where $\sigma(\cdot|\mathcal{I}) = (\frac{5}{9}, \frac{4}{9})$. For $q \in D_0$, $\hat{m} = \max\{\hat{v}^d, \hat{h}\}$ takes value 3 if $q(\omega_{tr}) < 0.5$ and 4 otherwise. Thus, we obtain $M(\sigma(\cdot|\mathcal{I}))$ by inducing q^c and q^d in Table 4, which map to the payoffs denoted by solid red circles in Figure 3. Importantly, q^c leads to payoff $h(\omega_{tf}) = 3$, which means that the scientist conceals ω_{tf} and induces a_{hr}. Again, $\gamma q^d + (1-\gamma)q^c = \sigma(\cdot|\mathcal{I})$ if and only if $\gamma = \frac{8}{9}$ and hence $\tau^*(q^c) = (1-\gamma)\sigma(\mathcal{I}) = 0.1$ and $\tau^*(q^d) = \gamma\sigma(\mathcal{I}) = 0.8$. Since $V^d(\sigma(\cdot|\mathcal{I})) < M(\sigma(\cdot|\mathcal{I}))$ (see Figure 3), Proposition 5 implies that τ^d is not maximal: Indeed, $\tau^d = \tau^*(q^d) < 0.9 = \sigma(\mathcal{I})$.

To translate τ^* into π^*, shown in Table 5, let s', s'', and s^d be the signals inducing q', q'', and q^d. Since q'' is the only posterior putting positive probability on ω_{hr}, we must have $\pi^*(s''|\omega_{hr}) = 1$. To get q'', s'' must have a 50-50 chance of arising from ω_{hf} and ω_{hr} according to σ, which requires $\pi^*(s''|\omega_{hf}) = \frac{3}{7}$. With the remaining probability, ω_{hf} must be the only state producing s'. As for s'', for s^d we must have $\pi^*(s^d|\omega_{tr}) = 1$ and a 50-50 chance that it arises from ω_{tf} and ω_{tr}, which requires $\pi^*(s^d|\omega_{tf}) = \frac{4}{5}$. With the remaining probability, ω_{tf} must be concealed to induce a_{hr}, which can be done using s''.

Proposition 5 addresses two other questions. First, is Sender ever forced to surprise Receiver in some states she would rather conceal in order to reach a specific goal in some other states? This is not obvious. The feasibility constraint $\mathbb{E}_\tau[q] = \sigma$ captures the fact that,

τ^*	0.04	0.06	0.1	0.8
Posterior	q'	q''	q^c	q^d
ω_{hf}	1	0.5	0	0
ω_{hr}	0	0.5	0	0
ω_{tf}	0	0	1	0.5
ω_{tr}	0	0	0	0.5

Table 4: Optimal distribution over posteriors

$\pi^*(s\|\omega)$	s'	s''	s^d
ω_{hf}	$\frac{4}{7}$	$\frac{3}{7}$	0
ω_{hr}	0	1	0
ω_{tf}	0	$\frac{1}{5}$	$\frac{4}{5}$
ω_{tr}	0	0	1

Table 5: Optimal experiment

for π to produce signal s in favor of some states, it must also produce signal s' against them, and the likelihood and intensity of s depends on those of s'. The answer is no, however: No optimal π has to induce $q \in D$ where the opportunity cost of surprising exceeds its benefit. The second question is whether Sender disproves ρ_0 whenever ω is inconsistent with it. Proposition 5 provides a necessary and sufficient condition, which involves comparing M and V^d *at the exogenously given* belief $\sigma(\cdot|\mathcal{I})$.[19] This can be used for finding ways to incentivize Sender to expand Receiver's worldview whenever possible: It suffices to change her final payoff in any way that leads to $M(\sigma(\cdot|\mathcal{I})) = V^d(\sigma(\cdot|\mathcal{I}))$.

Is it possible to identify in which states Sender will disprove Receiver, without having to characterize π entirely? The answer is given by the benefit and cost of *fully* revealing $\omega \in \mathcal{I}$ (i.e., of inducing posterior δ_ω), which can be easily derived from primitives of the model.

Corollary 5. *Fix $\omega \in \mathcal{I}$. If $\hat{v}^d(\delta_\omega) \geq h(\omega)$, then Sender surprises Receiver whenever ω obtains. If $\hat{v}^d(\delta_\omega) < h(\omega)$ and $\hat{v}^d$ is convex, then Sender never surprises Receiver in ω.*

Note that $\hat{v}^d(\delta_\omega) \geq h(\omega)$ does not imply that at the optimum Sender fully reveals ω. When $\hat{v}^d$ is not convex, even if $\hat{v}^d(\delta_\omega) < h(\omega)$, she may reveal ω with positive probability (even whenever ω occurs). As our running example illustrates, this is done to increase the total probability of some desirable outcome. By Corollary 5, one way to incentivize Sender to disprove ρ_0 whenever possible is to raise her payoff from *fully* revealing each $\omega \in \mathcal{I}$ to $h(\omega)$.

The last result shows that the model with $\varepsilon = 0$ is the limit of the model with $\varepsilon > 0$. Its proof uses the following minor condition.

Assumption 3. *If action a is optimal for Receiver at some $p \in \Delta(\mathcal{P})$, then there exists $\hat{p} \in \Delta(\mathcal{P})$ that renders a strictly optimal.*

Proposition 6. $\lim_{\varepsilon\to 0} V^c_\varepsilon(\sigma) = V^c_0(\sigma)$ *and* $\lim_{\varepsilon\to 0} V_\varepsilon(\sigma) = V_0(\sigma)$.

Two things are worth noting. First, this limit result does *not* hold if we consider a model without worldview changes where Receiver assigns $\varepsilon > 0$ probability to some states and we

[19]This condition can be used to obtain simpler ones which imply that Sender will not disprove ρ_0 for some $\omega \in \mathcal{I}$ (see Corollary 9 in Online Appendix B).

let $\varepsilon \to 0$ (see Section 7.2). Second, if for $\varepsilon = 0$ there is knowledge Sender would share (condition (5)), then it is optimal for her to disprove ρ_ε also for sufficiently small $\varepsilon > 0$.

5.1 Application: Lobbying

To illustrate the results and how optimal experiments depend on the agents' interests—also in comparison with standard Bayesian persuasion—this section considers a lobbying application similar to that in KG. The main difference is that the lobbyist (Sender) is an expert and can change the worldview of the politician (Receiver). This changes KG's predictions, also removing some of the discrepancies with reality pointed out by the authors.

The politician has to regulate a complex issue for society, a decision the lobbyist wants to influence. To do so, she commissions a study π to be submitted to his office. The policy space is $A = \mathbb{R}_+$.[20] Think of $a \in A$ as the level of regulation of a drug produced by the lobbyist's industry. Let ω be the socially optimal level. The politician is benevolent: $u(a, \omega) = -(a - \omega)^2$. Thus, $a(p) = \mathbb{E}_p[\omega]$ for every p. The lobbyist is biased: $v(a, \omega) = -(a - \beta(\omega))^2$, where $\beta(\omega)$ is *her* ideal policy given ω. The assumption here is that $\beta(\omega) < \omega$ for all ω, and that $\beta(\omega) = \kappa\omega + b$ for some $\kappa \in (0, 1]$ and $b < 0$. The industry is against regulation, but accepts that some restrictions may be necessary to prevent drug abuses ($\kappa > 0$).

Both agents know that the issue to be regulated is complex, but the lobbyist has a deeper and richer understanding of its many intricate aspects. In her worldview σ, the set of possible social optima is $\Omega = \{\omega_1, \dots, \omega_n\}$, where $0 < \omega_i < \omega_{i+1}$ for all i. By contrast, the coarser worldview of the inexpert politician ignores some policies that can be socially optimal: $\mathbf{supp}\, \rho_0 = \mathcal{P} \subset \Omega$. For simplicity, let $\rho_0 = \sigma(\cdot|\mathcal{P})$ and $\rho^1 = \sigma$. Also, we shall consider two cases separately: (1) $\mathcal{P}$ leaves out "extreme" regulations ($\mathcal{P} = \{\omega_i\}_{i=\underline{n}}^{\overline{n}}$ for $1 < \underline{n} < \overline{n} < n$); (2) $\mathcal{P}$ leaves out "moderate" regulations ($\mathcal{I} = \{\omega_i\}_{i=n'}^{n''}$ for $1 < n' < n'' < n$). We will analyze (1) and discuss (2) at the end.

As a first step, consider the lobbyist's payoffs from concealing states. For every $\omega \in \mathcal{I}$,

$$h(\omega) = \begin{cases} -(\omega_{\overline{n}} - \beta(\omega))^2 & \text{if } \beta(\omega) > \omega_{\overline{n}} \\ -(\omega_{\underline{n}} - \beta(\omega))^2 & \text{if } \beta(\omega) < \omega_{\underline{n}} \\ 0 & \text{if } \omega_{\underline{n}} \le \beta(\omega) \le \omega_{\overline{n}}. \end{cases}$$

In words, if her ideal policy at $\omega \in \mathcal{I}$ lies outside the range the politician would choose under his worldview, it is optimal to convince him to implement the most extreme in this range, depending on which is closest to $\beta(\omega)$. If $\beta(\omega)$ lies inside that range, she can induce him to implement exactly $\beta(\omega)$ by appropriately manipulating his beliefs: If $\omega_{\underline{n}} \le \beta(\omega) \le \omega_{\overline{n}}$, there exists $p \in \Delta(\mathcal{P})$ such that $\mathbb{E}_p[\omega] = \beta(\omega)$.

We can immediately see that, independently of the lobbyist's bias, there is knowledge

[20] As the proofs show, Propositions 3–5 hold unchanged even if A is infinite.

she would share and hence she will disprove the politician with positive probability (Proposition 4). She also *never* conceals any $\omega < \omega_{\underline{n}}$ (Corollary 5). Indeed, for $\omega < \omega_{\underline{n}}$, $h(\omega)$ is smaller than the payoff of fully revealing it: $\hat{v}^d(\delta_\omega) = -(\omega - \beta(\omega))^2$. In words, the anti-regulation lobbyist never benefits from letting the politician incorrectly think that it is socially optimal to regulate her industry more than it actually is. This is already a sharp difference from KG's predictions: In their model, for some biases (see below) the optimal π is fully uninformative and hence provides *no* information even when ω is smaller than the politician's prior expectation.

The lobbyist's bias plays a bigger and subtler role in how she designs π. To see this, let's derive her payoff $\hat{v}(q) = \mathbb{E}_q[v(\mathbb{E}_{\mathbf{p}(q)}[\omega], \omega)]$ for every posterior q:

$$\hat{v}(q) = -(\mathbb{E}_{\mathbf{p}(q)}[\omega])^2 + 2\kappa\mathbb{E}_{\mathbf{p}(q)}[\omega]\mathbb{E}_q[\omega] + 2b(\mathbb{E}_{\mathbf{p}(q)}[\omega] - \kappa\mathbb{E}_q[\omega]) - \kappa^2\mathbb{E}_q[\omega^2] - b^2.$$

By Proposition 5, we only need to worry about $\hat{v}(q)$ for $q \in \Delta(\mathcal{P})$ and $q \in D_0 = \Delta(\mathcal{I})$, as this is what matters for finding $V^c(\sigma(\cdot|\mathcal{P}))$ and $M(\sigma(\cdot|\mathcal{I}))$ in expression (6). The assumption that $\rho_0 = \sigma(\cdot|\mathcal{P})$ and $\rho^1 = \sigma$ implies that $\mathbf{p}(q) = q$ for every $q \in \Delta(\mathcal{P}) \cup \Delta(\mathcal{I})$.[21] This yields

$$\hat{v}(q) = (2\kappa - 1)(\mathbb{E}_q[\omega])^2 + 2b(1-\kappa)\mathbb{E}_q[\omega] - \kappa^2\mathbb{E}_q[\omega^2] - b^2, \quad q \in \Delta(\mathcal{P}) \cup \Delta(\mathcal{I}),$$

which is strictly convex (concave) if and only if $\kappa > (<) \frac{1}{2}$. If $\kappa \geq \frac{1}{2}$, both $V^c(\sigma(\cdot|\mathcal{P}))$ and $V^d(\sigma(\cdot|\mathcal{I}))$ are then achieved with a fully informative π (uniquely if $\kappa > \frac{1}{2}$); if instead $\kappa < \frac{1}{2}$, they are both achieved only with a fully uninformative π.[22]

Consider first a lobbyist who is sufficiently sensitive to society's needs: $\kappa \geq \frac{1}{2}$. Her optimal π always fully reveals ω, except when ω is above but close to $\omega_{\overline{n}}$. For $\omega > \omega_{\overline{n}}$, she faces a trade-off between pursuing her anti-regulation bias and letting the chosen policy adjust to the social optimum; the bias wins when concealing the truth leads to only mildly inefficient policies. Accordingly, she conceals more when her bias is stronger (i.e., κ *or* b are smaller). This is a second difference from KG's prediction: In their model, π is fully informative when $\kappa \geq \frac{1}{2}$ (independently of b).

Corollary 6. *If $\kappa \geq \frac{1}{2}$, the lobbyist's study π has the following properties:*
- *π fully reveals ω_i for all $i \leq \overline{n}$;*
- *there exist $b^*(\kappa) \leq 0$ and $i^*(b, \kappa)$ such that, if $b \geq b^*(\kappa)$, π fully reveals all states, while if $b < b^*(\kappa)$, π always conceals ω_i if $\overline{n} < i < i^*(b, \kappa)$ and fully reveals ω_i if $i \geq i^*(b, \kappa)$;*
- *$b^*(\kappa)$ decreases in κ (strictly if negative) and $i^*(b, \kappa)$ decreases in κ and b.*

For example, $b^*(1) = \frac{1}{2}(\omega_{\overline{n}} - \omega_{\overline{n}+1}) < 0$ (see the corollary's proof); so, even if the agents agree on how sensitive policies should be to ω, the lobbyist conceals no state only if her *systematic* anti-regulation bias is sufficiently weak.

[21] More generally, one could apply the techniques developed by AC for cases with different common-support priors to characterize $V^c(\sigma(\cdot|\mathcal{P}))$ and $M(\sigma(\cdot|\mathcal{I}))$.

[22] With regard to $V^c(\sigma(\cdot|\mathcal{P}))$, these properties are of course the same as in KG.

Consider now a lobbyist who cares relatively little about society's needs: $\kappa < \frac{1}{2}$. Unsurprisingly, this results in a much less informative π. For all $\omega \in \mathcal{P}$ and when disproving the politician for $\omega \in \mathcal{I}$, π produces the same evidence ($V^d(q^d) = \hat{v}^d(q^d)$ since v^d is concave). Thus, when π supports an unexpected policy, this is always the same policy and is either very low or very high. The optimal π can also support other policies within the range initially considered by the politician, which are as close as possible to the lobbyist's ideal policy for the concealed states. These properties can be inferred from the expression for $V(\sigma)$ in (8). They are notably different from KG's predictions: In their model, π is fully uninformative when $\kappa < \frac{1}{2}$ (again independently of b).

Corollary 7. *If $\kappa < \frac{1}{2}$, the lobbyist's expected payoff from her study π equals*

$$-\tau^d \mathbb{E}_{q^d}(a^d - \omega)^2 + (1 - \tau^d) \sum_{\omega > \omega_{\overline{n}}} h(\omega) q^c(\omega) - \sum_{\omega \in \mathcal{P}} (\mathbb{E}_\sigma[\omega|\mathcal{P}] - \omega)^2 \sigma(\omega), \tag{8}$$

where $\tau^d > 0$, $\tau^d q^d + (1 - \tau^d) q^c = \sigma$, and $a^d = \mathbb{E}_{q^d}[\omega]$. Either $\mathbb{E}_\sigma[\omega|\omega < \omega_{\underline{n}}] \leq a^d < \omega_{\underline{n}}$ or $a^d > \omega_{\overline{n}}$. If $\mathbb{E}_\sigma[\omega|\mathcal{I}] \leq \omega_{\overline{n}}$, π conceals some $\omega > \omega_{\overline{n}}$ with positive probability and $a^d < \omega_{\underline{n}}$.

The condition $\mathbb{E}_\sigma[\omega|\mathcal{I}] \leq \omega_{\overline{n}}$ can be interpreted as saying that the lobbyist does not assign too high a probability to large states. In this case, the only unexpected policy she induces will cater to her anti-regulation bias ($a^d < \omega_{\underline{n}}$).

Finally, suppose the politician ignores that moderate regulation levels can be socially optimal (case (2)). Focus on $\mathcal{I}$. If $\beta(\omega) \geq \omega_1$ for all $\omega \in \mathcal{I}$, there is no knowledge the lobbyist would share.[23] All levels ignored by the politician are more stringent than the lowest he contemplates. The lobbyist has then no reason to contradict him: She will conceal all states in $\mathcal{I}$ and obtain her ideal policies. If instead $\beta(\omega) < \omega_1$ for some $\omega \in \mathcal{I}$, she never conceal such ω, again because she never benefits by letting the politician overestimate the social optimum. If $\kappa \geq \frac{1}{2}$, she fully conceals all states for which $\beta(\omega) \geq \omega_1$. If $\kappa < \frac{1}{2}$, she again always produces the same evidence when disproving the politician and may reveal states with $\beta(\omega) \geq \omega_1$ to obtain a better policy a^d. Either way, a^d will cater to her anti-regulation bias ($a^d < \omega_1$).

In conclusion, the lobbyist's experiment tends to be biased in favor of the industry she represents, by leveraging the politician's different worldview. This seems consistent with KG's observation that "industry-funded studies often seem to produce results more favorable to the industry than independent studies," which they acknowledge is at odds with the predictions of their Bayesian model. This section's results suggest that incompatible worldviews may be as important as conflicting interests in shaping persuaders' strategies.

[23] Note that $h(\omega) = 0$ whenever $\beta(\omega) \geq \omega_1$.

6 Multiple Alternative Worldviews

In the baseline model, after unexpected evidence Receiver always adopts the same new worldview. This section allows for multiple alternative worldviews. This may capture agents who are willing to expand their worldview only gradually, or can handle only simple theories with a small number of possibilities due to cognitive limitations. This section shows that multiple alternative worldviews can help Sender as well as hurt her. Besides this, the main message of the paper does not change.

Lexicographic Belief System

The main question here is how to model a Receiver who may adopt alternative worldviews.[24] Perhaps the simplest way is to borrow Kreps and Wilson's (1982) idea of a "sequence of hypotheses." Imagine that Receiver treats $\rho^0 = \rho_\varepsilon$ as his "primary" worldview, which he always tries to apply first in order to process evidence. If he deems the evidence unexpected under ρ^0, however, he tries to use a second worldview, ρ^1, then a third, ρ^2, and so on, stopping at the first one in the sequence for which the evidence ceases to seem unexpected. Thus, Receiver "digs deeper" into hypotheses of how the world functions only when forced by the evidence.

To formalize this idea, we can let each hypothesis ρ^i assign some small probability $\varepsilon^i > 0$ to a subset of states and adopt across hypotheses the same definition of unexpected evidence as in the baseline model. For the sake of space, this section focuses on the case of $\varepsilon^i = 0$.

Assumption 4 (A4: Lexicographic-Belief-System (LBS) Model)**.** *Receiver is described by a finite sequence $\rho^0, \rho^1, \ldots, \rho^N \in \Delta(\Omega)$ such that (i) for each $\omega \in \Omega$, $\rho^i(\omega) > 0$ for some $i \in \{0, \ldots, N\}$ and (ii) $\mathbf{supp}\,\rho^i \not\subset \mathbf{supp}\,\rho^j$ if $i > j$. Given (s, π), he computes $p(\cdot|s, \pi)$ by applying Bayes' rule to the ρ^i of lowest index i that satisfies $\sum_{\omega\in\Omega} \pi(s|\omega)\rho^i(\omega) > 0$.*

Online Appendix B.2 presents the analysis of the feasible distributions over posteriors and optimal persuasion for the LBS model, whose conclusions are summarized here.

The feasible distributions change in two ways, but the substance is unaffected. A new function $\mathbf{p}_L$ maps Sender's posterior to Receiver's posteriors, which generalizes $\mathbf{p}_0$ while maintaining its key properties highlighted in Corollary 2–4. The first difference is that the range of $\mathbf{p}_L$ is a proper subset of that of $\mathbf{p}_0$. Therefore, when Receiver has more than one alternative worldview, persuasion may be even harder for Sender to the extent that there are even fewer posteriors she can induce Receiver to have.

A multiplicity of Receiver's worldviews does not mean only bad news for Sender, as it also expands the possibilities for concealment. Unlike in the baseline model, now some worldviews

[24] Recall that this model can simply reflect Sender's subjective view.

besides ρ_0 continue to deem some states impossible and may offer a better opportunity to conceal those states, depending on which actions Sender can induce Receiver to implement. This highlights another novel aspects of the present setting: Roughly speaking, an expert persuader has to decide "how deeply or hard" she wants to make her listener think in order to form a new worldview.

To illustrate these new constraints and opportunities, the next example shows that multiple alternative worldviews can help Sender, but can also hurt her.

Example 4 (Multiple alternative worldviews)**.** Modify the introduction example by letting the policymaker have an intermediate worldview which rules out ω_{tf} or ω_{tr}. That is, $\text{LBS}_{tr} = (\rho^0, \hat{\rho}^1, \sigma)$ and $\text{LBS}_{tf} = (\rho^0, \tilde{\rho}^1, \sigma)$, where $\rho^0 = \sigma(\cdot|\mathcal{P})$, $\omega_{tr} \notin \textbf{supp}\, \hat{\rho}^1$, and $\omega_{tf} \notin \textbf{supp}\, \tilde{\rho}^1$.

Under LBS_{tr}, the scientist can conceal ω_{tr} in two ways. The first uses evidence confirming ρ^0 and induces a_{hr}; the second uses evidence disproving ρ^0 but consistent with ω_{tf} and hence $\hat{\rho}^1$, so that the policymaker assigns to ω_{tf} probability 1 and chooses a_{tf}. Since the scientist prefers a_{tr} to a_{hr}, she never conceals ω_{tr} as before. However, she can no longer pool ω_{tf} and ω_{tr} in the same evidence to raise her chances of getting a_{tr}—any such evidence would result in a_{tf}. Therefore, she always conceals ω_{tf}, inducing a_{hr}. Figure 4 shows the optimal π for ω_{tf} and ω_{tr}.[25] Relative to the baseline example, the policymaker's reluctance to conceive the good state for the scientist, ω_{tr}, causes her to conceal the bad state, ω_{tf}, with strictly higher probability; it is easy to see that this lowers her expected profits.

Under LBS_{tf}, the scientist can conceal ω_{tf} in two ways. The first is as before and induces a_{hr}; the second uses evidence disproving ρ^0 but is consistent with ω_{tr} and hence $\tilde{\rho}^1$, so that the policymaker chooses a_{tr}. Concealing ω_{tf} is now even more attractive. Also, if the scientist pools ω_{tr} and ω_{tf} in the same evidence, the policymaker is surprised, but thinks that ω_{tr} obtained and always chooses a_{tr}. Therefore, now she always conceals ω_{tf} and induces a_{tr}, which leads to higher expected profits than in the baseline example. Figure 5 shows the optimal π for ω_{tf} and ω_{tr}.

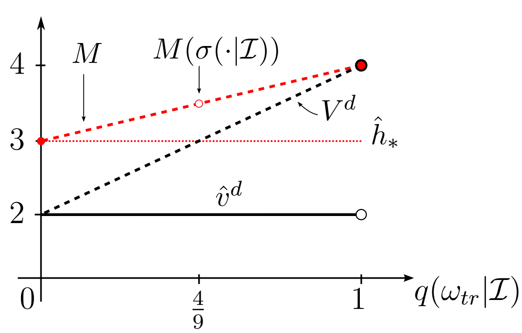

Figure 4: Concealing with LBS_{tr}

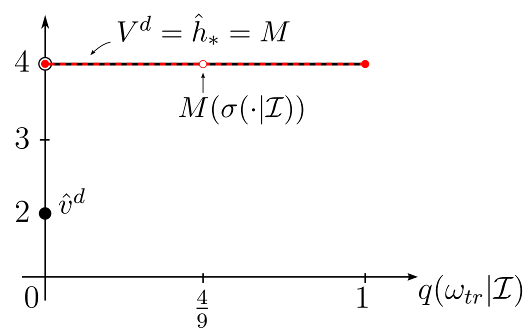

Figure 5: Concealing with LBS_{tf}

[25] In the figure, the function $\hat{h}_*$ represents the opportunity cost of surprising taking into account the possibilities of concealment offered by multiple alternative worldviews (see Online Appendix B.2).

Consistent with Kreps and Wilson (1982), one interpretation of A4 is that Receiver has a "meta-prior" over worldviews which assigns to lower-ranked elements of his LBS a higher likelihood of being correct. Receiver always adopts the worldview with highest likelihood under his *initial* meta-prior, after removing those disproven by the evidence. That is, the ranking of surviving worldviews is never updated. The next model relaxes this property.

Ortoleva's (2012) Model of Worldview Change

Ortoleva's (2012) model offers another *as-if* way of describing how Sender thinks that Receiver responds to evidence. Besides standard axioms of subjective expected utility, Ortoleva (2012) introduces behavior-based notions of coherence and dynamic consistency to discipline Receiver's response. Dynamic consistency is self-explanatory. Coherence says that given two pieces of evidence carrying the same informational content Receiver should have the same posterior.

The model works as follows. Receiver has a prior μ over worldviews. Initially, he selects the worldview with highest μ-likelihood (in our notation, ρ_ε). After evidence confirming ρ_ε, he updates ρ_ε. After unexpected evidence, he *first* updates μ to μ' and *then* adopts the worldview with highest μ'-likelihood and updates it. Updating always follows Bayes' rule.[26] Formally, let $\mu \in \Delta(\Delta(\Omega))$ satisfy (i) $\mathbf{supp}\,\mu$ is finite, (ii) $\rho_\varepsilon = \arg\max_\rho \mu(\rho)$, and (iii) for every $\omega \in \Omega$ there exists $\rho \in \mathbf{supp}\,\mu$ with $\rho(\omega) > 0$. Then, given (s, π), let

$$\mu'(\rho|s,\pi) = \frac{\left[\sum_{\omega\in\Omega} \pi(s|\omega)\rho(\omega)\right] \mu(\rho)}{\sum_{\tilde{\rho}\in\mathbf{supp}\,\mu} \left[\sum_{\omega\in\Omega} \pi(s|\omega)\tilde{\rho}(\omega)\right] \mu(\tilde{\rho})}, \quad \rho \in \Delta(\Omega). \tag{9}$$

The maximizer of $\mu'(\cdot|s,\pi)$ need not be unique. To avoid indeterminacies in Receiver's "choice" of a prior, following Ortoleva (2012), we endow him with a strict linear order $\succ$ over priors, which he applies when the maximum-likelihood criterion is inconclusive.

Assumption 5 (A5: Ortoleva's (2012) Model)**.** *If (s,π) confirms ρ_ε, Receiver updates it via Bayes' rule; otherwise, he updates $\rho(s,\pi)$ via Bayes' rule, where $\rho(s,\pi)$ is $\succ$-maximal in* $\arg\max_\rho\ \mu'(\rho|s,\pi)$.

Note that the notion of unexpected evidence used here differs from the one in Ortoleva (2012), where the information structure is exogenous. According to his notion, s is unexpected if ρ_ε assigns to it a sufficiently small probability under π. As explained in Section 3, such a notion needs to be modified to account for information endogeneity.

Under A5, the characterization of the feasible distributions over posteriors is similar to Proposition 2. A new function $\mathbf{r}_\mu$ replaces $\mathbf{r}$ as follows. If $q \in C_\varepsilon$, $\mathbf{r}_\mu(q) = \rho_\varepsilon$. If $q \in D_\varepsilon$, we

[26] The analysis does not change if Receiver chooses worldviews using any other criterion that continues to rely only on the priors μ and μ'. For example, he may select ρ's to minimize the expectation under μ (resp. μ') of some loss function that depends on his choice and the "true" ρ.

can express $\pi(s|\omega)$ in terms of $q(\omega|s,\pi)$ using AC's logic and write (9) as

$$\mu'(\rho;q) = \frac{\left[\sum_{\omega\in\Omega} q(\omega)\frac{\rho(\omega)}{\sigma(\omega)}\right]\mu(\rho)}{\sum_{\tilde{\rho}\in\mathbf{supp}\,\mu}\left[\sum_{\omega\in\Omega} q(\omega)\frac{\tilde{\rho}(\omega)}{\sigma(\omega)}\right]\mu(\tilde{\rho})}, \quad \rho\in\Delta(\Omega).$$

Now define $\mathbf{r}_\mu(q) = \rho$ where ρ is the $\succ$-maximal element of $\arg\max_\rho \ \mu'(\rho;q)$. Given every $\mathbf{r}_\mu(q)$, we can get Receiver's posterior from Sender's using AC's logic again, thereby obtaining a function $\mathbf{p}_\mu : \Delta(\Omega) \to \Delta(\Omega)$. This function maintains the properties in Corollaries 2–4. Depending on μ, $\mathbf{p}_\mu$ can be discontinuous over D_ε and allow Sender to conceal states after disproving ρ_ε as under A4 (see Online Appendix B.3).

Given this, the analysis in Section 5 remains valid. Under A5, finding the concealment payoff of each $\omega \in \mathcal{I}$ can be more intricate: Now Receiver may discard a new ρ that continues to deem ω implausible for one that does not when the evidence is *sufficiently* against ρ. Nonetheless, given those payoffs, Proposition 7 and Corollary 8 in Online Appendix B.2 apply unchanged.

7 Discussion

7.1 Origins of Different Worldviews

Many reasons can lead Sender and Receiver to have different worldviews, that is, prior beliefs. In Savage's (1972) words,

> the criteria incorporated in the personalistic view [of probability] do not guarantee agreement on all questions among all honest and freely communicating people, even in principle. That incompleteness does not distress me, for I think that at least some of the disagreement we see around us is due to neither dishonesty, to errors in reasoning, nor to friction in communication. (p. 67)

Echoing Savage, Gilboa et al. (2008) pointed out that the axiomatic foundations of individual beliefs "do not imply [...] common priors in multi-agent problems" and leave open the question of where those beliefs come from.[27] Researchers have been trying to answer this question for a long time. Agents form their beliefs by accumulating new experiences and observations as well as by consciously assessing and interpreting the data they already have.

Several papers have analyzed this process. Gilboa and Schmeidler (2000) argue that "agents learn from experience based on similarities between cases, which makes this process path dependent and can lead to different paradigms [i.e., worldviews] even in light of the same experience." Similarity judgments are also inherently subjective, possibly influenced

[27] See also Morris (1995).

by cultural background, political ideology, or religion. According to this theory, an expert in a given field (Sender, in our case) typically has a richer memory of cases, but also a greater ability to judge similarities. Building on these ideas, Billot et al. (2005) develop a theory of subjective probability in which an inexperienced agent can assign 0 probability to some state, while an experienced agent does not. Thus, due to her richer experience or "database," an expert can have a belief whose support contains that of the belief of an inexpert (Receiver, in our case), whose poorer database lacks observations of some states. Also, it seems plausible that agents cannot easily share their entire experiences, which may lead their beliefs to converge, and hence can agree to disagree based on the commonly known fact that they underwent different experiences. At the same time, the expert can provide the inexpert with new evidence which forces the latter to revise his worldview.

Learning, however, is not just accumulation of data. According to Aragones et al. (2005), "much of human learning has to do with making observations and finding regularities that, in principle, could have been determined using existing [facts], rather that with the acquisition of new facts." Given this, they examine the problem of an agent who wants to find the regularities hidden in a database so as to construct a theory, caring about its accuracy as well as simplicity. They view this desire for simplicity as "a natural tendency of human mind," which has normative foundations, dating back to William of Occam, and descriptive validity as recognized also by philosophers like Wittgenstein (1922). Aragones et al. (2005) show that the agent's problem is computationally complex. Therefore, agents "will generally not know all the regularities that exist in their [database]," especially when they face intricate and ever-changing environments as in real economic or political settings. Their "model suggests two reasons why people may have different beliefs, even if these beliefs are defined by rules that are derived from a shared [database]. First, two people may notice different regularities. Since finding the 'best' regularities is a hard problem, we should not be surprised if one person failed to see a regularity that another came up with. Second, even if the individuals shared the rules that they found, they may entertain different beliefs if they make different trade-offs between accuracy and simplicity." Again, an expert in a field may have a richer database as well as a greater ability to discover regularities or to handle complex theories.

As the example of the introduction illustrates, when Receiver stands for a complex bureaucratic organization (like a government's authority), it can have a poorer understanding of a subject matter because of the outdated models and practices it currently uses. By contrast, a private and dynamic company, constantly up-to-date with the scientific and technological progress, may have a richer understanding—as does Sender in the present paper.

Finally, the following arguments underlie the assumption that Receiver sticks to his worldview whenever the evidence is consistent with it. The premise that he agrees to disagree with Sender means that he trusts his worldview: If he thought that the world could follow another theory allowing for states outside $\mathcal{R}$, a probabilistic description of his worldview should assign positive probability to those states. Therefore, ex ante Receiver is confident that states outside $\mathcal{R}$ are impossible, so only definitive evidence supporting them can change his mind. A similar assumption appears, for instance, in Hong et al. (2007) on the basis of psychological studies suggesting that “rather than having the meta-understanding that the real world is in fact complex [...], people tend to behave as if their simple models provide an accurate depiction of reality,” and that “people tend to resist changing their models, even in the face of evidence that [...] appears to strongly contradict these models.”[28] This is consistent with two phenomena called *confirmatory bias*, namely, that people tend to interpret inconclusive evidence in favor of their initial hypothesis (see Rabin and Schrag (1999) and references therein), and *conservatism*, namely, that people tend to update their models slowly (Edwards (1968)). Science itself constantly attempts to fit new observations into existing theories, whereas their revision occurs only when a field undergoes a crisis triggered by clear evidence of anomalies which those theories cannot explain (Kuhn (1962)). Finally, our assumption is consistent with the notion of Perfect Bayesian Equilibrium, which require that Bayes' rule apply whenever possible, and the property that only the evidence produced by an experiment, not its design, can convey information.

7.2 Is this Model Just the Limit of a Standard Model?

Can we use a standard model in which Sender's and Receiver's priors are fixed and have common support to capture the phenomena discussed in this paper? The answer is no: In general, we may not be able to even approximate those phenomena using a standard setting in which the probability ε Receiver initially assigns to some states goes to 0. The intuition hinges on how Receiver reacts to evidence in the two cases. In a standard setting, no matter how small ε is, he always process the evidence content entirely and through the same lens. Therefore, Sender can fine-tune how strongly this content supports each state so as to induce Receiver to have any posterior over the entire state space (recall Proposition 1) and choose the correspondingly optimal action. In the present setting, by contrast, Receiver's resistance to abandon his worldview and tendency to ignore evidence inconsistent with it can prevent Sender from inducing some posteriors of Receiver and thus some actions that he would choose in the standard case. As a result, her optimal experiments, her expected payoffs, and Receiver's induced behavior in the two cases may exhibit differences which do not vanish as ε goes to 0.

The following example aims to illustrate the point in the simplest way. Let $\Omega = \{\omega_1, \omega_2\}$, $\sigma = (\frac{1}{2}, \frac{1}{2})$, and $\rho_0 = (1 - \varepsilon, \varepsilon)$ for $\varepsilon > 0$. Let $A = \{a, b, c\}$, and let the payoff functions be

[28]See Hong et al. (2007) for references to the psychology literature.

v	a	b	c
ω_1	1	0	-1
ω_2	1	0	-1

u	a	b	c
ω_1	1	2	-1
ω_2	1	-1	2

Abusing notation, let p and q be the posterior probabilities Receiver and Sender assign to ω_2, so that $\mathbf{p}(q) = \frac{q\varepsilon}{q\varepsilon+(1-q)(1-\varepsilon)}$. She can induce him to choose a if $p \in [\frac{1}{3}, \frac{2}{3}]$; he will choose b if $p < \frac{1}{3}$ and c if $p > \frac{2}{3}$. Note that $\mathbf{p}(q) \in [\frac{1}{3}, \frac{2}{3}]$ if and only if $q \in [\underline{q}_\varepsilon, \overline{q}_\varepsilon]$, where $\underline{q}_\varepsilon = \frac{1-\varepsilon}{1+\varepsilon}$ and $\overline{q}_\varepsilon = \frac{2(1-\varepsilon)}{2-\varepsilon}$. Also, $\underline{q}_\varepsilon$ and $\overline{q}_\varepsilon$ satisfy $0 < \underline{q}_\varepsilon < \overline{q}_\varepsilon < 1$, are strictly decreasing, and converge to 1 as $\varepsilon \to 0$. Therefore, no matter how small ε is, Sender can always persuade Receiver to choose a with positive probability. Focusing on the case of $\underline{q}_\varepsilon > \frac{1}{2}$, it is easy to check that her optimal τ generates two posteriors, $q = 0$ and $q = \underline{q}_\varepsilon$, with $\tau(\underline{q}_\varepsilon) = \frac{1}{2\underline{q}_\varepsilon}$. This τ always induces Receiver to choose both a and b with positive probability and yields expected payoff $\frac{1}{2\underline{q}_\varepsilon}$, which converges to $\frac{1}{2}$ as $\varepsilon \to 0$. When ε reaches 0, we obtain a case of $\varepsilon = 0$ and $\mathcal{P} = \{\omega_1\}$ in the baseline model of Section 3. Now Sender cannot induce any $p \in [\frac{1}{3}, \frac{2}{3}]$ and hence action a. Since she strictly prefers b to c, her best experiment is fully uninformative ($\tau = \delta_\sigma$), yielding payoff 0. Proposition 6 confirms that this substantive difference between the two models extends beyond the case of $\varepsilon = 0$.

Appendix

A Proofs of the Main Results

A.1 Proof of Corollary 2

Note that D_ε is closed. Let $\hat{q}$ be any element of $D_\varepsilon \cap \mathrm{cl}C_\varepsilon$, where $\mathrm{cl}C_\varepsilon$ is the closure of C_ε. Let $\{q^n\} \subset C_\varepsilon$ be any sequence that satisfies $q^n \to \hat{q}$ as $n \to \infty$. For all $\omega \in \Omega$, we have

$$\lim_{n\to\infty} \mathbf{p}_\varepsilon(\omega; q^n) = \frac{\frac{\hat{q}(\omega)}{\sigma(\omega)}\rho_\varepsilon(\omega)}{\sum_{\omega'\in\Omega} \frac{\hat{q}(\omega')}{\sigma(\omega')}\rho_\varepsilon(\omega')} \quad \text{and} \quad \mathbf{p}_\varepsilon(\omega; \hat{q}) = \frac{\frac{\hat{q}(\omega)}{\sigma(\omega)}\rho^1(\omega)}{\sum_{\omega'\in\Omega} \frac{\hat{q}(\omega')}{\sigma(\omega')}\rho^1(\omega')}.$$

In the case of $\varepsilon = 0$, $\hat{q}(\hat{\omega}) > 0$ for some $\hat{\omega} \in \mathcal{I}$ since $\hat{q} \in \Delta(\mathcal{I})$. Thus, $\lim_{n\to\infty} \mathbf{p}_0(\hat{\omega}; q^n) = 0 < \mathbf{p}_0(\hat{\omega}; \hat{q})$, which proves that $\mathbf{p}_0$ is not continuous at the boundary between C_0 and D_0.

Now consider the case of $\varepsilon > 0$. Since at $\hat{q}$ we have

$$\frac{\sum_{\omega\in\mathcal{P}} \frac{\hat{q}(\omega)}{\sigma(\omega)}\frac{\rho_\varepsilon(\omega)}{\rho_\varepsilon(\mathcal{P})}}{\sum_{\omega'\in\mathcal{I}} \frac{\hat{q}(\omega')}{\sigma(\omega')}\frac{\rho_\varepsilon(\omega')}{\rho_\varepsilon(\mathcal{I})}} = \theta_\varepsilon \frac{\varepsilon}{1-\varepsilon} > 0,$$

it follows that $(\mathbf{supp}\,\hat{q}) \cap \mathcal{P} \equiv \mathcal{P}_{\hat{q}} \neq \varnothing$ and $(\mathbf{supp}\,\hat{q}) \cap \mathcal{I} \equiv \mathcal{I}_{\hat{q}} \neq \varnothing$. Thus, for all $\hat{\omega}$, $\omega' \in \mathcal{P}_{\hat{q}} \cup \mathcal{I}_{\hat{q}}$, the following quantities are well defined:

$$\lim_{n\to\infty}\frac{\mathbf{p}_\varepsilon(\hat\omega;q^n)}{\mathbf{p}_\varepsilon(\omega';q^n)}=\frac{\frac{\hat q(\hat\omega)}{\sigma(\hat\omega)}\rho_\varepsilon(\hat\omega)}{\frac{\hat q(\omega')}{\sigma(\omega')}\rho_\varepsilon(\omega')}\quad\text{and}\quad\frac{\mathbf{p}_\varepsilon(\hat\omega;\hat q)}{\mathbf{p}_\varepsilon(\omega';\hat q)}=\frac{\frac{\hat q(\hat\omega)}{\sigma(\hat\omega)}\rho^1(\hat\omega)}{\frac{\hat q(\omega')}{\sigma(\omega')}\rho^1(\omega')}.$$

If $\mathbf{p}_\varepsilon$ is continuous, the limit on the left must equal the quantity on the right for all pairs of states in $\mathcal{P}_{\hat q}\cup\mathcal{I}_{\hat q}$. This implies $\rho^1(\omega)=\frac{\rho^1(\omega')}{\rho_\varepsilon(\omega')}\rho_\varepsilon(\omega)$ for all $\omega\in\mathcal{P}_{\hat q}\cup\mathcal{I}_{\hat q}$ and some fixed $\omega'\in\mathcal{P}_{\hat q}\cup\mathcal{I}_{\hat q}$. Using this, we get

$$\frac{\rho^1(\mathcal{P}_{\hat q})}{\rho^1(\mathcal{I}_{\hat q})}=\frac{\rho_\varepsilon(\mathcal{P}_{\hat q})}{\rho_\varepsilon(\mathcal{I}_{\hat q})}=\frac{\rho_\varepsilon(\mathcal{P}_{\hat q})/\rho_\varepsilon(\mathcal{P})}{\rho_\varepsilon(\mathcal{I}_{\hat q})/\rho_\varepsilon(\mathcal{I})}\times\frac{1-\varepsilon}{\varepsilon}.$$

By A1, $\frac{\rho_\varepsilon(\mathcal{P}_{\hat q})/\rho_\varepsilon(\mathcal{P})}{\rho_\varepsilon(\mathcal{I}_{\hat q})/\rho_\varepsilon(\mathcal{I})}$ does not change as $\varepsilon\to 0$, and by the finiteness of Ω this ratio can only take finitely many, strictly positive, values. Also, by the properties of ρ^1, we have $\frac{\rho^1(\mathcal{P}_{\hat q})}{\rho^1(\mathcal{I}_{\hat q})}\le\frac{1-\underline{\rho}^1}{\underline{\rho}^1}$ for all $\hat q\in D_\varepsilon\cap \mathrm{cl}C_\varepsilon$. It follows that there exists $\overline{\varepsilon}>0$ such that, if $\varepsilon<\overline{\varepsilon}$, then $\frac{\rho^1(\mathcal{P}_{\hat q})}{\rho^1(\mathcal{I}_{\hat q})}<\frac{\rho_\varepsilon(\mathcal{P}_{\hat q})}{\rho_\varepsilon(\mathcal{I}_{\hat q})}$ for all $\hat q\in D_\varepsilon\cap\mathrm{cl}C_\varepsilon$ and hence again $\mathbf{p}_\varepsilon$ cannot be continuous over this set.

A.2 Proof of Corollary 3

Consider $\mathbf{p}_\varepsilon(C_\varepsilon)$ first. A posterior p belongs to $\mathbf{p}_\varepsilon(C_\varepsilon)$ if and only if it results from expected evidence, which means that $\frac{p(\mathcal{P})}{p(\mathcal{I})}>\theta_\varepsilon$. By A2, θ_ε increases as ε decreases, which renders this condition more stringent. As a result, the set of beliefs that satisfy it shrinks, which implies that $\sup_{p\in\mathbf{p}_\varepsilon(C_\varepsilon)}p(\mathcal{I})$ decreases as ε decreases. Also, using $p(\mathcal{P})=1-p(\mathcal{I})$, we have that $p(\mathcal{I})<\frac{1}{1+\theta_\varepsilon}$ for every $p\in\mathbf{p}_\varepsilon(C_\varepsilon)$. By A2, $\lim_{\varepsilon\to0}\frac{1}{1+\theta_\varepsilon}=0$. Therefore, $p\in\lim_{\varepsilon\to0}\mathbf{p}_\varepsilon(C_\varepsilon)=\cap_{\varepsilon\ge0}\mathbf{p}_\varepsilon(C_\varepsilon)$ implies that $p(\mathcal{I})=0$, which means that $p\in\Delta(\mathcal{P})$. Now consider any $p'\in\Delta(\mathcal{P})$. Since $p'(\mathcal{I})=0$, it means that the evidence inducing p' rules out all states in $\mathcal{I}$ with certainty and must therefore be expected for Receiver: Indeed, $\frac{p'(\mathcal{P})}{p'(\mathcal{I})}=\infty$. Since $\theta_\varepsilon<\frac{1-\varepsilon}{\varepsilon}$, we have that $\frac{p'(\mathcal{P})}{p'(\mathcal{I})}>\theta_\varepsilon$ for all $\varepsilon\ge0$. Therefore, $p'\in\cap_{\varepsilon\ge0}\mathbf{p}_\varepsilon(C_\varepsilon)$. This proves that $\cap_{\varepsilon\ge0}\mathbf{p}_\varepsilon(C_\varepsilon)=\Delta(\mathcal{P})$.

Now consider $\mathbf{p}_\varepsilon(D_\varepsilon)$. Each p in this set satisfies

$$\mathbf{p}_\varepsilon(\omega;q)=\frac{q(\omega)\frac{\rho^1(\omega)}{\sigma(\omega)}}{\sum_{\omega'\in\Omega}q(\omega')\frac{\rho^1(\omega')}{\sigma(\omega')}},\quad\omega\in\Omega.$$

Since $\mathbf{supp}\,\rho^1=\Omega$, viewed as a function over the entire set $\Delta(\Omega)$, this function is an homeomorphism (see Corollary 1). Therefore, since $D_\varepsilon\subset D_{\varepsilon'}$ for $\varepsilon<\varepsilon'$, we have $\mathbf{p}_\varepsilon(D_\varepsilon)\subsetneq\mathbf{p}_\varepsilon(D_{\varepsilon'})$ and hence $\sup_{p\in\mathbf{p}_\varepsilon(D_\varepsilon)}p(\mathcal{P})$ decreases as ε decreases. Moreover, $\lim_{\varepsilon\to0}\mathbf{p}_\varepsilon(D_\varepsilon)=\cap_{\varepsilon\ge0}\mathbf{p}_\varepsilon(D_\varepsilon)=\mathbf{p}_0(D_0)=\mathbf{p}_0(\Delta(\mathcal{I}))$. Restricted to $\Delta(\mathcal{I})$, $\mathbf{p}_0$ can be written as

$$\mathbf{p}_0(\omega;q)=\frac{q(\omega)\frac{\rho^1(\omega|\mathcal{I})}{\sigma(\omega|\mathcal{I})}}{\sum_{\omega'\in\mathcal{I}}q(\omega')\frac{\rho^1(\omega'|\mathcal{I})}{\sigma(\omega'|\mathcal{I})}},\quad\omega\in\mathcal{I}.$$

Thus, applying Proposition 1 and Corollary 1 for the common-support priors $\rho^1(\cdot|\mathcal{I})$ and $\sigma(\cdot|\mathcal{I})$, we conclude that $\mathbf{p}_0:\Delta(\mathcal{I})\to\Delta(\mathcal{I})$ must be onto and hence $\mathbf{p}_0(\Delta(\mathcal{I}))=\Delta(\mathcal{I})$.

A.3 Proof of Corollary 4

Given q_η and the corresponding $q' \in \Delta(\mathcal{P})$ and $\hat{\omega} \in \mathcal{I}$, we have that

$$\begin{aligned} \ell(q_\eta) &= \frac{(1-\eta)\sum_{\omega\in\mathcal{P}} \frac{q'(\omega)}{\sigma(\omega)}\rho_\varepsilon(\omega)}{\eta\frac{\rho_\varepsilon(\hat{\omega})}{\sigma(\hat{\omega})}} = \frac{(1-\eta)(1-\varepsilon)}{\eta\varepsilon} \times \frac{\frac{1}{\sigma(\mathcal{P})}\sum_{\omega\in\mathcal{P}}\frac{q'(\omega)}{\sigma(\omega|\mathcal{P})}\rho_\varepsilon(\omega|\mathcal{P})}{\frac{\rho_\varepsilon(\hat{\omega}|\mathcal{I})}{\sigma(\hat{\omega})}} \\ &\geq \frac{(1-\eta)(1-\varepsilon)}{\eta\varepsilon} \times \frac{\min_{\omega\in\mathcal{I}}\sigma(\omega)}{\sigma(\mathcal{P})}, \end{aligned}$$

where the last inequality uses the fact that $\sum_{\omega\in\mathcal{P}}\frac{q'(\omega)}{\sigma(\omega|\mathcal{P})}\rho_\varepsilon(\omega|\mathcal{P}) = 1$ for all $q' \in \Delta(\mathcal{P})$. Also, recall that by A1, the quantity multiplying $\frac{(1-\eta)(1-\varepsilon)}{\eta\varepsilon}$ is independent of ε. Therefore, for every $\varepsilon > 0$, $\ell(q_\eta) > \theta_\varepsilon$ for all q' and $\hat{\omega}$ if and only if $\eta < \eta_\varepsilon$ for some unique $\eta_\varepsilon < 1$. Since by A2 $\theta_\varepsilon\frac{\varepsilon}{1-\varepsilon}$ decreases monotonically to zero, it follows that η_ε increases monotonically to 1.

To prove the last part of the result, suppose $\hat{q}, \tilde{q} \in C_\varepsilon$ and satisfy $\hat{q}(\cdot|\mathcal{P}) = \tilde{q}(\cdot|\mathcal{P}) = q'$. We can then write $\hat{q} = (1-\hat{q}(\mathcal{I}))q' + \hat{q}(\mathcal{I})\hat{q}'$ and $\tilde{q} = (1-\tilde{q}(\mathcal{I}))q' + \tilde{q}(\mathcal{I})\tilde{q}'$, where $\hat{q}', \tilde{q}' \in \Delta(\mathcal{I})$. Letting $\hat{\mathbf{p}} = \mathbf{p}_\varepsilon(\hat{q})$ and $\tilde{\mathbf{p}} = \mathbf{p}_\varepsilon(\tilde{q})$, we have

$$\begin{aligned} ||\hat{\mathbf{p}} - \tilde{\mathbf{p}}||^2 &= \sum_{\omega\in\Omega}\left[\hat{\mathbf{p}}(\mathcal{P})\hat{\mathbf{p}}(\omega|\mathcal{P}) + \hat{\mathbf{p}}(\mathcal{I})\hat{\mathbf{p}}(\omega|\mathcal{I}) - \tilde{\mathbf{p}}(\mathcal{P})\tilde{\mathbf{p}}(\omega|\mathcal{P}) - \tilde{\mathbf{p}}(\mathcal{I})\tilde{\mathbf{p}}(\omega|\mathcal{I})\right]^2 \\ &= \sum_{\omega\in\mathcal{P}}\left[\hat{\mathbf{p}}(\mathcal{P})\hat{\mathbf{p}}(\omega|\mathcal{P}) - \tilde{\mathbf{p}}(\mathcal{P})\tilde{\mathbf{p}}(\omega|\mathcal{P})\right]^2 + \sum_{\omega\in\mathcal{I}}\left[\hat{\mathbf{p}}(\mathcal{I})\hat{\mathbf{p}}(\omega|\mathcal{I}) - \tilde{\mathbf{p}}(\mathcal{I})\tilde{\mathbf{p}}(\omega|\mathcal{I})\right]^2 \\ &= \sum_{\omega\in\mathcal{P}}\left[\mathbf{p}'(\omega|\mathcal{P})\right]^2\left[\hat{\mathbf{p}}(\mathcal{I}) - \tilde{\mathbf{p}}(\mathcal{I})\right]^2 + \sum_{\omega\in\mathcal{I}}\left[\hat{\mathbf{p}}(\mathcal{I})\hat{\mathbf{p}}(\omega|\mathcal{I}) - \tilde{\mathbf{p}}(\mathcal{I})\tilde{\mathbf{p}}(\omega|\mathcal{I})\right]^2 \\ &\leq |\mathcal{P}|\left[\hat{\mathbf{p}}(\mathcal{I}) - \tilde{\mathbf{p}}(\mathcal{I})\right]^2 + 2|\mathcal{I}|\left[\max\{\hat{\mathbf{p}}(\mathcal{I}), \tilde{\mathbf{p}}(\mathcal{I})\}\right]^2 \end{aligned}$$

where the third step uses the fact that, for all $\omega \in \mathcal{P}$,

$$\mathbf{p}'(\omega|\mathcal{P}) \equiv \frac{q'(\omega)\frac{\rho_\varepsilon(\omega)}{\sigma(\omega)}}{\sum_{\omega'\in\mathcal{P}} q'(\omega')\frac{\rho_\varepsilon(\omega')}{\sigma(\omega')}} = \frac{\hat{\mathbf{p}}(\omega)}{\hat{\mathbf{p}}(\mathcal{P})} = \frac{\tilde{\mathbf{p}}(\omega)}{\tilde{\mathbf{p}}(\mathcal{P})},$$

and therefore $\hat{\mathbf{p}}(\omega|\mathcal{P}) = \tilde{\mathbf{p}}(\omega|\mathcal{P})$. Now recall that $\sup_{p\in\mathbf{p}_\varepsilon(C_\varepsilon)} p(\mathcal{I}) \to 0$ as $\varepsilon \to 0$ by Corollary 3, which implies that $\max\{\hat{\mathbf{p}}(\mathcal{I}), \tilde{\mathbf{p}}(\mathcal{I})\} \to 0$ and $|\hat{\mathbf{p}}(\mathcal{I}) - \tilde{\mathbf{p}}(\mathcal{I})| \to 0$ as well.

A.4 Proof of Proposition 3

Part I: Property (1). Note that

$$V^c(\sigma) = \sup_{\{\tau\in\mathcal{T}:\,\mathbf{supp}\,\tau\subseteq C\}} \mathbb{E}_\tau[\hat{v}(q)] = \sup_{\mathcal{T}^c}\sum_i \tau_i\hat{v}(q_i),$$

where

$$\mathcal{T}^c = \left\{(q_1,\tau_1;\ldots;q_N,\tau_N) : \sum_{i=1}^N \tau_i q_i = \sigma, \sum_{i=1}^N \tau_i = 1, \tau_i \geq 0, \text{ and } q_i \in C \text{ for all } i\right\}.$$

Take any $\tau \in \mathcal{T}^c$ with $q \in \mathbf{supp}\,\tau$ that satisfies $q(\omega) > 0$ and $q(\omega') > 0$ for some $\omega, \omega' \in \mathcal{I}$ with $\omega \neq \omega'$. We will show that there exists $\tau' \in \mathcal{T}^c$ which has the first part of property (1) and satisfies $\mathbb{E}_{\tau'}[\hat{v}(q)] \geq \mathbb{E}_{\tau}[\hat{v}(q)]$.

Since $\tau \in \mathcal{T}^c$, it is induced by some π, that is, every $q \in \mathbf{supp}\,\tau$ equals $q(\cdot|s,\pi)$ for some (s,π). For every $q \in \mathbf{supp}\,\tau$, let $\mathcal{I}(q) = \{\omega \in \mathcal{I} : q(\omega) > 0\}$. By assumption, $|\mathcal{I}(\hat{q})| > 1$ for some $\hat{q} \in \mathbf{supp}\,\tau$. Fix any such $\hat{q}$, and let $\hat{s}$ be the signal inducing it under π. Clearly, $\pi(\hat{s}|\omega) > 0$ if and only if $\omega \in \mathbf{supp}\, q(\cdot|\hat{s},\pi)$, which includes $\mathcal{I}(q(\cdot|\hat{s},\pi))$ as a proper subset because $q(\cdot|\hat{s},\pi) \in C$. Starting from π, construct π' as follows: For each $\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))$, create a piece of evidence (s^ω, π') with the following properties: (i) $\pi'(s^\omega|\omega) = \pi(\hat{s}|\omega)$, (ii) $\pi'(s^\omega|\omega') = 0$ for all $\omega' \in \mathcal{I}(q(\cdot|\hat{s},\pi)) \setminus \{\omega\}$, and (iii) for each $\tilde{\omega} \in \mathcal{P}$,

$$\pi'(s^\omega|\tilde{\omega}) = \pi(\hat{s}|\tilde{\omega}) \frac{\pi(\hat{s}|\omega)}{\sum_{\omega' \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \pi(\hat{s}|\omega')}.$$

It is easy to see that π' is well defined. For every $\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))$, by construction, $q(\omega'|s^\omega,\pi') = 0$ if $\omega' \in \mathcal{I}(q(\cdot|\hat{s},\pi)) \setminus \{\omega\}$, and for every $\tilde{\omega} \in \mathcal{P}$,

$$q(\tilde{\omega}|\mathcal{P}, s^\omega, \pi') = \frac{\pi'(s^\omega|\tilde{\omega})\sigma(\tilde{\omega}|\mathcal{P})}{\sum_{\omega' \in \mathcal{P}} \pi'(s^\omega|\omega')\sigma(\omega'|\mathcal{P})} = \frac{\pi(\hat{s}|\tilde{\omega})\sigma(\tilde{\omega}|\mathcal{P})}{\sum_{\omega' \in \mathcal{P}} \pi(\hat{s}|\omega')\sigma(\omega'|\mathcal{P})} = q(\tilde{\omega}|\mathcal{P}, \hat{s}, \pi).$$

This implies that $\mathcal{A}(\mathbf{p}(q(\cdot|s^\omega,\pi'))) = \mathcal{A}(\mathbf{p}(q(\cdot|\hat{s},\pi)))$ for every $\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))$. Let the total probability that $q(\cdot|s^\omega,\pi')$ arises under π' be $\beta(s^\omega,\pi') = \sum_{\omega' \in \Omega} \pi'(s^\omega|\omega')\sigma(\omega')$, and note that

$$\begin{aligned}
\sum_{\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \beta(s^\omega,\pi') &= \sum_{\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \pi(\hat{s}|\omega)\sigma(\omega) + \sum_{\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \left[\sum_{\tilde{\omega} \in \mathcal{P}} \pi'(s^\omega|\tilde{\omega})\sigma(\tilde{\omega})\right] \\
&= \sum_{\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \pi(\hat{s}|\omega)\sigma(\omega) + \sum_{\tilde{\omega} \in \mathcal{P}} \pi(\hat{s}|\tilde{\omega})\sigma(\tilde{\omega}) = \beta(\hat{s},\pi),
\end{aligned}$$

that is, the probability that $q(\cdot|\hat{s},\pi)$ arises under π. Since $q(\cdot|\hat{s},\pi) = \sum_{\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))} q(\cdot|s^\omega,\pi')\frac{\beta(s^\omega,\pi')}{\beta(\hat{s},\pi)}$, we have that $\hat{q}$ is the conditional expectation of posteriors $\{q(\cdot|s^\omega,\pi')\}_{\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))}$. Indeed, if $\omega \in \mathcal{I}(q(\cdot|\hat{s},\pi))$, then $q(\omega|s^{\omega'},\pi') = 0$ for all $\omega' \neq \omega$ and

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

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

$$\sum_{\omega' \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \frac{\pi'(s^{\omega'}|\omega)\sigma(\omega)}{\beta(\hat{s},\pi)} = \sum_{\omega' \in \mathcal{I}(q(\cdot|\hat{s},\pi))} \frac{\pi(\hat{s}|\omega)\sigma(\omega)\frac{\pi(\hat{s}|\omega')}{\sum_{\omega'' \in \mathcal{I}(q(\cdot|\hat{s},\pi))}\pi(\hat{s}|\omega'')}}{\beta(\hat{s},\pi)} = q(\omega|\hat{s},\pi).$$

In summary, π' replaces $\hat{q}$ by allocating its probability $\tau(\hat{q})$ across a collection of posteriors $q^\omega = q(\cdot|s^\omega,\pi')$, each with probability $\tau'(q^\omega)$, that satisfies $\hat{q} = \sum_{\omega \in \mathcal{I}(\hat{q})} q^\omega \frac{\tau'(q^\omega)}{\tau(\hat{q})}$. Note that $\tau'(q) = \tau(q)$ for all other $q \in \mathbf{supp}\,\tau$, hence Sender's payoff changes only when $\hat{q}$ arises. This change can only be an improvement. To see this, given $\hat{q}$ let $a(\hat{q}) \in \mathcal{A}(\mathbf{p}(\hat{q}))$ be Receiver's resulting action. Then

Sender's conditional expected payoff from the distribution over the q^{ω}'s is

$$\begin{aligned}\sum_{\omega\in\mathcal{I}(\hat{q})}\left\{\max_{a\in\mathcal{A}(\mathbf{p}(\hat{q}))}\sum_{\tilde{\omega}\in\Omega}v(a,\tilde{\omega})q^{\omega}(\tilde{\omega})\right\}\frac{\tau'(q^{\omega})}{\tau(\hat{q})} &\geq \sum_{\omega\in\mathcal{I}(\hat{q})}\left\{\sum_{\tilde{\omega}\in\Omega}v(a(\hat{q}),\tilde{\omega})q^{\omega}(\tilde{\omega})\right\}\frac{\tau'(q^{\omega})}{\tau(\hat{q})}\\ &= \sum_{\tilde{\omega}\in\Omega}v(a(\hat{q}),\tilde{\omega})\left\{\sum_{\omega\in\mathcal{I}(\hat{q})}q^{\omega}(\tilde{\omega})\frac{\tau'(q^{\omega})}{\tau(\hat{q})}\right\}\\ &= \sum_{\tilde{\omega}\in\Omega}v(a(\hat{q}),\tilde{\omega})\hat{q}(\tilde{\omega})=\hat{v}(\hat{q}).\end{aligned}$$

Replicating this construction for all $\hat{q}\in\mathbf{supp}\,\tau$ with $|\mathcal{I}(\hat{q})|>1$, we obtain a distribution $\tau'\in\Delta(C)$ such that $\mathbb{E}_{\tau'}[\hat{v}(q)]\geq\mathbb{E}_{\tau}[\hat{v}(q)]$ and for every $q\in\mathbf{supp}\,\tau'$ there exists at most one $\omega\in\mathcal{I}$ that satisfies $q(\omega)\in(0,1)$. For any such q, $q(\mathcal{P})=\sum_{\omega'\in\mathcal{P}}q(\omega')=1-q(\omega)$. Hence, $q(\omega')=(1-q(\omega))q(\omega'|\mathcal{P})$ for every $\omega'\in\mathcal{P}$, and $q(\omega)=q(\omega)\delta_{\omega}$, so that $q=(1-q(\omega))q(\cdot|\mathcal{P})+q(\omega)\delta_{\omega}$.

Part II: Property (2). Take any $\tau\in\mathcal{T}^c$ that has property (1). For every $\omega\in\mathcal{I}$, let $Q(\omega)=\{q\in\mathbf{supp}\,\tau : q(\omega)>0\}$. Suppose that $|Q(\omega^*)|>1$ for some $\omega^*\in\mathcal{I}$. Then let $T^*=\sum_{q\in Q(\omega^*)}\tau(q)$ and

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

where the second equality uses property (1). Thus, q^* arises with probability T^* and is the convex combination of the posteriors δ_{ω^*} and $\{q(\cdot|\mathcal{P})\}_{q\in Q(\omega^*)}$. Now consider Sender's expected payoff conditional on $Q(\omega^*)$, letting $a(q)$ be Receiver's action for each $q\in Q(\omega^*)$:

$$\begin{aligned}&\sum_{q\in Q(\omega^*)}\left\{\sum_{\omega\in\Omega}v(a(q),\omega)q(\omega)\right\}\tau(q|Q(\omega^*))\\ &=\sum_{q\in Q(\omega^*)}\left\{(1-q(\omega^*))\sum_{\omega\in\Omega}v(a(q),\omega)q(\omega|\mathcal{P})+q(\omega^*)v(a(q),\omega^*)\right\}\tau(q|Q(\omega^*))\\ &=\sum_{q\in Q(\omega^*)}\left\{\sum_{\omega\in\Omega}v(a(q),\omega)q(\omega|\mathcal{P})\right\}(1-q(\omega^*))\tau(q|Q(\omega^*))+\sum_{q\in Q(\omega^*)}v(a(q),\omega^*)q(\omega^*)\tau(q|Q(\omega^*))\\ &=(1-q^*(\omega^*))\left\{\sum_{q\in Q(\omega^*)}\left\{\sum_{\omega\in\Omega}v(a(q),\omega)q(\omega|\mathcal{P})\right\}\frac{(1-q(\omega^*))\tau(q|Q(\omega^*))}{1-q^*(\omega^*)}\right\}\\ &\quad+q^*(\omega^*)\left\{\sum_{q\in Q(\omega^*)}v(a(q),\omega^*)\frac{q(\omega^*)\tau(q|Q(\omega^*))}{q^*(\omega^*)}\right\}.\end{aligned}\tag{10}$$

Recall that $\mathcal{A}(\mathbf{p}(q))$ depends only on $q(\cdot|\mathcal{P})$, so any change in q which leaves $q(\cdot|\mathcal{P})$ unaffected does not change the actions Sender can induce Receiver to choose. The value of the quantity in (10) can only increase if, for every $q\in Q(\omega^*)$, we replace $v(a(q),\omega^*)$ with $v(a(\tilde{q}),\omega^*)\equiv\max_{q'\in Q(\omega^*)}v(a(q'),\omega^*)$—that is, if we shift the entire weight $q^*(\omega^*)$ to the largest $v(a(q),\omega^*)$.

This shift causes τ to undergo a modification to the distribution τ', which is defined as follows: $\tau'(q) = \tau(q)$ for every $q \notin Q(\omega^*)$, each $q \in Q(\omega^*)$ with $q \neq \tilde{q}$ is replaced by $q' = q(\cdot|\mathcal{P})$, and $\tilde{q}$ is replaced by

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

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 $\tau'(\tilde{q}'|Q') = (1-\tilde{q}(\omega^*))\tau(\tilde{q}|Q(\omega^*)) + q^*(\omega^*)$. By construction, $\sum_{q'\in Q'} q'\tau'(q'|Q') = q^*$ and $\sum_{q'\in Q'} \tau'(q') = T^*$, so that $\tau' \in \mathcal{T}^c$. Using this notation, by the previous argument the expression on the right-hand side of (10) is less than or equal to

$$\sum_{q\in Q(\omega^*)}\left\{\sum_{\omega\in\Omega} v(a(q),\omega)q(\omega|\mathcal{P})\right\}(1-q(\omega^*))\tau(q|Q(\omega^*)) + q^*(\omega^*)v(a(\tilde{q}),\omega^*)$$
$$= \sum_{q'\in Q'}\left\{\sum_{\omega\in\Omega} v(a(q'),\omega)q'(\omega)\right\}\tau(q'|Q') \leq \sum_{q'\in Q'}\left\{\max_{a\in\mathcal{A}(\mathbf{p}(q'))}\sum_{\omega\in\Omega} v(a,\omega)q'(\omega)\right\}\tau(q'|Q'),$$

where the inequality stems from the fact that $a(q) \in \mathcal{A}(\mathbf{p}(q'))$ for each q' and associated original q in the construction. We conclude that, for every $\omega \in \mathcal{I}$ such that $|Q(\omega)| > 1$ under the original $\tau \in \mathcal{T}^c$, it is possible to modify τ to obtain $\tau' \in \mathcal{T}^c$ such that $|Q'(\omega)| = 1$ and $\mathbb{E}_{\tau'}[\hat{v}(q)] \geq \mathbb{E}_{\tau}[\hat{v}(q)]$.

Part III: Property (3). Take any $\omega \in \mathcal{I}$ 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 there exists $q' \in \Delta(\mathcal{P})$ that satisfies $v(a(q^\omega),\omega) < \max_{a\in\mathcal{A}(\mathbf{p}(q'))} v(a,\omega) = v(a(q'),\omega)$. By property (1), for any sufficiently small $\gamma > 0$, we can write

$$q^\omega = t_\gamma q^\omega(\cdot|\mathcal{P}) + (1-t_\gamma)\left[\gamma q^\omega(\cdot|\mathcal{P}) + (1-\gamma)\delta_\omega\right],$$

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

$$q^{\omega\prime} = t_\gamma q^\omega(\cdot|\mathcal{P}) + (1-t_\gamma)[\gamma q' + (1-\gamma)\delta_\omega].$$

Given q^ω, there exist $z_\gamma \geq \tau(q^\omega)$ and $\alpha_\gamma \in (0,1)$ that satisfy $q_{z_\gamma} \in intC$, $q_{\alpha_\gamma} \in C$, and $q_{z_\gamma} = \alpha_\gamma q_{\alpha_\gamma} + (1-\alpha_\gamma)q^{\omega\prime}$. Moreover, as $\gamma \to 0$ we can choose z_γ and α_γ so that $z_\gamma \downarrow \tau(q^\omega)$ and $\alpha_\gamma \downarrow 0$. Hence, we can modify τ to obtain τ_γ given by

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

It can easily be checked that $\tau_\gamma \in \mathcal{T}^c$ for every $\gamma > 0$.

Now consider $\mathbb{E}_{\tau}[\hat{v}(q)]$ and $\mathbb{E}_{\tau_\gamma}[\hat{v}(q)]$. Letting $k = \sum_{q \neq q^\omega} v(q) \frac{\tau(q)}{1-\tau(q^\omega)}$, we have

$$\begin{aligned}
\mathbb{E}_{\tau}[\hat{v}(q)] &= \tau(q^\omega)\left\{(1-q^\omega(\omega))\sum_{\omega' \in \Omega} v(a(q^\omega),\omega')q^\omega(\omega'|\mathcal{P}) + q^\omega(\omega)v(a(q^\omega),\omega)\right\} + (1-\tau(q^\omega))k, \\
\mathbb{E}_{\tau_\gamma}[\hat{v}(q)] &= (1-z_\gamma)k + z_\gamma\alpha_\gamma \sum_{\omega'\in\Omega} v(a(q_{\alpha_\gamma}),\omega')q_{\alpha_\gamma}(\omega') + z_\gamma(1-\alpha_\gamma)t_\gamma \sum_{\omega'\in\Omega} v(a(q^\omega),\omega')q^\omega(\omega'|\mathcal{P}) \\
&\quad + z_\gamma(1-\alpha_\gamma)(1-t_\gamma)\left\{\gamma \sum_{\omega'\in\Omega} v(a^\omega,\omega')q'(\omega') + (1-\gamma)v(a^\omega,\omega)\right\},
\end{aligned}$$

where $a^\omega = \arg\max_{a\in\mathcal{A}(\mathbf{p}(q'))} v(a,\omega)$. Since Sender's expected payoff from any action and posterior q is finite, we have

$$\begin{aligned}
\lim_{\gamma\to 0} \mathbb{E}_{\tau_\gamma}[\hat{v}(q)] &= \tau(q^\omega)\left\{(1-q^\omega(\omega))\sum_{\omega'\in\Omega} v(a(q^\omega),\omega')q^\omega(\omega'|\mathcal{P}) + q^\omega(\omega)v(a^\omega,\omega)\right\} \\
&\quad + (1-\tau(q_\omega))k > \mathbb{E}_\tau[\hat{v}(q)].
\end{aligned}$$

Therefore, there exists $\gamma > 0$ such that $\mathbb{E}_{\tau_\gamma}[\hat{v}(q)] > \mathbb{E}_\tau[\hat{v}(q)]$. This shows that we can focus on distributions $\tau \in \mathcal{T}^c$ that satisfy $v(a(q^\omega),\omega) \geq \max_{q\in\Delta(\mathcal{P})}\{\max_{a\in\mathcal{A}(\mathbf{p}(q))} v(a,\omega)\}$ for every $\omega \in \mathcal{I}$. Since by varying $q \in \Delta(\mathcal{P})$ Sender can achieve every $p \in \Delta(\mathcal{P})$,[29] the right-hand side of the last inequality equals $h(\omega)$.

Part IV: $V^c(\sigma)$ satisfies (4). The proof consists of the following two claims. Let $V^{c*}(\sigma)$ be the expression for $V^c(\sigma)$ in (4), and let $\mathcal{T}^{c*} \subseteq \mathcal{T}^c$ contain all the τ's that satisfy properties (1)–(3).[30]

Claim 1. $\mathbb{E}_\tau[\hat{v}(q)] \leq V^{c*}(\sigma)$ for all $\tau \in \mathcal{T}^{c*}$.

Proof. Given $\tau \in \mathcal{T}^{c*}$, for every $\omega \in \mathcal{I}$ let q^ω be the posterior in $\mathbf{supp}\,\tau$ that assigns positive probability to ω and is such that $a(q^\omega) = a^\omega \in \arg\max_{a\in\mathcal{A}(\mathbf{p}(q^\omega))} v(a,\omega)$. Also, let $\overline{Q} = \{q^\omega\}_{\omega\in\mathcal{I}}$ and $Q = \mathbf{supp}\,\tau \setminus \overline{Q}$. Then

$$\begin{aligned}
\mathbb{E}_\tau[\hat{v}(q)] &= \sum_{q\in Q}\hat{v}(q)\tau(q) + \sum_{q^\omega\in\overline{Q}}\left[\sum_{\tilde{\omega}\in\Omega} v(a^\omega,\tilde{\omega})\{(1-q^\omega(\omega))q^\omega(\tilde{\omega}|\mathcal{P}) + q^\omega(\omega)v(a^\omega,\omega)\}\right]\tau(q^\omega) \\
&= \sum_{q\in Q}\hat{v}(q)\tau(q) + \sum_{q^\omega\in\overline{Q}}\left[\sum_{\tilde{\omega}\in\Omega} v(a^\omega,\tilde{\omega})q^\omega(\tilde{\omega}|\mathcal{P})\right](1-q^\omega(\omega))\tau(q^\omega) \\
&\quad + \sum_{q^\omega\in\overline{Q}} v(a^\omega,\omega)q^\omega(\omega)\tau(q^\omega) \\
&\leq \sum_{q\in Q}\hat{v}(q)\tau(q) + \sum_{q_\omega\in\overline{Q}}\hat{v}(q^\omega(\cdot|\mathcal{P}))(1-q^\omega(\omega))\tau(q^\omega) + \sum_{q^\omega\in\overline{Q}} h(\omega)q^\omega(\omega)\tau(q^\omega). \qquad (11)
\end{aligned}$$

[29] This follows from Proposition 1 and Corollary 1 with $\rho = \rho_0$ and $\sigma = \sigma(\cdot|\mathcal{P})$.

[30] Note that $\lim_{q'\to q} V^c(q') \geq V^c(q)$ for all $q \in C$. Indeed, being concave, V^c is continuous at every $q \in intC$ by Theorem 10.1 in Rockafellar (1997). By Theorem 10.3 in Rockafellar (1997), there exists only one way to extend V^c from $intC$ to a continuous finite concave function on $\Delta(\Omega)$. In fact, this extension equals $-\mathrm{cl}(-V^c)$ on $\Delta(\Omega)$, where $\mathrm{cl}(-V^c)$ is the closure of the convex function $-V^c$ (Rockafellar (1997), p. 52). Therefore, for any $q \in C_0 \setminus intC$, we have $\lim_{q'\to q} V^c(q') \geq V^c(q)$ since $\mathrm{cl}(-V^c) \leq -V^c$. Similar properties hold for V^d.

Since

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

we have $\sigma(\omega) = q^\omega(\omega)\tau(q^\omega)$ for every $\omega \in \mathcal{I}$ and $\sum_{q^\omega \in \overline{Q}} q^\omega(\omega)\tau(q^\omega) = \sigma(\mathcal{I})$. This implies that

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

Divided by $\sigma(\mathcal{P})$, the first two terms in (11) are then a convex combination of values of $\hat{v}^c(q)$ involving posteriors $q \in \Delta(\mathcal{P})$ whose average equals $\sigma(\cdot|\mathcal{P})$. This combination cannot exceed $V^c(\sigma(\cdot|\mathcal{P}))$, and hence (11) is bounded above by $V^{c*}(\sigma)$.

□

Claim 2. For every $\gamma > 0$ there exists $\tau_\gamma \in \mathcal{T}^{c*}$ that satisfies $\mathbb{E}_{\tau_\gamma}[\hat{v}(q)] \geq V^{c*}(\sigma) - \gamma$. Hence, $V^{c*}(\sigma)$ is the least upper bound for the values of $\mathbb{E}_\tau[\hat{v}(q)]$ over $\mathcal{T}^{c*}$.

Proof. Given any $\tau \in \mathcal{T}^{c*}$, we first construct a sequence $\{\tau_n\}_{n=1}^\infty \subseteq \mathcal{T}^{c*}$ starting 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 every $q^\omega \in \overline{Q}$ and $n \geq 1$, split $\tau_0(q^\omega)$ by replacing it with

$$\tau_n(q') = \begin{cases} \tau_0(q^\omega)z_n^\omega & \text{if } q' = q^\omega(\cdot|\mathcal{P}) \\ \tau_0(q^\omega)(1 - z_n^\omega) & \text{if } q' = q_n^\omega \equiv \frac{1}{[K^\omega]^n} q^\omega(\cdot|\mathcal{P}) + \left(1 - \frac{1}{[K^\omega]^n}\right)\delta_\omega \end{cases},$$

where $z_n^\omega \in (0,1)$ and $K^\omega > 1$ are chosen so that $q^\omega(\omega) = (1 - z_n^\omega)(1 - [K^\omega]^{-n})$ for all $n \geq 1$. By construction, $z_n^\omega \uparrow (1 - q^\omega(\omega))$ and $q_n^\omega \to \delta_\omega$ as $n \to \infty$; moreover, $q^\omega = z_n^\omega q^\omega(\cdot|\mathcal{P}) + (1 - z_n^\omega)q_n^\omega$ for every n. Thus, for every n the conditional distribution defined by z_{n+1}^ω is a mean-preserving spread around q^ω of that defined by z_n^ω. It follows that, for every $q^\omega \in \overline{Q}$,

$$\begin{aligned} \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q^\omega(\tilde{\omega}) \quad \leq \quad & z_1^\omega \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q^\omega(\tilde{\omega}|\mathcal{P}) \\ & + (1 - z_1^\omega) \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q_1^\omega(\tilde{\omega}), \end{aligned}$$

and for all $n \geq 1$,

$$\begin{aligned} & z_n^\omega \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q^\omega(\tilde{\omega}|\mathcal{P}) + (1 - z_n^\omega) \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q_n^\omega(\tilde{\omega}) \\ \leq & z_{n+1}^\omega \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q^\omega(\tilde{\omega}|\mathcal{P}) + (1 - z_{n+1}^\omega) \max_{a \in \mathcal{A}(\mathbf{p}(q^\omega))} \sum_{\tilde{\omega} \in \Omega} v(a, \tilde{\omega})q_{n+1}^\omega(\tilde{\omega}). \end{aligned}$$

For every n, letting $Q'_n = \mathbf{supp}\,\tau_n \setminus Q$, 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 $\sum_{q^\omega \in \overline{Q}} \tau_0(q^\omega) = \sum_{q' \in Q'_n} \tau_n(q')$ by construction—hence $\tau_n \in \mathcal{T}^{c*}$. We conclude that $\mathbb{E}_{\tau_n}[\hat{v}(q)] \leq \mathbb{E}_{\tau_{n+1}}[\hat{v}(q)]$ for every n.

Now, for each n, let $Z_n = \sum_{q \in Q} \tau_n(q) + \sum_{\omega \in \mathcal{I}} \tau_0(q^\omega) z_n^\omega$ and express $\mathbb{E}_{\tau_n}[\hat{v}(q)]$ as

$$\underbrace{Z_n \sum_{q \in Q} \hat{v}(q) \frac{\tau_0(q)}{Z_n} + \sum_{\omega \in \mathcal{I}} \hat{v}(q^\omega(\cdot|\mathcal{P})) \frac{\tau_0(q^\omega) z_n^\omega}{Z_n}}_{B_n} + \underbrace{\sum_{\omega \in \mathcal{I}} \hat{v}(q_n^\omega) \tau_0(q^\omega)(1 - z_n^\omega)}_{B_n'}.$$

Note that $\lim_{n\to\infty} Z_n = 1 - \lim_{n\to\infty} \sum_{\omega \in \mathcal{I}} \tau_0(q^\omega)(1 - z_n^\omega) = \sigma(\mathcal{P})$. Also, since $z_n^\omega \uparrow (1 - q^\omega(\omega))$ as $n \to \infty$ for every $\omega \in \mathcal{I}$ and for all n

$$\sigma(\omega) = \tau_0(q^\omega)(1 - z_n^\omega)\left(1 - [K^\omega]^{-n}\right) = \tau(q^\omega) q^\omega(\omega),$$

it follows that $\lim_{n\to\infty} B_n' = \sum_{\omega \in \mathcal{I}} h(\omega)\sigma(\omega)$. Regarding the term B_n, first observe that

$$\frac{1}{Z_n}\left[\sum_{q \in Q} q \tau_n(q) + \sum_{\omega \in \mathcal{I}} q^\omega(\cdot|\mathcal{P}) \tau_0(q^\omega) z_n^\omega\right] = \hat{q}_n \in \Delta(\mathcal{P}),$$

and hence $\lim_{n\to\infty} \hat{q}_n = \sigma(\cdot|\mathcal{P})$. This implies that, for every n, B_n is a convex combination of values of $\hat{v}^c(q)$ involving posteriors $q \in \Delta(\mathcal{P})$ whose average equals $\hat{q}_n$. Recall that $V^c(q)$ is continuous at every $q \in \Delta(\mathcal{P})$. Letting $\chi = V^c(\sigma(\cdot|\mathcal{P}))$ and $\zeta = \sum_{\omega \in \mathcal{I}} h(\omega)\sigma(\omega)$, we can write

$$V^{c*}(\sigma) - \mathbb{E}_{\tau_n}[\hat{v}(q)] \leq |\chi|\,|Z_n - \sigma(\mathcal{P})| + Z_n\,|B_n - \chi| + \left|B_n' - \zeta\right|.$$

Given any $\gamma > 0$, there exist N_1 and N_2 such that $|\chi|\,|Z_n - \sigma(\mathcal{P})| + |B_n' - \zeta| \leq \frac{\gamma}{2}$ for all $n \geq N_1$ and $|\chi - V^c(\hat{q}_n)| \leq \frac{\gamma}{2}$ for all $n \geq N_2$. Fix $n^* \geq \max\{N_1, N_2\}$, and consider a distribution $\hat{\tau} \in \Delta(\Delta(\mathcal{P}))$ that achieves $V^c(\hat{q}_{n^*})$. Define τ_γ as follows:

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

By construction, $\mathbb{E}_{\tau_\gamma}[\hat{v}(q)] \geq V^{c*}(\sigma) - \gamma$ and

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

□

A.5 Proof of Proposition 4

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

Proof. For $\tau \in \mathcal{T}$, let $D^\tau = \mathbf{supp}\,\tau \cap D$ and $C^\tau = \mathbf{supp}\,\tau \setminus D^\tau$. Define $\tau(q|C^\tau) = \frac{\tau(q)}{\tau(C^\tau)}$ and

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

Also, let $q^c = \sum_{q \in C^\tau} q\tau(q|C^\tau)$, $q^d = \sum_{q \in D^\tau} q\tau(q|D^\tau)$, and $\tau^c = \tau(C^\tau)$, so that $q^c \in C$, $q^d \in D$, and $\sigma = \tau^c q^c + (1-\tau^c)q^d$. Then

$$\begin{aligned} \mathbb{E}_\tau[\hat{v}(q)] &\leq \tau^c V^c(q^c) + (1-\tau^c)\sum_{q \in D^\tau} v^d(q)\tau(q|D^\tau) \leq \tau^c V^c(q^c) + (1-\tau^c)\sum_{q \in D^\tau} \hat{h}(q)\tau(q|D^\tau) \\ &= \tau^c V^c(q^c) + (1-\tau^c)\sum_{q \in D^\tau}\left[\sum_{\omega \in \mathcal{I}} h(\omega)q(\omega)\right]\tau(q|D^\tau) \\ &= \tau^c V^c(q^c) + \sum_{\omega \in \mathcal{I}} h(\omega)\beta(\omega), \end{aligned} \tag{12}$$

where $\beta(\omega) = (1-\tau^c)\sum_{q \in D^\tau} q(\omega)\tau(q|D^\tau)$. Note that $\mathbf{supp}\,q^c \supset \mathcal{P}$ 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

$$V^c(q^c) = q^c(\mathcal{P})V^c(q^c(\cdot|\mathcal{P})) + \sum_{\omega \in \mathcal{I}} h(\omega)q^c(\omega).$$

Moreover, we have

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

which implies that $\sigma(\omega) = \beta(\omega) + \tau^c q^c(\omega)$ for all $\omega \in \mathcal{I}$, and hence $q^c(\cdot|\mathcal{P}) = \sigma(\cdot|\mathcal{P})$. Therefore, the expression on the right-hand side of (12) is equal to

$$\sigma(\mathcal{P})V(\sigma(\cdot|\mathcal{P})) + \sum_{\omega \in \mathcal{I}} h(\omega)\sigma(\omega) = V^c(\sigma).$$

Using the definition of V^* and Lemma 2, we conclude that $V(\sigma) \leq V^c(\sigma)$.

□

Claim 4. If $\hat{v}^d(q) > \hat{h}(q)$ for some $q \in D$, then there exists τ such that $\mathbb{E}_\tau[\hat{v}(q)] > V^c(\sigma)$, and hence $V(\sigma) > V^c(\sigma)$.

Proof. Let $q^* \in D$ satisfy $\hat{v}^d(q^*) > \hat{h}(q^*)$. Since $\sigma \in int\Delta(\Omega)$, there exist $t \in (0,1)$ and $q^c \in int\Delta(\Omega)$ such that $\sigma = tq^c + (1-t)q^*$. By the same argument as in the proof of Claim 3,

$$\begin{aligned} (1-t)\hat{v}^d(q^*) + tV^c(q^c) &> (1-t)\sum_{\omega \in \mathcal{I}} h(\omega)q^*(\omega) + tV^c(q^c) \\ &= \sum_{\omega \in \mathcal{I}} h(\omega)\{(1-t)q^*(\omega) + tq^c(\omega)\} + tq^c(\mathcal{P})V^c(q^c(\cdot|\mathcal{P})) = V^c(\sigma), \end{aligned}$$

where the last equality follows from observing that $\sigma(\omega) = (1-t)q^*(\omega) + tq^c(\omega)$ for all $\omega \in \mathcal{I}$, and hence $q^c(\cdot|\mathcal{P}) = \sigma(\cdot|\mathcal{P})$. Therefore, there exists $\tau \in \mathcal{T}$ such that $\mathbb{E}_\tau[\hat{v}(q)] > V^c(\sigma)$.

□

A.6 Proof of Proposition 5

The next three claims prove the result for the LBS model (Assumption 4), since Proposition 5 corresponds to the case of a binary LBS (ρ_0, ρ^1). As explained before the statement of Proposition 7 in Online Appendix B.2, $\hat{v}^d_*$ denotes the smallest u.s.c. function that pointwise dominates $\hat{v}^d$. Recall that $\hat{v}^d_* = \hat{v}^d$ for every binary LBS.

Claim 5. If $V(\sigma) > V^c(\sigma)$, then $V(\sigma) = \tau^c V^c(q^c) + (1-\tau^c)V^d_*(q^d)$, where $q^c \in C$, $q^d \in D$, $\sigma(\mathcal{P}) \le \tau^c < 1$, and $\sigma = \tau^c q^c + (1-\tau^c)q^d$.

Proof. For every $\tau \in \mathcal{T}$ with $|\mathbf{supp}\,\tau| \le |\Omega|$ and $\tau^c = \tau(C^\tau) < 1$, we can write

$$\mathbb{E}_\tau[\hat{v}(q)] \quad = \quad \tau^c \sum_{q \in C^\tau} \hat{v}(q)\tau(q|C^\tau) + (1-\tau^c) \sum_{q \in D^\tau} \hat{v}(q)\tau(q|D^\tau).$$

Define q^c and q^d as in the proof of Claim 3. Since $\mathbb{E}_\tau[\hat{v}(q)] \le \tau^c V^c(q^c) + (1-\tau^c)V^d(q^d)$, we must have

$$V(\sigma) = V^* = \sup_{\tau \in \mathcal{T}} \mathbb{E}_\tau[\hat{v}(q)] \le \sup_{\hat{\mathcal{T}}} \{\tau^c V^c(q^c) + (1-\tau^c)V^d(q^d)\}, \tag{13}$$

where

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

moreover, the inequality in (13) must be an equality, otherwise there would be $\tau \in \mathcal{T}$ with $\mathbb{E}_\tau[\hat{v}(q)] > V^*$. Since $\tau^c \in [\sigma(\mathcal{P}), 1]$, we must have $q^c(\omega) = \frac{1}{\tau^c}\sigma(\omega) \ge \sigma(\omega)$ for all $\omega \in \mathcal{P}$. The function V^c is continuous over $\Delta(\mathcal{P})$; therefore, by expression (4), $V^c(q^c)$ is continuous over $Q^c = \{q \in C : q(\omega) \ge \sigma(\omega) \text{ for all } \omega \in \mathcal{P}\} = \{q \in \Delta(\Omega) : q(\omega) \ge \sigma(\omega) \text{ for all } \omega \in \mathcal{P}\}$, which is compact. Construct $\mathcal{T}'$ by replacing C with Q^c in $\hat{\mathcal{T}}$. Since $V^d \le \mathrm{cl}V^d = \mathrm{cl}V^d_* = V^d_*$ and V^d_* is continuous by Lemma 4 in Appendix B, the right-hand side of (13) equals

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

First, note that the maximum on the right-hand side must be attained at $\tau^c < 1$, because by assumption $V^c(\sigma) < V(\sigma)$ (recall Lemma 3). Second, the inequality must be an equality. This is immediate if $|\mathcal{I}| = 1$, since in this case $V^d = v^d = v^d_* = V^d_*$, so suppose $|\mathcal{I}| > 1$ and define $D_n = \{q \in D : q(\omega) \ge |\mathcal{I}|^{-n} \text{ for all } \omega \in \mathcal{I}\}$ for $n \ge 1$. Construct $\mathcal{T}'_n$ by replacing D with D_n in $\mathcal{T}'$ for every n. Note that, for all n, $D_n \subseteq \mathit{int}\, D$ and $D_n \subseteq D_{n+1}$, and that $D_n \to D$ as $n \to \infty$. Since $V^d = V^d_*$ over $\mathit{int}\, D$ by Lemma 4 in Appendix B, for all n we have

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

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

$\square$

Recall that if $f : X \to Y$ and F is its concavification, then

$$F(x) = \sup\{r : (x,r) \in \text{co}(\text{hyp}\, f)\}, \quad x \in X, \tag{14}$$

where $\text{hyp}\, f = \{(x,r) \in X \times \mathbb{R} : r \leq f(x)\}$ and $\text{co}(\text{hyp}\, f)$ is its convex hull (Rockafellar (1997)).

Claim 6. If $V(\sigma) > V^c(\sigma)$, then (6) holds. Moreover, if $\hat{\tau} \in \Delta(D)$ satisfies $\mathbb{E}_{\hat{\tau}}[\hat{v}^d_*(q)] = V^d_*(q^d)$ and $\mathbb{E}_{\hat{\tau}}[q] = q^d$, then $\hat{v}_*{}^d(q) \geq \hat{h}(q)$ for all $q \in \mathbf{supp}\,\hat{\tau}$.

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

$$V(\sigma) \;=\; \tau^c V^c(q^c) + (1-\tau^c)V^d_*(q^d) = \tau^c q^c(\mathcal{P})V^c(q^c(\cdot|\mathcal{P})) + \tau^c \sum_{\omega\in\mathcal{I}} h(\omega)q^c(\omega) + (1-\tau^c)V^d_*(q^d),$$

where

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

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

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

Hence, for every $\omega \in \mathcal{I}$,

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

Thus, we obtain

$$V(\sigma) = \sigma(\mathcal{P})V^c(\sigma(\cdot|\mathcal{P})) + \sigma(\mathcal{I})\underbrace{\left[\gamma V^d_*(q^d) + (1-\gamma)\hat{h}(q^c(\cdot|\mathcal{I}))\right]}_{\xi^*}, \tag{15}$$

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

Now consider any $\hat{\tau} \in \Delta(D)$ that satisfies $\mathbb{E}_{\hat{\tau}}[\hat{v}^d_*(q)] = V^d_*(q^d)$ and $\mathbb{E}_{\hat{\tau}}[q] = q^d$. Suppose $\hat{v}^d_*(q') < \hat{h}(q')$ for some $q' \in \mathbf{supp}\,\hat{\tau}$. Then

$$\mathbb{E}_{\hat{\tau}}[\hat{v}^d_*(q)] < \sum_{\{q:q\neq q'\}} \hat{v}^d_*(q)\hat{\tau}(q) + \hat{\tau}(q')\sum_{\omega\in\mathcal{I}} h(\omega)q'(\omega).$$

But this implies the existence of $\tau \in \mathcal{T}$ with $\mathbb{E}_\tau[\hat{v}(q)] > V^*$ (a contradiction). Thus, for all $q \in \mathbf{supp}\,\hat{\tau}$, $\hat{v}^d_*(q) \geq \hat{h}(q)$.

In (15) we must have $\hat{h}(q^c(\cdot|\mathcal{I})) \geq \hat{v}^d_*(q^c(\cdot|\mathcal{I}))$, because otherwise it would be possible to improve upon V^*. Hence, ξ^* in (15) belongs to the set $\{\xi : (\sigma(\cdot|\mathcal{I}), \xi) \in \text{co}(\text{hyp}\,\hat{m})\}$—where $\hat{m} = \max\{\hat{h}, \hat{v}^d_*\}$—and must be equal to its maximum, which exists and equals $M(\sigma(\cdot|\mathcal{I}))$.

□

Claim 7. $\tau^c = \sigma(\mathcal{P})$ if and only if $M(\sigma(\cdot|\mathcal{I})) = V^d(\sigma(\cdot|\mathcal{I}))$.

Proof. If $1 - \tau^c = \sigma(\mathcal{I})$, then $\gamma = 0$ in (15). This implies that $q^c(\mathcal{I}) = 0$, $q^d = \sigma(\cdot|\mathcal{I})$, and $\xi^* = V^d_*(q^d)$; moreover, since $\sigma(\cdot|\mathcal{I}) \in int\, D$, $V^d_*(\sigma(\cdot|\mathcal{I})) = V^d(\sigma(\cdot|\mathcal{I}))$ by Lemma 4. Conversely, suppose $M(\sigma(\cdot|\mathcal{I})) = V^d(\sigma(\cdot|\mathcal{I}))$. Then, $\xi^* = V^d_*(\sigma(\cdot|\mathcal{I}))$ in (15), and hence $\gamma = 0$, which implies that $q^c(\mathcal{I}) = 0$ and hence $\sigma(\mathcal{I}) = \tau^c q^c(\mathcal{I}) + 1 - \tau^c = 1 - \tau^c$.

□

A.7 Proof of Corollary 5

I again prove this result for the LBS model (Assumption 4).

Part I. By Proposition 5, conditional on $\mathcal{I}$, Sender's expected payoff equals $M(\sigma(\cdot|\mathcal{I})) = \gamma V^d_*(q_1) + (1-\gamma)\hat{h}(q_2)$, where $q_1, q_2 \in D$ and $\gamma q_1 + (1-\gamma) q_2 = \sigma(\cdot|\mathcal{I})$. Suppose that $\hat{v}^d(\delta_{\omega'}) > h(\omega')$ for some $\omega' \in \mathcal{I}$ and $\omega' \in \textbf{supp}\, q_2$. Letting $\hat{\tau} \in \Delta(D)$ satisfy $\mathbb{E}_{\hat{\tau}}[q] = q_1$ and $\mathbb{E}_{\hat{\tau}}[\hat{v}^d_*(q)] = V^d_*(q_1)$, we have

$$\begin{aligned} &\sum_{\omega \neq \omega'} h(\omega)(1-\gamma)q_2(\omega) + h(\omega')(1-\gamma)q_2(\omega') + \sum_q \hat{v}^d_*(q)\gamma\hat{\tau}(q) \\ &= \sum_{\omega \neq \omega'} \hat{h}(\delta_\omega)\tau'(\delta_\omega) + \hat{h}(\delta_{\omega'})\tau'(\delta_{\omega'}) + \sum_q \hat{v}^d_*(q)\tau'(q), \end{aligned} \tag{16}$$

where $\tau'(\delta_\omega) = (1-\gamma)q_2(\omega)$ for the first summation and $\tau'(q) = \gamma\hat{\tau}(q)$ for the second. Thus,

$$\mathbb{E}_{\tau'}[q(\omega)] = \tau'(\delta_\omega)\delta_\omega + \gamma \mathbb{E}_{\hat{\tau}}[q(\omega)] = \gamma q_1(\omega) + (1-\gamma)q_2(\omega) = \sigma(\omega|\mathcal{I}) \quad \text{for every } \omega \in \mathcal{I}.$$

Since by assumption $\tau'(\delta_{\omega'}) > 0$ and $\hat{h}(\delta_{\omega'}) < \hat{v}^d(\delta_{\omega'}) \leq \hat{v}^d_*(\delta_{\omega'})$, the expression on the right-hand side of (16) is strictly smaller than $\mathbb{E}_{\tau'}[\hat{m}(q)]$. But this leads to a contradiction, since $M(\sigma(\cdot|\mathcal{I})) = \max_{\{\tau : \mathbb{E}_\tau[q] = \sigma(\cdot|\mathcal{I})\}} \mathbb{E}_\tau[\hat{m}(q)]$.

When $\hat{v}^d(\delta_\omega) = h(\omega)$, we can say that Sender always surprises Receiver in ω. Indeed, for every τ that satisfies $\mathbb{E}_\tau[q] = \sigma(\cdot|\mathcal{I})$ which implies that with positive probability she does not surprise Receiver in ω, there exists τ' with $\mathbb{E}_{\tau'}[q] = \sigma(\cdot|\mathcal{I})$ such that she always surprises him in ω and $\mathbb{E}_{\tau'}[\hat{m}(q)] = \mathbb{E}_\tau[\hat{m}(q)]$. Thus, always surprising Receiver in ω is compatible with achieving $M(\sigma(\cdot|\mathcal{I}))$.

Part II. First, note that if $\hat{v}^d$ is convex, then $V^d_*(q) = \mathbb{E}_q[\hat{v}^d(\delta_\omega)]$ for every $q \in D$. This holds because convexity of $\hat{v}^d$ implies that $V^d(q) = \mathbb{E}_q[\hat{v}^d(\delta_\omega)]$ and $V^d(q) = V^d_*(q)$ for every $q \in int\, D$ by Lemma 4 in Appendix B; the equality extends to the boundary of D, because V^d_* is continuous. Now let $\mathcal{I}_< = \{\omega : \hat{v}^d(\delta_\omega) < h(\omega)\}$ and $\mathcal{I}_\geq = \{\omega : \hat{v}^d(\delta_\omega) \geq h(\omega)\}$. Suppose that, contrary to the claimed result, q^d in (7) satisfies $\textbf{supp}\, q^d \cap \mathcal{I}_< \neq \varnothing$. Then, as noted before, $M(\sigma(\cdot|\mathcal{I})) =$

$(1-\gamma)\hat{h}(q_1) + \gamma V_*^d(q^d)$, which can be written as

$$(1-\gamma)\mathbb{E}_{q_1}[h(\omega)] + \gamma\mathbb{E}_{q^d}[\hat{v}^d(\delta_\omega)] < (1-\gamma)\mathbb{E}_{q_1}[h(\omega)] + \gamma \sum_{\omega\in\mathcal{I}_<} h(\omega)q^d(\omega) + \gamma \sum_{\omega\in\mathcal{I}_\geq} \hat{v}^d(\delta_\omega)q^d(\omega)$$
$$= \sum_{\omega\in\mathcal{I}_<} h(\omega)\sigma(\omega|\mathcal{I}) + \sum_{\omega\in\mathcal{I}_\geq} \hat{v}^d(\delta_\omega)\sigma(\omega|\mathcal{I}),$$

where the equality follows because $(1-\gamma)q_1 + \gamma q^d$ must equal $\sigma(\cdot|\mathcal{I})$. However, the last expression cannot exceed $M(\sigma(\cdot|\mathcal{I}))$, which leads to a contradiction.

References

Alonso, R. and O. Câmara (2016). Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory, forthcoming*.

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

Aragones, E., I. Gilboa, I., A. Postlewaite, A., and D. Schmeidler (2005). Fact free learning. *American Economic Review 95*(5), 1355–1368.

Aumann, R. J. and M. Maschler (1995). *Repeated Games with Incomplete Information*. MIT Press.

Billot, A., I. Gilboa, D. Samet, and D. Schmeidler (2005). Probabilities as similarity-weighted frequencies. *Econometrica 73*(4), 1125–1136.

Blume, L., A. Brandenburger, and E. Dekel (1991a). Lexicographic probabilities and choice under uncertainty. *Econometrica 59*(1), 61–79.

Blume, L., A. Brandenburger, and E. Dekel (1991b). Lexicographic probabilities and equilibrium refinements. *Econometrica 59*(1), 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 50*(6), 1431–1451.

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

Edwards, W. (1968). Conservatism in human information processing. *In: Kleinmutz, B. (Ed.), Formal Representation of Human Judgment. John Wiley and Sons, New York*, 17–51.

Ely, J. (2016). Beeps. *American Economic Review, forthcoming.*

Ely, J., A. Frankel, and E. Kamenica (2015). Suspense and surprise. *Journal of Political Economy 123*(1), 215–260.

Fermé, E. and S. O. Hansson (2011). Agm 25 years. Twenty-five years of research in belief change. *Journal of Philosophical Logic 40*(2), 295–331.

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

Gardner, H. (2006). *Changing minds: The art and science of changing our own and other people's minds*. Harvard Business Review Press.

Gilboa, I., A. W. Postlewaite, and D. Schmeidler (2008). Probability and uncertainty in economic modeling. *Journal of Economic Perspectives 22*(3), 173–188.

Gilboa, I. and D. Schmeidler (2000). Case-based knowledge and induction. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 30*(2), 85–95.

Glazer, J. and A. Rubinstein (2012). A model of persuasion with boundedly rational agents. *Journal of Political Economy 120*(6), 1057–1082.

Glazer, J. and A. Rubinstein (2014). Complex questionnaires. *Econometrica 82*(4), 1529–1541.

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.

Hong, H., J. C. Stein, and J. Yu (2007). Simple forecasts and paradigm shifts. *Journal of Finance 62*(3), 1207–1242.

Horner, J. and A. Skrzypacz (2014). Selling information. *Journal of Political Economy, forthcoming.*

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Knowles, E. S. and J. A. Linn (2004). *Resistance and persuasion.* Psychology Press.

Kreps, D. M. and R. Wilson (1982). Sequential equilibria. *Econometrica 50*(4), 863–894.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

McCloskey, D. and A. Klamer (1995). One quarter of GDP is persuasion. *American Economic Review 85*(2), 191–195.

Milgrom, P. and J. Roberts (1986). Relying on the information of interested parties. *The RAND Journal of Economics 17*(1), 18–32.

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

Morgan, J. and P. C. Stocken (2003). An analysis of stock recommendations. *The RAND Journal of Economics 34*(1), 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. *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 114*(1), 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.

Savage, L. J. (1972). *The foundations of statistics*. Dover Publications, New York.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. Reprint, London: Routledge, 1961.

B Online Appendix (For Online Publication Only)

B.1 Nonexistence of Optimal Experiments

Let $\Omega = \{\omega_1, \omega_2\}$ with $\sigma = (\frac{1}{2}, \frac{1}{2})$, $\varepsilon = 0$, $\mathcal{P} = \{\omega_1\}$, and $A = \{a, b, c\}$. Sender's and Receiver's utility functions are as follows:

v	a	b	c
ω_1	1	0	-1
ω_2	0	1	-1

u	a	b	c
ω_1	1	1	1
ω_2	0	0	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 $q \neq (0,1)$, $\mathbf{p}(q) = (1,0)$ and $\mathcal{A}((1,0)) = A$. Therefore, Sender can make Receiver choose a or b, depending on q. For $q = (0,1)$, $\mathbf{p}_0(q) = (0,1)$ and $\mathcal{A}((0,1)) = \{c\}$. Note that $\mathbf{p}_0$ is not continuous at $q = (0,1)$.

Consider experiment π, where $\pi(s_1|\omega_1) = 1-\gamma$, $\pi(s_1|\omega_2) = \gamma$, $\pi(s_2|\omega_2) = 1-\gamma$, and $\pi(s_2|\omega_1) = \gamma$. Thus, $q(\cdot|s_1, \pi) = q_1 = (1-\gamma, \gamma)$ and $q(\cdot|s_2, \pi) = q_2 = (\gamma, 1-\gamma)$, each arising with probability $\frac{1}{2}$. For any $\gamma > 0$, Sender's expected payoff is then

$$\frac{1}{2}\hat{v}(q_1) + \frac{1}{2}\hat{v}(q_2) = \frac{1-\gamma}{2} + \frac{1-\gamma}{2} = 1-\gamma. \tag{17}$$

However, for $\gamma = 0$ we have

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

Clearly, the supremum of Sender's expected payoff over all π's is 1, which is also the maximum she can hope for. But no π can achieve 1. This would require Sender to learn the true state in order to induce Receiver to choose her preferred action accordingly; but then Receiver must also learn the true state, and hence will choose c in ω_2. Nonetheless, expression (17) says that any π which allows Sender to almost perfectly learn ω but never disproves ρ_0 is virtually optimal.

B.2 Analysis of the LBS Model

With regard to the feasible distributions over posteriors, the statement of Proposition 2 for $\varepsilon = 0$ changes in two ways. First, the function $\mathbf{r}$ is replaced by a new function $\mathbf{r}_L$ that, given LBS $(\rho_0, \ldots, \rho_N)$, says which prior Receiver updates for every $q \in \Delta(\Omega)$. By A4, for every (s, π) this is the ρ^i with lowest index that satisfies $\rho^i(\mathbf{supp}\, q(\cdot|s,\pi)) > 0$; therefore, for every $q \in \Delta(\Omega)$,

$$\mathbf{r}_L(q) = \rho^{i(q)}, \text{ where } i(q) = \min\{i : \rho^i(\mathbf{supp}\, q) > 0\}. \tag{18}$$

The second change is in the range of Receiver's feasible posterior. To describe this, for $i = 1, \ldots, N$ let $\mathcal{P}^i$ contain all the states which are deemed possible by prior ρ^i but impossible by all ρ^j of lower index:

$$\mathcal{P}^i = \mathbf{supp}\,\rho^i \setminus (\cup_{j<i} \mathbf{supp}\,\rho^j).$$

Also, let $\mathcal{R} = \cup_{i=0}^N \Delta(\mathcal{P}^i)$, where $\mathcal{P}^0 = \mathcal{P}$. Then, under A4, the new function $\mathbf{p}_L$ from Sender's to Receiver's posterior has range $\mathcal{R}$ rather than $\Delta(\mathcal{P}) \cup \Delta(\mathcal{I})$. To see this, note first that if $\mathbf{r}_L(q) = \rho^i$, then $\mathbf{supp}\,q \cap (\cup_{j<i} \mathbf{supp}\,\rho^j) = \varnothing$, and hence $\mathbf{supp}\,\mathbf{p}(q) = \mathbf{supp}\,q \cap \mathbf{supp}\,\mathbf{r}_L(q) \subseteq \mathcal{P}^i$. Conversely, for any $i \in \{0, \ldots, N\}$ and $q \in \Delta(\mathcal{P}^i)$, we have $\mathbf{r}_L(q) = \rho^i$, and hence for all $\omega \in \mathcal{P}^i$,

$$\mathbf{p}_L(\omega; q) = q(\omega) \frac{\rho^i(\omega|\mathcal{P}^i)}{\sigma(\omega|\mathcal{P}^i)} \left[\sum_{\omega' \in \mathcal{P}^i} q(\omega') \frac{\rho^i(\omega'|\mathcal{P}^i)}{\sigma(\omega'|\mathcal{P}^i)} \right]^{-1}.$$

Then Proposition 1 applied to priors $\rho^i(\omega|\mathcal{P}^i)$ and $\sigma(\omega|\mathcal{P}^i)$ implies that by varying $q \in \Delta(\mathcal{P}^i)$ we obtain every $p \in \Delta(\mathcal{P}^i)$. The function $\mathbf{p}_L$ continues to have the three key properties highlighted after Proposition 2, which drive the analysis in Section 5.

Formally, consider any LBS $(\rho_0, \ldots, \rho_N)$ with $N \geq 2$. For every $\omega \in \mathcal{I}$, let $i(\omega) = \max\{i : \omega \notin \mathbf{supp}\,\rho^i\}$. Then for every $i \leq i(\omega)$ Sender can conceal ω in any q that satisfies $\mathbf{r}_L(q) = \rho^i$. Since the range of $\mathbf{p}_L$ is $\mathcal{R} = \cup_{i=0}^N \Delta(\mathcal{P}^i)$, for $i = 0, \ldots, N$ let $\mathcal{A}^i$ be the set of all actions Receiver could choose if he applied ρ^i: $\mathcal{A}^i = \cup_{p \in \Delta(\mathcal{P}^i)} \mathcal{A}(p)$. Sender's best payoff from concealing each $\omega \in \mathcal{I}$ across *all ways* allowed by Receiver's LBS is then

$$h_*(\omega) = \max_{a \in \cup_{i=0}^{i(\omega)} \mathcal{A}^i} v(a, \omega).$$

Her expected payoff from concealment is given by the function $\hat{h}_* : D_0 \to \mathbb{R}$ that satisfies

$$\hat{h}_*(q) = \mathbb{E}_q[h_*(\omega)], \quad q \in D_0.$$

Now define $\hat{m}_* : D_0 \to \mathbb{R}$ by $\hat{m}_* = \max\{\hat{v}_*^d, \hat{h}_*\}$, where $\hat{v}_*^d$ is the smallest upper-semicontinuous (u.s.c.) function that satisfies $\hat{v}_*^d(q) \geq \hat{v}^d(q)$ for all q.[31] Let M_* denote the concavification of $\hat{m}_*$. The reason for this slight modification of $\hat{v}^d$ is that, if $\hat{v}^d$ is u.s.c., then V^d is continuous—by Lemma 4 in Appendix B—and $V^d(q)$ is achieved for all $q \in D_0$. But $\hat{v}^d$ can fail to be u.s.c. when $\mathbf{p}_L$ is discontinuous on D_0. However, we can render $\hat{v}^d$ u.s.c. using $\hat{v}_*^d$ without loss of generality (see the next result).

The next result characterizes the payoff from an optimal π that possibly exploits Receiver's multiple worldviews.

Proposition 7. *Sender's expected payoff for states outside Receiver's worldview—$M(\sigma(\cdot|\mathcal{I}))$ in (6)—satisfies*

$$M(\sigma(\cdot|\mathcal{I})) = M_*(\sigma(\cdot|\mathcal{I})) = \max_{\alpha, q^h, q^s} \{\alpha \hat{h}_*(q^h) + (1-\alpha) V_*^d(q^s)\}, \tag{19}$$

[31]Note that $\hat{v}^d$ is always u.s.c. for LBSs with only one alternative prior as in the baseline model. Also, since $\hat{v}_*^d$ and $\hat{h}_*$ are u.s.c., so is $\hat{m}_*$. Indeed, a function is u.s.c. if and only if its hypograph is closed (Theorem 7.1, Rockafellar [1997]). In our case, $\operatorname{hyp} \hat{m}_* = \operatorname{hyp} \hat{h}_* \cup \operatorname{hyp} \hat{v}_*^d$.

subject to $\alpha \in [0,1]$, $q^h, q^s \in D_0$, and $\alpha q^h + (1-\alpha)q^s = \sigma(\cdot|\mathcal{I})$. Given an optimal α^*, Sender conceals states with ex-ante probability $\alpha^*\sigma(\mathcal{I})$.

Note that $\mathbf{supp}\, q^h$ tells us which states are concealed, while $h_*(\omega)$ tells us which evidence *and* worldview are used to conceal ω—hence, how deeply Receiver is induced to think. Similarly, if $\omega \in \mathbf{supp}\, q^s \cap \mathcal{P}^i$, it means that he is induced to think up to (at least) level i in his LBS.

Proof. We need to show only that $M(\sigma(\cdot|\mathcal{I})) = M_*(\sigma(\cdot|\mathcal{I}))$, because the second equality in (19) follows immediately from the definition of $M_*(\sigma(\cdot|\mathcal{I}))$. Since $h_* \geq h$, we have that $\hat{m}_* \geq \hat{m}$, and hence it suffices to show that $M_*(\sigma(\cdot|\mathcal{I})) \leq M(\sigma(\cdot|\mathcal{I}))$.

Since by definition $\hat{v}^d_* \geq \hat{v}^d$, we have $\hat{m} \geq \max\{\hat{h}, \hat{v}^d\}$, and hence

$$M(\sigma(\cdot|\mathcal{I})) \geq \sup_{\{\tau:\mathbb{E}_\tau[q]=\sigma(\cdot|\mathcal{I})\}} \mathbb{E}_\tau[\max\{\hat{h}(q), \hat{v}^d(q)\}].$$

By the same logic as in the proof of Proposition 3, without loss of generality each $q \in D_0$ conceals at most one $\omega \in \mathcal{I}$ and hence can be written as $q = (1-q(\omega))q(\cdot|\mathcal{P}^i) + q(\omega)\delta_\omega$, where $\mathcal{P}^i = \mathbf{supp}\, q \cap \mathbf{supp}\, \mathbf{r}_L(q)$; moreover, Sender can always conceal ω in some $q \in D_0$ that satisfies $v(a(\mathbf{p}_L(q)), \omega) = h_*(\omega)$. Let $\mathcal{T}^h_{\sigma(\cdot|\mathcal{I})}$ be the subset of $\{\tau : \mathbb{E}_\tau[q] = \sigma(\cdot|\mathcal{I})\}$ that contains only the $\tau \in \Delta(D_0)$ for which she conceals states in this way. Then for every $\tau \in \mathcal{T}^h_{\sigma(\cdot|\mathcal{I})}$,

$$\begin{aligned}
\mathbb{E}_\tau[\max\{\hat{h}(q), \hat{v}^d(q)\}] &= \mathbb{E}_\tau[\max\{\hat{h}(q), (1-q(\omega))\hat{v}^d(q(\cdot|\mathcal{P}^i)) + q(\omega)\delta_\omega \hat{h}_*(\delta_\omega)\}] \\
&= \mathbb{E}_{\tau'}[\max\{\hat{h}_*(q'), \hat{v}^d(q')\}]
\end{aligned}$$

for some $\tau' \in \{\tau : \mathbb{E}_\tau[q] = \sigma(\cdot|\mathcal{I})\}$. It follows that

$$M(\sigma(\cdot|\mathcal{I})) \geq \sup_{\{\tau:\mathbb{E}_\tau[q]=\sigma(\cdot|\mathcal{I})\}} \mathbb{E}_\tau[\max\{\hat{h}_*(q), \hat{v}^d(q)\}].$$

Now let $\hat{m}^d = \max\{\hat{h}_*, \hat{v}^d\}$. Note that $\hat{m}^d \leq \max\{\hat{h}_*, \hat{v}^d_*\} = \hat{m}_*$, and that $\hat{m}_*$ is the smallest u.s.c. function that pointwise dominates $\hat{m}^d$. Therefore, $M^d \leq M_*$; moreover, by the same logic as in the proof of Lemma 4 in Appendix B, we must have $M^d(\sigma(\cdot|\mathcal{I})) = M_*(\sigma(\cdot|\mathcal{I}))$ because $\sigma(\cdot|\mathcal{I}) \in \mathit{int}\, D_0$. This implies that $M(\sigma(\cdot|\mathcal{I})) \geq M_*(\sigma(\cdot|\mathcal{I}))$, as desired.

□

We can also identify which states Sender will conceal, before characterizing her entire experiment.

Corollary 8. *Fix $\omega \in \mathcal{I}$. If $\hat{v}^d(\delta_\omega) \geq h_*(\omega)$, then Sender never conceals ω, that is, $\omega \notin \mathbf{supp}\, q^h$ in (19). If $\hat{v}^d(\delta_\omega) < h_*(\omega)$ and $\hat{v}^d$ is convex, then Sender always conceals ω, that is, $\omega \notin \mathbf{supp}\, q^s$ in (19).*[32]

Even if $\hat{v}^d(\delta_\omega) \geq h_*(\omega)$, Sender need not fully reveal ω. If $\hat{v}^d$ is not convex, even if $\hat{v}^d(\delta_\omega) < h_*(\omega)$, she may reveal $\omega \in \mathcal{I}$ with positive probability (even whenever ω occurs).

[32]The proof of Corollary 8 follows the same logic as that of Corollary 5 and is therefore omitted.

B.3 Discontinuity of Receiver's Posterior over D_0 under A5

Let $\varepsilon = 0$. 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 $\mu(\rho^0) = \frac{1}{2}$, $\mu(\rho^1) = \mu(\rho^2) = \frac{1}{4}$, $\mathcal{P} = \{\omega_1\}$, $\rho^1 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0)$, and $\rho^2 = \sigma$. Consider Sender's posterior $q_z = (0, \frac{1-z}{2}, \frac{1-z}{2}, z) \in D_0$ for $z \in (0,1)$. Then

$$\mu'(\rho^1; q_z) = \frac{\sum_{\omega\in\Omega} q_z(\omega)\rho^1(\omega)}{\sum_{\omega\in\Omega} q_z(\omega)\rho^1(\omega) + \sum_{\omega\in\Omega} q_z(\omega)\rho^2(\omega)} = \frac{1}{1+\frac{2}{3(1-z)}}, \quad \text{and} \quad \mu'(\rho^2; q_z) = 1 - \mu'(\rho^1; q_z).$$

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 AC's logic, 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

$$\mathbf{p}_1(\omega_3; q_z) = \frac{(1-z)\frac{1}{2}}{(1-z)\frac{1}{2} + (1-z)\frac{1}{4}} = \frac{2}{3},$$

$$\mathbf{p}_2(\omega_3; q_z) = \frac{(1-z)\frac{1}{4}}{(1-z)\frac{1}{4} + (1-z)\frac{1}{4} + 2z\frac{1}{4}} = \frac{1-z}{2}.$$

Thus, Receiver's posterior must vary discontinuously in q_z at $z = \frac{1}{3}$. Note also that for z with $0 < z < \frac{1}{3}$ we have $\mathbf{p}_1(\omega_4; q_z) = 0$, even though $q_z(\omega_4) > 0$. Therefore, Sender is hiding ω_4 in posterior q_z after disproving ρ^0.

B.4 Proof of Lemma 1

The case of $\varepsilon = 0$ is proven in the main text. Let $\varepsilon > 0$. Since $\mathbf{supp}\,\rho_0 = \mathbf{supp}\,\sigma$, by Proposition 1 there exists an homeomorphism from Sender's to Receiver's posteriors that describes the Bayes' updating of ρ_0 corresponding to the (s, π) inducing each q. Using this, we can calculate the ratio ℓ defining unexpected evidence directly in terms of q as

$$\ell(q) = \frac{\sum_{\omega\in\mathcal{P}} \frac{q(\omega)}{\sigma(\omega)}\rho_0(\omega)}{\sum_{\omega'\in\mathcal{I}} \frac{q(\omega')}{\sigma(\omega')}\rho_0(\omega')}.$$

Since the condition $\ell(q) \leq \theta_\varepsilon$ partitions the whole set of Receiver's posteriors $\Delta(\Omega)$, it implies (via the homeomorphism) the partition of Sender's posterior given by

$$D_\varepsilon = \{q : \ell(q) \leq \theta_\varepsilon\} \quad \text{and} \quad C_\varepsilon = \{q : \ell(q) > \theta_\varepsilon\} = \Delta(\Omega) \setminus D_\varepsilon.$$

As ε falls, D_ε shrinks in the sense of set inclusion. Indeed,

$$\ell(q) \leq \theta_\varepsilon \quad \text{iff} \quad \frac{\sum_{\omega\in\mathcal{P}} \frac{q(\omega)}{\sigma(\omega)}\frac{\rho_0(\omega)}{\rho_0(\mathcal{P})}}{\sum_{\omega'\in\mathcal{I}} \frac{q(\omega')}{\sigma(\omega')}\frac{\rho_0(\omega')}{\rho_0(\mathcal{I})}} \leq \theta_\varepsilon \frac{\varepsilon}{1-\varepsilon};$$

also, by A1 the left-hand side of the second inequality is independent of ε, and by A2 the right-hand side decreases to zero as $\varepsilon \to 0$. Using continuity of ℓ in q, we get $D_{\varepsilon'} \subset D_\varepsilon$ for all $\varepsilon' < \varepsilon$. In the limit $q \in D_0$ if and only if $q(\omega) = 0$ for all $\omega \in \mathcal{P}$. Hence, $\lim_{\varepsilon\to 0} D_\varepsilon = \cap_{\varepsilon \geq 0} D_\varepsilon = \Delta(\mathcal{I})$. Finally, since $C_\varepsilon = \Delta(\Omega) \setminus D_\varepsilon$, C_ε grows in the sense of set inclusion as ε falls. Also, $\lim_{\varepsilon\to 0} C_\varepsilon = \Delta(\Omega) \setminus (\lim_{\varepsilon\to 0} D_\varepsilon) = \Delta(\Omega) \setminus \Delta(\mathcal{I})$.

B.5 Proof of Lemma 2

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

$$V_\varepsilon(\sigma) = \sup_{T_\sigma} \sum_i \tau_i \hat{v}_\varepsilon(q_i),$$

where

$$T_\sigma = \left\{ (q_1, \tau_1; \ldots; q_{|\Omega|}, \tau_{|\Omega|}) : \sum_{i=1}^{|\Omega|} \tau_i q_i = \sigma, \sum_{i=1}^{|\Omega|} \tau_i = 1, \tau_i \geq 0, \text{ and } q_i \in \Delta(\Omega) \text{ for all } i \right\}.$$

Since $T_\sigma \subseteq \mathcal{T}$, it follows that $V_\varepsilon(\sigma) \leq V_\varepsilon^*$. By definition of V_ε^*, for every $\gamma > 0$ there exists $\tau_\gamma \in \mathcal{T}$ such that $\mathbb{E}_{\tau_\gamma}[\hat{v}_\varepsilon(q)] \geq V^* - \gamma$. However, $\mathbb{E}_{\tau_\gamma}[\hat{v}_\varepsilon(q)] \in \{\xi : (\sigma, \xi) \in \text{co}(\text{hyp}\, v)\}$, and hence $\mathbb{E}_{\tau_\gamma}[\hat{v}_\varepsilon(q)] \leq V_\varepsilon(\sigma)$. Thus, for every $\gamma > 0$, $V_\varepsilon(\sigma) \geq V_\varepsilon^* - \gamma$, which implies that $V_\varepsilon(\sigma) \geq V_\varepsilon^*$.

B.6 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,

$$V_\varepsilon^c(\sigma) = \sup_{T_\sigma^c} \sum_i \tau_i \hat{v}_\varepsilon(q_i),$$

where

$$T_\sigma^c = \{(q_1, \tau_1; \ldots; q_N, \tau_N) : N \geq 1, \sum_{i=1}^{N} \tau_i q_i = \sigma, \sum_{i=1}^{N} \tau_i = 1, \tau_i \geq 0, \text{ and } q_i \in C_\varepsilon \text{ for all } i\}.$$

Suppose $\mathbb{E}_\tau[\hat{v}_\varepsilon(q)] > V_\varepsilon^c(\sigma)$, but $\tau^d = 0$. Since $\tau \in \mathcal{T}$, $|\mathbf{supp}\, \tau| = N$ for some finite N, and hence $\mathbf{supp}\, \tau \cap D_\varepsilon = \varnothing$. Therefore, $\tau \in T_\sigma^c$, and hence $\mathbb{E}_\tau[\hat{v}_\varepsilon(q)] \leq V_\varepsilon^c(\sigma)$, a contradiction.

B.7 Lemma 4

Lemma 4. *The function V_*^d has the following properties:*
(i) For every $q \in D_0$, there exists $\tau \in \Delta(D_0)$ such that $V_*^d(q) = \mathbb{E}_\tau[\hat{v}_*^d(q')]$, where $q = \mathbb{E}_\tau[q']$ and $|\mathbf{supp}\, \tau| \leq |\mathcal{I}|$.
(ii) $V^d \leq V_*^d$, with equality over $\text{int}\, D_0$.

(iii) $V_*^d = \text{cl}V_*^d$ *and hence is continuous.*

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

$$V_*^d(q) = \sup_{T(q)} \sum_{i=1}^{|\mathcal{I}|} \hat{v}_*^d(q_i)\tau_i,$$

where

$$T(q) = \left\{ (q_1, \tau_1; \ldots; q_{|\mathcal{I}|}, \tau_{|\mathcal{I}|}) : \sum_{i=1}^{|\mathcal{I}|} q_i \tau_i = q, \sum_{i=1}^{|\mathcal{I}|} \tau_i = 1, \tau_i \geq 0, \text{ and } q_i \in \Delta(\mathcal{I}) \text{ for all } i \right\}.$$

Since $\hat{v}_*^d$ is u.s.c. and $T(q)$ is compact, by standard arguments $V_*^d(q)$ is achieved for every $q \in D_0$.

Part (ii): Given a function $f : D_0 \to \mathbb{R}$, let hyp$f$ be the hypograph of f: $\text{hyp}f = \{(q, \xi) : q \in D_0, \xi \in \mathbb{R}, \xi \leq f(q)\}$. Note that $\text{hyp}\, \hat{v}_*^d = \overline{\text{hyp}\, \hat{v}^d}$. Therefore, for all $q \in D_0$,

$$V^d(q) \quad = \quad \sup\{\xi : (q, \xi) \in \text{co}(\text{hyp}\, \hat{v}^d)\} \leq \sup\{\xi : (q, \xi) \in \text{co}(\overline{\text{hyp}\, \hat{v}^d})\} = V_*^d(q).$$

Now consider the closure of V^d, $\text{cl}V^d$, which is the unique continuous extension of V^d to D_0 by Theorem 10.3 in Rockafellar (1997), is concave, and satisfies $\text{cl}V^d \geq V^d \geq \hat{v}^d$. Thus, for every $q \in D_0$,

$$\hat{v}_*^d(q) = \limsup_{q' \to q} \hat{v}^d(q') \leq \limsup_{q' \to q} \text{cl}V^d(q') = \text{cl}V^d(q).$$

Hence, $\text{cl}V^d$ is a concave function majorizing $\hat{v}_*^d$. Since V_*^d is the smallest of such functions, $\text{cl}V^d \geq V_*^d$. Finally, since $\text{cl}V^d = V^d$ over $\textit{int}\, D_0$, property (ii) follows.

Part (iii): We already know that $V_*^d = \text{cl}V_*^d$ over $\textit{int}\, D_0$. By definition, $\text{hyp}\,\text{cl}V_*^d = \overline{\text{hyp}\, V_*^d}$. If $\text{hyp}\, V_*^d$ is closed, then $\text{hyp}\, V_*^d = \text{hyp}\,\text{cl}V_*^d$ and we are done. Indeed, by definition, $V_*^d \leq \text{cl}V_*^d$, so suppose there exists $q \in \partial D_0$ such that $V_*^d(q) < \text{cl}V_*^d(q)$. Then there exists $\xi \in \mathbb{R}$ such that $V_*^d(q) < \xi \leq \text{cl}V_*^d$, which is a contradiction. Hence, all that remains is to prove that $\text{hyp}\, V_*^d$ is closed.

First, for every $q \in D_0$, by property (i) we have $V_*^d(q) = \max\{\xi : (q, \xi) \in \text{co}(\text{hyp}\, \hat{v}_*^d)\}$, and therefore $\text{hyp}\, V_*^d = \text{co}(\text{hyp}\, \hat{v}_*^d)$. Second, define $\underline{v}_*^d = \inf_{q \in D_0} \hat{v}_*^d(q)$, so that we can express $\text{hyp}\, \hat{v}_*^d$ as $\mathcal{G} \cup \mathcal{H}$, where

$$\mathcal{G} = \{(q, \xi) : q \in D_0, \underline{v}_*^d - 1 \leq \xi \leq \hat{v}_*^d(q)\} \quad \text{and} \quad \mathcal{H} = \{(q, \xi) : q \in D_0, \xi \leq \underline{v}_*^d - 1\}.$$

Now we will show that $\text{co}(\text{hyp}\, \hat{v}_*^d) = (\text{co}\mathcal{G}) \cup (\text{co}\mathcal{H}) = (\text{co}\mathcal{G}) \cup \mathcal{H}$. The inclusion from right to left is trivial, so consider $(q, \xi) \in \text{co}(\text{hyp}\, \hat{v}_*^d)$. Then, by Theorem 2.3 in Rockafellar (1997), (q, ξ) is a convex combination of points (q_n, ξ_n) in $\text{hyp}\, \hat{v}_*^d$. Therefore, $q \in D_0$, as the latter is a convex set, and $\xi = \sum_n \alpha_n \xi_n \leq \sum_n \alpha_n \hat{v}_*^d(q_n)$ since $\alpha_n \geq 0$ for all n. But $(\text{co}\mathcal{G}) \cup \mathcal{H}$ contains all convex combinations of points in $\text{hyp}\, \hat{v}_*^d$ that satisfy this property, proving the inclusion from left to right. Finally, note that $\mathcal{H}$ is closed, and that $\mathcal{G}$ is bounded and closed, since $\hat{v}_*^d$ is upper semicontinuous. Therefore, $\text{co}(\mathcal{G})$ is also closed by Theorem 17.2 in Rockafellar (1997). We conclude that $\text{co}(\text{hyp}\, \hat{v}_*^d) = (\text{co}\mathcal{G}) \cup \mathcal{H}$

is closed, as desired.

□

B.8 Corollary 9

The next result provides a simpler conditions for Sender not to disprove Receiver's worldview whenever the true ω is in $\mathcal{I}$. Intuitively, one has to consider the optimal ways in which Sender can disprove ρ_0 whenever $\omega \in \mathcal{I}$, and check if *any* induces posteriors at which her expected payoff falls short of the opportunity cost of surprising.

Corollary 9. *Suppose $V_0(\sigma) > V_0^c(\sigma)$. Let $T(\sigma(\cdot|\mathcal{I}))$ be the set of distributions τ that satisfy $\mathbb{E}_\tau[q] = \sigma(\cdot|\mathcal{I})$ and $\mathbb{E}_\tau[\hat{v}_*^d(q)] = V_*^d(\sigma(\cdot|\mathcal{I}))$. If there exist $\hat{\tau} \in T(\sigma(\cdot|\mathcal{I}))$ and $q \in \mathbf{supp}\,\hat{\tau}$ such that $\hat{v}_*^d(q) < \hat{h}(q)$, then $\tau^d < \sigma(\mathcal{I})$.*

Example. In the running example of the paper, as Figure **??** illustrates, the only optimal way to disprove Receiver in both s_r and s_b involves fully revealing s_r with positive probability. However, Sender would rather conceal this state: $\hat{v}^d(\delta_{s_r}) = 2 < 6 = \hat{h}(\delta_{s_r})$. Therefore, her optimal experiment must involve some concealing.

B.9 Proof of Corollary 6

For $\kappa \geq \frac{1}{2}$, we already know that the lobbyist fully reveals every $\omega \leq \omega_{\overline{n}}$. For $i > \overline{n}$, since $\hat{v}^d$ is convex, by Corollary 2 we have that ω_i is concealed with probability 1 if $\hat{v}^d(\delta_{\omega_i}) < \hat{h}(\delta_{\omega_i})$ and is never concealed otherwise. So fix $i > \overline{n}$. For each value of κ there exists a value $b_i(\kappa)$ such that $\hat{v}^d(\delta_{\omega_i}) \geq \hat{h}(\delta_{\omega_i})$ if and only if $b \geq b_i(\kappa)$: This threshold is given by

$$b_i(\kappa) = \min\left\{\left(\frac{1}{2} - \kappa\right)\omega_i + \frac{1}{2}\omega_{\overline{n}}, 0\right\}.$$

Each $b_i(\kappa)$ is decreasing in κ (strictly when negative), and $b_{i+1}(\kappa) \leq b_i(\kappa)$ (with $<$ if either threshold is negative). If $b \geq b^*(\kappa) \equiv b_{\overline{n}+1}(\kappa)$, we have that $V^d(\sigma(\cdot|\mathcal{I})) = M(\sigma(\cdot|\mathcal{I}))$, and hence $\tau^d = \sigma(\mathcal{I})$ by Proposition 5. On the other hand, if $b < b^*(\kappa)$, let $i^*(b,\kappa) = \min\{i > \overline{n} : b \geq b_i(\kappa)\}$, which is non-increasing in both κ and b. Thus it is optimal to conceal with probability 1 every ω_i with $\overline{n} < i < i^*(b,\kappa)$ and to fully reveal all other states in $\mathcal{I}$.

B.10 Proof of Corollary 7

By Proposition 5, $M(\sigma(\cdot|\mathcal{I}))$ is given by

$$\max_{\gamma\in[0,1],q_1,q_2\in D_0} \left\{\gamma\mathbb{E}_{q_1}[h(\omega)] + (1-\gamma)\mathbb{E}_{q_2}[-(\mathbb{E}_{q_2}[\omega] - \beta(\omega))^2]\right\},$$

subject to $\gamma q_1 + (1-\gamma)q_2 = \sigma(\cdot|\mathcal{I})$. By continuity of $\hat{h}$ and $\hat{v}^d$, a solution (γ, q_1, q_2) to this problem exists. Recall that $\omega_i \notin \operatorname{supp} q_1$ for all $i < \underline{n}$. Suppose that (γ, q_1, q_2) implies $\omega_{\underline{n}} \leq \mathbb{E}_{q_2}[\omega] \leq \omega_{\overline{n}}$.

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 $\omega_j \in \mathbf{supp}\, q_2$ for some $j > \overline{n}$. Suppose first that $\beta(\omega) > \omega_{\underline{n}}$ for some $\omega > \omega_{\overline{n}}$. Then, for any $\xi \in [\omega_{\underline{n}}, \omega_{\overline{n}}]$,

$$-\mathbb{E}_{q_2}[(\xi-\omega)^2] < -\mathbb{E}_{q_2}[(\omega_{\underline{n}} - \beta(\omega))^2 \mathbf{1}\{\beta(\omega) < \omega_{\underline{n}}\}] - \mathbb{E}_{q_2}[(\omega_{\overline{n}} - \beta(\omega))^2 \mathbf{1}\{\beta(\omega) > \omega_{\overline{n}}\}] = \mathbb{E}_{q_2}[h(\omega)].$$

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

Finally, if σ is such that $\omega_{\underline{n}} \leq \mathbb{E}_{\sigma(\cdot|\mathcal{I})}[\omega] \leq \omega_{\overline{n}}$, then $\tau^d = \sigma(\mathcal{I})$ implies that $q^d = \sigma(\cdot|\mathcal{I})$, and hence $\omega_{\underline{n}} \leq \mathbb{E}_{q^d}[\omega] \leq \omega_{\overline{n}}$, which cannot be optimal as we have just argued. Moreover, concealing can occur only for $\omega > \omega_{\overline{n}}$ and never occurs for $\omega < \omega_{\underline{n}}$, so $\mathbb{E}_{q^d}[\omega]$ must be lower than $\omega_{\underline{n}}$. A fortiori, the same holds if $\mathbb{E}_{\sigma(\cdot|\mathcal{I})}[\omega] < \omega_{\underline{n}}$.

B.11 Proof of Proposition 6

Part I: $\lim_{\varepsilon \to 0} V_\varepsilon^c(\sigma) = V_0^c(\sigma)$. The proof consists of the following three claims.

Claim 8. There exists $\overline{\varepsilon} > 0$ such that, if $\varepsilon < \overline{\varepsilon}$, then $a(\mathbf{p}_\varepsilon(q)) \in \mathcal{A}_0$ for all $q \in C_\varepsilon$.

Proof. For every $p \in \Delta(\Omega)$, let $a(p) \in \arg\max_{a \in A} \mathbb{E}_p[u(a, \omega)]$. Also, define

$$\zeta = \max_{a \notin \mathcal{A}_0, \hat{a} \in \mathcal{A}_0} \left\{ \max_{\omega \in \mathcal{I}} u(a, \omega) - \min_{\omega \in \mathcal{I}} u(\hat{a}, \omega) \right\},$$

which is bounded by finiteness of $\mathcal{I}$ and A, and

$$\chi = \max_{p \in \Delta(\mathcal{P})} \left\{ \max_{\tilde{a} \notin \mathcal{A}_0} \mathbb{E}_p[u(\tilde{a}, \omega)] - \mathbb{E}_p[u(a(p), \omega)] \right\},$$

which is strictly negative because A is finite. Indeed, otherwise, there would be $p \in \Delta(\mathcal{P})$ and $\tilde{a} \notin \mathcal{A}_0$ such that $\mathbb{E}_p[u(\tilde{a}, \omega)] = \mathbb{E}_p[u(a(p), \omega)]$, which contradicts the premise that $\tilde{a} \notin \mathcal{A}_0$. Consider any $a' \notin \mathcal{A}_0$ and $p \in \mathbf{p}_\varepsilon(C_\varepsilon)$. We have

$$\begin{aligned}
\mathbb{E}_p[u(a', \omega)] &= \mathbb{E}_p[u(a(p(\cdot|\mathcal{P})), \omega)] \\
&\quad + p(\mathcal{P}) \left\{ \mathbb{E}_{p(\cdot|\mathcal{P})}[u(a', \omega)] - \mathbb{E}_{p(\cdot|\mathcal{P})}[u(a(p(\cdot|\mathcal{P})), \omega)] \right\} \\
&\quad + p(\mathcal{I}) \left\{ \mathbb{E}_{p(\cdot|\mathcal{I})}[u(a', \omega)] - \mathbb{E}_{p(\cdot|\mathcal{I})}[u(a(p(\cdot|\mathcal{P})), \omega)] \right\} \\
&\leq \mathbb{E}_p[u(a(p(\cdot|\mathcal{P})), \omega)] \\
&\quad + p(\mathcal{P}) \left\{ \mathbb{E}_{p(\cdot|\mathcal{P})}[u(a', \omega)] - \mathbb{E}_{p(\cdot|\mathcal{P})}[u(a(p(\cdot|\mathcal{P})), \omega)] \right\} \\
&\quad + p(\mathcal{I}) \left\{ \max_{\omega \in \mathcal{I}} u(a', \omega) - \min_{\omega \in \mathcal{I}} u(a(p(\cdot|\mathcal{P})), \omega) \right\} \\
&\leq \mathbb{E}_p[u(a(p(\cdot|\mathcal{P})), \omega)]
\end{aligned}$$

$$
\begin{aligned}
&+p(\mathcal{P})\left\{\mathbb{E}_{p(\cdot|\mathcal{P})}[u(a',\omega)]-\mathbb{E}_{p(\cdot|\mathcal{P})}[u(a(p(\cdot|\mathcal{P})),\omega)]\right\}+p(\mathcal{I})\zeta \\
\leq\; & \mathbb{E}_p[u(a(p(\cdot|\mathcal{P})),\omega)] \\
&+p(\mathcal{P})\left\{\max_{\tilde{a}\notin\mathcal{A}_0}\mathbb{E}_{p(\cdot|\mathcal{P})}[u(\tilde{a},\omega)]-\mathbb{E}_{p(\cdot|\mathcal{P})}[u(a(p(\cdot|\mathcal{P})),\omega)]\right\}+p(\mathcal{I})\zeta \\
\leq\; & \mathbb{E}_p[u(a(p(\cdot|\mathcal{P})),\omega)]+\chi+p(\mathcal{I})[\zeta-\chi],
\end{aligned}
$$

Now recall that $\sup_{p\in\mathbf{p}_\varepsilon(C_\varepsilon)} p(\mathcal{I})\to 0$ as $\varepsilon\to 0$ by Corollary 3. Thus, there exists $\varepsilon_{a'}>0$ so that if $\varepsilon<\varepsilon_{a'}$, then $\mathbb{E}_p[u(a',\omega)]<\mathbb{E}_p[u(a(p(\cdot|\mathcal{P})),\omega)]$ for every $p\in\mathbf{p}_\varepsilon(C_\varepsilon)$. Let $\overline{\varepsilon}=\min_{a'\in\mathcal{A}_0}\varepsilon_{a'}>0$. It follows that, if $\varepsilon<\overline{\varepsilon}$, then $\mathbb{E}_p[u(a',\omega)]<\mathbb{E}_p[u(a(p(\cdot|\mathcal{P})),\omega)]$ for every $a\notin\mathcal{A}_0$ and $p\in\mathbf{p}_\varepsilon(C_\varepsilon)$.

□

By A3, if $a\in\mathcal{A}_0$, there exists $p_a\in\Delta(\mathcal{P})$ such that $\mathbb{E}_{p_a}[u(a,\omega)]-\gamma\geq\mathbb{E}_{p_a}[u(a',\omega)]$ for some $\gamma>0$ and all other $a'\in\mathcal{A}_0$. Since $\sup_{p\in\mathbf{p}_\varepsilon(C_\varepsilon)} p(\mathcal{I})\to 0$ as $\varepsilon\to 0$, it follows that if $p\in\Delta(\Omega)$ is such that $p(\cdot|\mathcal{P})=p_a$, then $a=\arg\max_{a'\in A}\mathbb{E}_p[u(a',\omega)]$ for ε small enough. If we let a^ω be a best action in $\mathcal{A}_0$ for $\omega\in\mathcal{I}$, then any $p\in\Delta(\Omega)$ that satisfies $p(\cdot|\mathcal{P})=p_{a^\omega}$ guarantees that Receiver chooses a^ω provided that ε is smaller than some ε_{a^ω}. Since $\mathcal{I}$ is finite, if we let $\underline{\varepsilon}=\min_{\omega\in\mathcal{I}}\varepsilon_{a^\omega}$, then the previous conclusions holds for every $\omega\in\mathcal{I}$ provided that $\varepsilon<\underline{\varepsilon}$. Hereafter, $\varepsilon<\underline{\varepsilon}$.

Claim 9. $\liminf_{\varepsilon\to 0} V^c_\varepsilon(\sigma)\geq V^c_0(\sigma)$.

Proof. Consider any $\tau\in\mathcal{T}$ that satisfies properties (1)–(3) of Proposition 3 and $\mathbf{supp}\,\tau\subseteq C_\varepsilon$. Then, for every $q^\omega\in\mathbf{supp}\,\tau$ that conceals an $\omega\in\mathcal{I}$, we can write q^ω as

$$q^\omega=(1-q^\omega(\omega))p_{a^\omega}+q^\omega(\omega)\delta_\omega,$$

letting $q^\omega(\omega)=\frac{1}{1+\varepsilon}\eta_\varepsilon$ for all $\omega\in\mathcal{I}$, where η_ε is as in Corollary 4. It is also the case that $q^\omega(\omega)\tau(q^\omega)=\sigma(\omega)$. So, by the same argument as in Part IV of the proof of Proposition 3,

$$\mathbb{E}_\tau[\hat{v}_\varepsilon(q)]\leq\sigma(\mathcal{P})V^c_\varepsilon(\sigma(\cdot|\mathcal{P}))+\sum_{\omega\in\mathcal{I}}h(\omega)\sigma(\omega)=V^c_0(\sigma),$$

where the equality follows from observing $\mathbf{p}_\varepsilon$ is independent of ε for all $q\in\Delta(\mathcal{P})$: Indeed, for $q\in\Delta(\mathcal{P})$,

$$\mathbf{p}_\varepsilon(\omega;q)=\frac{\frac{q(\omega)}{\sigma(\omega)}\rho_0(\omega)}{\sum_{\omega'\in\mathcal{P}}\frac{q(\omega')}{\sigma(\omega')}\rho_0(\omega')}=\frac{\frac{q(\omega)}{\sigma(\omega|\mathcal{P})}\rho_0(\omega|\mathcal{P})}{\sum_{\omega'\in\mathcal{P}}\frac{q(\omega')}{\sigma(\omega'|\mathcal{P})}\rho_0(\omega'|\mathcal{P})},$$

which does not depend on ε by A1. This implies that

$$
\begin{aligned}
V^c_\varepsilon(\sigma(\cdot|\mathcal{P})) &= \sup_{\{\tau:\mathbb{E}_\tau[q]=\sigma(\cdot|\mathcal{P})\}}\mathbb{E}_\tau[\mathbb{E}_q[v(a(\mathbf{p}_\varepsilon(q)),\omega)]] \\
&= \sup_{\{\tau:\mathbb{E}_\tau[q]=\sigma(\cdot|\mathcal{P})\}}\mathbb{E}_\tau[\mathbb{E}_q[v(a(\mathbf{p}_0(q)),\omega)]]=V^c_0(\sigma(\cdot|\mathcal{P})).
\end{aligned}
$$

Now note that by the same observation

$$\mathbb{E}_\tau[\hat{v}^c_\varepsilon(q)]=\sigma(\mathcal{P})\mathbb{E}_{\tau_{\Delta(\mathcal{P})}}[\hat{v}^c_0(q)]+\sum_{\omega\in\mathcal{I}}h(\omega)\sigma(\omega),$$

where $\tau_{\Delta(\mathcal{P})}$ is the restriction of τ to $\Delta(\mathcal{P})$ which belongs to the set $\{\tau : \mathbb{E}_\tau[q] = \sigma(\cdot|\mathcal{P})\}$. Thus, if we construct τ so that $\mathbb{E}_{\tau_{\Delta(\mathcal{P})}}[\hat{v}_0^c(q)] \to V_0^c(\sigma(\cdot|\mathcal{P}))$ as in Part IV of the proof of Proposition 3, we get that $\mathbb{E}_\tau[\hat{v}_\varepsilon^c(q)] \to V_0^c(\sigma)$ from below.

We conclude that, for every sequence $\{\varepsilon_n\}$ converging to 0, we can construct a sequence $\{\tau_{\varepsilon_n}\}$ such that each $\tau_{\varepsilon_n} \in \mathcal{T}$ and satisfies $\mathbf{supp}\,\tau_{\varepsilon_n} \subseteq C_{\varepsilon_n}$, $\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \le V_{\varepsilon_n}^c(\sigma)$ (by definition), $\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \le V_0^c(\sigma)$ (by construction), and $\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \to V_0^c(\sigma)$. The claim follows.

□

Claim 10. $\limsup_{\varepsilon\to 0} V_\varepsilon^c(\sigma) \le V_0^c(\sigma)$.

Proof. The proof is by contradiction. Suppose $\limsup_{\varepsilon\to 0} V_\varepsilon^c(\sigma) = \overline{V} > V_0^c(\sigma)$. Fix $\gamma > 0$ so that $\overline{V} - \gamma > V_0^c(\sigma)$. There must exists a sequence $\{\varepsilon_n\}$ with $\varepsilon_n \to 0$ and a corresponding sequence $\{\tau_{\varepsilon_n}\}$ with each $\tau_{\varepsilon_n} \in \mathcal{T}$ satisfying $\mathbf{supp}\,\tau_{\varepsilon_n} \subseteq C_{\varepsilon_n}$ such that $\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \in [\overline{V} - \gamma, \overline{V} + \gamma]$ for all $n \ge N$ sufficiently large.

Note that without loss of generality we can assume that, for every τ_{ε_n} in the sequence, there must exists $q \in \mathbf{supp}\,\tau_{\varepsilon_n}$ such that $a(\mathbf{p}_{\varepsilon_n}(q)) \notin \mathcal{A}(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})))$—otherwise, we would have $\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \le V_0^c(\sigma)$. To see this, fix ε_n and note that

$$\begin{aligned}
\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] &= \sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} \left[\sum_{\omega\in\mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega) + \sum_{\omega\in\mathcal{I}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega)\right]\tau_{\varepsilon_n}(q) \\
&\le \sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} \left[\sum_{\omega\in\mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega) + \sum_{\omega\in\mathcal{I}} h(\omega)q(\omega)\right]\tau_{\varepsilon_n}(q) \\
&= \sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} \left[\sum_{\omega\in\mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega)\right]\tau_{\varepsilon_n}(q) + \sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} \left[\sum_{\omega\in\mathcal{I}} h(\omega)q(\omega)\right]\tau_{\varepsilon_n}(q) \\
&= \sigma(\mathcal{P}) \sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} \left[\sum_{\omega\in\mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega|\mathcal{P})\right]\frac{q(\mathcal{P})\tau_{\varepsilon_n}(q)}{\sigma(\mathcal{P})} + \sum_{\omega\in\mathcal{I}} h(\omega)\sigma(\omega),
\end{aligned}$$

where we used that $\sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} q(\omega)\tau(q) = \sigma(\omega)$ for every $\omega \in \Omega$. If we define $\tau'(q(\cdot|\mathcal{P})) = \frac{q(\mathcal{P})\tau_{\varepsilon_n}(q)}{\sigma(\mathcal{P})}$ for every $q(\cdot|\mathcal{P})$ corresponding to a $q \in \mathbf{supp}\,\tau_{\varepsilon_n}$, then τ' defines a distribution over $\Delta(\mathcal{P})$ with

$$\mathbb{E}_{\tau'}[q(\omega)] = \frac{1}{\sigma(\mathcal{P})} \sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}} q(\omega)\tau_{\varepsilon_n}(q) = \sigma(\omega|\mathcal{P}), \quad \omega \in \mathcal{P}.$$

Recall that for ε small enough, $a(\mathbf{p}_{\varepsilon_n}(q)) \in \mathcal{A}_0$ for all $q \in \mathbf{supp}\,\tau$. Hence, if $a(\mathbf{p}_{\varepsilon_n}(q)) \in \mathcal{A}(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})))$ for all $q \in \mathbf{supp}\,\tau'$, we would have

$$\begin{aligned}
\sum_{q\in\mathbf{supp}\,\tau'} \left[\sum_{\omega\in\mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega|\mathcal{P})\right]\tau'(q(\cdot|\mathcal{P})) &\le \sum_{q\in\mathbf{supp}\,\tau'} \hat{v}_{\varepsilon_n}^c(q(\cdot|\mathcal{P}))\tau'(q(\cdot|\mathcal{P})) \\
&\le V_{\varepsilon_n}^c(\sigma(\cdot|\mathcal{P})) = V_0^c(\sigma(\cdot|\mathcal{P})).
\end{aligned}$$

We will proceed by constructing for every τ_{ε_n} as 'twin' distribution $\hat{\tau}_{\varepsilon_n}$ and by showing that both $\lim_{n\to\infty} \left|\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] - \mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)]\right| = 0$ and $\lim_{n\to\infty} \mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \le V_0^c(\sigma)$, which delivers the desired contradiction.

Focussing on q's such that $a(\mathbf{p}_{\varepsilon_n}(q)) \notin \mathcal{A}(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})))$, consider the following observations. By optimality $a(\mathbf{p}_{\varepsilon_n}(q))$ of for Receiver, $\frac{\mathbf{p}_{\varepsilon_n}(\mathcal{I};q)}{\mathbf{p}_{\varepsilon_n}(\mathcal{P};q)}\mathbb{E}_{\mathbf{p}_{\varepsilon_n}(\cdot|\mathcal{I};q)}[u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)-u(a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))),\omega)]$ must exceed $\mathbb{E}_{\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))}[u(a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))),\omega)-u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)]$, where we used

$$\mathbf{p}_{\varepsilon_n}(\omega|\mathcal{P};q) = \frac{\frac{q(\omega)}{\sigma(\omega)}\rho_o(\omega)}{\sum_{\omega'\in\mathcal{P}}\frac{q(\omega')}{\sigma(\omega')}\rho_o(\omega')} = \frac{\frac{q(\omega|\mathcal{P})}{\sigma(\omega)}\rho_o(\omega)}{\sum_{\omega'\in\mathcal{P}}\frac{q(\omega'|\mathcal{P})}{\sigma(\omega')}\rho_o(\omega')} = \mathbf{p}_{\varepsilon_n}(\omega;q(\cdot|\mathcal{P})).$$

Letting $M=\max_{\omega\in\mathcal{I}}[\max_{a\in A}u(a,\omega)-\min_{a\in A}u(a,\omega)]>0$, which is bounded, it follows that

$$B_{\varepsilon_n} = M \sup_{p\in\mathbf{p}_{\varepsilon_n}(C_{\varepsilon_n})}\frac{p(\mathcal{I})}{1-p(\mathcal{I})} \geq \mathbb{E}_{\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))}[u(a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})),\omega)-u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)]>0.$$

Recall that $\sup_{p\in\mathbf{p}_{\varepsilon_n}(C_{\varepsilon_n})}\frac{p(\mathcal{I})}{1-p(\mathcal{I})}\to 0$ as $\varepsilon_n\to 0$. Therefore, for every ε_n and $q\in C_{\varepsilon_n}$, there is a uniform upper abound $B_{\varepsilon_n}>0$ in the payoff difference for Receiver between $a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})))$ and $a(\mathbf{p}_{\varepsilon_n}(q))$, which both belong to $\mathcal{A}_0$, and $B_{\varepsilon_n}\to 0$ as $\varepsilon_n\to 0$.

Next, for every $a\in\mathcal{A}_0$, let $\Delta^a(\mathcal{P})=\{p\in\Delta(\mathcal{P}): a\in\arg\max_{a\in A}\mathbb{E}_p[u(a,\omega)]\}$. Clearly, $\Delta^a(\mathcal{P})$ is closed and convex and $int\Delta^a(\mathcal{P})\neq\varnothing$ for every $a\in\mathcal{A}_0$ by A3. Therefore, if for every $a\in\mathcal{A}_0$ we fix $p_a\in int\Delta^a(\mathcal{P})$ so that $\min_{a'\in A}\mathbb{E}_{p_a}[u(a,\omega)-u(a',\omega)]>0$ and define

$$p^\beta(a(\mathbf{p}_{\varepsilon_n}(q))) = \beta\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))+(1-\beta)p_{a(\mathbf{p}_{\varepsilon_n}(q))},$$

then there exists a $\tilde{\beta}_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))<1$ such that $\mathbb{E}_{p^\beta(a(\mathbf{p}_{\varepsilon_n}(q)))}[u(a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})),\omega)-u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)]\geq 0$ if and only if $\beta\geq\tilde{\beta}_{\varepsilon_n}$ and with equality if and only if $\beta=\tilde{\beta}_{\varepsilon_n}$. Note that, by definition,

$$\begin{aligned}\frac{\tilde{\beta}_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))}{1-\tilde{\beta}_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))} &= \frac{\mathbb{E}_{p_{a(\mathbf{p}_{\varepsilon_n}(q))}}[u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)-u(a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})),\omega)]}{\mathbb{E}_{\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))}[u(a(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})),\omega)-u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)]}.\\ &\geq \frac{\min_{a\in A}\mathbb{E}_{p_{a(\mathbf{p}_{\varepsilon_n}(q))}}[u(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)-u(a,\omega)]}{B_{\varepsilon_n}} \geq \frac{K}{B_{\varepsilon_n}}\end{aligned}$$

where $K=\min_{a'\in\mathcal{A}_0}\left\{\min_{a\in A}\mathbb{E}_{p_{a'}}[u(a',\omega)-u(a,\omega)]\right\}>0$. This implies that

$$\tilde{\beta}_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))\geq\frac{K}{K+B_{\varepsilon_n}}.$$

Now let $\beta_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))=\frac{K}{(1+\varepsilon_n)(K+B_{\varepsilon_n})}$. By construction,

$$\begin{aligned}||\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))-p^{\beta_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))}(a(\mathbf{p}_{\varepsilon_n}(q)))|| &= (1-\beta_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))||\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P}))-p_{a(\mathbf{p}_{\varepsilon_n}(q))}||\\ &\leq \left[1-\frac{K}{(1+\varepsilon_n)(K+B_{\varepsilon_n})}\right]|\Omega|. \qquad (20)\end{aligned}$$

□

Since $\mathbf{p}_{\varepsilon_n}$ is an homeomorphism independent of ε_n when restricted to $\Delta(\mathcal{P})$, there exists a unique $q'(a(\mathbf{p}_{\varepsilon_n}(q)))\in\Delta(\mathcal{P})$ such that $\mathbf{p}_{\varepsilon_n}(q'(a(\mathbf{p}_{\varepsilon_n}(q))))=p^{\beta_{\varepsilon_n}(a(\mathbf{p}_{\varepsilon_n}(q)))}(a(\mathbf{p}_{\varepsilon_n}(q)))$. Using $q'(a(\mathbf{p}_{\varepsilon_n}(q)))$, it is possible to construct $\hat{q}(a(\mathbf{p}_{\varepsilon_n}(q))) \in \Delta(\Omega)$ in the form $\hat{q}(a(\mathbf{p}_{\varepsilon_n}(q))) = q(\mathcal{P})q'(a(\mathbf{p}_{\varepsilon_n}(q))) + (1 - q(\mathcal{P}))q(\cdot|\mathcal{I})$. Now, note that $\mathbf{p}_{\varepsilon_n}(\cdot|\mathcal{I}; \hat{q}(a(\mathbf{p}_{\varepsilon_n}(q)))) = \mathbf{p}_{\varepsilon_n}(\cdot|\mathcal{I}; q)$ by A1 and

$$\frac{\mathbf{p}_{\varepsilon_n}(\mathcal{P};\hat{q})}{\mathbf{p}_{\varepsilon_n}(\mathcal{I};\hat{q})} = \frac{\mathbf{p}_{\varepsilon_n}(\mathcal{P};q)}{\mathbf{p}_{\varepsilon_n}(\mathcal{I};q)}.$$

It follows that $a(\mathbf{p}_{\varepsilon_n}(q)) = \arg\max_{a\in A} \mathbb{E}_{\mathbf{p}_{\varepsilon_n}(\hat{q})}[u(a,\omega)]$ and $\hat{q}(a(\mathbf{p}_{\varepsilon_n}(q))) \in C_{\varepsilon_n}$. This is because Receiver's expected payoff conditional on $\mathcal{I}$ is the same for $a(\mathbf{p}_{\varepsilon_n}(q))$ and every other strictly worse action under both $\mathbf{p}_{\varepsilon_n}(q)$ and $\mathbf{p}_{\varepsilon_n}(\hat{q})$, his expected payoff conditional on $\mathcal{P}$ strictly favors $a(\mathbf{p}_{\varepsilon_n}(q))$ to every other action under $\mathbf{p}_{\varepsilon_n}(\hat{q})$ compared to under $\mathbf{p}_{\varepsilon_n}(q)$, and the relative likelihood of $\mathcal{P}$ and $\mathcal{I}$ is the same under both beliefs.

We can now construct a twin distribution $\hat{\tau}_{\varepsilon_n}$ for every τ_{ε_n} in the sequence as follows. For every $q \in \mathbf{supp}\,\tau_{\varepsilon_n}$ such that $a(\mathbf{p}_{\varepsilon_n}(q)) \notin \mathcal{A}(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})))$, construct $\hat{q}(a(\mathbf{p}_{\varepsilon_n}(q)))$ as before and assign to it probability $\hat{\tau}_{\varepsilon_n} = \tau_{\varepsilon_n}(q)$; if $a(\mathbf{p}_{\varepsilon_n}(q)) \in \mathcal{A}(\mathbf{p}_{\varepsilon_n}(q(\cdot|\mathcal{P})))$, simply let $\hat{q}(a(\mathbf{p}_{\varepsilon_n}(q))) = q$ and assign to it probability $\hat{\tau}_{\varepsilon_n} = \tau_{\varepsilon_n}(q)$. In general, $\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}] \neq \sigma$. As we will show shortly, however, $\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}] \to \sigma$ as $n \to \infty$. Thus, since $\sigma \in int\Delta(\Omega)$ and $\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}] \in \Delta(\Omega)$ for every ε_n, we can always augment each $\hat{\tau}_{\varepsilon_n}$ by adding a realization $\tilde{q} \in \Delta(\Omega)$ so that $\tilde{q}\hat{\tau}_{\varepsilon_n}(\tilde{q}) + (1-\hat{\tau}_{\varepsilon_n}(\tilde{q}))\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}] = \sigma$ and choose $\tilde{q}$ so that $\hat{\tau}_{\varepsilon_n}(\tilde{q}) \to 0$ as $n \to \infty$.

Now, $\left|\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}^c_{\varepsilon_n}(q)] - \mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{v}^c_{\varepsilon_n}(q)]\right|$ is bounded above by

$$\begin{aligned}
&\left|\sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\left[\sum_{\omega\in\mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)[q(\omega|\mathcal{P}) - (1-\hat{\tau}_{\varepsilon_n}(\tilde{q}))q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))]\right] q(\mathcal{P})\tau_{\varepsilon_n}(q)\right| \\
&\quad + \hat{\tau}_{\varepsilon_n}(\tilde{q})\left|\sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\left[\sum_{\omega\in\mathcal{I}} v(a(\mathbf{p}_{\varepsilon_n}(q)),\omega)q(\omega|\mathcal{I})\right] q(\mathcal{I})\tau_{\varepsilon_n}(\hat{q}) - \sum_{\omega\in\Omega} v(a(\mathbf{p}_{\varepsilon_n}(\tilde{q})),\omega)\tilde{q}(\omega)\right| \\
\leq &\sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\left[M'\sum_{\omega\in\mathcal{P}}\left|q(\omega|\mathcal{P}) - (1-\hat{\tau}_{\varepsilon_n}(\tilde{q}))q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))\right|\right] q(\mathcal{P})\tau_{\varepsilon_n}(q) + \hat{\tau}_{\varepsilon_n}(\tilde{q})M'' \\
\leq &\overline{M}\left\{\sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\sum_{\omega\in\mathcal{P}}\left[\left|q(\omega|\mathcal{P}) - q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))\right| + \hat{\tau}_{\varepsilon_n}(\tilde{q})q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))\right] q(\mathcal{P})\tau_{\varepsilon_n}(q) + \hat{\tau}_{\varepsilon_n}(\tilde{q})\right\} \\
\leq &\overline{M}\left\{\sum_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\left[\max_{\omega\in\mathcal{P}}\left|q(\omega|\mathcal{P}) - q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))\right| q(\mathcal{P})\tau_{\varepsilon_n}(q)\right] + \hat{\tau}_{\varepsilon_n}(\tilde{q})[\sigma(\mathcal{P})+1]\right\} \\
\leq &\overline{M}\left\{\max_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\left[\max_{\omega\in\mathcal{P}}\left|q(\omega|\mathcal{P}) - q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))\right|\right] + \hat{\tau}_{\varepsilon_n}(\tilde{q})[\sigma(\mathcal{P})+1]\right\},
\end{aligned}$$

where $\overline{M} = \max\{M', M''\}$ is bounded. Given this, consider the program

$$\Phi_{\varepsilon_n} = \max_{q\in\mathbf{supp}\,\tau_{\varepsilon_n}}\max_{\omega\in\mathcal{P}}\left|q(\omega|\mathcal{P}) - q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))\right|$$

subject to

$$||\mathbf{p}_0(q(\cdot|\mathcal{P})) - \mathbf{p}_0(q'(a(\mathbf{p}_{\varepsilon_n}(q)))|| \leq \left[1 - \frac{K}{(1+\varepsilon_n)(K+B_{\varepsilon_n})}\right]|\Omega|,$$

where we replaced $\mathbf{p}_{\varepsilon_n}$ with $\mathbf{p}_0$ because $q'(a(\mathbf{p}_{\varepsilon_n}(q))) \in \Delta(\mathcal{P})$ for all considered q's. For each ε_n, we have $\Phi_{\varepsilon_n} \leq \hat{\Phi}_{\varepsilon_n}$, where

$$\hat{\Phi}_{\varepsilon_n} = \max_{q,q' \in \Delta(\mathcal{P})} \max_{\omega \in \mathcal{P}} \left| q(\omega) - q'(\omega) \right|$$

subject to

$$||\mathbf{p}_0(q) - \mathbf{p}_0(q')|| \leq \left[1 - \frac{K}{(1+\varepsilon_n)(K + B_{\varepsilon_n})} \right] |\Omega|.$$

By continuity of $\mathbf{p}_0$ and (20), it follows that $\hat{\Phi}_{\varepsilon_n} \to 0$ as $\varepsilon_n \to 0$ and hence $\Phi_{\varepsilon_n} \to 0$.

Finally, note that

$$\begin{aligned}
\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{v}^c_{\varepsilon_n}(q)] &= (1-\hat{\tau}_{\varepsilon_n}(\tilde{q})) \sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} \left[\sum_{\omega \in \mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(\hat{q})), \omega) \hat{q}(\omega|\mathcal{P}) \right] q(\mathcal{P}) \tau_{\varepsilon_n}(\hat{q}) \\
&\quad + (1-\hat{\tau}_{\varepsilon_n}(\tilde{q})) \sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} \left[\sum_{\omega \in \mathcal{I}} v(a(\mathbf{p}_{\varepsilon_n}(\hat{q})), \omega) \hat{q}(\omega|\mathcal{I}) \right] q(\mathcal{I}) \tau_{\varepsilon_n}(\hat{q}) \\
&\quad + \hat{\tau}_{\varepsilon_n}(\tilde{q}) \sum_{\omega \in \Omega} v(a(\mathbf{p}_{\varepsilon_n}(\tilde{q})), \omega) \tilde{q}(\omega) \\
&\leq (1-\hat{\tau}_{\varepsilon_n}(\tilde{q})) \sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} \left[\sum_{\omega \in \mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(\hat{q})), \omega) \hat{q}(\omega|\mathcal{P}) \right] q(\mathcal{P}) \tau_{\varepsilon_n}(\hat{q}) \\
&\quad + (1-\hat{\tau}_{\varepsilon_n}(\tilde{q})) \sum_{\omega \in \mathcal{I}} h(\omega) \sigma(\omega) \\
&\quad + \hat{\tau}_{\varepsilon_n}(\tilde{q}) \sum_{\omega \in \Omega} v(a(\mathbf{p}_{\varepsilon_n}(\tilde{q})), \omega) \tilde{q}(\omega).
\end{aligned}$$

Thus, in order to show that $\lim_{n\to\infty} \mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{v}^c_{\varepsilon_n}(q)] \leq V^c_0(\sigma)$, it remains to argue that

$$\sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} \left[\sum_{\omega \in \mathcal{P}} v(a(\mathbf{p}_{\varepsilon_n}(q)), \omega) q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q))) \right] (1-\hat{\tau}_{\varepsilon_n}(\tilde{q})) q(\mathcal{P}) \tau_{\varepsilon_n}(q) \to V \leq V^c_0(\sigma(\cdot|\mathcal{P})).$$

Note that $\sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} (1-\hat{\tau}_{\varepsilon_n}(\tilde{q})) q(\mathcal{P}) \tau_{\varepsilon_n}(q) = (1-\hat{\tau}_{\varepsilon_n}(\tilde{q}))\sigma(\mathcal{P})$ and therefore, if we define τ'_{ε_n} by $\tau'_{\varepsilon_n}(q) = \frac{(1-\hat{\tau}_{\varepsilon_n}(\tilde{q}))q(\mathcal{P})\tau_{\varepsilon_n}(q)}{(1-\hat{\tau}_{\varepsilon_n}(\tilde{q}))\sigma(\mathcal{P})}$, we have that τ'_{ε_n} is a distribution over $\Delta(\mathcal{P})$. Define $\mu_{\varepsilon_n} = \mathbb{E}_{\tau'_{\varepsilon_n}}[q]$. Then, recalling that $a(\mathbf{p}_{\varepsilon_n}(q)) = a(\mathbf{p}_{\varepsilon_n}(q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))))$, we have

$$\sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} \hat{v}^c_0(q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))) \frac{q(\mathcal{P})\tau_{\varepsilon_n}(q)}{\sigma(\mathcal{P})} \leq V^c_0(\mu_{\varepsilon_n}).$$

If we prove that $\mu_{\varepsilon_n} \to \sigma(\cdot|\mathcal{P})$, we are done.[33] For every $\omega \in \mathcal{P}$, we have

$$|\mathbb{E}_{\tau'_{\varepsilon_n}}[q'(\omega)] - \sigma(\omega|\mathcal{P})| = \left| \sum_{q \in \mathbf{supp}\, \tau_{\varepsilon_n}} [q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q))) - q(\omega|\mathcal{P})] \frac{q(\mathcal{P})\tau_{\varepsilon_n}(q)}{\sigma(\mathcal{P})} \right| \leq \Phi_{\varepsilon_n},$$

which converges to zero as $\varepsilon_n \to 0$ as we saw before.

[33] Continuity of V^c_0 over $\Delta(\mathcal{P})$ can be established along the lines of Lemma 4 in Appendix B.

To conclude, we show that $\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}] \to \sigma$ as $n \to \infty$. Note that, for every $\omega \in \Omega$,

$$\begin{aligned}
\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}(\omega)] &= \sum_{q \in \mathbf{supp}\,\tau_{\varepsilon_n}} [q(\mathcal{P})q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q))) + q(\mathcal{I})q(\omega|\mathcal{I})]\tau_{\varepsilon_n}(q) \\
&= \sum_{q \in \mathbf{supp}\,\tau_{\varepsilon_n}} q'(\omega; a(\mathbf{p}_{\varepsilon_n}(q)))q(\mathcal{P})\tau_{\varepsilon_n}(q) + \sum_{q \in \mathbf{supp}\,\tau_{\varepsilon_n}} q(\omega|\mathcal{I})q(\mathcal{I})\tau_{\varepsilon_n}(q) \\
&= \sigma(\mathcal{P})\mu_{\varepsilon_n}(\omega) + \sigma(\mathcal{I})\sigma(\omega|\mathcal{I}).
\end{aligned}$$

Since we established that $\mu_{\varepsilon_n}(\omega) \to \sigma(\omega|\mathcal{P})$ for every $\omega \in \mathcal{P}$, it follows that $\mathbb{E}_{\hat{\tau}_{\varepsilon_n}}[\hat{q}(\omega)] \to \sigma(\omega)$ for every $\omega \in \Omega$.

Part II: $\lim_{\varepsilon \to 0} V_\varepsilon(\sigma) = V_0(\sigma)$**.**

Claim 11. $\liminf_{\varepsilon \to 0} V_\varepsilon(\sigma) \geq V_0(\sigma)$.

Proof. By Claim 5, we can write $V(\sigma) = \tau^c V^c(q^c) + (1-\tau^c)V^d(q^d)$, where $q^c \in C$, $q^d \in D$, $\sigma(\mathcal{P}) \leq \tau^c \leq 1$, and $\sigma = \tau^c q^c + (1-\tau^c)q^d$. Also, since $\tau^c \in [\sigma(\mathcal{P}), 1]$, we must have $q^c(\omega) = \frac{1}{\tau^c}\sigma(\omega) \geq \sigma(\omega)$ for all $\omega \in \mathcal{P}$. This implies that for every ε,

$$\mathbf{p}_\varepsilon(\mathcal{P}; q^c) = \sum_{\omega \in \mathcal{P}} \mathbf{p}_\varepsilon(\omega; q^c) = \frac{\sum_{\omega \in \mathcal{P}} \frac{q^c(\omega)}{\sigma(\omega)}\rho_0(\omega)}{\sum_{\omega' \in \Omega} \frac{q^c(\omega')}{\sigma(\omega')}\rho_0(\omega')} \geq \frac{\rho_0(\mathcal{P})}{\sum_{\omega' \in \Omega} \frac{q^c(\omega')}{\sigma(\omega')}\rho_0(\omega')} \geq \rho_0(\mathcal{P}),$$

because $\sum_{\omega' \in \Omega} \frac{q^c(\omega')}{\sigma(\omega')}\rho_0(\omega') \leq 1$. It follows that $\mathbf{p}_\varepsilon(\mathcal{I}; q^c) \leq \rho_0(\mathcal{I})$ and hence

$$\frac{\mathbf{p}_\varepsilon(\mathcal{P}; q^c)}{\mathbf{p}_\varepsilon(\mathcal{I}; q^c)} = \ell(q^c) \geq \frac{1-\varepsilon}{\varepsilon} > \theta_\varepsilon.$$

We conclude that $q^c \in C_\varepsilon$ for all ε.

Therefore, for the distribution $\tau \in \mathcal{T}$ defined by τ^c, q^c, and q^d, for all ε we have that

$$\mathbb{E}_\tau[\hat{v}_\varepsilon(q)] = \tau^c \hat{v}^c(q^c) + (1-\tau^c)\hat{v}^d(q^d) \leq \tau^c V_\varepsilon^c(q^c) + (1-\tau^c)V_\varepsilon^d(q^d) \leq V_\varepsilon(\sigma).$$

Since $q^d \in D_0 \subseteq D_\varepsilon$ for all ε, $V_\varepsilon^d(q^d) = V_0^d(q^d)$ for all ε. Now consider $V_\varepsilon^c(q^c)$. If $q^c \in \Delta(\mathcal{P})$, then $q^c = \sigma(\cdot|\mathcal{P})$ and hence $V_\varepsilon^c(q^c) = V_\varepsilon^c(\sigma(\cdot|\mathcal{P})) = V_0^c(\sigma(\cdot|\mathcal{P}))$. If instead $q^c \notin \Delta(\mathcal{P})$, then $\mathbf{supp}\, q^c \cap \mathcal{I} \neq \varnothing$. Letting $\Omega' = \mathbf{supp}\, q^c$, $\sigma' = q^c$, and $\rho_0' = \rho_0(\cdot|\Omega')$, we obtain a fictitious model equivalent to the main model, except for being characterized by $\varepsilon'(\varepsilon) < \varepsilon$ for every ε with $\varepsilon'(\varepsilon) \to 0$ as $\varepsilon \to 0$. Therefore, $V_\varepsilon^c(q^c) = V_{\varepsilon'(\varepsilon)}^c(q^c)$ for all ε. Since $V_{\varepsilon'(\varepsilon)}^c(q^c)$ must satisfy $\lim_{\varepsilon \to 0} V_{\varepsilon'(\varepsilon)}^c(q^c) = V_0^c(q^c)$ in the fictitious model with Sender's prior q^c, we have that $\lim_{\varepsilon \to 0} V_\varepsilon^c(q^c) = V_0^c(q^c)$.

Given this, we have

$$\begin{aligned}
\liminf_{\varepsilon \to 0} V_\varepsilon(\sigma) &\geq \liminf_{\varepsilon \to 0} [\tau^c V_\varepsilon^c(q^c) + (1-\tau^c)V_\varepsilon^d(q^d)] \\
&= \tau^c \lim_{\varepsilon \to 0} V_\varepsilon^c(q^c) + (1-\tau^c)V_0^d(q^d) = V_0(\sigma).
\end{aligned}$$

□

Claim 12. $\limsup_{\varepsilon\to 0} V_\varepsilon(\sigma) \le V_0(\sigma)$.

Proof. The proof is by contradiction. Suppose that $\limsup_{\varepsilon\to 0} V_\varepsilon(\sigma) = \overline{V} > V_0(\sigma)$. Fix $\gamma > 0$ so that $\overline{V} - \gamma > V_0^c(\sigma)$. There must exists a sequence $\{\varepsilon_n\}$ with $\varepsilon_n \to 0$ and a corresponding sequence $\{\tau_{\varepsilon_n}\}$ with each $\tau_{\varepsilon_n} \in \mathcal{T}^{|\Omega|}$ satisfying $\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}(q)] \in [\overline{V} - \gamma, \overline{V} + \gamma]$ for all $n \ge N$ sufficiently large, where $\mathcal{T}^{|\Omega|} = \{\tau \in \mathcal{T} : |\mathbf{supp}\, \tau_{\varepsilon_n}| = |\Omega|\}$ (see Lemma 2). Since $\lim_{n\to\infty} V_{\varepsilon_n}^c(\sigma) = V_0^c(\sigma) \le V_0(\sigma)$. There exists N' such that $V_{\varepsilon_n}^c(\sigma) < \mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}(q)] \le V_{\varepsilon_n}(\sigma)$ for all $n \ge N'$. Therefore, for all $n \ge N'$, τ_{ε_n} involves disproving ρ_0 and $\sup_{n\ge N'} \tau_{\varepsilon_n}^c = \overline{\tau}^c < 1$.

For every ε_n we can write

$$\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}(q)] = \tau_{\varepsilon_n}^c \mathbb{E}_{\tau_{\varepsilon_n}(\cdot|C^{\tau_{\varepsilon_n}})}[\hat{v}_{\varepsilon_n}(q)] + (1 - \tau_{\varepsilon_n}^c)\mathbb{E}_{\tau_{\varepsilon_n}(\cdot|D^{\tau_{\varepsilon_n}})}[\hat{v}_{\varepsilon_n}(q)],$$

where $C^{\tau_{\varepsilon_n}} = C_{\varepsilon_n} \cap \mathbf{supp}\, \tau_{\varepsilon_n}$ and $D^{\tau_{\varepsilon_n}} = D_{\varepsilon_n} \cap \mathbf{supp}\, \tau_{\varepsilon_n}$, which is non-empty for all $n \ge N'$. Let $q_{\varepsilon_n}^c = \mathbb{E}_{\tau_{\varepsilon_n}(\cdot|C^{\tau_{\varepsilon_n}})}[q]$ and $q_{\varepsilon_n}^d = \mathbb{E}_{\tau_{\varepsilon_n}(\cdot|D^{\tau_{\varepsilon_n}})}[q]$. Therefore, for all ε_n, we have

$$\mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}(q)] \le \tau_{\varepsilon_n}^c \mathbb{E}_{\tau_{\varepsilon_n}(\cdot|C^{\tau_{\varepsilon_n}})}[\hat{v}_{\varepsilon_n}^c(q)] + (1 - \tau_{\varepsilon_n}^c) V_{\varepsilon_n}^d(q_{\varepsilon_n}^d),$$

where

$$V_{\varepsilon_n}^d(q_{\varepsilon_n}^d) = \max_{\{\tau:\mathbb{E}_\tau[q]=q_{\varepsilon_n}^d, \mathbf{supp}\, \tau \subseteq D_{\varepsilon_n}\}} \mathbb{E}_\tau[\hat{v}_{\varepsilon_n}^d(q)].$$

Note that in the last expression $\hat{v}_{\varepsilon_n}^d(\cdot)$ is independent of ε_n (since Receiver always updates ρ^1 for all $q \in D_{\varepsilon_n}$ and all ε_n) and u.s.c. (which explains the "max"). In particular, for every ε_n, we can replace $\hat{v}_{\varepsilon_n}^d$ with the u.s.c. function $\hat{w}$ defined by

$$\hat{w}(q) = \max_{a \in \mathcal{A}(\mathbf{p}^*(q))} \mathbb{E}_q[v(a, \omega)], \quad q \in \Delta(\Omega),$$

where $\mathbf{p}^*$ is the function in Proposition 1 with priors σ and ρ^1.

Since $\mathcal{T}^{|\Omega|}$ is compact, $\{\tau_{\varepsilon_n}\}$ must have a converging subsequence. To avoid complicating notation, we will continue to use $\{\tau_{\varepsilon_n}\}$ for the subsequence. Denote its limit by τ_∞ with the corresponding objects τ_∞^c, q_∞^c, and q_∞^d. Note that $\tau_\infty^c \le \overline{\tau}^c < 1$, $q_\infty^d \in D_0$, and $q_\infty^c \in C_0$. The latter property holds because, if $q_\infty^c \in D_0$, then $\tau_\infty^c q_\infty^c + (1 - \tau_\infty^c) q_\infty^d \ne \sigma$.

Consider the sequence of values $V_{\varepsilon_n}^d(q_{\varepsilon_n}^d)$. For every n,

$$V_{\varepsilon_n}^d(q_{\varepsilon_n}^d) \le \max_{\{\tau:\mathbb{E}_\tau[q]=q_{\varepsilon_n}^d, \mathbf{supp}\, \tau \subseteq \Delta(\Omega)\}} \mathbb{E}_\tau[\hat{w}(q)] = W(q_{\varepsilon_n}^d).$$

By an argument similar to that establishing Lemma 4 in Appendix B, one can conclude that W is continuous. Thus, $\lim_{n\to\infty} W(q_{\varepsilon_n}^d) = W(q_\infty^d)$. But since $q_\infty^d \in D_0$, we have $W(q_\infty^d) = V_0^d(q_\infty^d)$.

Now consider the sequence of values $\mathbb{E}_{\tau_{\varepsilon_n}(\cdot|C^{\tau_{\varepsilon_n}})}[\hat{v}_{\varepsilon_n}^c(q)]$. Using an argument similar to that establishing Claim 10, we can construct a 'twin' sequence $\{\tau'_{\varepsilon_n}\}$, where $\mathbf{supp}\, \tau'_{\varepsilon_n} \subseteq C_{\varepsilon_n}$ for all n, which satisfies $\lim_{n\to\infty} \left|\mathbb{E}_{\tau_{\varepsilon_n}(\cdot|C^{\tau_{\varepsilon_n}})}[\hat{v}_{\varepsilon_n}^c(q)] - \mathbb{E}_{\tau'_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)]\right| = 0$ and $\lim_{n\to\infty} \mathbb{E}_{\tau'_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}^c(q)] \le V_0^c(q_\infty^c)$.

Combining these observations, we have that

$$\lim_{n\to\infty} \mathbb{E}_{\tau_{\varepsilon_n}}[\hat{v}_{\varepsilon_n}(q)] \leq \tau^c_\infty V^c_0(q^c_\infty) + (1-\tau^c_\infty)V^d_0(q^d_\infty) \leq V_0(\sigma),$$

which delivers the desired contradiction.

$\square$