Online Appendix for "Credible Persuasion"

B.1 Relationship to Rochet (1987)

The u_S-cyclical monotonicity condition in our characterization closely resembles the cyclical monotonicity condition for implementing transfers in Rochet (1987). The reader might wonder why cyclical monotonicity arises in our setting despite the lack of transfers. The connection is best summarized by the following three equivalent conditions from optimal transport theory (see, for example, Theorem 5.10 of Villani (2008)).

Kantorovich Duality. *Suppose X and Y are both finite sets, and $u : X \times Y \to \mathbb{R}$ is a real-valued function. Let μ be a probability measure on X and ν be a probability measure on Y, and $\Pi(\mu, \nu)$ be the set of probability measures on $X \times Y$ such that the marginals on X and Y are μ and ν, respectively. Then for any $\pi^* \in \Pi(\mu, \nu)$, the following three statements are equivalent:*

1. $\pi^* \in \arg\max_{\pi \in \Pi(\mu,\nu)} \sum_{x,y} \pi(x, y)u(x, y)$;
2. *π^* is u-cyclically monotone. That is, for any n and $(x_1, y_1), ..., (x_n, y_n) \in supp(\pi^*)$,*

$$\sum_{i=1}^{n} u(x_i, y_i) \geq \sum_{i=1}^{n} u(x_i, y_{i+1}).$$

3. *There exists $\psi : Y \to \mathbb{R}$ such that for any $(x, y) \in supp(\pi^*)$ and any $y' \in Y$,*[23]

$$u(x, y) - \psi(y) \geq u(x, y') - \psi(y').$$

Our Theorem 1 builds on the equivalence between 1 and 2 in the Kantorovich duality theorem above to show the equivalence between credibility and u_S-cyclical monotonicity.

Rochet (1987)'s classic result on implementation with transfers follows from the equivalence between 2 and 3. To see this, consider a principal-agent problem where the agent's private type space is Θ with full-support prior μ_0, and the principal's action space is A. The agent's payoff is $u(\theta, a) - t$, where t is the transfer she makes to the principal. Given an allocation rule $q : \Theta \to A$, let $v_q(\theta, \theta') \equiv u(\theta, q(\theta'))$ denote the payoff that a type-θ agent obtains from the allocation intended for type θ'. Let $X = Y = \Theta$ and $\mu = \nu = \mu_0$ in the Kantorovich

[23] This statement can also be equivalently written as: there exists $\phi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$, such that $\phi(x) + \psi(y) \geq u(x, y)$ for all x and y, with equality for (x, y) in the support of π^*.

duality theorem above, and consider the distribution $\pi^* \in \Pi(\mu, \nu)$ defined by

$$\pi^*(\theta, \theta') = \begin{cases} \mu_0(\theta) & \text{if } \theta = \theta' \\ 0 & \text{otherwise} \end{cases}$$

By the equivalence of 2 and 3 in the Kantorovich duality theorem, π^* is v_q-cyclically monotone if and only if there exists $\psi : \Theta \to \mathbb{R}$ such that for all $\theta, \theta' \in \Theta$, $v_q(\theta, \theta) - \psi(\theta) \geq v_q(\theta, \theta') - \psi(\theta')$. That is,

$$u(\theta, q(\theta)) - \psi(\theta) \geq u(\theta, q(\theta')) - \psi(\theta'),$$

so the allocation rule q can be implemented by the transfer rule $\psi : \Theta \to \mathbb{R}$. The v_q-cyclical monotonicity condition says that for every sequence $\theta_1, ..., \theta_n \in \Theta$ with $\theta_{n+1} \equiv \theta_1$,

$$\sum_{i=1}^{n} u(\theta_i, q(\theta_i)) \geq \sum_{i=1}^{n} u(\theta_i, q(\theta_{i+1})).$$

This is exactly the cyclical monotonicity condition in Rochet (1987).

When $X = \Theta$ is interpreted as the set of an agent's true types and $Y = \Theta$ interpreted as the set of reported types, the distribution π^* constructed in the previous paragraph can be interpreted as the agent's truthful reporting strategy. Based on this interpretation, Rahman (2010) uses the duality between 1 and 3 to show that the incentive compatibility of truthful reporting subject to quota constraints is equivalent to implementability with transfers.

B.2 Finite-Sample Approximation

As discussed in Section 2.1, we interpret our model as one where the Sender designs an information structure that assigns scores to a large population of realized θ's; in particular, our model abstracts away from sampling variation, so there is no uncertainty in the population's realized type distribution. In this section we explicitly allow sampling variation by considering a finite-sample model where the Sender observes N random i.i.d. draws from Θ, and assigns each realized θ a message $m \in M$, subject to certain message quotas—in particular, these message quotas substitute for the Sender's commitment to message distributions in the continuum model. We will show that credible and R-IC profiles in our continuum model are approximated by credible and R-IC profiles in the discrete model when the sample size is large.

Consider a finite i.i.d sample of size N drawn from the type space Θ according to the prior distribution μ_0. The set of all possible empirical distributions over Θ in this N-sample captures the sampling variations in the realized type distribution, and can be written as

$$\mathcal{F}_\Theta^N = \Big\{ f/N : f \in \mathbb{N}^{|\Theta|}, \sum_{\theta\in\Theta} f(\theta) = N \Big\}.$$

The Sender assigns each realized θ in the N-sample a message $m \in M$, which leads to an N-sample of messages. Let

$$\mathcal{F}_M^N = \Big\{ f/N : f \in \mathbb{N}^{|M|}, \sum_{m\in M} f(m) = N \Big\}$$

denote the set of N-sample empirical distributions over messages. Lastly, for a pair of state and message distributions (μ, ν), let

$$X^N(\mu,\nu) = \Big\{ f/N : f \in \mathbb{N}^{|\Theta|\times|M|}, \sum_{\theta} f(\theta,\cdot) = N\nu(\cdot), \sum_{m} f(\cdot, m) = N\mu(\cdot) \Big\}$$

denote the set of N-sample empirical joint distributions over states and messages that have marginals μ and ν. Notice that $X^N(\mu,\nu) \neq \emptyset$ if and only if $\mu \in \mathcal{F}_\Theta^N$ and $\nu \in \mathcal{F}_M^N$.[24]

Let us now define the N-sample analogue of credible and R-IC profiles. We consider a Sender who assigns a message $m \in M$ to each realized $\theta \in \Theta$ subject to a message quota $\nu^N \in \mathcal{F}_M^N$. An N-sample profile is therefore a triple $(\nu^N, \phi^N, \sigma^N)$, where $\phi^N : \mathcal{F}_\Theta^N \to \Delta(\Theta \times M)$ is a Sender's strategy that takes every realized empirical distribution over states $\mu^N \in \mathcal{F}_\Theta^N$ to a joint distribution $\phi^N(\mu^N) \in X^N(\mu^N, \nu^N)$; meanwhile, $\sigma^N : M \to A$ is a Receiver's strategy that assigns an action to each observed message.[25]

The definitions of Sender credibility and Receiver incentive compatibility in the N-sample setting mirror those in our continuum model. In particular, we say an N-sample profile $(\nu^N, \phi^N, \sigma^N)$ is credible if for each realized empirical distribution over Θ, the Sender always chooses an optimal assignment of messages subject to the message quotas specified in ν^N: $(\nu^N, \phi^N, \sigma^N)$ is credible if for every $\mu^N \in \mathcal{F}_\Theta^N$,

$$\phi(\mu^N) \in \operatorname*{arg\,max}_{\lambda^N \in X(\mu^N,\nu^N)} \sum_{\theta,m} \lambda^N(\theta,m) u_S(\theta, \sigma^N(m)).$$

We say the N-sample profile $(\nu^N, \phi^N, \sigma^N)$ is Receiver incentive compatible (R-IC) if σ^N best-responds to the Sender's strategy ϕ^N. In particular, let P^N denote the probability distribution

[24]Notice that for any $f \in \mathbb{N}^{|\Theta|\times|M|}$, the sum of any row or column has to be integer, so $X^N(\mu,\nu) = \emptyset$ if either $\mu \notin \mathcal{F}_\Theta^N$ or $\nu \notin \mathcal{F}_M^N$. On the other hand if $\mu \in F_\Theta^N$ and $\nu \in F_M^N$, a $\lambda \in X^N(\mu,\nu)$ can be constructed from the so-called Northwest corner rule.

[25]Note that our formulation of the Sender's strategy assumes that the Sender conditions her strategy only on the empirical distribution of the realized N samples, and ignores the identity of each individual sample point.

over $\mathcal{F}_\Theta^N$ induced by i.i.d. draws from the prior distribution $\mu_0 \in \Delta(\Theta)$, and let $\phi^N(\theta, m|\mu^N)$ be the probability assigned to (θ, m) in the joint distribution $\phi^N(\mu^N)$ chosen by the Sender. The profile $(\nu^N, \phi^N, \sigma^N)$ is R-IC if

$$\sigma^N \in \underset{\sigma':M\to A}{\arg\max} \sum_{\mu^N \in \mathcal{F}_\Theta^N} P^N(\mu^N) \sum_{\theta,m} \phi(\theta, m|\mu^N) u_R(\theta, \sigma'(m)).$$

Proposition 7 below shows that credible and R-IC profiles in the continuum model are approximated by credible and R-IC profiles in the N-sample model, provided N is sufficiently large. Note that in the second statement in Proposition 7, we distinguish a strictly credible profile (λ^*, σ^*) in the continuum model as one where λ^* is the unique maximizer in Definition 1; similarly, (λ^*, σ^*) is strictly R-IC if σ^* is the unique maximizer in Definition 2.

Proposition 7. *1. Let (λ^*, σ^*) be a profile in the continuum model. If for every $\varepsilon > 0$, there exists a finite credible profile $(\nu^N, \phi^N, \sigma^N)$ for some sample size N, such that $|\nu^N - \lambda^*_M| < \varepsilon$, $|\sigma^N - \sigma^*| < \varepsilon$ and $P(|\phi^N(F_\Theta^N) - \lambda^*| < \varepsilon) > 1 - \varepsilon$, then (λ^*, σ^*) is credible and R-IC.*

2. Suppose (λ^*, σ^*) is a strictly credible and strictly R-IC profile in the continuum model, then for each $\varepsilon > 0$ there exists a finite-sample credible and R-IC profile $(\nu^N, \phi^N, \sigma^N)$ such that $|\nu^N - \lambda^*_M| < \varepsilon$, $|\sigma^N - \sigma^*| < \varepsilon$ and $P(|\phi^N(F_\Theta^N) - \lambda^*| < \varepsilon) > 1 - \varepsilon$.

The first statement in Proposition 7 is analogous to the upper-hemicontinuity of Nash equilibrium correspondences: if a profile (λ^*, σ^*) in the continuous model can be arbitrarily approximated by credible and R-IC profiles in the finite model, then profile (λ^*, σ^*) must itself be credible and R-IC. Conversely, the second statement in Proposition 7 can be interpreted in a way similar to the lower-hemicontinuity of strict Nash equilibria: if a profile (λ^*, σ^*) in the continuous model is strictly credible and strictly R-IC, then it can be arbitrarily approximated by credible and R-IC profiles in the finite model.[26]

B.2.1 Proof of Proposition 7

For $\mu \in \Delta(\Theta)$ and $\nu \in \Delta(A)$, let $\Lambda(\mu, \nu) \equiv \{\lambda \in \Delta(\Theta \times A) : \lambda_\Theta = \mu, \lambda_A = \nu\}$ denote the set of joint distributions on $\Theta \times A$ that with marginals given by μ and ν. The following lemmas will be useful in our proofs.

[26]We conjecture that for generic payoffs, an outcome distribution induced by a credible and R-IC profile can be approximated by their strict counterparts, but we have not been able to identify a proof.

Lemma 10. *The correspondence*

$$B(\mu,\nu) \equiv \arg\max_{\lambda\in\Lambda(\mu,\nu)} \sum_{\theta,m} \lambda(\theta,m)u_S(\theta,\sigma(m))$$

is upper hemi-continuous with respect to (μ,ν)*. Thus, the value function*

$$V(\mu,\nu) \equiv \max_{\lambda\in\Lambda(\mu,\nu)} \sum_{\theta,m} \lambda(\theta,m)u_S(\theta,\sigma(m))$$

is continuous.

Proof. The first statement follows directly from Theorem 1.50 of Santambrogio (2015). For any sequence $(\lambda_k,\mu_k,\nu_k) \to (\lambda,\mu,\nu)$ so that $\lambda_k \in B(\mu_k,\nu_k)$ for all k, we have $\lambda \in B(\mu,\nu)$. Then $V(\mu,\nu) = \sum_{\theta,m} \lambda(\theta,m)u_S(\theta,\sigma(m)) = \lim_{k\to\infty} \sum_{\theta,m} \lambda_k(\theta,m)u_S(\theta,\sigma(m)) = \lim_{k\to\infty} V(\nu_k,\nu_k)$, which proves the second statement. □

Lemma 11. *Suppose* $\mu \in \mathcal{F}_\Theta^N$ *and* $\nu \in \mathcal{F}_M^N$*, then the extreme points of* $\Lambda(\mu,\nu)$ *are contained in* $X^N(\mu,\nu)$*.*

Proof. Consider the set $Y^N(\mu,\nu) = \{f \in \mathbb{R}_+^{|\Theta|\times|M|} : \sum_\theta f(\theta,\cdot) = N\nu(\cdot), \sum_m f(\cdot,m) = N\mu(\cdot)\}$. From Corollary 8.1.4 of Brualdi (2006), the extreme points of $Y^N(\mu,\nu)$ are contained in $Z^N(\mu,\nu) = \{f \in \mathbb{N}^{|\Theta|\times|M|} : \sum_\theta f(\theta,\cdot) = N\nu(\cdot), \sum_m f(\cdot,m) = N\mu(\cdot)\}$. Since $\Lambda(\mu,\nu) = \{\frac{f}{N} : f \in Y^N(\mu,\nu)\}$ and $X^N(\mu,\nu) = \{\frac{f}{N} : f \in Z^N(\mu,\nu)\}$, the extreme points of $\Lambda(\mu,\nu)$ are contained in $X^N(\mu,\nu)$. □

Lemma 12. *Let* X, Y *be metric spaces and* $\Gamma : X \rightrightarrows Y$ *be a correspondence. If* Γ *is upper hemi-continuous at* $x_0 \in X$*, and* $\Gamma(x_0) = \{y_0\}$ *for some* $y_0 \in Y$*, then* Γ *is continuous at* x_0*.*

Proof. For any $\varepsilon > 0$, let $B(y_0,\varepsilon) \subseteq Y$ denote the ε-ball centered at y_0. We will show that there exists $\delta > 0$ such that for all $|x - x_0| < \delta$, $\Gamma(x) \cap B(y_0,\varepsilon) \neq \emptyset$, which implies that Γ is lhc and therefore continuous.

Now since $\Gamma(x) = \{y_0\} \subseteq B(y_0,\varepsilon)$ and Γ is uhc at x_0, it follows that there exists $\delta > 0$ such that $\Gamma(x) \subseteq B(y_0,\varepsilon)$ for all $|x - x_0| < \delta$, so $\Gamma(x) \cap B(y_0,\varepsilon) \neq \emptyset$ for all $|x - x_0| < \delta$, which completes the proof. □

Proof of Proposition 7 statement 1. First suppose (λ^*,σ^*) is not credible. Then there exists $\lambda' \in \Lambda(\mu_0,\lambda_M^*)$ (recall μ_0 is the prior distribution on Θ) and $\varepsilon_0 > 0$ such that

$$\sum_{\theta,m} \lambda^*(\theta,m)u_S(\theta,\sigma^*(m)) < \sum_{\theta,m} \lambda'(\theta,m)u_S(\theta,\sigma^*(m)) - \varepsilon_0$$

By continuity, there exists $\varepsilon_1 > 0$ such that for all $|\lambda - \lambda^*| < \varepsilon_1$ we have

$$\sum_{\theta,m} \lambda(\theta,m) u_S(\theta,\sigma^*(m)) < \sum_{\theta,m} \lambda'(\theta,m) u_S(\theta,\sigma^*(m)) - \varepsilon_0/2 \tag{22}$$

By Lemma 10, there exists $\varepsilon_2 > 0$ such that for all $|\mu - \mu_0| < \varepsilon_2$ and $|\nu - \lambda^*_M| < \varepsilon_2$, there exists $\lambda \in \Lambda(\mu,\nu)$ with

$$\sum_{\theta,m} \lambda(\theta,m) u_S(\theta,\sigma^*(m)) > \sum_{\theta,m} \lambda'(\theta,m) u_S(\theta,\sigma^*(m)) - \varepsilon_0/2. \tag{23}$$

Moreover, since the Receiver is choosing only pure strategies, there exists ε_3 such that for any σ where $|\sigma - \sigma^*| < \varepsilon_3$, $\sigma = \sigma^*$.

Now let $\varepsilon = \min\{\varepsilon_1, \frac{\varepsilon_2}{|\Theta|\times|M|}, \varepsilon_3\}$, By assumption, there exists a finite-sample, credible and R-IC profile $(\nu^N, \phi^N, \sigma^N)$ such that $|\nu^N - \lambda^*_M| < \varepsilon$, $|\sigma^N - \sigma| < \varepsilon$ and $P(|\phi^N(F^N_\Theta) - \lambda| < \varepsilon) > 1 - \varepsilon$.

Under such a finite-sample profile, $\sigma^N = \sigma^*$, and there exists $F^N_\Theta \in \mathcal{F}^N_\Theta$, realized with positive probability, such that $\tilde{\lambda}^* = \phi^N(F^N_\Theta)$ satisfies $|\tilde{\lambda}^* - \lambda^*| < \min\{\varepsilon_1, \frac{\varepsilon_2}{|\Theta|\times|M|}\}$.

Now since $|\tilde{\lambda}^* - \lambda^*| < \varepsilon_1$, by (22) we know that

$$\sum_{\theta,m} \tilde{\lambda}^*(\theta,m) u_S(\theta,\sigma^*(m)) < \sum_{\theta,m} \lambda'(\theta,m) u_S(\theta,\sigma^*(m)) - \varepsilon_0/2 \tag{24}$$

In addition, since $|\tilde{\lambda}^* - \lambda^*| < \frac{\varepsilon_2}{|\Theta|\times|M|}$, we know $F^N_\Theta = \tilde{\lambda}^*_\Theta$ satisfies $|F^N_\Theta - \mu_0| < \varepsilon_2$, and $\nu^N = \tilde{\lambda}^*_M$ satisfies $|\nu^N - \lambda^*_M| < \varepsilon_2$, so by (23) there exists $\tilde{\lambda}' \in \Lambda(F^N_\Theta, \nu^N)$ such that

$$\sum_{\theta,m} \tilde{\lambda}'(\theta,m) u_S(\theta,\sigma^*(m)) > \sum_{\theta,m} \lambda'(\theta,m) u_S(\theta,\sigma^*(m)) - \varepsilon_0/2 \tag{25}$$

Combining (24) and (25), we have

$$\sum_{\theta,m} \tilde{\lambda}'(\theta,m) u_S(\theta,\sigma^*(m)) > \sum_{\theta,m} \tilde{\lambda}^*(\theta,m) u_S(\theta,\sigma^*(m)),$$

Note that $\tilde{\lambda}' \in \Lambda(F^N_\Theta, \nu^N)$, but by Lemma 11, we can replace $\tilde{\lambda}'$ with an extreme point in $X^N(F^N_\Theta, \nu^N)$, and the above inequality still holds. That is, there exists $\hat{\lambda}' \in X^N(F^N_\Theta, \nu^N)$ such that

$$\sum_{\theta,m} \hat{\lambda}'(\theta,m) u_S(\theta,\sigma^*(m)) > \sum_{\theta,m} \tilde{\lambda}^*(\theta,m) u_S(\theta,\sigma^*(m)),$$

Notice that $\tilde{\lambda}^*$ and $\hat{\lambda}'$ are both in $X^N(F^N_\Theta, \nu^N)$, which is a contradiction since by the credibility of $(\nu^N, \phi^N, \sigma^N)$

$$\tilde{\lambda}^* = \phi^N(F_\Theta^N) = \underset{\lambda \in X^N(F_\Theta^N, \nu^N)}{\arg\max} \sum_{\theta,m} \lambda(\theta, m) u_S(\theta, \sigma^*(m)).$$

Second, suppose (λ^*, σ^*) violates R-IC. Then there exists σ' such that

$$\sum_{\theta,m} \lambda^*(\theta, m) u_R(\theta, \sigma'(m)) > \sum_{\theta,m} \lambda^*(\theta, m) u_R(\theta, \sigma^*(m))$$

By continuity, there exist $\eta > 0$ and $\varepsilon_4 > 0$ such that for all λ' satisfying $|\lambda^* - \lambda'| < \varepsilon_4$, we have

$$\sum_{\theta,m} \lambda'(\theta, m) u_R(\theta, \sigma'(m)) - \sum_{\theta,m} \lambda'(\theta, m) u_R(\theta, \sigma^*(m)) \geq \eta > 0$$

Let $d \equiv \max_{\theta,a} u_R(\theta, a) - \min_{\theta,a} u_R(\theta, a)$ denote the gap between the Receiver's highest and lowest payoffs. Let $\varepsilon_5 \leq \frac{\eta}{d+\eta}$ and $\varepsilon = \min\{\varepsilon_3, \varepsilon_4, \varepsilon_5\}$. By assumption, there exists a credible and R-IC finite-sample profile $(\nu^N, \phi^N, \sigma^N)$ such that $Pr(|\phi^N(F_\Theta^N) - \lambda^*| \leq \varepsilon) > 1 - \varepsilon$, and $\sigma^N = \sigma^*$. We will show that in the finite sample profile $(\nu^N, \phi^N, \sigma^N)$, the Receiver can profitably deviate from $\sigma^N = \sigma^*$ to σ', which contradicts $(\nu^N, \phi^N, \sigma^N)$ being R-IC.

By choosing σ^* the Receiver obtains payoff

$$\sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N) \sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma^*(m))$$

By contrast, the Receiver obtains

$$\sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N) \sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma'(m))$$

from choosing σ'. Denote $E^N \equiv \{F_\Theta^N : |\phi^N(F_\Theta^N) - \lambda^*| \leq \delta\}$ so $Pr(E^N) > 1 - \varepsilon$. By switching from σ^* to σ', the Receiver obtains an extra payoff of

$$\begin{aligned}
&\sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N) \left[\sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma^*(m)) - \sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma'(m))\right] \\
=& \sum_{F_\Theta^N \in E^N} P^N(F_\Theta^N) \left[\sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma^*(m)) - \sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma'(m))\right] \\
&+ \sum_{F_\Theta^N \notin E^N} P^N(F_\Theta^N) \left[\sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma^*(m)) - \sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma'(m))\right]
\end{aligned}$$

Note that $\sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma^*(m)) - \sum_{\theta,m} \phi^N(\theta, m | F_\Theta^N) u_R(\theta, \sigma'(m)) \geq \eta$ for all $F_\Theta^N \in$

E^N, while for all $F_\Theta^N \notin E^N$, $\sum_{\theta,m}\phi^N(\theta,m|F_\Theta^N)u_R(\theta,\sigma^*(m)) - \sum_{\theta,m}\phi^N(\theta,m|F_\Theta^N)u_R(\theta,\sigma'(m)) \geq -d$. Together they imply,

$$\begin{aligned}&\sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N)\left[\sum_{\theta,m}\phi^N(\theta,m|F_\Theta^N)u_R(\theta,\sigma^*(m)) - \sum_{\theta,m}\phi^N(\theta,m|F_\Theta^N)u_R(\theta,\sigma'(m))\right]\\ \geq&\eta P^N(E^N) - d(1-P^N(E^N)).\end{aligned}$$

Since $P^N(E^N) > 1-\varepsilon$, we have

$$\begin{aligned}&\sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N)\left[\sum_{\theta,m}\phi^N(\theta,m|F_\Theta^N)u_R(\theta,\sigma^*(m)) - \sum_{\theta,m}\phi^N(\theta,m|F_\Theta^N)u_R(\theta,\sigma'(m))\right]\\ >&(1-\varepsilon)\eta - \varepsilon d = \eta - \varepsilon(\eta + d) \geq 0\end{aligned}$$

This contradicts the R-IC of $(\nu^N, \phi^N, \sigma^N)$.

□

Proof of Proposition 7 statement 2. For each $N \geq 1$, define $\sigma^N = \sigma^*$, $\nu^N \in \arg\min_{\nu \in \mathcal{F}_M^N} |\lambda_M^* - \nu|$, and $\phi^N : \mathcal{F}_\Theta^N \to \cup_{F_\Theta^N \in \mathcal{F}_\Theta^N} X(F_\Theta^N, \nu^N)$ by

$$\phi(F_\Theta^N) \in \underset{\lambda \in X^N(F_\Theta^N,\nu^N)}{\arg\max} \sum_{\theta,m}\lambda(\theta,m)u_S(\theta,\sigma^*(m)).$$

By construction, for every N, $(\nu^N, \phi^N, \sigma^N)$ is credible and $|\sigma^N - \sigma^*| = 0$. It remains to show that for every $\varepsilon > 0$, there exists large enough N, such that

1. $|\nu^N - \lambda_M^*| < \varepsilon$;
2. $P(|\phi^N(F_\Theta^N) - \lambda^*| < \varepsilon) > 1 - \varepsilon$;
3. $(\nu^N, \phi^N, \sigma^N)$ is R-IC.

From the denseness of rational numbers, we know that $\nu^N \to \lambda_M^*$ as $N \to \infty$ so the first statement follows.

To prove the second statement, note that since (λ^*, σ^*) is strictly credible, λ^* is the unique maximizer to

$$\max_{\lambda \in \Lambda(\mu,\nu)} \sum_{\theta,m}\lambda(\theta,m)u_S(\theta,\sigma^*(m)).$$

From Lemma 10, the best response correspondence $B(\mu,\nu) \equiv \arg\max_{\lambda\in\Lambda(\mu,\nu)}\sum_{\theta,m}\lambda(\theta,m)u_S(\theta_i,\sigma^*(m_j))$ is upper hemi-continuous. Since $B(\mu,\nu) = \{\lambda^*\}$ is a singleton, from Lemma 12, B is continuous at (μ, ν). Therefore, there exists $\delta > 0$, so that for any (μ', ν') such $|\mu - \mu'| < \delta$ and $|\nu - \nu'| < \delta$, we have $|\lambda' - \lambda^*| < \varepsilon$ for every $\lambda' \in B(\mu', \nu')$.

From the Glivenko–Cantelli theorem, for large N, $P(|F_\Theta^N - \mu_0| < \delta) > 1 - \varepsilon$. Pick N large enough so that $P(|F_\Theta^N - \mu_0| < \delta) > 1 - \varepsilon$ and $|\nu^N - \mu_M^*| < \delta$. Follows from the definition of ϕ and Lemma 11, $\phi(F_\Theta^N) \in \arg\max_{\lambda \in X^N(F_\Theta^N, \nu^N)} \sum_{\theta,m} \lambda(\theta, m) u_S(\theta, \sigma^*(m)) \subset \arg\max_{\lambda \in \Lambda(F_\Theta^N, \nu^N)} \sum_{\theta,m} \lambda(\theta, m) u_S(\theta, \sigma^*(m)$. So with at least $1-\varepsilon$ probability, $|\phi(F_\Theta^N) - \lambda^*| < \varepsilon$.

Lastly, we show that $(\nu^N, \phi^N, \sigma^N)$ is R-IC for large N. Since (λ^*, σ^*) is strictly R-IC, for any $\sigma \neq \sigma^*$,

$$\sum_{\theta,m} \lambda^*(\theta, m) u_R(\theta, \sigma^*(m)) > \sum_{\theta,m} \lambda^*(\theta, m) u_R(\theta, \sigma(m)).$$

From continuity, there exists $\eta > 0$ such that for any λ such that $|\lambda^* - \lambda| < \varepsilon$,

$$\sum_{\theta,m} \lambda(\theta, m) u_R(\theta, \sigma^*(m)) - \sum_{\theta,m} \lambda(\theta, m) u_R(\theta, \sigma(m)) \geq \eta > 0.$$

As we have shown, for any $\varepsilon > 0$, for large enough N, $Pr(|\phi(F_\Theta^N) - \lambda^*| \leq \varepsilon) \geq 1 - \varepsilon$. Pick $\varepsilon \leq \frac{\eta}{d+\eta}$, then follow from the same argument above, we have

$$\sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N) \sum_{\theta,m} \phi(\theta, m | F_\Theta^N) u_R(\theta, \sigma^*(m)) > \sum_{F_\Theta^N \in \mathcal{F}_\Theta^N} P^N(F_\Theta^N) \sum_{\theta,m} \phi(\theta, m | F_\Theta^N) u_R(\theta, \sigma(m))$$

for any $\sigma \neq \sigma^*$. So $(\nu^N, \phi^N, \sigma^N)$ is R-IC.

□

B.3 Cycle Length in Theorem 1

The following claim formalizes the observation made after Theorem 1: when verifying u_S-cyclical monotonicity, it is without loss to focus on cycles no longer than $\min\{|\Theta|, |A|\}$.

Claim. *An outcome distribution π is u_S-cyclically monotone if and only if for each sequence $(\theta_1, a_1), \ldots, (\theta_n, a_n) \in supp(\pi)$ where $n \leq \min\{|\Theta|, |A|\}$ and $a_{n+1} \equiv a_1$, we have*

$$\sum_{i=1}^{n} u_S(\theta_i, a_i) \geq \sum_{i=1}^{n} u_S(\theta_i, a_{i+1}).$$

Proof. We will prove that if there exists a sequence $(\theta_1, a_1), ..., (\theta_n, a_n) \in supp(\pi)$ with $n > \min\{|\Theta|, |A|\}$ such that

$$u_S(\theta_1, a_1) + ... + u_S(\theta_n, a_n) < u_S(\theta_1, a_2) + ... + u_S(\theta_n, a_1),$$

then there exists a sequence $(\theta_1', a_1'), ..., (\theta_k', a_k') \in supp(\pi)$ with $k < n$ such that

$$u_S(\theta_1', a_1') + ... + u_S(\theta_k', a_k') < u_S(\theta_1', a_2') + ... + u_S(\theta_k', a_1').$$

The statement of the claim then follows from iteration.

Suppose by contradiction that there exists a sequence $(\theta_1, a_1), ..., (\theta_n, a_n)$ with $n > \min\{|\Theta|, |A|\}$ such that

$$u_S(\theta_1, a_1) + ... + u_S(\theta_n, a_n) < u_S(\theta_1, a_2) + ... + u_S(\theta_n, a_1),$$

and that for all sequences with length $k < n$,

$$u_S(\theta_1', a_1') + ... + u_S(\theta_k', a_k') \geq u_S(\theta_1', a_2') + ... + u_S(\theta_k', a_1'). \tag{26}$$

Suppose $\min\{|\Theta|, |A|\} = |A|$ (a similar argument works for $\min\{|\Theta|, |A|\} = |\Theta|$), then there exists a^* that appears twice in the sequence $(\theta_1, a_1), ..., (\theta_n, a_n)$. Without loss let $a_1 = a_l = a^*$ with $1 < l \leq n$. Then

$$\begin{aligned} u_S(\theta_1, a_1) + ... + u_S(\theta_n, a_n) &= [u_S(\theta_1, a_1) + ... + u_S(\theta_{l-1}, a_{l-1})] + [u_S(\theta_l, a_l) + ... + u_S(\theta_n, a_n)] \\ &\geq [u_S(\theta_1, a_2) + ... + u_S(\theta_{l-1}, a_1)] + [u_S(\theta_l, a_{l+1}) + ... + u_S(\theta_n, a_l)] \\ &= [u_S(\theta_1, a_2) + ... + u_S(\theta_{l-1}, a_l)] + [u_S(\theta_l, a_{l+1}) + ... + u_S(\theta_n, a_1)] \end{aligned}$$

where the inequality follows from (26), and the second equality holds because $a_l = a_1$. This is a contradiction. □

B.4 The Benefit of Credible Persuasion: An Intermediate Example

In this section, we provide an example in which the Sender can benefit from credible persuasion, but cannot achieve her optimal full-commitment payoff. This example also corresponds to the first case of Proposition 3.

The prior belief is μ_0 with $\mu_0(\theta_H) = 0.4$. Note that both the Sender's and the Receiver's payoffs are supermodular. The horizontal axis μ in the graph represents the probability assigned to θ_H by the posterior belief.

According to the concavification, the optimal full-commitment information structure λ^* induces two posterior beliefs $\mu = 1/3$ and $\mu = 2/3$, with the Receiver's strategy σ^* playing a_2 when $\mu = 1/3$ and a_3 when $\mu = 2/3$. However in this case the support of the outcome distribution is $\{(\theta_L, a_2), (\theta_L, a_3), (\theta_H, a_2), (\theta_H, a_3)\}$. This outcome distribution is not comonotone, so (λ^*, σ^*) is not credible.

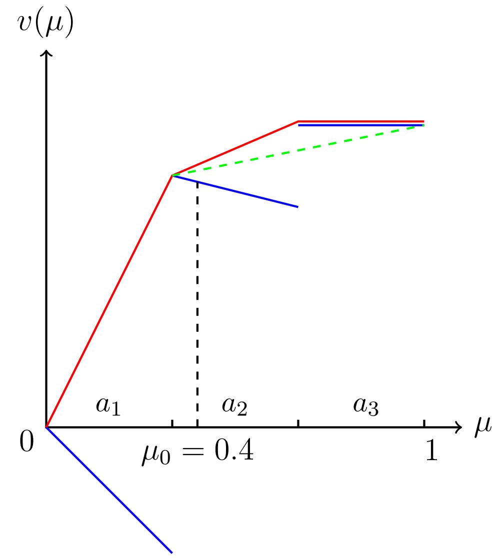
The optimal credible information structure λ° is represented by the green dashed line:

it induces two posteriors, $\mu = 1/3$ and $\mu = 1$, with the Receiver strategy σ° playing a_2 when $\mu = 1/3$ and a_3 when $\mu = 1$. In particular, the support of the outcome distribution is $\{(\theta_L, a_2), (\theta_H, a_2), (\theta_H, a_3)\}$ which is comonotone, so $(\lambda^\circ, \sigma^\circ)$ is credible.

u_S	a_1	a_2	a_3
$\theta = H$	-1	0.5	0.8
$\theta = L$	0	0.75	0.8

u_R	a_1	a_2	a_3
$\theta = H$	0	2	3
$\theta = L$	3	2	0

Sender and Receiver's payoffs



Concavification

B.5 Comparative Statics: Examples

B.5.1 A Class of Aligned Preferences

Let $u_S(\theta, a)$ be a strictly supermodular payoff function; in addition, assume that u_S favors higher actions: $u_S(\theta, a') \geq u_S(\theta, a)$ for all θ and $a' \geq a$. Let $\{u_R^\kappa\}_{\kappa \in K}$ denote a collection of Receiver's payoff functions defined by $u_R^\kappa(\theta, a) \equiv w(\theta, a, \kappa)$, where $w : \Theta \times A \times K \to \mathbb{R}$ is a strictly supermodular function, and $K \subseteq \mathbb{R}$ represents a parameter space.

It's straightforward to see that for each $\kappa \in K$, the Receiver payoff function $u_R^\kappa : \Theta \times A \to \mathbb{R}$ is strictly supermodular. Furthermore, preferences $(u_S, u_R^{\kappa'})$ are more aligned than (u_S, u_R^κ) whenever $\kappa' \geq \kappa$. To see why, for each $\kappa \in K$ and $\mu \in \Delta(\Theta)$, let $\hat{a}^\kappa(\mu) \equiv \max\{\arg\max_{a \in A} \sum_\theta \mu(\theta) u_R^\kappa(\theta, a)\}$ denote the Receiver's highest best response to μ when the payoff function is u_R^κ (note that since the Sender favors higher actions, selecting the highest best response is equivalent to breaking ties in the Sender's favor). By Lemma 2.8.1 of Topkis (2011), $\hat{a}^{\kappa'}(\mu) \geq \hat{a}^\kappa(\mu)$ for $\kappa' \geq \kappa$. Since the Sender favors higher actions, for any $a \in A$, $\mu \in \Delta(\Theta)$, and $\kappa' \geq \kappa$, we have

$$E_\mu\big[u_S(\theta, \hat{a}^\kappa(\mu))\big] \geq E_\mu\big[u_S(\theta, a)\big] \Rightarrow E_\mu\big[u_S(\theta, \hat{a}^{\kappa'}(\mu))\big] \geq E_\mu\big[u_S(\theta, a)\big].$$

This implies that $(u_S, u_R^{\kappa'})$ are more aligned than (u_S, u_R^{κ}) whenever $\kappa' \geq \kappa$. So according to Proposition 5, the Sender obtains a higher payoff from the Sender-optimal stable outcome distribution under $(u_S, u_R^{\kappa'})$ than from that under (u_S, u_R^{κ}).[27]

B.5.2 Set of Stable Outcome Distributions

The following example illustrates that even in a binary-state, binary-action setting, more aligned preferences do not necessarily lead to a larger set of stable outcomes.

Suppose the state space is $\Theta = \{0, 1\}$ with equal prior probabilities, and the action space is $A = \{0, 1\}$. Players' payoffs are given by

u_S	$a = 0$	$a = 1$
$\theta = 0$	0	1
$\theta = 1$	0	2

u_R	$a = 0$	$a = 1$
$\theta = 0$	0	-1
$\theta = 1$	0	k

where the parameter $k \in (0, \infty)$ captures the alignment between the players preferences. A higher k implies players' preferences are more aligned under the alignment notion in Kamenica and Gentzkow (2011).

We will characterize the set of stable outcome distributions; that is, $\pi \in \Delta(\Theta \times A)$ satisfying $\pi_\Theta = \mu_0$, u_R-obedience, and u_S-cyclical monotonicity (comonotonicity in this case due to the supermodularity of Sender's payoff).

Since states and actions are binary, an outcome distribution can be represented by a vector in $[0, 1]^2$. This is because specifying $\pi(a = 1|\theta = 1)$ and $\pi(a = 1|\theta = 0)$ pins down $\pi(a = 0|\theta = 1) = 1 - \pi(a = 1|\theta = 1)$ and $\pi(a = 0|\theta = 0) = 1 - \pi(a = 1|\theta = 0)$.

The obedient constraint for action $a = 1$ is

$$\pi(a = 1|\theta = 1)u_R(\theta = 1, a = 1) + \pi(a = 1|\theta = 0)u_R(\theta = 0, a = 1) \geq \\ \pi(a = 1|\theta = 1)u_R(\theta = 1, a = 0) + \pi(a = 1|\theta = 0)u_R(\theta = 0, a = 0)$$

By defining vectors $\boldsymbol{\pi} = \big(\pi(a = 1|\theta = 1), \pi(a = 1|\theta = 0)\big)$, $\boldsymbol{u}_1 = \big(u_R(\theta = 1, a = 1), u_R(\theta = 0, a = 1)\big)$ and $\boldsymbol{u}_0 = \big(u_R(\theta = 1, a = 0), u_R(\theta = 0, a = 0)\big)$, the constraint can be re-written in vector form:

$$\boldsymbol{\pi} \cdot (\boldsymbol{u}_1 - \boldsymbol{u}_0) = \boldsymbol{\pi} \cdot \begin{pmatrix} k \\ -1 \end{pmatrix} \geq 0. \tag{OB-1}$$

[27] Note also that the following variant of the alignment notion in Gentzkow and Kamenica (2017) is a further special case of this class of preferences: $u_S(\theta, a) = f(\theta, a)$, $u_R^{\kappa}(\theta, a) = f(\theta, a) + \kappa g(\theta, a)$ with $\kappa \in [0, \infty)$, where both f and g are strictly supermodular and $f(\theta, a') \geq f(\theta, a)$ for all θ and $a' \geq a$.

Similarly, the obedient constraint for action $a = 0$ is:

$$(\mathbf{1} - \boldsymbol{\pi}) \cdot (\boldsymbol{u}_1 - \boldsymbol{u}_0) = (\mathbf{1} - \boldsymbol{\pi}) \cdot \begin{pmatrix} k \\ -1 \end{pmatrix} \leq 0. \tag{OB-0}$$

The credibility (comonotonicity) constraint is

$$\pi(a = 1|\theta = 0) > 0 \quad \Rightarrow \quad \pi(a = 1|\theta = 1) = 1 \tag{CO}$$

To visualize how these constraints vary with k, let us represent them in a two-dimensional figure. In Fig. 4, notice that for any k, the hyperplane defined by (OB-1) always crosses $(0, 0)$, and the hyperplane defined by (OB-0) always crosses $(1, 1)$. These two parallel hyperplanes both have normal vector $(k, -1)$, and rotate at $(0, 0)$ and $(1, 1)$ respectively as k varies.

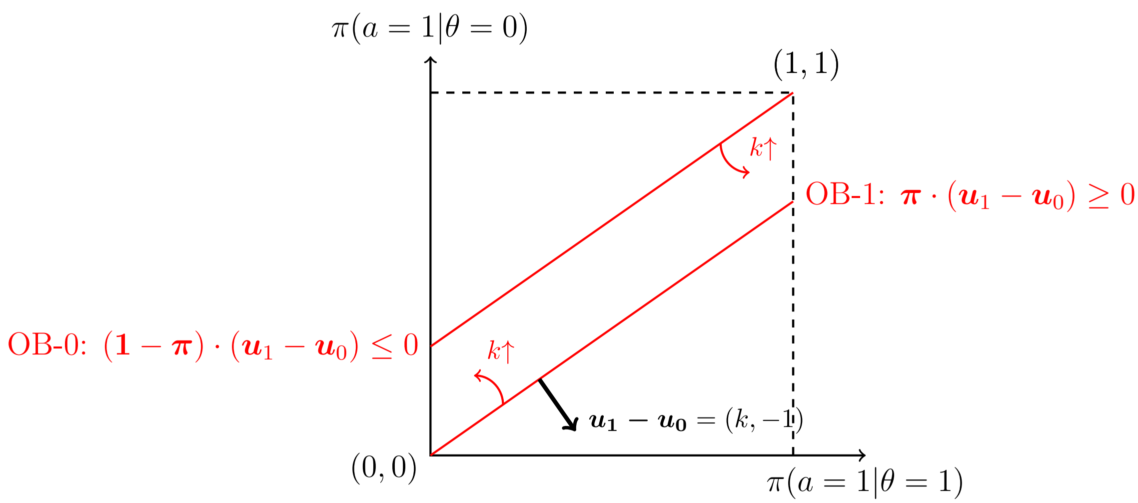


Figure 4: Hyperplanes defined by OB-1 and OB-0

When $k < 1$, the two obedience constraints are represented in the pink areas in the left panel of Fig. 5. In this scenario, (OB-1) is the binding obedience constraint. By contrast, when $k > 1$, the binding obedience constraint is (OB-0), as shown in the right panel of Fig. 5. The comonotonicity constraint, which requires that $\pi(a = 1|\theta = 1) = 1$ whenever $\pi(a = 1|\theta = 0) > 0$, is represented by the blue line segments in both cases.

The set of stable outcome distributions is thus the intersection of the pink and blue areas, which is represented by the purple line segments in Fig. 6. When $k < 1$, the set of stable outcome distributions is $\{(x, 0)|x \in [0, 1]\} \cup \{(1, x)|x \in [0, k]\}$, which expands when k increases. By contrast, after the hyperplane defined by OB-1 crosses the 45 degree line, i.e., when $k > 1$, the binding obedience constraint becomes OB-0, which starts to rotate at (1,1). So the set of stable outcome distributions, $\{(x, 0)|x \in [1 - \frac{1}{k}, 1]\} \cup \{(1, x)|x \in [0, 1]\}$, shrinks when k

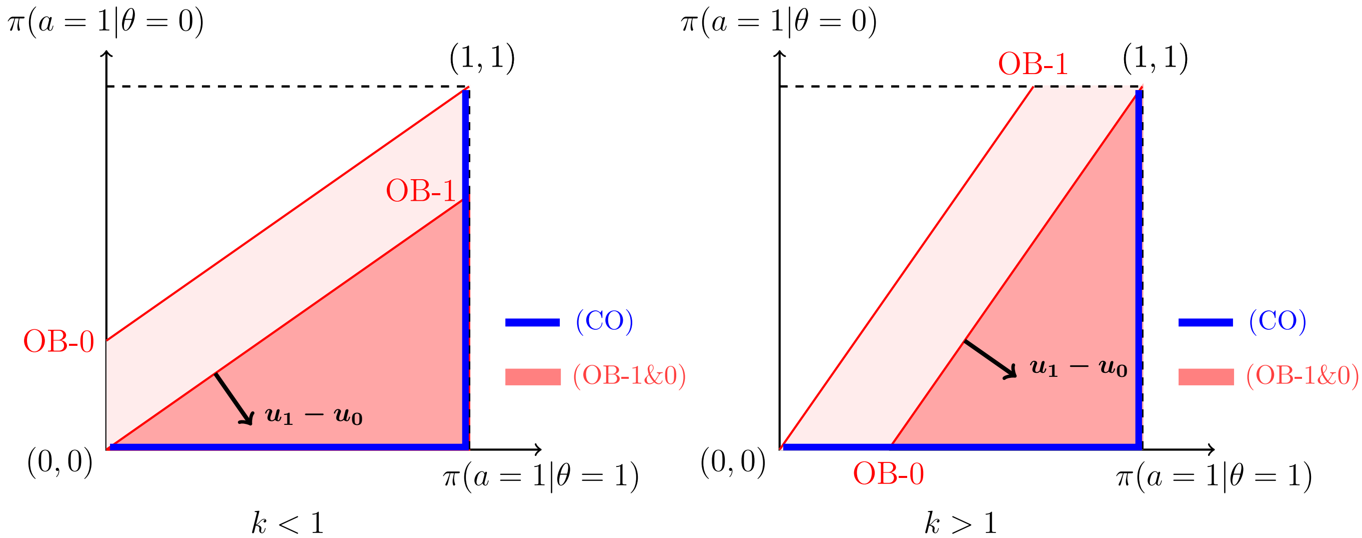


Figure 5: Stable outcome distributions

increases.

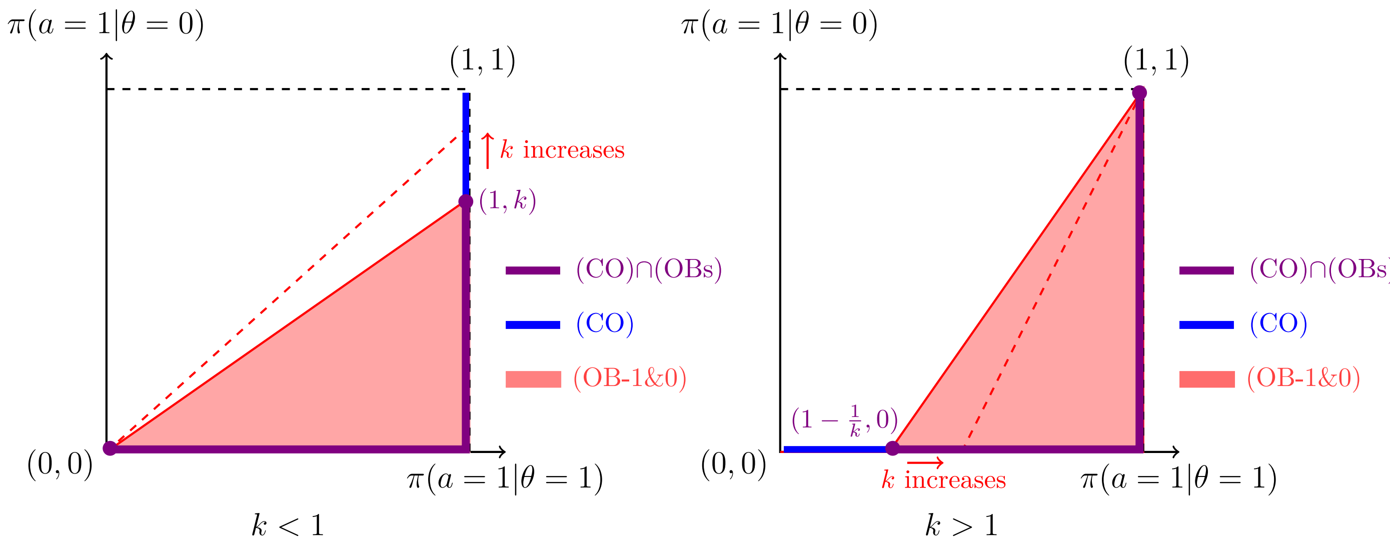


Figure 6: How the set of stable outcome distributions varies with k.

A final remark is that, even though the set of stable outcome distributions changes non-monotonically as k increases, according to Proposition 5, the Sender's optimal payoff is always (weakly) increasing. This can be seen in Fig. 7. Let $\boldsymbol{u_S} = \big(u_S(\theta=1,a=1), u_S(\theta=0,a=1)\big) = (2,1)$. The Sender's objective is to maximize $\boldsymbol{\pi}\cdot\boldsymbol{u_S}$ among the set of $\boldsymbol{\pi}$ represented by the purple line segment. The vector $\boldsymbol{u_S}$ points in the northeast direction, so the value is maximized at the northeast corner of the purple line segment.

When $k<1$, increasing k expands the stable outcome distributions and the Sender's optimal stable outcome distribution $\boldsymbol{\pi}^* = (1,k)$ changes accordingly, which strictly increases the Sender's value. When $k > 1$, increasing k shrinks the stable outcome distributions but the optimal outcome distribution, $\boldsymbol{\pi}^* = (1,1)$ remains feasible, and thus the Sender's value is unchanged.

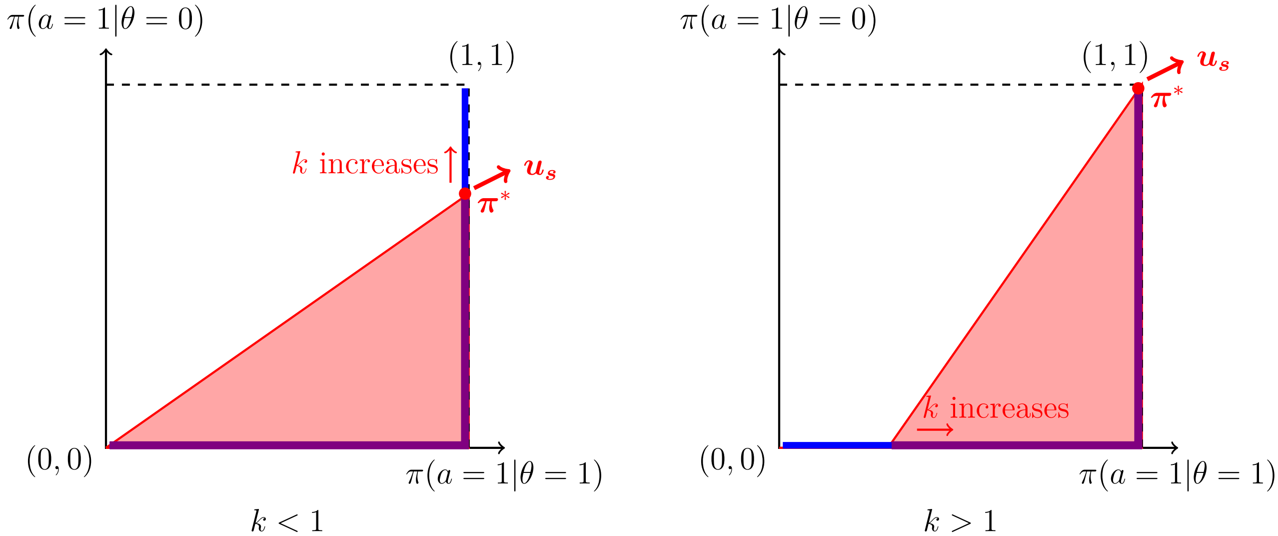


Figure 7: How the optimal stable outcome distribution varies with k.

B.6 Extension to Infinite Spaces

Suppose Θ and A are compact Polish spaces, and let M be a Polish space containing A. An information structure $\lambda \in \Delta(\Theta \times M)$ is a Borel probability measure on $\Theta \times M$. A strategy $\sigma : M \to A$ is a measurable function from M to A.

An outcome distribution is a Borel measure $\pi \in \Delta(\Theta \times A)$. The outcome distribution π is induced by the profile (λ, σ) if π is the pushforward measure of λ obtained from the function $\rho : (\theta, m) \to (\theta, \sigma(m))$. That is, for any $S \in \mathcal{B}(\Theta \times A)$, $\pi(S) = \lambda(\rho^{-1}(S))$.

Definition 1*. A profile (λ, σ) is credible if

$$\lambda \in \arg\max_{\lambda' \in D(\lambda)} \int u_S(\theta, \sigma(m))\, d\lambda',$$

where $D(\lambda) = \{\lambda' \in \Delta(\Theta \times M) \mid \lambda'_\Theta = \mu_0, \lambda'_M = \lambda_M\}$.

Definition 2*. A profile (λ, σ) is R-IC if for any Receiver's strategy σ', we have

$$\int u_R(\theta, \sigma(m))\, d\lambda \geq \int u_R(\theta, \sigma'(m))\, d\lambda.$$

Definition 3*. An outcome distribution is stable if it can be induced by a profile that is both credible and R-IC.

B.6.1 Extension of Theorem 1

Let $\{\pi(\cdot|a)\}_{a\in A} \subseteq \Delta(\Theta)$ denote a system of regular conditional probabilities obtained from disintegrating π with π_A (see, for example, Chang and Pollard, 1997). The following result is an extension of Theorem 1.

Theorem 1*. An outcome distribution $\pi \in \Delta(\Theta \times A)$ is stable if and only if there exists a Borel set $E^\circ \subseteq \Theta \times A$ with $\pi(E^\circ) = 1$ such that

1. π is u_S-cyclically monotone on E°: for any sequence $(\theta^1, a^1), \ldots, (\theta^n, a^n) \in E^\circ$ and $a^{n+1} \equiv a^1$,
$$\sum_{i=1}^{n} u_S\left(\theta^i, a^i\right) \geq \sum_{i=1}^{n} u_S\left(\theta^i, a^{i+1}\right).$$

2. π is u_R-obedient on E°: for each $a \in A^\circ \equiv \mathrm{proj}_A(E^\circ)$, let $E^\circ_a \equiv \{\theta : (\theta, a) \in E^\circ\}$, then
$$\int_{E^\circ_a} u_R(\theta, a)\, d\pi(\theta|a) \geq \int_{E^\circ_a} u_R(\theta, a')\, d\pi(\theta|a)$$
for all $a \in A^\circ$ and all $a' \in A$.

Proof. The "only if" direction: Suppose that π is induced by a credible and R-IC profile (λ, σ). The profile (λ, σ) being credible implies

$$\lambda \in \underset{\lambda' \in D(\lambda)}{\arg\max} \int u_S\left(\theta, \sigma(\theta, m)\right) d\lambda'(\theta, m)$$

Let $\tilde{u}(\theta, m) \equiv u_S\left(\theta, \sigma(m)\right)$. Since $\tilde{u}(\theta, m)$ is Borel measurable and $|\tilde{u}(\theta, m)| < \infty$, by Beiglböck et al. (2009), λ is $\tilde{u}$-cyclically monotone: i.e. there exists a Borel set $F \subseteq \Theta \times M$ such that $\lambda(F) = 1$ and for every sequence $(\theta_1, m_1), ..., (\theta_n, m_n) \in F$,

$$\sum_{i=1}^{n} u_S\left(\theta_i, \sigma\left(m_i\right)\right) \geq \sum_{i=1}^{n} u_S\left(\theta_i, \sigma\left(m_{i+1}\right)\right).$$

Consider the function $\rho : (\theta, m) \to (\theta, \sigma(m))$, and define $E \equiv \rho(F)$. Since $\lambda(F) = 1$ and π is the pushforward measure of λ obtained from ρ, it follows that $\pi(E) = 1$. In addition, for any sequence $(\theta_1, a_1), \ldots, (\theta_n, a_n) \in E$, there exists sequence $(\theta_1, m_1), \ldots, (\theta_n, m_n) \in F$ such that $a_i = \sigma(m_i)$. So

$$\sum_{i=1}^{n} u_S(\theta_i, a_i) = \sum_{i=1}^{n} u_S(\theta_i, \sigma(m_i)) \geq \sum_{i=1}^{n} u_S(\theta_i, \sigma(m_{i+1})) = \sum_{i=1}^{n} u_S(\theta_i, a_{i+1}),$$

which implies that π is u_S-cyclically monotone on the set E.

Now for each $a \in A$, let $E_a \equiv \{\theta : (\theta, a) \in E\}$. Note that $\pi(E_a|a) = 1$ for π_A-almost all $a \in A$, since otherwise there exists $\tilde{A} \subseteq A$ with $\pi_A(\tilde{A}) > 0$, such that for all $a \in \tilde{A}$, $\pi(E_a|a) < 1$. This would then imply

$$\begin{aligned}
\pi(\Theta \times \tilde{A}) &= \pi(E \cap (\Theta \times \tilde{A})) \\
&= \int_A \left[\int_\Theta \mathbb{1}_E \times \mathbb{1}_{\Theta \times \tilde{A}} \;\; d\pi(\theta|a) \right] d\pi_A(a) \\
&< \int_{\tilde{A}} 1 \, d\pi_A(a) \\
&= \pi(\Theta \times \tilde{A}),
\end{aligned}$$

which is a contradiction. So $\pi(E_a|a) = 1$ for π_A-almost all $a \in A$. As a result, for all measurable functions $\phi : \Theta \to \mathbb{R}$ and all $a \in A$, we have

$$\int_\Theta g \, d\pi(\theta|a) = \int_{E_a} g \, d\pi(\theta|a).$$

Next we establish that for π_A-almost all $a \in A$,

$$\int_{E_a} u_R(\theta, a) \, d\pi(\theta|a) \geq \int_{E_a} u_R(\theta, a') \, d\pi(\theta|a) \tag{27}$$

for all $a' \in A$. We prove this by proving its contraposition: suppose this is not true, we will show that this implies (λ, σ) is not R-IC. Specifically, if (27) does not hold for π_A-almost all $a \in A$ and all $a' \in A$, then there exists $\hat{A} \in \mathcal{B}(A)$ with $\pi_A(\hat{A}) > 0$, and for each $a \in \hat{A}$, we can find $d(a) \in A$ that satisfies

$$\int_{E_a} u_R(\theta, d(a)) \, d\pi(\theta|a) > \int_{E_a} u_R(\theta, a) \, d\pi(\theta|a).$$

Since $u_R(\theta, a)$ is a bounded Carathéodory function, the function

$$g(a, a') \equiv \int_{E_a} u_R(\theta, a') \, d\pi(\theta|a)$$

is measurable in a and continuous in a', and therefore also Carathéodory. For each $a \in \hat{A}$, let

$\phi(a) \equiv \arg\max_{a' \in A} g(a, a')$ denote the maximizers of the Receiver's interim expected payoff. Since A is compact, by the Measurable Maximum Theorem (see, for example, Theorem 18.19 in Aliprantis and Border, 2006), the correspondence $\phi(a)$ admits a measurable selection $d^* : \hat{A} \to A$, such that for all $a \in \hat{A}$,

$$\int_{E_a} u_R(\theta, d^*(a))\, d\pi(\theta|a) \geq \int_{E_a} u_R(\theta, d(a))\, d\pi(\theta|a) > \int_{E_a} u_R(\theta, a)\, d\pi(\theta|a).$$

Now define $f^* = f$ for $a \in \hat{A}$ and $f^* = I$ for $a \notin \hat{A}$. Clearly $f^* : A \to A$ is measurable. In addition,

$$\int_{E_a} u_R(\theta, f^*(a))\, d\pi(\theta|a) > \int_{E_a} u_R(\theta, a)\, d\pi(\theta|a).$$

for all $a \in \hat{A}$. Since $\pi_A(\hat{A}) > 0$, we have that

$$\begin{aligned}
\int_{\Theta \times A} u_R(\theta, f^*(a))\, d\pi(\theta, a) &= \int_A \left[\int_\Theta u_R(\theta, f^*(a))\, d\pi(\theta|a)\right] d\pi_A(a) \\
&= \int_A \left[\int_{E_a} u_R(\theta, f^*(a))\, d\pi(\theta|a)\right] d\pi_A(a) \\
&> \int_A \left[\int_{E_a} u_R(\theta, a)\, d\pi(\theta|a)\right] d\pi_A(a) \\
&= \int_A \left[\int_\Theta u_R(\theta, a)\, d\pi(\theta|a)\right] d\pi_A(a) \\
&= \int_{\Theta \times A} u_R(\theta, a)\, d\pi(\theta, a),
\end{aligned} \tag{28}$$

Now since π is the pushforward measure of λ, we have

$$\int_{\Theta \times A} u_R(\theta, a) d\pi(\theta, a) = \int_{\Theta \times M} u_R(\theta, \sigma(m)) d\lambda(\theta, m). \tag{29}$$

In addition, let $\sigma' \equiv f^* \circ \sigma$, then $\sigma' : M \to \mathbb{R}$ is a Borel measurable function on M, and

$$\begin{aligned}
\int_{\Theta \times A} u_R(\theta, f^*(a))\, d\pi(\theta, a) &= \int_{\Theta \times M} u_R(\theta, f^* \circ \sigma(m))\, d\lambda(\theta, m) \\
&= \int_{\Theta \times M} u_R(\theta, \sigma'(m)) d\lambda(\theta, m).
\end{aligned} \tag{30}$$

Plugging (29) and (30) into (28), we have

$$\int_{\Theta \times M} u_R(\theta, \sigma'(m)) d\lambda(\theta, m) > \int_{\Theta \times A} u_R(\theta, \sigma(m))\, d\lambda(\theta, m),$$

which is a contradiction to (λ, σ) being R-IC. So there exists $\overline{A} \subseteq A$ with $\pi_A(\overline{A}) = 1$, such that

$$\int_{E_a} u_R(\theta, a)\, d\pi(\theta|a) \geq \int_{E_a} u_R(\theta, a')\, d\pi(\theta|a)$$

for all $a \in \overline{A}$ and all $a' \in A$.

Define $E^\circ \equiv E \cap (\Theta \times \overline{A})$. Note that $\pi(E^\circ) = 1$, and π is u_R-obedient on E°. In addition, since π is u_S-cyclically monotone on E and $E^\circ \subset E$, we have that π is u_S-cyclically monotone on E°. This completes the proof of the "only if" direction.

The "if" direction: Suppose there exists a Borel set $E^\circ \subseteq \Theta \times A$ with $\pi(E^\circ) = 1$, where the outcome distribution $\pi \in \Delta(\Theta \times A)$ is both u_S-cyclical monotone and u_R-obedient. Let the message space $M = A$, and consider the profile (λ, σ) where $\lambda \equiv \pi$ and σ is the identity mapping. Clearly, (λ, σ) induces π. We will show that (λ, σ) is both credible and R-IC.

To see that (λ, σ) is R-IC, first note that following a similar argument as the one in the "only if" direction, we have

$$\int_\Theta g\, d\pi(\theta|a) = \int_{E_a^\circ} g\, d\pi(\theta|a).$$

for all measurable functions $\phi : \Theta \to \mathbb{R}$ and π_A-almost all $a \in A$. So for all $\sigma' : A \to A$,

$$\begin{aligned}
\int_{\Theta \times A} u_R(\theta, a) d\pi(\theta, a) &= \int_A \int_\Theta u_R(\theta, a)\, d\pi(\theta|a)\, d\pi_A(a) \\
&= \int_A \int_{E_a^\circ} u_R(\theta, a)\, d\pi(\theta|a)\, d\pi_A(a) \\
&\geq \int_A \int_{E_a^\circ} u_R(\theta, \sigma'(a))\, d\pi(\theta|a)\, d\pi_A(a) \\
&= \int_{\Theta \times A} u_R(\theta, \sigma'(a)) d\pi(\theta, a),
\end{aligned}$$

which implies that (λ, σ) is R-IC.

Next we show (λ, σ) is credible. Since π is u_S-cyclically monotone on E°, every sequence $(\theta_1, a_1), \ldots, (\theta_n, a_n) \in E^\circ$ satisfies

$$\sum_{i=1}^n u_S(\theta_i, a_i) \geq \sum_{i=1}^n u_S(\theta_i, a_{i+1}),$$

where $a_{n+1} \equiv a_1$. Since $\lambda = \pi$ and σ is the identity mapping, this is equivalent to the existence of a set $F \subseteq \Theta \times M$ with $\lambda(F) = 1$, such that

$$\sum_{i=1}^n u_S(\theta_i, \sigma(m_i)) \geq \sum_{i=1}^n u_S(\theta_i, \sigma(m_{i+1}));$$

for every sequence $(\theta_1, m_1), \ldots, (\theta_n, m_n) \in F$ with $m_{n+1} = m_1$. By Beiglböck et al. (2009), λ satisfies

$$\lambda \in \underset{\lambda' \in D(\lambda)}{\arg\max} \int_{\Theta \times M} u_S(\theta, \sigma(m)) \, d\lambda'$$

which means (λ, σ) is credible. □

B.6.2 Extension of Proposition 2

Next, we extend Proposition 2 to infinite spaces. Let A, Θ, and M be compact subsets of $\mathbb{R}$, and $A \subseteq M$.

Definition 4. *An outcome distribution $\pi \in \Delta(\Theta \times A)$ is a no-information outcome if there exists $\hat{a} \in A$ such that $\pi(\Theta \times \{\hat{a}\}) = 1$.*

For each pair of actions $a, a' \in A$, let $\Theta_0(a, a') \equiv \{\theta : u_R(\theta, a) = u_R(\theta, a')\}$ denote the set of states under which the Receiver is indifferent between a and a'.

Proposition 2*. Suppose $\mu_0(\Theta_0(a, a')) < 1$ for all distinct $a, a' \in A$. In addition, suppose $u_S : \Theta \times A \to \mathbb{R}$ is strictly supermodular and $u_R : \Theta \times A \to \mathbb{R}$ is submodular, then any stable outcome distribution must be a no-information outcome.

Proof. Let π be a stable outcome distribution. By Theorem 1 and Lemma 1, there exists a Borel set $E^\circ \subseteq \Theta \times A$ with $\pi(E^\circ) = 1$, such that

1. π is comonotone on E°: for all $(\theta, a), (\theta', a') \in E$, $a < a'$ implies $\theta \leq \theta'$; and

2. π is u_R-obedient on E°: for each $a \in A^\circ \equiv \text{proj}_A(E^\circ)$, let $I_a \equiv \{\theta : (\theta, a) \in E^\circ\}$, then

$$\int_{I_a} u_R(\theta, a) \, d\pi(\theta|a) \geq \int_{I_a} u_R(\theta, a') \, d\pi(\theta|a)$$

   for all $a \in A^\circ$ and $a' \in A$.

From u_R-obedience, we know that for all $a, a' \in A^\circ$ with $a < a'$, we have

$$\int_{I_a} [u_R(\theta, a) - u_R(\theta, a')] \, d\pi(\theta|a) \geq 0 \geq \int_{I_{a'}} [u_R(\theta, a) - u_R(\theta, a')] \, d\pi(\theta|a'). \tag{31}$$

In addition, comonotonicity implies that $\theta \leq \theta'$ for all $\theta \in I_a$, $\theta' \in I_{a'}$. Since u_R is submodular, we have $u_R(\theta, a) - u_R(\theta, a') \leq u_R(\theta', a) - u_R(\theta', a')$ for all $\theta \in I_a$ and $\theta' \in I_{a'}$, which implies

$$\sup_{\theta \in I_a} \{u_R(\theta, a) - u_R(\theta, a')\} \leq \inf_{\theta' \in I_{a'}} \{u_R(\theta', a) - u_R(\theta', a')\},$$

and therefore

$$\begin{aligned}\int_{I_a} [u_R(\theta,a)-u_R(\theta,a')]\,d\pi(\theta|a) &\le \int_{I_a} \sup_{\theta\in I_a}[u_R(\theta,a)-u_R(\theta,a')]\,d\pi(\theta|a)\\ &= \sup_{\theta\in I_a}\{u_R(\theta,a)-u_R(\theta,a')\}\\ &\le \inf_{\theta'\in I_{a'}}\{u_R(\theta',a)-u_R(\theta',a')\}\\ &\le \int_{I_{a'}} [u_R(\theta,a)-u_R(\theta,a')]\,d\pi(\theta|a').\end{aligned} \tag{32}$$

Combining (31) and (32), we have for all $a, a' \in A^\circ$, $a < a'$,

$$\int_{I_a} [u_R(\theta,a)-u_R(\theta,a')]\,d\pi(\theta|a) = \sup_{\theta\in I_a}\{u_R(\theta,a)-u_R(\theta,a')\} = 0$$

and

$$\int_{I_{a'}} [u_R(\theta,a)-u_R(\theta,a')]\,d\pi(\theta|a') = \inf_{\theta'\in I_{a'}}\{u_R(\theta',a)-u_R(\theta',a')\} = 0.$$

This implies that for all $a, a' \in A^\circ$, $a < a'$, we have

$$\begin{gathered} u_R(\theta,a)-u_R(\theta,a') \le 0 \text{ for all } \theta \in I_a,\\ \text{with } u_R(\theta,a) = u_R(\theta,a') \text{ for } \pi(.|a)\text{-almost all } \theta\in I_a; \end{gathered} \tag{33}$$

and

$$\begin{gathered} u_R(\theta',a)-u_R(\theta',a') \ge 0 \text{ for all } \theta' \in I_{a'},\\ \text{with } u_R(\theta',a) = u_R(\theta',a') \text{ for } \pi(.|a')\text{-almost all } \theta'\in I_{a'}, \end{gathered} \tag{34}$$

For each $a \in A^\circ$, let $N(a) \equiv \{\theta \in I_a : u_R(\theta,a) \neq u_R(\theta,a') \text{ for some } a' \in A^\circ\}$ denote the set of states in I_a under which the Receiver is not indifferent towards all actions in A°. We want to show that $\pi(N(a)|a) = 0$ for each $a \in A^\circ$. Note that this does not follow directly from (33) and(34) since A° may be uncountably infinite, and an uncountable union of $\pi(.|a)$-null sets may no longer be a $\pi(.|a)$-null set.

However, note that since u_R is submodular, for any $a' > a$, if $u_R(\theta,a) - u_R(\theta,a') < 0$, then $u_R(\theta',a) - u_R(\theta',a') < 0$ for all $\theta' < \theta$. Similarly, for any $a' < a$, if $u_R(\theta,a') - u_R(\theta,a) > 0$, then $u_R(\theta',a') - u_R(\theta',a) > 0$ for all $\theta' > \theta$. This means that $N(a)$ is the union of nested sets that are located at either the lower or upper ends of I_a. We will exploit this structure to show $\pi(N(a)|a) = 0$.

For each $a' \in A^\circ$, $a' > a$, let us define

$$\hat{N}(a'|a) \equiv \Big\{\theta \in I_a : u_R(\theta, a) - u_R(\theta, a') < 0\Big\},$$

and

$$\hat{\theta}(a'|a) \equiv \sup \Big\{\theta \in I_a : u_R(\theta, a) - u_R(\theta, a') < 0\Big\}.$$

It follows that

$$\big(-\infty,\, \hat{\theta}(a'|a)\big) \cap I_a \subseteq \hat{N}(a'|a) \subseteq \big(-\infty,\, \hat{\theta}(a'|a)\big] \cap I_a \tag{35}$$

and $\pi(\hat{N}(a'|a)|a) = 0$.

Analogous, for each $a' \in A^\circ$, $a' < a$, define

$$\tilde{N}(a'|a) \equiv \Big\{\theta \in I_a : u_R(\theta, a') - u_R(\theta, a) > 0\Big\},$$

and

$$\tilde{\theta}(a'|a) \equiv \inf \Big\{\theta \in I_a : u_R(\theta, a') - u_R(\theta, a) > 0\Big\},$$

then

$$\big(\tilde{\theta}(a'|a),\, \infty\big) \cap I_a \subseteq \tilde{N}(a'|a) \subseteq \big[\tilde{\theta}(a'|a),\, \infty\big) \cap I_a \tag{36}$$

and $\pi(\tilde{N}(a'|a)|a) = 0$.

Let $\hat{N}(a) \equiv \cup_{a' \in A^\circ, a' > a} \hat{N}(a'|a)$ and $\tilde{N}(a) \equiv \cup_{a' \in A^\circ, a' < a} \tilde{N}(a'|a)$, then we have $N(a) = \hat{N}(a) \cup \tilde{N}(a)$. In order to show $\pi(N(a)|a) = 0$, it suffices to show both $\pi(\hat{N}(a)|a) = 0$ and $\pi(\tilde{N}(a)|a) = 0$. Below we will show $\pi(\hat{N}(a)|a) = 0$. The fact that $\pi(\tilde{N}(a)|a) = 0$ follows from similar arguments.

Let $\hat{\theta}(a) \equiv \sup_{a' \in A^\circ, a' > a} \hat{\theta}(a'|a)$. By (35) and the definition of $\hat{N}(a)$, we have

$$(-\infty, \hat{\theta}(a)) \cap I_a \subseteq \hat{N}(a) \subseteq (-\infty, \hat{\theta}(a)] \cap I_a.$$

However, note that if $\hat{\theta}(a) \in \hat{N}(a)$, then $\hat{\theta}(a) \in \hat{N}(a'|a)$ for some $a' \in A^\circ$ with $a' > a$, and this would imply $\pi(\{\hat{\theta}(a)\} \,|\, a) = 0$ since $\pi(\hat{N}(a'|a) \,|\, a) = 0$ for all $a' \in A^\circ$ with $a' > a$. Therefore, in order to prove $\pi(\hat{N}(a) \,|\, a) = 0$, it suffices to prove that $\pi\big((-\infty, \hat{\theta}(a)) \cap I_a \,|\, a\big) = 0$.

To this end, note that $(-\infty, \hat{\theta}(a)) = \cup_{n=1}^{\infty} (-\infty, \hat{\theta}(a) - 1/n)$. Since $\hat{\theta}(a) \equiv \sup_{a' \in A^\circ, a' > a} \hat{\theta}(a'|a)$, for each $n \geq 1$, $(-\infty, \hat{\theta}(a) - 1/n) \subseteq (-\infty, \hat{\theta}(a'|a))$ for some $a' \in A^\circ$ with $a' > a$. So for each $n \geq 1$,

$$\pi\Big((-\infty, \hat{\theta}(a) - 1/n) \cap I_a \,|\, a\Big) \leq \pi\Big(\,(-\infty, \hat{\theta}(a'|a)) \cap I_a \,|\, a\Big) \leq \pi\big(N(a'|a) \,|\, a\big) = 0.$$

As a result, we have

$$\pi\Big(\,(-\infty,\hat\theta(a))\cap I_a\,|\,a\Big) = \pi\left(\cup_{n=1}^{\infty}\Big((-\infty,\ \hat\theta(a)-1/n)\cap I_a\Big)\,\Big|\,a\right)$$
$$\le \sum_{n=1}^{\infty}\pi\Big((-\infty,\hat\theta(a)-1/n)\cap I_a\,\Big|\,a\Big) = 0,$$

so $\pi(\hat N(a)\,|\,a) = 0$.

Using similar arguments as above, we can establish that $\pi(\tilde N(a)\,|\,a) = 0$ as well, so $\pi(N(a)|a) = \pi(\,\hat N(a)\cup\tilde N(a)\,|\,a) = 0$. For each $a\in A^\circ$, let

$$\hat\Theta_0(a) \equiv \big\{\theta\in I_a : u_R(\theta,a) = u_R(\theta,a') \text{ for all } a'\in A^\circ\big\} = [N(a)]^c,$$

so $\pi(\,\hat\Theta_0(a)\,|\,a) = 1-\pi(N(a)|a) = 1$.

Let $\hat\Theta_0 \equiv \{\theta\in\Theta : u_R(\theta,a) = u_R(\theta,a') \text{ for all } a,a'\in A^\circ\}$. We have

$$\mu_0(\hat\Theta_0) = \pi(\hat\Theta_0\times A) = \pi(\,(\hat\Theta_0\times A)\cap E\,),$$

so

$$\begin{aligned}
\mu_0(\hat\Theta_0) &= \int_{\Theta\times A}\mathbb{1}_{\hat\Theta_0\times A}\cdot\mathbb{1}_E\, d\pi(\theta,a)\\
&= \int_A\left[\int_\Theta \mathbb{1}_{\hat\Theta_0\times A}\cdot\mathbb{1}_E\, d\pi(\theta|a)\right] d\pi_A(a)\\
&= \int_A\left[\int_{I_a} \mathbb{1}_{\hat\Theta_0\times A}\, d\pi(\theta|a)\right] d\pi_A(a)\\
&= \int_A \pi(\,\hat\Theta_0(a)|\,a\,)\, d\pi_A(a)\\
&= \int_A 1\, d\pi_A(a) = 1.
\end{aligned}$$

Recall that $\Theta_0(a,a') \equiv \{\theta\in\Theta : u_R(\theta,a) = u_R(\theta,a')\}$ and by our assumption $\mu_0(\Theta_0(a,a')) < 1$ for all distinct $a,a'\in A$. Since $\hat\Theta_0 \equiv \{\theta\in\Theta : u_R(\theta,a) = u_R(\theta,a') \text{ for all } a,a'\in A^\circ\}$ and we have established that $\mu_0(\hat\Theta_0) = 1$, A° must be a singleton set. Since $\pi(\Theta\times A^\circ)\ge\pi(E^\circ) = 1$, it follows that π is a no-information outcome. □

B.7 Credible Persuasion in Games

In this section, we generalize the framework in Section 2.1 to a setting with multiple Receivers, where the Sender can also take actions after information is disclosed. We also allow the state space and action space to be infinite.

Consider an environment with a single Sender (she) and r Receivers (each of whom is a he). The Sender has action set A_S while each Receiver $i \in \{1, \ldots, r\}$ has action set A_i. Let $A = A_S \times A_1 \times \ldots \times A_r$ denote the set of action profiles. Each player has payoff function $u_i : \Theta \times A \to \mathbb{R}$, $i = S, 1, \ldots, r$, respectively. The state space Θ and action spaces A_i are Polish spaces endowed with their respective Borel sigma-algebras. Players hold full-support common prior $\mu_0 \in \Delta(\Theta)$. We refer to $G = (\Theta, \mu_0, A_S, u_S, \{A_i\}_{i=1}^r, \{u_i\}_{i=1}^r)$ as the base game.

Let M be a Polish space that contains A. The Sender chooses an information structure $\lambda \in \Delta(\Theta \times M)$ where $\lambda_\Theta = \mu_0$: note that this formulation implies that the information structure generates *public* messages observed by all Receivers. Together the information structure and the base game constitute a Bayesian game $\mathcal{G} = \langle G, \lambda \rangle$, where:[28]

1. At the beginning of the game a state-message pair (θ, m) is drawn from the information structure λ;
2. The Sender observes (θ, m) while the Receivers observe only m; and
3. All players choose an action simultaneously.

A strategy profile $\sigma : \Theta \times M \to A$ in $\mathcal{G}$ consists of a Sender's strategy $\sigma_S : \Theta \times M \to A_S$ and Receivers' strategies $\sigma_i : M \to A_i$, $i = 1, \ldots, r$. For each profile of Sender's information structure and players' strategies (λ, σ), players' expected payoffs are given by

$$U_i(\lambda, \sigma) = \int_{\Theta \times M} u_i(\theta, \sigma(\theta, m)) \, d\lambda(\theta, m) \quad \text{for } \; i = S, 1, \ldots, r.$$

We now generalize the notion of credibility and incentive compatibility in Section 2 to the current setting. For each λ, let $D(\lambda) \equiv \left\{ \lambda' \in \Delta(\Theta \times M) : \lambda'_\Theta = \mu_0, \lambda'_M = \lambda_M \right\}$ denote the set of information structures that induce the same distribution of messages as λ. Definition 5 is analogous to Definition 1, which requires that given the players' strategy profile, no deviation in $D(\lambda)$ can be profitable for the Sender.

Definition 5. *A profile* (λ, σ) *is* ***credible*** *if*

$$\lambda \in \underset{\lambda' \in D(\lambda)}{\arg\max} \int u_S(\theta, \sigma(\theta, m)) \, d\lambda'(\theta, m). \tag{37}$$

In addition, Definition 6 generalizes Definition 2, and requires players' strategies to form a Bayesian Nash equilibrium of the game $\langle G, \lambda \rangle$.

[28] The information structure λ can be viewed as "additional information" observed by both the Sender and the Receivers, on top of the base information structure where the Sender observes the state and the Receivers do not observe any signal.

Definition 6. *A profile* (λ, σ) *is* ***incentive compatible*** *(IC) if* σ *is a Bayesian Nash equilibrium in* $\mathcal{G} = \langle G, \lambda \rangle$. *That is,*

$$\sigma_S \in \underset{\sigma'_S:\Theta\times M\to A_S}{\arg\max}\; U_S(\lambda, \sigma'_S, \sigma_{-S}) \quad \textit{and} \quad \sigma_i \in \underset{\sigma'_i:M\to A_i}{\arg\max}\, U_i(\lambda, \sigma'_i, \sigma_{-i}) \textit{ for } i = 1, \ldots, r. \tag{38}$$

Note that in Definition 5, when the Sender deviates to a different information structure, say λ', we use the original strategy profile $\sigma(\theta, m)$ to predict players' actions in the ensuing Bayesian game $\langle G, \lambda' \rangle$. One might worry that the Sender may simultaneously change not only her information structure but also her strategy $\sigma_S(\theta, m)$ in $\langle G, \lambda' \rangle$. This, however, is unnecessary since the Sender's optimal strategy in $\langle G, \lambda' \rangle$ will remain unchanged: the Sender knows θ perfectly, her best response in $\langle G, \lambda' \rangle$ depends only on θ and the Receivers' actions (and not on her own information structure).