Perfect Bayesian Persuasion*

Elliot Lipnowski†
Columbia University

Doron Ravid
University of Chicago

Denis Shishkin
UC San Diego

February 8, 2022

Abstract

A sender commits to an experiment to persuade a receiver. We study attainable sender payoffs, accounting for sender incentives for experiment choice, and not presupposing a receiver tie-breaking rule when indifferent. We characterize when the sender's equilibrium payoff is unique and so coincides with her value in Kamenica and Gentzkow (2011). A sufficient condition is that every action which is receiver-optimal at some belief over a set of states is a uniquely optimal at some other such belief—a generic property for finite models. In an extension, this uniqueness generates robustness to imperfect sender commitment.

*Lipnowski and Ravid acknowledge support from the National Science Foundation (grant SES-1730168).
†e.lipnowski@columbia.edu, dravid@uchicago.edu, dshishkin@ucsd.edu

In this paper, we concern ourselves with the model of communication with commitment from Kamenica and Gentzkow (2011), hereafter KG . A receiver (R, he) must choose an action $a \in A$, but a sender (S, she) controls R's available information about a payoff-relevant state $\theta \in \Theta$, which is distributed according to some prior probability distribution $\mu_0 \in \Delta\Theta$. Specifically, S first announces a Blackwell experiment concerning the state, that is, a measurable function $\psi : \Theta \to \Delta M$ for a given space M of messages. Then, after observing both ψ and a realized message $m \in M$, R chooses an action. Each player $i \in \{S, R\}$ seeks to maximize the expectation of an objective $u_i(a, \theta)$. Finally, we maintain the technical assumptions that both A and Θ are nontrivial compact metrizable spaces; that M is a sufficiently large Polish space;[1] and that the objectives $u_S, u_R : A \times \Theta \to \mathbb{R}$ are continuous.

The main result of KG is a characterization of the S-optimal equilibrium payoff. We defer a formal definition of equilibrium to the following section, but a brief overview of KG's analysis is in order. They adopt a belief-based approach, casting S's optimization problem as one of directly choosing $p \in \Delta\Delta\Theta$, the ex-ante distribution of R's posterior belief μ concerning the state. Because R is Bayesian (and S's experiment choice is made in ignorance of the state), it must be that p belongs to $\mathcal{I}(\mu_0)$, the set of belief distributions with barycenter equal to the prior; KG term this condition Bayes plausibility. But what payoff does S derive from a given Bayes-plausible belief distribution? By R rationality, R will choose from his best responses $A_R^*(\mu) \subseteq A$ whenever her posterior belief is μ. S's expected payoff from such an action $a \in A_R^*(\mu)$ is then equal to $\int u_S(a, \cdot)\, \mathrm{d}\mu$. Given KG's focus on S-optimal equilibrium, they can assume without loss that R always breaks any indifferences in S's favor. Thus, KG can summarize S's payoff from inducing belief μ as $v(\mu) := \max_{a \in A_R^*(\mu)} \int u_S(a, \cdot)\, \mathrm{d}\mu$. We call v the *value function*. Hence, S's best equilibrium value is given by $\hat{v}(\mu_0) = \max_{p \in \mathcal{I}(\mu_0)} \int v\, \mathrm{d}p$.

But what happens if R may choose best responses that differ from those S would prefer? If, in the worst case, R always chooses S's least favorite of his best responses, then from inducing R belief μ, S can only expect a payoff of $w(\mu) := \min_{a \in A_R^*(\mu)} \int u_S(a, \cdot)\, \mathrm{d}\mu$. Accordingly, S would have a profitable deviation if her payoff were ever strictly below $\hat{w}(\mu_0) = \sup_{p \in \mathcal{I}(\mu_0)} \int w\, \mathrm{d}p$. In what follows, we verify that this payoff lower bound is the only additional constraint imposed by S's experiment-choice incentives. Using this result, we go on to fully characterize when S has a unique equilibrium payoff, and to provide meaningful sufficient conditions for the same.

The remainder of the paper proceeds as follows. First, we briefly survey some related

[1] Specifically, we assume either that M is uncountable or that Θ is finite and $|M| \geq |\Theta|^2$.

literature. Then, Section 1 formalizes our equilibrium concept and characterizes the S equilibrium payoff set. Then, Section 2 develops sufficient conditions for this equilibrium payoff to be unique, that is, for the Bayesian persuasion value to be robust to equilibrium selection. Finally, Section 3 introduces a model in which S may have the ability manipulate the experiment's results ex-post, and applies the above uniqueness results to show robustness to imperfect credibility.

Related literature We contribute to the Bayesian persuasion literature (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011; Kamenica, 2019), which studies sender-receiver games in which a sender commits to an information-transmission strategy. Our main goal is to understand the KG model's range of equilibrium payoffs for the sender, taking her experiment-choice incentives into account, and allowing the receiver to choose actions in a manner unfavorable to the sender, even when he is indifferent. Our extension to the case in which the sender's commitment is imperfect extends the analysis of Lipnowski, Ravid, and Shishkin (2022) in the same manner.[2]

Our work is related to multiple distinct strands of the Bayesian persuasion literature that explicitly account for sender incentives in choosing an experiment. The first is the literature on competition in persuasion, in which multiple senders design flexible information either simultaneously (Gentzkow and Kamenica, 2016, 2017; Ravindran and Cui, 2020) or sequentially (Li and Norman, 2021; Wu, 2021) and must make individually rational experiment choices in equilibrium. The most related paper from this set is (Wu, 2021), which proves (a finite-state version of) Proposition 1. The second strand studies how experiment constraints shape chosen information (e.g., Ichihashi, 2019; Perez-Richet and Skreta, 2021), whereas the third strand studies signaling and informed principal problems for information design settings in which an experiment choice can reveal private information (e.g., Perez-Richet, 2014; Hedlund, 2017; Alonso and Câmara, 2018; Koessler and Skreta, 2021).[3] Finally, beyond the Bayesian persuasion literature, work on sequential mechanism design under limited commitment (e.g., Skreta, 2006; Doval and Skreta, 2020) accounts for a principal's incentives while her future beliefs play a prominent role in the analysis.

[2] See also Min (2021), which develops a generalization of the limited-commitment model and shows some credibility is better than no credibility in a leading example; and Fréchette, Lizzeri, and Perego (2019), which studies communication outcomes in a laboratory experiment.

[3] In a sense, the literature on verifiable disclosure can be seen as a combination of the second and third strand, with sender incentives being a key object of study and verifiable information being limited in a type-dependent manner.

1. The Equilibrium Payoff Set

We now formally define an equilibrium concept for the persuasion game. Note two features of the definition. First, while R must respond optimally to his belief, we make no direct assumption concerning which of his best responses he chooses in the case that he is indifferent.[4] Second we explicitly include an optimality condition for S at the experiment-choice stage. Note that the latter condition would have no bite under S-favorable tie-breaking, and so is rarely included in the literature.

Definition 1. *Let Ψ denote the set of all measurable functions $\tilde{\psi} : \Theta \to \Delta M$ (a.k.a. experiments), which we view as a measurable space endowed with the discrete σ-algebra. A sender strategy is an experiment $\psi \in \Psi$; a receiver strategy is a measurable function $\alpha : M \times \Psi \to \Delta A$; and a receiver belief map is a measurable function $\pi : M \times \Psi \to \Delta\Theta$. A **(perfect Bayesian) equilibrium** is a triple of such maps $\langle \psi, \alpha, \pi \rangle$ such that*

1. *The sender's choice satisfies*

$$\psi \in \operatorname{argmax}_{\tilde{\psi} \in \Psi} \int_\Theta \int_M \int_A u_S(a, \theta)\, d\alpha(a|\tilde{\psi}, m)\, d\tilde{\psi}(m|\theta)\, d\mu_0(\theta);$$

2. *Every $\tilde{\psi} \in \Psi$ and $m \in M$ have*

$$\alpha\left(\operatorname{argmax}_{a \in A} \int_\Theta u_R(a, \theta)\, d\pi(\theta|m, \tilde{\psi}) \,\middle|\, m, \tilde{\psi}\right) = 1;$$

3. *Every $\tilde{\psi} \in \Psi$, Borel $\hat{M} \subseteq M$, and Borel $\hat{\Theta} \subseteq \Theta$ have*

$$\int_\Theta \int_{\hat{M}} \pi(\hat{\Theta}|m, \tilde{\psi})\, d\tilde{\psi}(m|\theta)\, d\mu_0(\theta) = \int_{\hat{\Theta}} \tilde{\psi}(\hat{M}|\theta)\, d\mu_0(\theta).$$

*In such a case, we say the induced **equilibrium sender payoff** is*

$$\int_\Theta \int_M \int_A u_S(a, \theta)\, d\alpha\,(a|m, \psi)\;\, d\psi(m|\theta)\, d\mu_0(\theta).$$

The interpretation is as follows. First, S publicly chooses an experiment $\tilde{\psi} \in \Psi$.[5] The

[4] Still, as we shall see, it will often be the case that *mutual* best response requires that R break indifferences in S's favor on the path of play, just as a recipient of a zero offer accepts the offer in the unique subgame-perfect equilibrium of the ultimatum game.

[5] One could easily extend the model to allow S to mix over experiment choice. Doing so would entail

experiment then produces a message $m \in M$ that R observes. Then, R updates his beliefs according to the message and the chosen experiment, and chooses an action $a \in A$. We require that S only choose experiments that maximize her expected payoffs, that R (having seen the realized experiment and message) only choose actions that maximize his expected payoffs with respect to his belief about the payoff state, and that R's beliefs conform to Bayesian updating.[6]

In what follows, we document the set of attainable equilibrium S payoffs, with a particular focus on understanding when it is unique.

Remark 1. *Although our focus is on equilibrium S payoffs rather than behavior, our results have natural implications for behavior as well. In particular, when S's equilibrium payoff is unique, our results imply that R breaks indifferences in S's favor with probability 1 on path in every equilibrium. Hence, in this case, the results of KG (and many subsequent papers surveyed in Kamenica, 2019) are robust to allowing arbitrary tie-breaking for R and to accounting for S's experiment-choice incentives.*

1.1. Characterizing equilibrium payoffs

We begin by stating a characterization of the equilibrium S payoff set as a function of the parameters of our game. This set is a compact interval, with highest value equal to KG's commitment solution, and lowest value equal to the supremum value S can guarantee when R breaks her indifferences adversarially. In the special case in which the state space is finite, this result is exactly Proposition 1 from Wu (2021). Although no substantive new arguments are required for the general case, we include a proof for the sake of completeness.

Proposition 1 (Payoff set)**.** *The set of equilibrium S payoffs is* $[\hat{w}(\mu_0), \hat{v}(\mu_0)]$.

Necessity is essentially immediate, and the proof of sufficiency is constructive. By degrading information from an S-optimal (under favorable tie-breaking) experiment and allowing for R to mix among optimal choices in the degraded experiment, one can find an experiment for S to choose and R best response to target any payoff in the given interval. Then, having R break indifference adversarially to S following off-path experiment choices ensures that this experiment choice is indeed optimal for S.

added notational burden but would have no effect on the resulting S payoff set because the experiment choice is public, not informed by private information, and not simultaneous to any other decisions.

[6] Moreover, we assume that S cannot signal what she does not know. Indeed, our Bayesian condition implies that every $\tilde{\psi} \in \Psi$ and Borel $\hat{\Theta} \subseteq \Theta$ have $\int_\Theta \int_M \pi(\hat{\Theta}|m, \tilde{\psi})\, \mathrm{d}\tilde{\psi}(m|\theta)\, \mathrm{d}\mu_0(\theta) = \mu_0(\hat{\Theta})$, so that the chosen experiment alone does not cause belief updating by R about the payoff state.

Remark 2. *It is apparent that Proposition 1 depends only on the value correspondence $V = [w, v]$, and moreover (as is clear from our proof) the only substantive property required of the environment is that the attainable S payoffs from R responding optimally to a given belief be convex.*[7] *In addition to making Proposition 1 more tractable to apply, this feature also expands its applicability beyond the basic model we have considered. For example, the proposition can be applied to settings in which a receiver is subject to independent private payoff shocks. Additionally, the proposition applies to public persuasion of a set of agents who play a game, so long as the set of induced payoffs for the sender is convex for every public belief. The latter condition holds, for instance, if the receivers have a unique equilibrium of their induced game, or if the receivers observe a rich public randomization device after the experiment choice but before their gameplay.*

2. Equilibrium Payoff Uniqueness

In this section we ask, when does S have a unique equilibrium payoff? Whenever she does, the traditional analysis that focuses on S-optimal equilibrium (and so assumes S-favorable tie-breaking by R) is essentially without loss.

As a starting observation, because uniqueness follows directly from Proposition 1 whenever $v = w$, a sufficient condition for S to have a unique equilibrium payoff is immediate.

Corollary 1 (No relevant ties)**.** *S has a unique equilibrium payoff if, at any belief, S is indifferent between between all of R's best responses.*

Although restrictive, the above condition nevertheless captures many cases of interest. For example, if the action space is a convex subset of some linear space with R's payoff being strictly concave in his action (e.g., Crawford and Sobel, 1982; Chakraborty and Harbaugh, 2010), then he has a unique best response to every belief, and so the corollary applies.

The following result gives an alternative sufficient condition for S to attain her Bayesian persuasion value in all equilibria. It says such uniqueness holds if information can serve as a stand-in for favorable tie-breaking at all relevant posterior beliefs. To state the result, the following definition is useful.

[7] Our analysis also uses the fact that V is nonempty-compact-valued and upper hemicontinuous, and that the set of optimal R choices is a weakly measurable correspondence of his belief. These features are satisfied in all past applications of which we are aware.

Definition 2. *Say a set $D \subseteq \Delta\Theta$ is **persuasion sufficient** if*

$$\sup_{p \in \mathcal{I}(\mu_0):\ p(D')=1 \text{ for some Borel } D' \subseteq D} \int v \, dp = \hat{v}(\mu_0).$$

As the following proposition says, if (on the relevant set of beliefs) information can be used to replace favorable selection, then S guarantees her KG payoff.

Proposition 2 (Information as selection)**.** *S has a unique equilibrium payoff if $\hat{w}|_D \geq v|_D$ for some persuasion-sufficient $D \subseteq \Delta\Theta$.*

The proof is based on a standard measurable selection result, augmenting an experiment with additional information to replicate favorable tie-breaking.

Following directly from the above proposition, the following corollary exactly characterizes when equilibrium S payoff uniqueness holds independent of the prior: such uniqueness is equivalent to information always replicating favorable tie-breaking.

Corollary 2 (Global uniqueness)**.** *S has a unique equilibrium payoff for every prior (holding other parameters fixed) if and only if $\hat{w} \geq v$.*

2.1. Direct sufficient conditions for uniqueness

The above conditions for payoff uniqueness were expressed in terms of the derived objects v, w, $\hat{v}$, and $\hat{w}$. While these functions are primitive to the environment, it is desirable to find more interpretable sufficient conditions on the players' preferences. We now develop sufficient conditions "directly" on R's preferences that ensure S has a unique equilibrium payoff. The following condition, satisfied in many applications of interest, is our key such condition.

Definition 3. *Given a set $D \subseteq \Delta\Theta$, say **the perturbed unique best response property (PUBR) holds on** D if, for any $\mu \in D$ and $a \in A^*_R(\mu)$, some $\mu' \in \Delta\Theta$ exists such that $A^*(\mu') = \{a\}$ and $supp(\mu') \subseteq supp(\mu)$.*

The above property says that any action that is optimal for R at some belief in D is in fact uniquely optimal at some alternative belief, where the alternative belief can be assumed to rule out any open neighborhood of states that the original belief rules out.[8] Note the

[8]Note, this condition is strictly stronger than the requirement that R has no duplicate actions. The latter condition is insufficient for ensuring uniqueness, as witnessed by $A = \Theta = \{0, 1\}$, $\mu_0 = \frac{1}{2}$, $u_S(a, \theta) = -a$, and $u_R(a, \theta) = a\theta$.

global PUBR property depends only on R's preferences (together with the spaces of states and actions), not on the prior or on S's preferences.

The next result shows the PUBR property is sufficient to guarantee equilibrium selection.

Theorem 1 (Sufficient condition for uniqueness)**.** *If PUBR holds on some persuasion-sufficient $D \subseteq \Delta\Theta$, then S has a unique equilibrium payoff.*[9]

In light of Proposition 2, it is enough to show that information can serve as a stand-in for selection. The key observation is that, for any given belief that appears in a solution to KG's program, the PUBR property helps locate nearby beliefs (by replacing a unique-best-response belief with a weighted average of the same and the original belief) at which *every* best response gives S an expected payoff nearly as high as favorable tie-breaking would at the belief itself. Using such beliefs, we can construct slightly more informative experiments that similarly give S a high payoff under all R best responses.

As a consequence of Theorem 1, the following result states that unique equilibrium S payoffs are a generic feature of finite environments.[10] The proof shows that global PUBR is generic, which suffices by the previous result. Intuitively, a failure of PUBR at some belief implies that, for some fixed action and fixed set of states, R's highest possible expected payoff gain from using said action rather than another is exactly zero—a knife-edge condition.

Proposition 3 (Generic uniqueness)**.** *If A and Θ are finite, then an open dense $\mathcal{U}_R \subseteq \mathbb{R}^{A\times\Theta}$ of full Lebesgue measure exists such that S has a unique equilibrium payoff (for any prior and any S preferences) as long as $u_R \in \mathcal{U}_R$.*

The following result gives a structured infinite environment in which payoff uniqueness follows easily from Theorem 1—cases with ordered states, finitely many ordered actions, preferences that respect these orders, and sufficient smoothness conditions. Even though the result makes use of a persuasion-sufficient proper subset, it makes no use of the specific form of the S objective u_S.

Proposition 4 (Uniqueness in a one-dimensional model)**.** *Suppose $A, \Theta \subseteq \mathbb{R}$ with A finite; the function $u_R(\cdot, \theta)$ has a unique maximizer for every μ_0-atom $\theta \in \Theta$;*[11] *and u_R is strictly concave*

[9] A converse trivially holds in the case that A is finite. If the PUBR property fails for u_R, as witnessed by some $a \in A$ and $\mu \in \Delta\Theta$, then equilibrium uniqueness fails for prior μ and S preferences $u_S(\tilde{a}, \tilde{\theta}) := \mathbf{1}_{\tilde{a}=a}$. Indeed, such preferences generate $v(\mu) = 1$ and $\hat{w} = 0$, so that S can get any payoff in $[0, 1]$ in equilibrium.

[10] A related result follows directly from Proposition 3 of Li and Norman (2021) in the generic finite case. That result implies that all equilibria with S-favorable tie-breaking result in the same state-contingent action distribution. Hence, *assuming* R breaks indifferences in S's favor generically yields a behavioral uniqueness property, not just payoff uniqueness.

[11] Note, this condition holds vacuously if μ_0 is atomless.

in its first argument with strictly increasing differences. Then, S has a unique equilibrium payoff.

To prove the result, we first note that (given finite action spaces) only finitely many beliefs are ever needed. Hence, R's posterior belief need only have an atom where the prior does, and so we need only check the PUBR property at such beliefs. The condition holds by fiat for beliefs that are degenerate on a prior atom, as R is assumed to have a unique best response at such beliefs. In the complementary case, we use the strict monotonicity hypothesis to show any belief to which R has two best responses can be split to nearby beliefs that ensure uniqueness.

3. Limited Commitment

In this section, we extend the model to one in which S has only limited commitment power, as modeled in Lipnowski, Ravid, and Shishkin (2022)—hereafter LRS1. To do so, we first specialize the existing parameters of our environment for tractability:

Assumption 1. *The spaces A, Θ, and M are all finite, and u_S is state independent (i.e., constant in its second argument).*

In mild abuse of notation, write $u_S : A \to \mathbb{R}$. In addition to the parameters of our baseline model, we parameterize our limited-commitment model by $\chi \in [0, 1]$, which denotes the sender's *credibility*.

The augmented game begins with S provisionally choosing an experiment, $\psi \colon \Theta \to \Delta M$, an "official" report. The state $\theta \in \Theta$ then realizes and, independent of the state, one of two possibilities occurs. With probability χ, the message $m \in M$ is sent in accordance with the report, and so is distributed according to $\psi(\cdot|\theta)$. With complementary probability $1 - \chi$, reporting is "influenced" and so S observes θ and can freely choose $m \in M$. Crucially, R observes the message m but observes neither θ nor whether reporting was influenced.

In what follows, we formalize an appropriate solution concept for the game with compromised reporting (again taking experiment-choice incentives into account), provide a result comparing the S equilibrium payoff set to that under perfect commitment, and apply our uniqueness results to address robustness to imperfect credibility.

3.1. The equilibrium payoff set

We formalize the solution concept for the augmented game as follows.

Definition 4. *A sender strategy is a pair (ψ, σ) consisting of an experiment $\psi \in \Psi$ and a measurable function $\sigma : \Theta \times \Psi \to \Delta M$; a receiver strategy is a measurable function $\alpha : M \times \Psi \to \Delta A$; and a receiver belief map is a measurable function $\pi : M \times \Psi \to \Delta\Theta$. A* ***perfect Bayesian χ-equilibrium (χ-PBE)*** *is a quadruple of such maps $\langle \psi, \sigma, \alpha, \pi \rangle$ such that*

1. *The sender's experiment choice satisfies*

$$\psi \in \operatorname{argmax}_{\tilde{\psi} \in \Psi} \int_\Theta \int_M \int_A u_S \; d\alpha(\cdot|m, \tilde{\psi}) \left[\chi \; d\tilde{\psi}(m|\theta) + (1 - \chi) \; d\sigma(m|\theta, \tilde{\psi}) \right] \; d\mu_0(\theta);$$

2. *Every $\tilde{\psi} \in \Psi$ and $m \in M$ have*

$$\alpha \left(\operatorname{argmax}_{a \in A} \int_\Theta u_R(a, \theta) \; d\pi(\theta|m, \tilde{\psi}) \;\middle|\; m, \tilde{\psi} \right) = 1;$$

3. *Every $\tilde{\psi} \in \Psi$, Borel $\hat{M} \subseteq M$, and Borel $\hat{\Theta} \subseteq \Theta$ have*

$$\int_\Theta \int_{\hat{M}} \pi(\hat{\Theta}|\cdot, \tilde{\psi}) \; d\left[\chi\tilde{\psi}(\cdot|\theta) + (1 - \chi)\sigma(\cdot|\theta, \tilde{\psi}) \right] \; d\mu_0(\theta) = \int_{\hat{\Theta}} \left[\chi\tilde{\psi}(\hat{M}|\cdot) + (1 - \chi)\sigma(\hat{M}|\cdot, \tilde{\psi}) \right] \; d\mu_0;$$

4. *Every $\tilde{\psi} \in \Psi$ and $\theta \in \Theta$ have*

$$\sigma \left(\operatorname{argmax}_{m \in M} \int_A u_S(a) \; d\alpha(a|\tilde{\psi}, m) \;\middle|\; \theta, \tilde{\psi} \right) = 1.$$

In such a case, we say the induced ***χ-PBE payoff*** *is*

$$\int_\Theta \int_M \int_A u_S(a) \; d\alpha \, (a|m, \psi) \left[\chi \; d\psi(m|\cdot) + (1 - \chi) \; d\sigma(m|\cdot, \psi) \right] \; d\mu_0.$$

Let $w^*_\chi(\mu_0)$ denote the infimum χ-PBE payoff for S.

We begin our analysis of this richer model with a partial description of the range of χ-PBE payoff for S. With full credibility, the range of S payoffs is naturally identical to the game without an influencing stage (as influence occurs with zero probability): one completes the equilibrium by incorporating some best response for an influencing S. But which payoffs can S attain when unable to perfectly commit? With S-favorable selection, and ignoring experiment-choice incentives, it is nearly immediate that higher credibility can only ever help: she could always use her additional commitment power to replicate the way should would behave without it. But the effect of credibility is less clear when one considers S

incentives and the full range of equilibria. First, the above replication argument is only available if S finds it optimal to mimic her hypothetical influencing behavior. Second, because influencing S is subject to incentive constraints more often under lower credibility, it is natural to wonder whether her incentive constraints would refine away some bad equilibria.

The following result shows that, in spite of these concerns, lower credibility is in fact worse for S than full commitment power in a set-valued sense.[12]

Proposition 5 (Payoff set under partial credibility)**.** *The nonempty set of χ-PBE payoffs for S is weakly below $[\hat{w}(\mu_0), \hat{v}(\mu_0)]$ in the strong set order, coinciding with it for $\chi = 1$.*

The part of the proposition for $\chi = 1$ is very straightforward, but let us briefly discuss the results for general χ (and take χ-PBE existence for granted for the sake of exposition). Because additional incentive considerations can only constrain S, it is easy to see that the highest χ-PBE payoff is no greater than $\hat{v}(\mu_0)$. The strong set ranking would then follow directly if the highest χ-PBE payoff were below $\hat{w}(\mu_0)$. In the complementary case, when the highest χ-PBE payoff is above $\hat{w}(\mu_0)$, we need to show that every payoff between the two is in fact compatible with χ-PBE. We can think of this result as having two constituent pieces. First, the set of S payoffs attainable *ignoring* experiment-choice incentives is an interval. And second, an appropriate continuation equilibrium can be chosen adversarially to make sure S obtains a lower payoff if she deviates to a different experiment choice. For the first feature, we can use the fact (proven in LRS1) that the set of attainable S payoffs is an interval when ignoring experiment-choice incentives, and pair this fact with the observation that the low payoff $w(\mu_0)$ could be implemented by an uninformative official report with a babbling equilibrium. The second step—showing S can be held to a payoff upper bound of $\hat{w}(\mu_0)$ if she chooses the wrong experiment—is more involved.

How does one show that, for any experiment S could initially choose, some continuation equilibrium gives S a payoff of no more than $\hat{w}(\mu_0)$? This feature would follow immediately from the definition of $\hat{w}$ if R were to choose adversarially to S whenever he is indifferent between multiple actions, but we do not know a continuation equilibrium with the latter feature exists—the wrinkle being that w is not continuous (hence, not upper hemicontinuous). To resolve this issue, we search for a continuation equilibrium in which R gives S a continuation payoff in $[w(\mu), z(\mu)]$ whenever his posterior belief is μ, where z is the smallest upper semicontinuous function above w (and so $z \leq v$). Because the correspondence $[w, z]$ is upper hemicontinuous, an appropriate application of Kakutani's fixed point theorem delivers such continuation play and beliefs. This continuation play generates an S payoff of at most $\hat{z}(\mu_0)$

[12] Note, the proposition also establishes that a χ-PBE exists.

by KG's results, where $\hat{z}$ is the concave envelope of z. Finally, because the concave function $\hat{w}$ on a finite-dimensional simplex is automatically continuous on the interior, we can show that the concave envelopes $\hat{z}$ and $\hat{w}$ agree at the prior.

3.2. Strong robustness

LRS1 establishes a robustness result (Proposition 3 of that paper) concerning the best S payoff attainable when credibility is only slightly imperfect ($\chi \approx 1$). That result implies her highest attainable payoff converges to the Bayesian persuasion value for most payoff specifications—in particular, whenever the PUBR property holds globally. In this subsection, we demonstrate a tight link between two conceptually distinct notions of robustness. We show that S payoffs are robust to slightly imperfect credibility *and* equilibrium selection if and only if they are robust to equilibrium selection in the full-credibility case. With this result in hand, the results of the previous sections apply directly to address this stronger form of robustness in persuasion models.

Proposition 6 (Strong robustness)**.** *The lowest χ-PBE values satisfy $\lim_{\chi \nearrow 1} w^*_\chi(\mu_0) = \hat{w}(\mu_0)$. In particular, the Bayesian persuasion value is strongly robust to partial credibility ($\lim_{\chi \nearrow 1} w^*_\chi(\mu_0) = \hat{v}(\mu_0)$) if and only if it robust to equilibrium selection ($\hat{w}(\mu_0) = \hat{v}(\mu_0)$).*

The proof constructs lower payoff bounds for S in any χ-PBE by computing her payoff following viable deviations, and shows that these payoff bounds can be made arbitrarily close to $\hat{w}(\mu_0)$ as credibility becomes arbitrarily close to perfect. In brief, we consider the deviation to an experiment that would be approximately optimal under full S commitment with S-adversarial R tie-breaking. Using lower semicontinuity of w, we show S attains a payoff approximating her full-commitment payoff from this experiment in as her probability of influencing vanishes, because the belief a message induces is very nearby the full-commitment version of the same.

Finally, Propositions 6 and 3 immediately yield the following conclusion.

Corollary 3 (Generic strong robustness)**.** *For any finite A and Θ, for all but a Lebesgue-null (and nowhere dense) set of R objectives $u_R \in \mathbb{R}^{A\times\Theta}$, and every S objective $u_S \in \mathbb{R}^A$, the full commitment value is strongly robust to partial credibility.*

Thus, for most payoff specifications, S obtains essentially her Bayesian persuasion payoff in any equilibrium, even if her ability to commit to the information she provides is slightly imperfect.

References

Aliprantis, Charalambos D and Kim Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Science & Business Media.

Alonso, Ricardo and Odilon Câmara. 2018. "On the value of persuasion by experts." *Journal of Economic Theory* 174:103–123.

Aumann, Robert J and Michael Maschler. 1995. *Repeated games with incomplete information*. MIT press.

Chakraborty, Archishman and Rick Harbaugh. 2010. "Persuasion by cheap talk." *American Economic Review* 100 (5):2361–2382.

Crawford, Vincent P and Joel Sobel. 1982. "Strategic information transmission." *Econometrica* 50 (6):1431–1451.

de Clippel, Geoffroy. 2008. "An axiomatization of the inner core using appropriate reduced games." *Journal of Mathematical Economics* 44 (3-4):316–323.

Doval, Laura and Vasiliki Skreta. 2020. "Mechanism design with limited commitment." *Available at SSRN 3281132* .

Fréchette, Guillaume, Alessandro Lizzeri, and Jacopo Perego. 2019. "Rules and Commitment in Communication: An Experimental Analysis." Working paper.

Gentzkow, Matthew and Emir Kamenica. 2016. "Competition in persuasion." *The Review of Economic Studies* 84 (1):300–322.

———. 2017. "Bayesian persuasion with multiple senders and rich signal spaces." *Games and Economic Behavior* 104:411–429.

Hedlund, Jonas. 2017. "Bayesian persuasion by a privately informed sender." *Journal of Economic Theory* 167:229–268.

Ichihashi, Shota. 2019. "Limiting Sender's information in Bayesian persuasion." *Games and Economic Behavior* 117:276–288.

Kamenica, Emir. 2019. "Bayesian Persuasion and Information Design." *Annual Review of Economics* 11 (1).

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

Koessler, Frédéric and Vasiliki Skreta. 2021. "Information design by an informed designer." .

Li, Fei and Peter Norman. 2021. "Sequential persuasion." *Theoretical Economics* 16 (2):639–675.

Lipnowski, Elliot and Laurent Mathevet. 2018. "Disclosure to a psychological audience." *American Economic Journal: Microeconomics* 10 (4):67–93.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin. 2022. "Persuasion via Weak Institutions." *Available at SSRN 3168103* .

Min, Daehong. 2021. "Bayesian persuasion under partial commitment." *Economic Theory* 72 (3):743–764.

Perez-Richet, Eduardo. 2014. "Interim bayesian persuasion: First steps." *American Economic Review* 104 (5):469–474.

Perez-Richet, Eduardo and Vasiliki Skreta. 2021. "Test Design under Falsification." Working paper.

Ravindran, Dilip and Zhihan Cui. 2020. "Competing persuaders in zero-sum games." *arXiv preprint arXiv:2008.08517* .

Rockafellar, R Tyrrell. 1970. *Convex Analysis*. 28. Princeton University Press.

Skreta, Vasiliki. 2006. "Sequentially Optimal Mechanisms." *Review of Economic Studies* 73 (4):1085–1111.

Wu, Wenhao. 2021. "Sequential Bayesian Persuasion." .

A. Proofs

Before proceeding to formal proofs, we review for convenience several key notations.

$$\begin{aligned}
A_R^* : \Delta\Theta &\rightrightarrows A \\
\mu &\mapsto \operatorname{argmax}_{a\in A} \int u_R(a,\cdot)\,\mathrm{d}\mu \\
V : \Delta\Theta &\to \mathbb{R} \\
\mu &\mapsto \operatorname{co}\left\{\int u_S(a,\cdot)\,\mathrm{d}\mu :\ a \in A_R^*(\mu)\right\} \\
v : \Delta\Theta &\to \mathbb{R} \\
\mu &\mapsto \max V(\mu) \\
w : \Delta\Theta &\to \mathbb{R} \\
\mu &\mapsto \min V(\mu) \\
\mathcal{I} : \Delta\Theta &\rightrightarrows \Delta\Delta\Theta \\
\mu &\mapsto \left\{p \in \Delta\Delta\Theta :\ \int \tilde{\mu}\,\mathrm{d}p(\tilde{\mu}) = \mu\right\} \\
\hat{v} : \Delta\Theta &\to \mathbb{R} \\
\mu &\mapsto \max_{p\in\mathcal{I}(\mu)} \int v\,\mathrm{d}p \\
\hat{w} : \Delta\Theta &\to \mathbb{R} \\
\mu &\mapsto \sup_{p\in\mathcal{I}(\mu)} \int w\,\mathrm{d}p.
\end{aligned}$$

A.1. Proofs for Section 1

The following lemma applies Carathéodory's theorem to show information policies can be replaced with payoff-equivalent ones of small support.

Lemma 1. *If Θ is finite, then any $\bar{\mu} \in \Delta\Theta$, Borel $D \subseteq \Delta\Theta$, bounded measurable $f : \Delta\Theta \to \mathbb{R}$ and $p \in \mathcal{I}(\bar{\mu}) \cap \Delta D$ admit some $q \in \mathcal{I}(\bar{\mu}) \cap \Delta D$ with $supp(q)$ affinely independent (hence of cardinality no greater than $|\Theta|$) and $\int f\,dq \geq \int f\,dp$.*

Proof. The set $\mathcal{I}(\bar{\mu})$ is convex compact metrizable, and so Choquet's theorem yields some $Q \in \Delta[\mathcal{I}(\bar{\mu})]$ with barycenter p such that Q is supported on the extreme points of $\mathcal{I}(\bar{\mu})$. But this set $\operatorname{ext}[\mathcal{I}(\bar{\mu})]$ consists exactly of those $q \in \mathcal{I}(\bar{\mu})$ with affinely independent support.

By definition of the barycenter, $\int\int g\,\mathrm{d}q\,\mathrm{d}Q(q) = \int g\,\mathrm{d}p$ for every continuous $g : \Delta\Theta \to \mathbb{R}$. However, because the barycenter of this Q is unique, it follows that $p(B) = \int q(B)\,\mathrm{d}Q(q)$ for every Borel $B \subseteq \Delta\Theta$. Hence, $Q(\Delta D) = 1$ because $p(D) = 1$, and $\int\int g\,\mathrm{d}q\,\mathrm{d}Q(q) = \int g\,\mathrm{d}p$ for every bounded measurable $g : \Delta\Theta \to \mathbb{R}$—in particular for $g = f$.

Letting $\mathcal{J} := \operatorname{ext}[\mathcal{I}(\bar{\mu})] \cap \Delta D$, we have $Q(\mathcal{J}) = 1$. Therefore,

$$0 = \int\int f\,\mathrm{d}q\,\mathrm{d}Q(q) - \int f\,\mathrm{d}p = \int_{\mathcal{J}} \left[\int f\,\mathrm{d}q - \int f\,\mathrm{d}p\right] \mathrm{d}Q(q).$$

So the integrand is somewhere nonnegative: some $q \in \mathcal{J}$ has $\int f \, \mathrm{d}q \geq \int f \, \mathrm{d}p$. □

Now, we prove the characterization of all S equilibrium payoffs.

Proof of Proposition 1. To begin, we recall some well-known facts about experiments and Bayesian updating—which collectively tell us that S choosing from Ψ and choosing from $\mathcal{I}(\mu_0)$ are equivalent formalisms. First, any experiment $\tilde{\psi} \in \Psi$ admits some compatible belief map, that is, some measurable $\tilde{\pi} = \tilde{\pi}_{\tilde{\psi}} : M \to \Delta\Theta$ such that every Borel $\hat{M} \subseteq M$ and Borel $\hat{\Theta} \subseteq \Theta$ have $\int_\Theta \int_{\hat{M}} \tilde{\pi}(\hat{\Theta}|m, \tilde{\psi}) \, \mathrm{d}\tilde{\psi}(m|\theta) \, \mathrm{d}\mu_0(\theta) = \int_{\hat{\Theta}} \tilde{\psi}(\hat{M}|\theta) \, \mathrm{d}\mu_0(\theta)$. Second, given $\tilde{\psi} \in \Psi$ if we define the belief distribution $p_{\tilde{\psi},\tilde{\pi}} \in \Delta\Delta\Theta$ via $p_{\tilde{\psi},\tilde{\pi}}(D) := \int_\Theta \tilde{\psi}\left(\tilde{\pi}^{-1}(D) \,\middle|\, \theta\right) \mathrm{d}\mu_0(\theta)$ for each Borel $D \subseteq \Delta\Theta$, then $p_{\tilde{\psi},\tilde{\pi}} = p_{\tilde{\psi},\tilde{\pi}'}$ for any two such compatible $\tilde{\pi}$ and $\tilde{\pi}'$. We therefore refer to the associated belief distribution simply as $p_{\tilde{\psi}}$. Third, every $\tilde{\psi} \in \Psi$ has $p_{\tilde{\psi}} \in \mathcal{I}(\mu_0)$. Fourth, every $p \in \mathcal{I}(\mu_0)$ with $|\mathrm{supp}(p)| \leq |M|$ admits some $\tilde{\psi}_p \in \Psi$ such that $p_{\tilde{\psi}_p} = p$.

Now we proceed to show $s \in [\hat{w}(\mu_0), \hat{v}(\mu_0)]$ is necessary and sufficient for s to be an equilibrium S payoff.

First, to see the condition is necessary, fix an arbitrary equilibrium $\langle \psi, \alpha, \pi \rangle$, and let $s \in \mathbb{R}$ be the induced S payoff; we will show $s \in [\hat{w}(\mu_0), \hat{v}(\mu_0)]$. For any $\tilde{\psi} \in \Psi$ and $m \in M$ the R optimality condition implies $\alpha\left(A_R^*\left(\pi(\cdot|m, \tilde{\psi})\right) \,\middle|\, m, \tilde{\psi}\right) = 1$, so that $\int_A u_S(a, \theta) \, \mathrm{d}\alpha\left(a \middle| m, \tilde{\psi}\right) \in V\left(\pi(\cdot|m, \tilde{\psi})\right)$. Therefore, any $\tilde{\psi} \in \Psi$ has

$$\begin{aligned}
\int_\Theta \int_M \int_A u_S(a, \theta) \, \mathrm{d}\alpha(a|m, \tilde{\psi}) \, \mathrm{d}\tilde{\psi}(m|\theta) \, \mathrm{d}\mu_0(\theta) &= \int_\Theta \int_M V\left(\pi(\cdot|m, \tilde{\psi})\right) \mathrm{d}\tilde{\psi}(m|\theta) \, \mathrm{d}\mu_0(\theta) \\
&= \int_{\Delta\Theta} V \, \mathrm{d}p_{\tilde{\psi}} \\
&= \left[\int_{\Delta\Theta} w \, \mathrm{d}p_{\tilde{\psi}}, \int_{\Delta\Theta} v \, \mathrm{d}p_{\tilde{\psi}}\right] \\
&\subseteq \left[\int_{\Delta\Theta} w \, \mathrm{d}p_{\tilde{\psi}}, \hat{v}(\mu_0)\right].
\end{aligned}$$

Hence, $s \leq \hat{v}(\mu_0)$. Moreover, for any $p \in \mathcal{I}(\mu_0)$ if Θ is infinite, and for any $p \in \mathcal{I}(\mu_0)$ such that $|\mathrm{supp}(p)| \leq |\Theta|$ if Θ is finite, taking $\tilde{\psi} = \tilde{\psi}_p$ implies (by S rationality)

$$\begin{aligned}
&\int_\Theta \int_M \int_A u_S(a, \theta) \, \mathrm{d}\alpha(a|m, \psi) \, \mathrm{d}\psi(m|\theta) \, \mathrm{d}\mu_0(\theta) \\
\geq &\int_\Theta \int_M \int_A u_S(a, \theta) \, \mathrm{d}\alpha(a|m, \tilde{\psi}) \, \mathrm{d}\tilde{\psi}(m|\theta) \, \mathrm{d}\mu_0(\theta) \\
\geq &\int_{\Delta\Theta} w \, \mathrm{d}p.
\end{aligned}$$

Applying this observation to every such p (and applying Lemma 1 if Θ is finite) implies $s \geq \hat{w}(\mu_0)$.

Conversely, take any $s \in [\hat{w}(\mu_0), \hat{v}(\mu_0)]$. Letting $p_1 \in \mathcal{I}(\mu_0)$ with $\int_{\Delta\Theta} v \, \mathrm{d}p_v = \hat{v}(\mu_0)$ and $|\mathrm{supp}(p_v)| \leq |\Theta|$ if Θ is finite—which exists by Lemma 1—define $p_\lambda := \int \delta_{\lambda\mu + (1-\lambda)\mu_0} \, \mathrm{d}p_1(\mu) \in$

$\mathcal{I}(\mu_0)$ for each $\lambda \in [0,1]$; observe $|\text{supp}(p_\lambda)| \leq |\text{supp}(p_1)| \leq |M|$. As $\lambda \mapsto p_\lambda$ is continuous, it follows that the $\lambda \mapsto \int V \, \mathrm{d}p_\lambda$ is nonempty-compact-convex-valued and upper hemicontinuous because V is. Moreover,

$$\int V \, \mathrm{d}p_0 = V(\mu_0) \ni w(\mu_0) \leq s \leq \hat{v}(\mu_0) \in \int V \, \mathrm{d}p_1.$$

The intermediate value theorem for correspondences (e.g., Lemma 2 from de Clippel, 2008) therefore delivers some $\lambda \in [0,1]$ such that $s \in \int V \, \mathrm{d}p_\lambda$. Some measurable $\zeta : \Delta\Theta \to [0,1]$ then exists such that $s = \int [(1-\zeta)w + \zeta v] \, \mathrm{d}p_\lambda$. By the measurable maximum theorem (Theorem 18.19 from Aliprantis and Border, 2006), a pair of measurable functions $\alpha_w, \alpha_v : \Delta\Theta \to \Delta A$ exist such that, each $\mu \in \Delta\Theta$ has $\int_{A\times\Theta} u_S \, \mathrm{d}\left[\alpha_w(\cdot|\mu) \otimes \mu\right] = w(\mu)$, $\int_{A\times\Theta} u_S \, \mathrm{d}\left[\alpha_v(\cdot|\mu) \otimes \mu\right] = v(\mu)$, and $\alpha_w\left(A_R^*(\mu)\middle|\mu\right) = \alpha_v\left(A_R^*(\mu)\middle|\mu\right) = 1$. With these objects in hand, we can define our candidate $\psi : \Theta \to \Delta M$, $\alpha : M \times \Psi \to \Delta A$, and $\pi : M \times \Psi \to \Delta\Theta$ via

$$\begin{aligned}
\psi(\cdot|\theta) &:= \tilde{\psi}_{p_\lambda} \\
\pi(\cdot|m,\tilde{\psi}) &:= \tilde{\pi}_{\tilde{\psi}}(\cdot|m) \\
\alpha(\cdot|m,\tilde{\psi}) &:= \begin{cases} (1-\zeta)\alpha_w\left(\cdot \,\middle|\, \pi(\cdot|m,\tilde{\psi})\right) + \zeta\alpha_v\left(\cdot \,\middle|\, \pi(\cdot|m,\tilde{\psi})\right) & : \tilde{\psi} = \tilde{\psi}_{p_\lambda} \\ \alpha_w\left(\cdot \,\middle|\, \pi(\cdot|m,\tilde{\psi})\right) & : \tilde{\psi} \neq \tilde{\psi}_{p_\lambda}. \end{cases}
\end{aligned}$$

It is immediate from the construction that all three maps are measurable and that R rationality and the Bayesian property are both satisfied. Moreover, direct computation shows that choosing experiment $\tilde{\psi} \in \Psi$ gives S a continuation payoff of

$$\int_\Theta \int_M \int_A u_S(a,\theta) \, \mathrm{d}\alpha(a|m,\tilde{\psi}) \, \mathrm{d}\tilde{\psi}(m|\theta) \, \mathrm{d}\mu_0(\theta) = \begin{cases} s & : \tilde{\psi} = \tilde{\psi}_{p_\lambda} \\ \int_{\Delta\Theta} w \, \mathrm{d}p_{\tilde{\psi}} & : \tilde{\psi} \neq \tilde{\psi}_{p_\lambda} \end{cases}$$

Therefore, S gets payoff s if the triple is an equilibrium. Finally, the triple is indeed an equilibrium: In particular, S rationality is confirmed because $s \geq \hat{w}(\mu_0) \geq \int w \, \mathrm{d}p_{\tilde{\psi}}$ for every alternative $\tilde{\psi} \in \Psi$. □

A.2. Proofs for Section 2

The following proposition shows how information can be augmented to replicate favorable tie-breaking.

Proof of Proposition 2. Supposing $\hat{w}|_D \geq v|_D$, let us show, for arbitrary $\epsilon > 0$, that $\hat{w}(\mu_0) > \hat{v}(\mu_0) - \epsilon$. Replacing D with a subset, we may assume D is Borel.

Define the set $O := \left\{p \in \Delta\Delta\Theta : \int w \, \mathrm{d}p - v\left(\int \mu \, \mathrm{d}p(\mu)\right) > -\epsilon\right\}$ and the correspondence $\mathcal{I}_\Theta : \Delta\Theta \rightrightarrows \Delta\Delta\Theta$ given by $\mathcal{I}_\Theta(\mu) := \mathcal{I}(\mu)$ if Θ is infinite and $\mathcal{I}_\Theta(\mu) := \{p \in \mathcal{I}(\mu) : |\text{supp}(p)| \leq |\Theta|\}$ if Θ is finite. Let us show that the restriction to D of the correspondence $O \cap \mathcal{I}_\Theta : \Delta\Theta \rightrightarrows \Delta\Delta\Theta$ admits a measurable selector. First, because the barycenter map is continuous, the correspondence $\mathcal{I}$ is upper hemicontinuous and compact-valued. Therefore—the

measures in $\Delta\Delta\Theta$ of support size no greater than $|\Theta|$ being a closed subset if Θ is finite—$\mathcal{I}_\Theta$ is upper hemicontinuous and compact-valued too, hence weakly measurable. Next, because the barycenter map is continuous and w and v are lower and upper semicontinuous, respectively, the set O is open. But then, for every open $Q \subseteq \Delta\Delta\Theta$, the set $\{\mu \in \Delta\Theta : \ \mathcal{I}_\Theta(\mu) \cap O \cap Q\}$ is measurable, implying $O \cap \mathcal{I}_\Theta$ is weakly measurable. Moreover, that $\hat{w}|_D \geq v|_D$ implies $(O \cap \mathcal{I})|_D$ is nonempty-valued, and so Lemma 1 implies $(O \cap \mathcal{I}_\Theta)|_D$ is nonempty-valued as well. Therefore, the Kuratowski and Ryll-Nardzewski measurable selection theorem delivers a measurable selector φ of $(O \cap \mathcal{I}_\Theta)|_D$.

We can now establish that $\hat{w}(\mu_0) > \hat{v}(\mu_0) - \epsilon$. For any $p \in \mathcal{I}_\Theta(\mu_0) \cap \Delta D$, that $q_p := \int \varphi \, \mathrm{d}p \in \mathcal{I}(\mu_0)$, with $|\mathrm{supp}(q)| \leq |M|$, implies

$$\hat{w}(\mu_0) \geq \int w \, \mathrm{d}q_p = \int \int_D w \, \mathrm{d}\varphi(\cdot|\mu) \, \mathrm{d}p(\mu) > \int [v(\mu) - \epsilon] \, \mathrm{d}p(\mu) = \int v \, \mathrm{d}p - \epsilon.$$

Given Lemma 1, maximizing over $p \in \mathcal{I}_\Theta(\mu_0) \cap \Delta D$ yields $\hat{w}(\mu_0) > \hat{v}(\mu_0) - \epsilon$, establishing the claim. □

Now, we establish that the equilibrium S payoff is unique for every prior if and only information can always replace favorable tie-breaking.

Proof of Corollary 2. By Proposition 1, it suffices to show that $\hat{w} = \hat{v}$ if and only if $\hat{w} \geq v$. As it is immediate that $\hat{v} \geq \hat{w}$ and $\hat{v} \geq v$, we need only see that $\hat{w} \geq \hat{v}$ if $\hat{w} \geq v$. But this result follows directly from Proposition 2, because $\Delta\Theta$ is vacuously persuasion sufficient. □

Now, we show the PUBR property is sufficient to guarantee S has a unique equilibrium payoff. Observe the proof meaningfully uses the linear structure of R's problem—specifically that if $A_R^*(\mu') = \{a\} \subseteq A_R^*(\bar{\mu})$, then a is a unique best response to any proper convex combination of μ' and $\bar{\mu}$.

Proof of Theorem 1. Suppose the PUBR property holds on a persuasion-sufficient $D \subseteq \Delta\Theta$, and take an arbitrary $\bar{\mu} \in D$ and $\epsilon > 0$. We will show that $\hat{w}(\bar{\mu}) \geq s := v(\bar{\mu}) - \epsilon$, which will establish the result by Proposition 2. To that end, let $a \in A_R^*(\bar{\mu})$ be such that $\int u_S(a, \cdot) \, \mathrm{d}\bar{\mu} = v(\bar{\mu})$. Letting $\hat{\Theta} := \mathrm{supp}(\bar{\mu})$, the PUBR property then delivers some $\mu' \in \Delta\hat{\Theta}$ with $A^*(\mu') = \{a\}$. From linearity of expected utility in beliefs, it follows that $A_R^*(\mu) = \{a\}$ for any proper convex combination μ of μ' and $\bar{\mu}$. We may therefore assume without loss, replacing μ' with such a convex combination close enough to $\bar{\mu}$, that $\int u_S(a, \cdot) \, \mathrm{d}\mu' > s$.

Define now the sets of beliefs,

$$\hat{D} := \mathrm{co}\{\mu \in \Delta\hat{\Theta} : \ w(\mu) > s\} \text{ and } \tilde{D} := \left\{\tilde{\mu} \in \Delta\hat{\Theta} : \ [\mathrm{co}\{\bar{\mu}, \tilde{\mu}\}] \setminus \{\bar{\mu}\} \subseteq D\right\}.$$

Let us establish that $\tilde{D}$ contains a nonempty set that is relatively open in $\Delta\hat{\Theta}$. Our choice of μ' ensures that $\mu' \in \tilde{D}$. We will now show that, in the relative topology on $\Delta\hat{\Theta}$, the belief μ' is in fact interior in $\tilde{D}$. To that end, observe first that $\hat{D}$ is open in $\Delta\hat{\Theta}$, because w is lower semicontinuous and the convex hull of an open set is open. As $\mu' \in \tilde{D} \subseteq \hat{D}$, some open neighborhood $N \subseteq \mathrm{ca}(\Theta)$ of the zero measure therefore exists such that $(\mu' + N) \cap \Delta\hat{\Theta} \subseteq \hat{D}$.

To see that $(\mu' + \frac{1}{2}N) \cap \Delta\hat{\Theta} \subseteq \tilde{D}$, consider an arbitrary $\eta \in N$ such that $\mu' + \frac{1}{2}\eta \in \Delta\hat{\Theta}$ and an arbitrary $\lambda \in (0, 1]$. Observe that

$$(1-\lambda)\bar{\mu} + \lambda(\mu' + \tfrac{1}{2}\eta) = (1 - \tfrac{\lambda}{2})\left(\frac{1-\lambda}{1-\frac{\lambda}{2}}\,\bar{\mu} + \frac{\frac{\lambda}{2}}{1-\frac{\lambda}{2}}\,\mu'\right) + \tfrac{\lambda}{2}\,(\mu' + \eta),$$

which is in $\hat{D}$ because $\mu' \in \tilde{D}$ and $\hat{D}$ is convex.

Above, we showed that $\tilde{D}$ has nonempty interior, in the relative topology on $\Delta\hat{\Theta}$. Because the set $\{\mu \in \Delta\hat{\Theta} : \gamma\mu \leq \bar{\mu} \text{ for some } \gamma \in (0, 1)\}$ is dense in $\Delta\hat{\Theta}$ by Lemma 2 from Lipnowski and Mathevet (2018), it follows that some $\tilde{\mu} \in \tilde{D}$ and $\gamma \in (0, 1)$ exist with $\gamma\tilde{\mu} \leq \bar{\mu}$. Finally, because $\hat{w}$ is concave, $\tilde{\mu} \in \tilde{D}$, and

$$\bar{\mu} = \tfrac{\gamma}{\gamma+\lambda(1-\gamma)}\left[(1-\lambda)\bar{\mu} + \lambda\tilde{\mu}\right] + \tfrac{\lambda(1-\gamma)}{\gamma+\lambda(1-\gamma)}\left[\frac{\bar{\mu} - \gamma\tilde{\mu}}{1-\gamma}\right],$$

it follows that every $\lambda \in (0, 1]$ has

$$\begin{aligned}
\hat{w}(\bar{\mu}) &\geq \tfrac{\gamma}{\gamma+\lambda(1-\gamma)}\hat{w}\left((1-\lambda)\bar{\mu} + \lambda\tilde{\mu}\right) + \tfrac{\lambda(1-\gamma)}{\gamma+\lambda(1-\gamma)}\hat{w}\left(\tfrac{\bar{\mu}-\gamma\tilde{\mu}}{1-\gamma}\right) \\
&\geq \tfrac{\gamma}{\gamma+\lambda(1-\gamma)}s + \tfrac{\lambda(1-\gamma)}{\gamma+\lambda(1-\gamma)}\min u_S(A \times \Theta) \\
&\to s \text{ as } \lambda \to 0.
\end{aligned}$$

Therefore $\hat{w}(\bar{\mu}) \geq s$, as desired. □

The next lemma shows the global PUBR property is generic in finitary environments.

Lemma 2. *Given finite A and Θ, the set $\mathcal{U}_R \subseteq \mathbb{R}^{A\times\Theta}$ of R objectives satisfying PUBR on $\Delta\Theta$ is open, dense, and of full Lebesgue measure.*

Proof. Toward showing these properties, let us note an algebraic characterization of $\mathcal{U}_R$. Define the finite index set $\mathbb{I} := \{(a, \hat{\Theta}) : \ a \in A, \ \emptyset \neq \hat{\Theta} \subseteq \Theta\}$ and, for each $i = (a, \hat{\Theta}) \in \mathbb{I}$, define

$$\begin{aligned}
\varphi_i : \mathbb{R}^{A\times\Theta} &\to \mathbb{R} \\
u_R &\mapsto \max_{\mu\in\Delta\hat{\Theta}} \min_{a'\in A\setminus\{a\}} \int \left[u_R(a,\cdot) - u_R(a',\cdot)\right] \mathrm{d}\mu.
\end{aligned}$$

Then, clearly, $\mathcal{U}_R = \{u_R \in \mathbb{R}^{A\times\Theta} : \ \varphi_i(u_R) \text{ is nonzero for every } i \in \mathbb{I}\} = \bigcap_{i\in\mathbb{I}} \varphi_i^{-1}(\mathbb{R} \setminus \{0\})$.

We can therefore show $\mathcal{U}_R \subseteq \mathbb{R}^{A\times\Theta}$ is open, dense, and of full Lebesgue measure by establishing that $\varphi_i^{-1}(\mathbb{R} \setminus \{0\})$ enjoys these properties for every $i = (a, \hat{\Theta}) \in \mathbb{I}$. First, note it is open because (by Berge's theorem) φ_i is continuous. To show it is of full measure (hence also dense), define $\bar{z} := [\mathbf{1}_{\tilde{a}=a}]_{\tilde{a}\in A,\ \tilde{\theta}\in\Theta} \in \mathbb{R}^{A\times\Theta}$, and observe that $\varphi_i(u_R + \lambda\bar{z}) = \varphi_i(u_R) + \lambda$ for any $u_R \in \mathbb{R}^{A\times\Theta}$ and any $\lambda \in \mathbb{R}$. Now, fixing some $\bar{\theta} \in \Theta$, observe that we can decompose the vector space of all R objectives as the direct sum $\mathbb{R}^{A\times\Theta} = Y \oplus Z$, where

$$Y := \{u_R \in \mathbb{R}^{A\times\Theta} : \ u_R(a, \bar{\theta}) = 0\} \text{ and } Z := \{\lambda\bar{z} : \ \lambda \in \mathbb{R}\}.$$

As $\varphi_i(y + \lambda\bar{z}) = \varphi(y) + \lambda$ for any $y \in Y$ and $\lambda \in \mathbb{R}$, and a singleton is Lebesgue-null in $\mathbb{R} \cong Z$, it follows from the law of iterated expectations that $\varphi_i^{-1}(0)$ is Lebesgue-null as well. The proposition follows. □

Given the above results, generic uniqueness is immediate:

Proof of Proposition 3. The result follows directly from Theorem 1 and Lemma 2. □

Now, we show S has a unique payoff in finite-action ordered environments.

Proof of Proposition 4. That A is countable implies (by the revelation principle) that some $p \in \mathcal{I}(\mu_0)$ has $\int v \, \mathrm{d}p = \hat{v}(\mu_0)$ and puts full measure on a countable set $D \subseteq \Delta\Theta$. Removing from D any beliefs on which p puts zero mass, every μ-atom for $\mu \in D$ is also a μ_0-atom. Using this fact, let us show the PUBR property holds on D, delivering payoff uniqueness by Theorem 1.

Let $\mu \in D$ and $a \in A_R^*(\mu)$. We want to find $\mu' \in \Delta\Theta$ with $A^*(\mu') = \{a\}$ and $\mathrm{supp}(\mu') \subseteq \mathrm{supp}(\mu)$. Strict concavity of $u_R(\cdot, \theta)$ for every $\theta \in \Theta$ implies strict concavity of $\int u_R(\cdot, \theta) \, \mathrm{d}\mu(\theta)$, so that $|A_R^*(\mu)| \leq 2$. As we can take $\mu' = \mu$ if $|A_R^*(\mu)| = 1$, focus without loss on the case that $A_R^*(\mu) = \{a, a'\}$ for some $a' \in A \setminus \{a\}$. By symmetry, we may assume without loss that $a' > a$.

By hypothesis, we cannot have $A_R^*(\delta_\theta) = \{a, a'\}$ for any μ_0-atom $\theta \in \Theta$, and that $\mu \in D$ implies we cannot have $\mu = \delta_\theta$ for any $\theta \in \Theta$ that is not a μ_0-atom. Hence, μ is nondegenerate. We can therefore express it as $\mu = \frac{1}{2}(\mu_L + \mu_R)$ for some $\mu_L, \mu_R \in \Delta\Theta$ such that μ_R strictly first-order-stochastically dominates μ_L. Hence, for any $\epsilon \in (0, 1)$, the distribution $\mu^\epsilon := \frac{1+\epsilon}{2}\mu_L + \frac{1-\epsilon}{2}\mu_R$ is strictly first-order-stochastically dominated by μ, implying (by strictly increasing differences)

$$\int \left[u_R(a, \cdot) - u_R(a', \cdot)\right] \, \mathrm{d}\mu^\epsilon > \int \left[u_R(a, \cdot) - u_R(a', \cdot)\right] \, \mathrm{d}\mu = 0.$$

Moreover, A_R^* is upper hemicontinuous (by Berge's theorem) and A is discrete (being finite), so that sufficiently small ϵ has $A_R^*(\mu^\epsilon) \subseteq A_R^*(\mu)$. Hence, $\mu' = \mu^\epsilon$ is as required for small enough ϵ. □

A.3. Proofs for Section 3

To better understand the set of χ-PBE and the S payoffs they can generate, connecting this solution concept to the analysis of Lipnowski, Ravid, and Shishkin (2022) is useful. That paper defines a notion of a **χ-equilibrium** and characterizes the S payoffs such a solution can generate.[13] Roughly, a χ-equilibrium specifies an experiment $\psi \in \Psi$, together with continuation play and continuation beliefs that satisfy the incentive and Bayesian properties in the partial-credibility game, but with no requirement that the initial experiment ψ be chosen optimally. From the definitions in the present paper and in LRS1, the following is immediate.

Fact 1. *The quadruple $\langle \psi, \sigma, \alpha, \pi \rangle$ is a χ-PBE if and only if:*

1. *For every $\tilde{\psi} \in \Psi$, the quadruple $\langle \tilde{\psi}, \sigma(\cdot, \tilde{\psi}), \alpha(\cdot, \tilde{\psi}), \pi(\cdot, \tilde{\psi}) \rangle$ is a χ-equilibrium.*

[13] The model in LRS1 assumes the message space to be uncountable. However, given that $|M| \geq 2|\Theta|$, it follows readily from Carathéodory's theorem and Lemma 1 of LRS1 that the χ-equilibrium payoff set is unchanged.

2. *We have* $\psi \in \operatorname{argmax}_{\tilde{\psi}\in\Psi} s_{\tilde{\psi}}$*, where* $s_{\tilde{\psi}}$ *is the S payoff induced by* $\langle \tilde{\psi}, \sigma(\cdot,\tilde{\psi}), \alpha(\cdot,\tilde{\psi}), \pi(\cdot,\tilde{\psi})\rangle$.

In particular, every χ*-PBE payoff is a* χ*-equilibrium payoff.*

We will use the following notation, for an S payoff that LRS1 characterizes, throughout.[14]

Notation 1. *Let* $v^*_\chi(\mu_0)$ *denote the highest* χ*-equilibrium S payoff (given prior* μ_0*).*

Toward constructing adversarial χ-PBE that give S an undesirable payoff in response to off-path experiment choices, we begin with a technical lemma showing any reporting protocol comprises part of a χ-equilibrium in which R always chooses from a given restricted set of best responses.

Lemma 3. *If* $\tilde{V} \subseteq V$ *is a Kakutani correspondence and* ψ *is any official reporting protocol, then some* χ*-equilibrium* $(\psi, \tilde{\sigma}, \tilde{\alpha}, \tilde{\pi})$ *exists such that* $u_S(\tilde{\alpha}) \in \tilde{V}(\tilde{\pi})$.

Proof. Let $\Pi := (\Delta\Theta)^M$ be the set of all R belief mappings and define correspondences

$$\begin{aligned}
\hat{S}\colon \Pi &\rightrightarrows \mathbb{R}\\
\tilde{\pi} &\mapsto \left[\max_{m\in M} \min \tilde{V}(\tilde{\pi}(m)), \max_{m\in M} \max \tilde{V}(\tilde{\pi}(m))\right],\\
\hat{M}\colon \Pi &\rightrightarrows M\\
\tilde{\pi} &\mapsto \left\{m \in M\colon\ \tilde{V}(\tilde{\pi}(m)) \cap \hat{S}(\tilde{\pi}) \neq \varnothing\right\}.
\end{aligned}$$

Observe that $\hat{S}$ is Kakutani, since $\tilde{V}$ is Kakutani and a finite maximum or minimum of upper or lower semicontinuous functions inherits the same semicontinuity. Therefore, $\hat{M}$ is nonempty-valued with closed graph. Now let $\Sigma := (\Delta M)^\Theta$ and consider the correspondence mapping belief maps into S-IC influencing strategies (assuming R's strategy delivers S values from $\tilde{V}$)

$$\begin{aligned}
\hat{\Sigma}\colon\ \Pi &\rightrightarrows \Sigma\\
\tilde{\pi} &\mapsto \left\{\tilde{\sigma} \in \Sigma\colon\ \cup_{\theta\in\Theta} \operatorname{supp}(\tilde{\sigma}(\theta)) \subseteq \hat{M}(\tilde{\pi})\right\},
\end{aligned}$$

and the correspondence mapping influencing strategies into consistent belief maps

$$\begin{aligned}
\hat{\Pi}\colon\ &\Sigma \rightrightarrows \Pi,\\
&\tilde{\sigma} \mapsto \Big\{\tilde{\pi} \in \Pi\colon\ \tilde{\pi}(\theta|m) \int_\Theta \left[\chi\, \mathrm{d}\psi(m|\cdot) + (1-\chi)\tilde{\sigma}(m|\cdot)\right] \mathrm{d}\mu_0\\
&\qquad\qquad = \left[\chi(\theta)\psi(m|\theta) + (1-\chi(\theta))\tilde{\sigma}(m|\theta)\right]\mu_0(\theta), \forall \theta \in \Theta, m \in M\Big\}.
\end{aligned}$$

It then follows that $\hat{\Sigma}$ and $\hat{\Pi}$ are both Kakutani. Therefore, the Kakutani fixed point theorem delivers some $\tilde{\sigma} \in \Sigma$ and $\tilde{\pi} \in \Pi$ such that $\tilde{\sigma} \in \hat{\Sigma}(\tilde{\pi})$ and $\tilde{\pi} \in \hat{\Pi}(\tilde{\sigma})$. Now, take any $s_i \in$

[14] The value's characterization (Theorem 1 of LRS1) remains valid in the present setting in light of Corollary 1 of LRS1.

$\hat{S}(\tilde{\pi})$ and let $D := \tilde{\pi}(M)$. Note that $s_i \wedge \tilde{V}|_D$ is nonempty-valued and so admits a selector $\phi\colon D \to \mathbb{R}$.[15] Therefore, some $\hat{\alpha}\colon D \to \Delta A$ exists such that $u_S(\hat{\alpha}(m)) = \phi(m)$. Next, define $\tilde{\alpha} := \hat{\alpha} \circ \tilde{\pi}\colon M \to \Delta(A)$. It is then easy to verify that $(\psi, \tilde{\sigma}, \tilde{\alpha}, \tilde{\pi})$ is a χ-equilibrium. □

The following lemma shows the payoff $\hat{w}(\mu_0)$ dominates some χ-equilibrium payoff.

Lemma 4. *Every official reporting protocol ψ admits some χ-equilibrium $\langle \psi, \tilde{\sigma}, \tilde{\alpha}, \tilde{\pi} \rangle$ with ex-ante S payoff weakly below $\hat{w}(\mu_0)$.*

Proof. Without loss, we can focus on the case that μ_0 is of full support. Indeed, if we construct a χ-equilibrium as desired (for official reporting protocol $\psi|_{\Theta_0}$) in the restricted model with state space $\Theta_0 := \text{supp}(\mu_0)$, then this equilibrium can be extended to a χ-equilibrium in the true model, by fixing any $\theta_0 \in \Theta_0$ and extending σ to Θ via $\sigma(\theta) := \sigma(\theta_0)$ for $\theta \in \Theta \setminus \Theta_0$.

Note the lemma follows directly from Lemma 3 if we can find a Kakutani subcorrespondence $\tilde{V} \subseteq V$ such that the concave envelope of its upper selection satisfies $\text{cav}[\max \tilde{V}](\mu_0) \leq \hat{w}(\mu_0)$. Let us show $\tilde{V} := [w, z]$ has this property, where z is the upper semicontinuous envelope of w, given by

$$\begin{aligned} z\colon \Delta\Theta &\to \mathbb{R} \\ \mu &\mapsto \limsup_{\mu' \to \mu} w(\mu). \end{aligned}$$

First, $\tilde{V}$ is a Kakutani subcorrespondence of V since z is upper semicontinuous and lies above the lower semicontinuous function w (and hence lies below v). All that remains, then, is to show that $\hat{w}(\mu_0) \geq \text{cav}[\max \tilde{V}](\mu_0) = \hat{z}(\mu_0)$. To do so, let us establish the stronger claim that $\hat{z}|_D = \hat{w}|_D$, where $D \subseteq \Delta\Theta$ is the set of full-support beliefs.

Define

$$\begin{aligned} \tilde{z}\colon \Delta\Theta &\to \mathbb{R} \\ \mu &\mapsto \limsup_{\mu' \to \mu} \hat{w}(\mu). \end{aligned}$$

It follows from concavity of $\hat{w}$ that $\tilde{z}$ is concave too. Hence, because $\tilde{z} \geq z$ and $\tilde{z}$ is upper semicontinuous by construction, it follows that $\tilde{z} \geq \hat{z}$.[16] Moreover, Theorem 10.4 from Rockafellar (1970) implies the concave function $\hat{w}|_D$ is continuous. Hence, $\tilde{z}|_D = \hat{w}|_D$ by the definition of $\tilde{z}$. That $\tilde{z} \geq \hat{z} \geq \hat{w}$ then implies $\hat{z}|_D = \hat{w}|_D$. □

Here, we provide a sufficient condition for a payoff to be compatible with χ-PBE for an arbitrary credibility level.

Lemma 5. *If $s \in \left[\hat{w}(\mu_0) \wedge v^*_\chi(\mu_0),\ v^*_\chi(\mu_0)\right]$, then s is a χ-PBE payoff for S.*

[15] That is, some $\phi\colon D \to \mathbb{R}$ has $\phi \leq s_i$ and $\phi(\mu) \in \tilde{V}(\mu)$ for every $\mu \in D$.

[16] In fact, one can show $\tilde{z} = \hat{z}$, but this fact is immaterial to the present argument.

Proof. First, we argue a χ-equilibrium exists with ex-ante S payoff s. To that end, observe that Lemma 1 from LRS1 implies $(\delta_{\mu_0}, w(\mu_0), w(\mu_0))$ is a χ-equilibrium outcome (as witnessed by $k = \chi$ and $g = b = \delta_{\mu_0}$). But then, as Theorem 1 from LRS1 says $v^*_\chi(\mu_0)$ is the highest χ-equilibrium S payoff, it follows from Lemma 7 of LRS1 that every payoff in $[w(\mu_0), v^*_\chi(\mu_0)]$ is a χ-equilibrium S payoff. Thus, s is a χ-equilibrium S payoff because $\hat{w}(\mu_0) \wedge v^*_\chi(\mu_0) \geq w(\mu_0)$. So let $(\psi, \tilde{\sigma}, \tilde{\alpha}, \tilde{\pi})$ be some χ-equilibrium generating S payoff s.

Finally, to build a χ-PBE, construct σ, α, and π as follows. First, let $\sigma(\cdot, \psi) := \tilde{\sigma}$, $\alpha(\cdot, \psi) := \tilde{\alpha}$, and $\pi(\cdot, \psi) := \tilde{\pi}$. Second, given any $\tilde{\psi} \in \Psi \setminus \{\psi\}$, let $\langle \tilde{\psi}, \sigma(\cdot, \tilde{\psi}), \alpha(\cdot, \tilde{\psi}), \pi(\cdot, \tilde{\psi}) \rangle$ be some χ-equilibrium—which Lemma 4 exists—with ex-ante S value of at most $\hat{w}(\mu_0) \wedge v^*_\chi(\mu_0)$; this χ-equilibrium necessarily yields S payoff no greater than $v^*_\chi(\mu_0)$ by definition of the latter. Since $s \geq \hat{w}(\mu_0) \wedge v^*_\chi(\mu_0)$, the quadruple $\langle \psi, \sigma, \alpha, \pi \rangle$ is a χ-PBE as desired. □

Next, we characterize the highest χ-PBE payoff S can attain; it coincides with her highest χ-equilibrium payoff.

Lemma 6. *The highest χ-PBE payoff for S is $v^*_\chi(\mu_0)$, her highest χ-equilibrium payoff. This payoff is weakly increasing in χ.*

Proof. By Fact 1, no χ-PBE payoff is strictly higher than $v^*_\chi(\mu_0)$. Meanwhile, Lemma 5 implies $v^*_\chi(\mu_0)$ is a χ-PBE payoff. The last statement follows from LRS1's Corollary 3. □

Now, we characterize the set of all 1-PBE payoffs S can attain; it coincides with the 1-equilibrium payoffs.

Lemma 7. *The set of all 1-PBE payoffs for S is $[\hat{w}(\mu_0), \hat{v}(\mu_0)]$.*

Proof. Theorem 1 from LRS1 tells us $\hat{v}^*_1 = \hat{v}$, and so Lemma 5 says all payoffs in $[\hat{w}(\mu_0), \hat{v}(\mu_0)]$ are 1-PBE payoffs for S. Conversely, every 1-PBE generates a 1-equilibrium by Fact 1, and so generates an equilibrium (in the sense of Definition 1) by throwing away the influencing S behavior, the set of 1-PBE payoffs is a subset of the set of equilibrium payoffs. Hence, the other containment follows from Proposition 1. □

Note now that our partial characterization of χ-PBE follows readily from our intermediate results.

Proof of Proposition 5. Let S_χ denote the set of χ-PBE payoffs for S, and let $\bar{S} := [\hat{w}(\mu_0), \hat{v}(\mu_0)]$. First, Lemma 6 says $v^*(\mu_0) = \max S_\chi$, and $v^*(\mu_0) \leq \hat{v}(\mu_0) = \max \bar{S}$ by Theorem 1 from LRS1. Meanwhile, Lemma 5 implies $\bar{S} \cap (-\infty, \max S_\chi] \subseteq S_\chi$. Hence S_χ is weakly below $\bar{S}$ in the strong set order. Finally, that $S_1 = \bar{S}$ is exactly Lemma 7. □

Finally, we prove our main imperfect-credibility result.

Proof of Proposition 6. Fix any full-support prior $\mu_0 \in \Delta\Theta$, and define $\underline{s}_\chi := w^*_\chi(\mu_0)$ for each $\chi \in [0, 1]$; in particular, $\underline{s}_1 = \hat{w}(\mu_0)$ by Lemma 7. To prove the equivalence, it suffices to show that $\lim_{\chi \nearrow 1} \underline{s}_\chi = \underline{s}_1$. Further, Lemma 5 tells us $\underline{s}_\chi \leq \underline{s}_1$ for every $\chi \in [0, 1]$, so we need only show $\liminf_{\chi \nearrow 1} \underline{s}_\chi \geq \underline{s}_1$, which we do below.

Take an arbitrary $\epsilon > 0$. By definition of $\underline{s}_1$, some $p \in \mathcal{I}(\mu_0)$ exists such that $\int w \, \mathrm{d}p > \underline{s}_1 - \epsilon$. Moreover, by Lemma 1, we may further assume $|\mathrm{supp}(p)| \leq |\Theta| \leq |M|$. For each $\mu \in \mathrm{supp}(p)$, let $N(\mu) \subseteq \Delta\Theta$ be some open neighborhood of μ on which $w > w(\mu) - \epsilon$, which exists because w is lower semicontinuous. Because $\mathrm{supp}(p)$ is finite, which in particular implies $p(\mu) > 0$ for every $\mu \in \mathrm{supp}(p)$, some $\underline{\chi} \in (0, 1)$ is such that

$$\tfrac{1}{\underline{\chi} p(\mu) + (1 - \underline{\chi})} \left[\underline{\chi} p(\mu) \mu + (1 - \underline{\chi}) \Delta\Theta \right] \subseteq N(\mu)$$

for each $\mu \in \mathrm{supp}(p)$, and so (since $\Delta\Theta$ is convex) the containment holds as well when we replace $\underline{\chi}$ with any $\chi \in (\underline{\chi}, 1)$.

Consider now, any $\chi \in (\underline{\chi}, 1)$, and fix some χ-PBE $\langle \psi, \sigma, \alpha, \pi \rangle$ generating S payoff $s \in \mathbb{R}$. Let $\tilde{\psi}_p \in \Psi$ be as defined in Proposition 1's proof, so that $p_{\tilde{\psi}_p} = p$. Modifying $\tilde{\psi}_p$ if necessary, we may assume without loss that any two distinct messages from $M_p := \{m \in M : \int_\Theta \tilde{\psi}_p(m|\cdot) \, \mathrm{d}\mu_0 > 0\}$ would generate distinct beliefs. Hence, every belief $\mu \in \mathrm{supp}(p)$ admits a unique $m_\mu \in M_p$ such that every $\theta \in \Theta$ has $\frac{\tilde{\psi}_p(m|\theta)\mu_0(\theta)}{\int_\Theta \tilde{\psi}_p(m|\cdot) \, \mathrm{d}\mu_0} = \mu(\theta)$. If S chooses official reporting protocol $\tilde{\psi}_p$ and sends message m_μ for some $\mu \in \mathrm{supp}(p)$, the Bayesian property implies $\pi(m_\mu, \tilde{\psi}_p) \in \frac{1}{\chi p(\mu) + (1-\chi)} \left[\underline{\chi} p(\mu) \mu + (1 - \underline{\chi}) \Delta\Theta \right] \subseteq N(\mu)$, so that R's best response property implies S has continuation value exceeding $w(\mu) - \epsilon$. But because S chooses $\psi \in \Psi$ optimally, and has the option to choose $\tilde{\psi}_p$, it must be that

$$\begin{aligned}
s &\geq \int_\Theta \left(\int_M \left[\int_A u_S(a) \, \mathrm{d}\alpha(a|m, \tilde{\psi}_p) \right] \mathrm{d} \left[\chi \tilde{\psi}_p(m|\theta) + (1 - \chi) \, \sigma(m|\theta, \tilde{\psi}_p) \right] \right) \mathrm{d}\mu_0(\theta) \\
&\geq \chi \int_\Theta \int_M \left[\int_A u_S(a) \, \mathrm{d}\alpha(a|m, \tilde{\psi}_p) \right] \mathrm{d}\tilde{\psi}_p(m|\theta) \, \mathrm{d}\mu_0(\theta) + (1 - \chi) \min w(\Delta\Theta) \\
&\geq \chi \int_{\Delta\Theta} (w - \epsilon) \, \mathrm{d}p + (1 - \chi) \min w(\Delta\Theta) \\
&\geq \chi(\underline{s}_1 - \epsilon) + (1 - \chi) \min w(\Delta\Theta)
\end{aligned}$$

Because s was the payoff from an arbitrary χ-PBE, we learn that every $\chi \in (\underline{\chi}, 1)$ has $\underline{s}_\chi \geq \chi(\underline{s}_1 - \epsilon) + (1 - \chi) \min w(\Delta\Theta)$, which converges to $\underline{s}_1 - \epsilon$ as χ converges to 1. Hence, $\liminf_{\chi \nearrow 1} \underline{s}_\chi \geq \underline{s}_1 - \epsilon$. But ϵ was itself arbitrary, so that $\liminf_{\chi \nearrow 1} \underline{s}_\chi \geq \underline{s}_1$, as desired. □