A belief-based approach to signaling*

Frédéric Koessler† Marie Laclau‡ Tristan Tomala§

July 8, 2025

Abstract

We provide a geometric characterization of the set of interim equilibrium payoffs in the general class of costly signaling games. Our characterization offers a unified, belief-based framework to study both cheap talk and costly signaling, with or without transparent motives. The key ingredient is the analysis of Bayes-plausible belief distributions and signal-contingent interim values that are incentive-compatible for the sender. Geometrically, this leads to a constrained convexification of the graphs of the interim value correspondences. We apply and illustrate the results in a class of intimidation games. We also derive the sender's best equilibrium payoff under transparent motives. Finally, we compare the equilibrium outcomes to those arising when the sender can commit to a signaling strategy.

KEYWORDS: belief-based approach, cheap talk, information transmission, incomplete information, intimidation games, signaling.

1 Introduction

Information asymmetries are a well-documented and significant source of market failures. A key economic concept for addressing such situations is signaling, wherein an informed party undertakes a costly action to credibly convey information to an uninformed decision maker. While job market signaling is a well-known example, signaling can also serve as a means of deterrence in various strategic settings.

A notable application is litigation and settlement negotiations, where plaintiffs often possess private information about the strength of their case and may attempt to intimidate the defendant by investing heavily in legal preparation or issuing aggressive demands—potentially leading the defendant to forgo confrontation. Similar forms of strategic intimidation arise in trade policy, where tariff threats can discourage retaliation; in procurement, where costly certification may persuade a buyer not to conduct verification; and in speculative trade, where a trader signals the quality of their information about future price movements through the size of their position, potentially deterring a counterparty from taking

*We thank Raphael Boleslavky, Françoise Forges, Laurent Mathevet, and Vasiliki Skreta for useful comments. The authors gratefully acknowledge the support of the ANR through the program Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047).

†HEC Paris and GREGHEC-CNRS, 1 rue de la Libération, 78351 Jouy-en-Josas, France; koessler@hec.fr

‡HEC Paris and GREGHEC-CNRS, 1 rue de la Libération, 78351 Jouy-en-Josas, France; laclau@hec.fr

§HEC Paris and GREGHEC, 1 rue de la Libération, 78351 Jouy-en-Josas, France; tomala@hec.fr

the opposite side. Key theoretical questions in these environments concern the situations under which different types of senders can credibly exert such pressure and the conditions under which receivers are willing to back down. We refer to these settings as *intimidation games*—a specific class of signaling games.

More broadly, signaling games are foundational across classical economic domains. In job market signaling, candidates communicate their qualifications to potential employers; in reputation analysis, individuals build and manage their reputations in social and professional contexts; in financial markets, firms signal their financial health to attract investors. Signaling theory also underpins advertising strategies, where firms convey product quality through costly promotional efforts. Beyond economics, signaling games have applications in biology, notably exemplified by the handicap principle, where individuals signal their fitness through costly traits.

In this paper, we provide a geometric characterization of the set of all (perfect Bayesian) interim equilibrium payoffs for the general class of signaling games, including all those discussed in the applications above. We consider a finite set of sender types, but otherwise make no assumptions on preferences: utility functions may violate single-peakedness or single-crossing; Sets of types, signals, and actions need not be one-dimensional. Within this framework, the sender may use cheap talk messages in addition to costly signals. We adopt a tractable belief-based approach, inspired by techniques from the literatures on repeated games with incomplete information, cheap talk, and Bayesian persuasion. This approach allows us to characterize equilibrium outcomes without explicitly specifying the strategies of the sender and receiver or the associated belief system.

Unlike in models of cheap talk and Bayesian persuasion, information is conveyed through signals that have payoff consequences, so the sender's interim payoff (i.e., the expected payoff conditional on her type) depends not only on the belief induced to the receiver, but also on the specific signal used. Considering the collection of correspondences of interim values indexed by the sender's signal, a *splitting* of the prior belief (i.e., a Bayes-plausible distribution of posterior beliefs) must associate each posterior belief with an interim value under some signal. Taking into account the sender's equilibrium conditions, our method is based on incentive-compatible splittings of the prior type distribution, resulting in a *constrained convexification* of the graphs of the interim value correspondences. Notably, our characterization extends the existing equilibrium description for sender-receiver cheap talk games, regardless of any assumptions about the sender's preference, such as state-independent payoffs.

Our characterization implies that the set of interim equilibrium payoffs does not depend on the set of cheap talk messages, provided it is at least as large as the set of the sender's types: we show that regardless of the sizes of the action, signal and message sets, any interim equilibrium payoff of the sender can be obtained using a strategy that employs no more pairs of signals and messages than there are sender types. We deduce the existence of a perfect Bayesian equilibrium in signaling games with such sets of cheap talk messages.

We use our main equilibrium characterization to study intimidation games, a class of signaling games in which the sender first decides whether to engage in the interaction, then, if opting in, signals her strength to the receiver. The receiver can then choose to challenge the sender or to forgo the interaction. A strong sender is eager to be challenged, whereas a weak sender is better off when the receiver forgoes.

As an application, consider litigation and settlement negotiations: a plaintiff may have either a strong or a weak case and can decide whether to engage in costly actions—such as preparing extensive evidence or hiring a prominent legal team—that serve as signals to the defendant. After observing the plaintiff's actions, the defendant chooses whether to forgo litigation by settling or to challenge the case in court. Other examples include tariffs on imports negotiation or speculative trade.[1] Our geometric approach allows us to find all possible equilibrium outcomes in this environment. We show that two types of equilibria can arise: non-revealing (i.e., pooling) equilibria where both strong and weak plaintiffs behave similarly and settlement occurs, and partially revealing (i.e., semi-separating) equilibria where stronger plaintiffs separate themselves from weaker ones. We characterize which equilibria arise depending on the prior probability that the case is strong. A key element in the analysis is the *signaling pressure ratio*, which captures how attractive it is for a strong plaintiff to get a challenge compared to how costly it is for a weak plaintiff to face one. The signal that maximizes this pressure ratio—the *maximal signaling pressure point*—plays a central role, as it defines the most effective way for a strong plaintiff to partially separate herself from weaker ones.

We also consider signaling games with *transparent motives* (Lipnowski and Ravid, 2020), where the sender's utility does not depend on her type. We characterize the maximal equilibrium payoff of the sender via the quasi-concave envelope of the sender's best value function, and provide sufficient conditions under which cheap talk messages can be dispensed with.

Finally, we consider the model where the sender has full commitment power. We characterize the set of all feasible interim values for this scenario, and study the impact of commitment in our main application to intimidation games.

Related literature The literature on signaling games, initiated by Spence (1973) in economics and by Zahavi (1975) and Grafen (1990) in biology, has led to significant research in applied work across various fields and in the analysis of equilibrium refinements in games. For a literature review, see, e.g., Kreps and Sobel (1994) and Sobel (2020). The combinaison of cheap talk (costless) and costly signals have been studied by Austen-Smith and Banks (2002), Wu (2022), Reny (2024), and Lizzeri, Lou, and Perego (2024).

Our belief-based approach to signaling games shares significant connections with the approach employed in the literature on repeated games with incomplete information, cheap talk games, and Bayesian persuasion. Roughly, the idea revolves around viewing the informed player's strategy as a means to split the prior belief of the uninformed player into a convex combination of posteriors beliefs. These splittings enable the informed player to generate new (ex-ante and interim) payoffs and write these payoffs as convex combinations of sender's (ex-ante and interim) values at posterior beliefs. For a comprehensive review and comparison of how this technique is applied across these areas, refer to Forges (2020).

In the context of zero-sum repeated games where only one player possesses private information, Aumann and Maschler (1966, 1995) and Ponssard and Zamir (1973) showed that the equilibrium payoff

[1] A previous version of the paper (Koessler, Laclau, and Tomala, 2024) also includes applications of our geometric characterization to more classical signaling games such as cheap-talk with lying cost and job market signaling.

for the informed player is given by the concave closure (or concavification) of the minmax value of the one-shot game with no information and common belief, evaluated at the prior. The minmax theorem for those zero-sum games implies that the informed player's commitment to her strategy is irrelevant.

The approach can be applied to one-shot non-zero-sum scenarios when the informed player is able to commit to an information disclosure strategy. For instance, in the Bayesian persuasion setting introduced by Kamenica and Gentzkow (2011), the sender commits to her messaging strategy before learning the state. The maximum ex-ante expected payoff she can obtain is given by the concave closure $\operatorname{cav} w(p_0)$ at the prior p_0 of the highest ex-ante expected payoff $w(p)$ for the sender, when the uninformed receiver selects an optimal action based on belief p. Doval and Smolin (2024) characterize the set of interim payoffs of the sender, for all possible messaging strategies, and its Pareto frontier, without imposing incentives conditions for the sender. See Section 6 for a more detailed comparison between our approach and commitment in Bayesian persuasion.

Hart (1985) expands upon Aumann and Maschler's analysis to encompass non-zero-sum repeated games and provides a geometric characterization of all Nash equilibrium payoffs in undiscounted games. Since no commitment is assumed, the characterization needs to account for the informed player's incentive compatibility conditions, and it keeps track of the interim payoff for each type. These ideas have been adapted and extended by Forges (1984, 1990) and Aumann and Hart (2003) for characterizing equilibrium payoffs in cheap talk games where both players send messages over multiple periods before taking action with payoff consequences. Similar techniques have been employed by Chakraborty and Harbaugh (2010) and Lipnowski and Ravid (2020) to characterize equilibrium payoffs in cheap talk games with transparent motives. A belief-based approach characterization of the sender's ex-ante preferred mediated cheap talk equilibrium under general utility functions can be found in Salamanca (2021). Relatedly Corrao and Dai (2023) consider games with transparent motives and study the value of mediation for the sender by comparing the sender preferred mediated equilibrium with the Bayesian persuasion and cheap talk benchmarks.

Unlike the cheap talk literature, in addition to costless messages, the sender can use payoff-relevant signals (also seen as actions observable by the receiver) from an exogenously given set. These signals are payoff-relevant either for the sender, for the receiver, or for both. Technically, this implies that the sender can split the prior belief and the corresponding interim payoffs on multiple graphs of interim value correspondences, one for each signal. In cases where all signals of the sender are payoff-irrelevant, the characterization reduces to the one mentioned above for sender-receiver cheap talk games.[2] Our characterization can also be related to disclosure games with hard information, as in Forges and Koessler (2008), by considering signals that are payoff-irrelevant for the receiver and, either payoff-irrelevant or very costly for the sender depending on her type.

Lastly, our result is related to the recent contribution of Boleslavsky and Shadmehr (2023), who extend Bayesian persuasion by including payoff-relevant signals. They characterize the maximal ex-ante expected payoff the sender can achieve in signaling games in which the sender ex-ante commits to her signaling strategy. Like our work, their approach revolves around the graphs of value correspondences

[2]The action stage that follows cheap talk in Aumann and Hart (2003) is slightly more general, as it allows both the sender and the receiver to take an action.

for each signal. However, in line with the cheap talk literature, we do not assume commitment, keep track of interim payoffs and maintain incentive-compatible conditions for the sender.

Organization of the paper Section 2 presents the model. Section 3 contains the main results: the geometric characterization of equilibrium payoffs and its corollaries. In Section 4, we apply the characterization to a class of intimidation games that feature signaling pressure. Section 5 focuses on the special case of transparent motives, for which we derive additional results. In Section 6, we compare the equilibrium payoffs obtained in our setting with those that arise when the sender can commit ex ante to a signaling strategy, as in the literature on Bayesian persuasion. All proofs not included in the main text are provided in the Appendix.

2 Model

Setup. There are two players, a sender and a receiver. The sender is privately informed of her type t drawn from a nonempty finite set T, according to a prior probability distribution p_0. The sender has a set of costly signals S and a set of costless (cheap talk) messages M, the set of actions of the receiver is denoted by A.[3] We assume that A, S, and M are nonempty compact metric spaces. Except if stated otherwise, we assume that M has at least $|T|$ elements. Given action $a \in A$, signal $s \in S$, and type $t \in T$, the utility of the sender is $u(a, s, t)$, and the utility of the receiver is $u_R(a, s, t)$, the utility functions u, u_R are continuous on $A \times S \times T$. Utility functions are extended to the set of mixed actions $\Delta(A)$ in the usual way by taking expectation.[4]

Signaling game. The timeline is the following:

1. The sender's type t is drawn from T according to p_0.
2. The sender observes the type t, chooses a signal $s \in S$ and a cheap talk message $m \in M$.
3. The receiver observes the signal s and the cheap talk message m, then chooses an action $a \in A$.

A strategy for the sender is $\sigma : T \to \Delta(S \times M)$, and a strategy for the receiver is a Borel measurable $\tau : S \times M \to \Delta(A)$. The interim payoff for the sender induced by the strategy profile (σ, τ) is given by $v = (v_t)_{t \in T}$ with

$$v_t = \int_{S \times M} u(\tau(s, m), s, t) d\sigma(s, m \mid t), \text{ for every } t \in T.^5 \tag{1}$$

A belief system is given by a Borel measurable map $\mu : S \times M \to \Delta(T)$, where $\mu(t \mid s, m)$ denotes the belief assigned by the receiver to type t when observing (s, m).

[3]We can allow the set of actions of the receiver to depend on the signal, i.e., let $A(s)$ denote the set of actions of the receiver given s.

[4]For a metric compact X, $\Delta(X)$ denotes the set of Borel probability measures over X.

[5]We focus on the interim payoffs of the sender for simplicity, the expected payoff of the receiver can be easily included in the analysis.

Perfect Bayesian equilibrium. For every $p \in \Delta(T)$ and $s \in S$, let $Y(s,p)$ be the set of mixed actions that are optimal for the receiver given signal s and belief p. That is:

$$Y(s,p) := \arg\max_{y \in \Delta(A)} u_R(y,s,p),$$

where $u_R(y,s,p) = \sum_t p(t) u_R(y,s,t) = \sum_t p(t) \int_A u_R(a,s,t) dy(a)$ with some abuse of notation.

A strategy profile (σ,τ), with interim payoff $v = (v_t)_{t \in T}$, is a perfect Bayesian equilibrium (PBE) of the signaling game if there exists a belief system μ such that the following three conditions are satisfied. First, for each type, the sender cannot increase her interim expected payoff by deviating to any pair of signal and message. Second, for every pair of signal and message, the receiver chooses optimal actions given the signal and her belief about the sender's type. Finally, the receiver's belief is computed using Bayes' rule whenever possible. Formally:

(i) Sequential rationality for the sender: for every $t \in T$ and $(s,m) \in S \times M$,

$$v_t \geq u(\tau(s,m),s,t);$$

(ii) Sequential rationality for the receiver: for every $(s,m) \in S \times M$,

$$\tau(s,m) \in Y(s,\mu(s,m));$$

(iii) Belief consistency: for every $t \in T$ and every Borel set $B \subseteq S \times M$,

$$p_0(t)\sigma(B \mid t) = \sum_{\tilde{t} \in T} p_0(\tilde{t}) \int_B \mu(t \mid s,m) d\sigma(s,m \mid \tilde{t}).$$

Such a triple (σ,τ,μ) is called a PBE assessment. We say that an interim payoff for the sender $v \in \mathbb{R}^T$ is a *interim PBE payoff* if it is induced by some PBE strategy profile.

3 Geometric characterization of equilibrium payoffs

For every $s \in S$ and $p \in \Delta(T)$, let $\mathcal{E}(s,p)$ be the set of all feasible interim payoffs of the sender when the receiver chooses an optimal (mixed) action given s and p:

$$\mathcal{E}(s,p) := \left\{ v \in \mathbb{R}^T : \exists\, y \in Y(s,p) \text{ s.t. } v = (u(y,s,t))_t \right\}.$$

We call this the set of *interim values* for the sender at p given s.

Observe that $\mathcal{E}(s,p)$ depends on the signal s through two distinct channels: through $Y(s,p)$, when the receiver's optimal action varies with s for a given posterior p, and through the direct impact of s on the utility of the sender for a given action of the receiver. We say that the signal s is cheap talk if it does not affect the utilities of the sender and of the receiver; in this case, $\mathcal{E}(s,p)$ is independent of s, as in Forges (1984, 1990, 1994) and Aumann and Hart (2003).

We define then the *modified* set of interim values for the sender at p given s as:

$$\mathcal{E}^+(s,p) := \big\{v \in \mathbb{R}^T : \exists\, y \in Y(s,p) \text{ s.t. } \forall t, v_t \geq u(y,s,t) \text{ and } v_t = u(y,s,t) \text{ if } p(t) > 0\big\}.$$

To obtain the modified interim value correspondence, consider for any p in $\Delta(T)$, the interim values given the signal s and, at the boundary of the simplex, adjust these values by granting higher values to the sender's types with vanishing probabilities. Notice that $\mathcal{E}(s,p) \subseteq \mathcal{E}^+(s,p)$, and $\mathcal{E}(s,p) = \mathcal{E}^+(s,p)$ if p has full support.[6]

The graph of the modified interim value correspondence of the sender given s is denoted by:

$$G_s := \operatorname{gr} \mathcal{E}^+(s,\cdot) := \big\{(v,p) \in \mathbb{R}^T \times \Delta(T) : v \in \mathcal{E}^+(s,p)\big\}.$$

We denote by $G = \bigcup_{s \in S} G_s$ the union of the graphs of the interim value correspondences for all s, that is,

$$\begin{aligned} G \;=\; & \big\{(v,p) \in \mathbb{R}^T \times \Delta(T) : \exists s \in S, y \in Y(s,p), \\ & \text{s.t. } \forall t, v_t \geq u(y,s,t) \text{ and } v_t = u(y,s,t) \text{ if } p(t) > 0\}. \end{aligned}$$

We let $\mathrm{co_f}\,(G)$ be the "flat" convex hull of G, defined as the set of interim payoffs obtained by convexifying G with respect to p, when the interim payoff v is kept constant. Formally:

$$\mathrm{co_f}\,(G) := \Big\{\Big(v, \sum_{k=1}^{|T|} \lambda_k p_k\Big) \in \mathbb{R}^T \times \Delta(T) : (\lambda_k)_{k=1}^{|T|} \in \Delta(\{1,\dots,|T|\}),\ \forall k, (v,p_k) \in G\Big\},$$

see Figure 1 for an illustration.

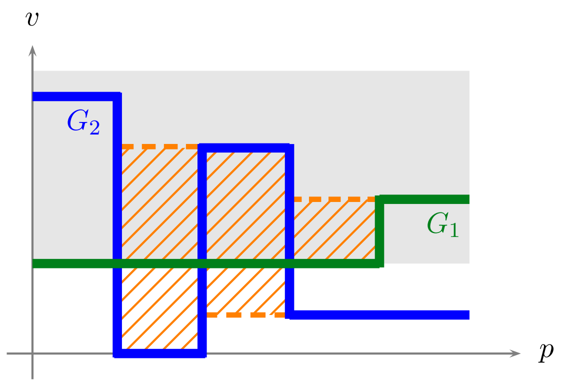

Figure 1: The flat convex hull of $G = G_1 \cup G_2$ is the dashed-orange area, the grey area is the INTIR set.

Since p lies in the $(|T|-1)$–dimensional simplex, only convex combination of cardinality at most $|T|$ are considered, thanks to Carathéodory's theorem. A family $(\lambda_k, p_k)_{k=1}^{|T|}$ such that $p_0 = \sum_{k=1}^{|T|} \lambda_k p_k$ and (v, p_k) belongs to G for all k, will be called an *incentive compatible splitting* of p_0.

[6]In Pęski (2014), the correspondance $\mathcal{E}^+(s,\cdot)$ is called the *enhancement* of $\mathcal{E}(s,\cdot)$.

Next, we introduce an individual rationality condition. Let INTIR_s be the set of all interim payoffs $v \in \mathbb{R}^T$ such that there exists $p \in \Delta(T)$ and $y \in Y(s,p)$ with $v_t \geq u(y,s,t)$ for every $t \in T$. Finally, let

$$\text{INTIR} := \bigcap_{s \in S} \text{INTIR}_s,$$

be the set of all interim individually rational payoffs $v \in \mathbb{R}^T$. These are the payoffs such that for every $s \in S$, there exists $p \in \Delta(T)$ and $y \in Y(s,p)$ such that $v_t \geq u(y,s,t)$ for every $t \in T$. The interim payoffs in INTIR are called *interim individually rational.*[7] The set INTIR is illustrated by the grey area in Figure 1. Observe that INTIR_s is directly obtained from G_s: this is the set of interim payoffs which are coordinate-wise above the projection of G_s on the set of payoff space, i.e., $\text{INTIR}_s = \text{proj}_{\mathbb{R}^T}(G_s) + \mathbb{R}^T_+$.

Our main result is that the set of interim PBE payoffs is fully pinned down by the interim individual rationality condition INTIR, together with the flat convex hull of modified interim values $\text{co}_\text{f}\,(G)$. This is stated in the following theorem.

Theorem 1. *The interim payoff $v \in \mathbb{R}^T$ is an interim PBE payoff of the sender in the signaling game if and only if $v \in \text{INTIR}$ and $(v, p_0) \in \text{co}_\text{f}\,(G)$.*

This theorem shows that geometric methods are useful to find all the equilibrium outcomes of the signaling game via two operations: taking the convex hull with respect to the beliefs while keeping the sender's payoff constant (flat convexification) and deleting all payoffs outside a comprehensive set (INTIR condition). This geometric characterization is used and illustrated in Section 4, where we provide all equilibrium payoffs, for all possible priors, in a class of signaling games called intimidation games. This method also allows direct constructions of perfect Bayesian equilibria in a belief-based way: from a convex combination of posteriors, we back out a signaling strategy. The flat convexification condition guarantees the incentive conditions for the sender, the sequential rationality for the receiver on equilibrium path follows from the construction of the interim value correspondence. The interim individual rationality condition ensures sequential rationality off-equilibrium path. This is a key difference with cheap talk models where PBE outcomes and Nash outcomes are the same, the receiver just has to ignore off-equilibrium cheap talk messages, treating them as on-path messages. When signals are payoff relevant, this is no longer possible and the receiver has to "punish" off-path signals.

The "if" part of the proof is constructive, it uses the cardinality of the set of types to construct a sender equilibrium strategy with finite support. The equilibrium strategy of the receiver is dictated on path by the interim value correspondence and given by the INTIR condition off path. For the "only if" direction, the necessity of the INTIR condition is easy to see. For the flat convexification part, we use a technical lemma which says that the prior belief belongs to the convex hull of the posteriors. This is well known for distributions of posteriors with finite support. With arbitrary sets of signals and messages, it is easy to see that the prior belongs to the *closure* of the convex hull. Yet, we show that it always belongs to the convex hull, no matter the strategy of the sender.

[7] This individual rationality condition requires the receiver to choose mixed actions which are optimal for some belief. A weaker condition requires that for every $s \in S$, there exists $y \in \Delta(A)$ such that $v_t \geq u(y,s,t)$ for every $t \in T$. This is the relevant condition for the study of Nash equilibria instead of PBE of the signaling game.

Theorem 1 has several theoretical consequences. First, the next corollary shows that to characterize all interim PBE payoffs of the signaling game with enough cheap talk messages, it suffices to focus on the sender's strategies with finite support, of cardinality at most $|T|$, regardless of the sizes of the sets of actions, signals and messages.[8] This result also implies that the set of interim PBE payoffs does not depend on the message space M, provided that M contains at least $|T|$ elements. Related results appear in Heumann (2020) and Reny (2024).

Corollary 1. *The interim payoff $v \in \mathbb{R}^T$ is an interim PBE payoff of the sender in the signaling game if and only if v is an interim payoff of a PBE in which at most $|T|$ pairs of signal and message are used with positive probability.*

What about the existence of PBE? If the game is finite (i.e., both sets of signals S and of actions A are finite), then there exists at least one sequential equilibrium (Kreps and Wilson, 1982, Proposition 1), and any sequential equilibrium is a PBE. Under the compactness and continuity assumptions of our model, Manelli (1996) shows that the signaling game has a PBE as long as the message space is large enough, precisely for $M = \Delta(A)$.[9] Corollary 1 shows that for any such PBE, the same interim payoff can be obtained through a PBE of the signaling game with any message set of cardinality at least $|T|$. Hence, due to our maintained assumption that M contains at least $|T|$ elements, Corollary 2 in Manelli (1996) and Corollary 1 imply the existence of a PBE in our signaling game. Summarizing:

Corollary 2. *The signaling game has at least one PBE and the set of interim payoffs:*

$$\big\{v \in \mathbb{R}^T : v \in \text{INTIR} \textit{ and } (v, p_0) \in \text{co}_{\text{f}}\,(G)\big\},$$

is non-empty.

A cheap talk game is a signaling game where signals are payoff-irrelevant, i.e., $u(a, s, t) = u(a, s', t)$ and $u_R(a, s, t) = u_R(a, s', t)$ for all signals s and s'. Theorem 1 characterizes interim PBE payoffs for a cheap talk game as a particular case. In such a game, we have $G = G_s$ for every s, and the condition $v \in \text{INTIR}$ is irrelevant as it is straightforwardly implied by $(v, p_0) \in \text{co}_{\text{f}}\,(G)$. Therefore, we have the following corollary of Theorem 1.

Corollary 3. *Assume that signals are payoff-irrelevant. The interim payoff $v \in \mathbb{R}^T$ is an interim PBE payoff of the sender in the cheap talk game iff $(v, p_0) \in \text{co}_{\text{f}}\,(G)$.*

Corollary 3 is well known in the cheap talk literature.[10] Notice that in a cheap talk game, every interim value $v \in \mathcal{E}(s, p_0)$ is a "babbling" interim PBE payoff. However, in a signaling game with payoff-relevant messages, interim values may fail to be interim individually rational and thus a non-revealing PBE may not exist. That is, in a signaling game, the set $\mathcal{E}(s, p_0) \cap \text{INTIR}$ may be empty for every s (see Section 4 for some examples). We deduce the following.

[8] Taking into account the receiver's ex ante expected payoff would require $|T| + 1$ pairs of signal and message instead of $|T|$.

[9] A PBE may not exist for continuous sets of signals if the message space is too small, see the examples in Manelli (1996).

[10] The characterization of Corollary 3 is known for Nash equilibria of cheap talk games where PBE and Nash equilibrium payoffs coincide, see Forges (1984, 1990) and Aumann and Hart (2003).

Corollary 4. *The interim payoff v is a non-revealing interim PBE payoff if and only if there exists a signal s such that $v \in \mathcal{E}(s, p_0) \cap \text{INTIR}$.*

4 Intimidation and Signaling Pressure

Setting and assumptions. We consider a class of signaling games in which the sender can either drop out immediately or opt in and signal her strength to the receiver, who then decides whether to challenge the sender. A strong sender welcomes a challenge, while a weak sender prefers the receiver to back down. We call these games *intimidation games*, as a weak sender aims to scare off the receiver. To be concrete, consider a litigation scenario where the plaintiff's case may be weak or strong. The plaintiff can signal strength through investments in legal representation and evidence preparation. If the case proceeds to court, the outcome is highly beneficial for a strong plaintiff but disastrous for a weak one, who would instead prefer to settle.

Here is the formal description. There are two types high and low, denoted $\{t_H, t_L\}$, with prior probabilities $p_0(t_H) = p_0$ and $p_0(t_L) = 1 - p_0$. The set of signals is $S = \{s_O\} \cup S_I$, where by choosing signal s_O, the sender drops out and concludes the game. Otherwise, the sender can opt in by choosing a signal $s \in S_I$, where S_I is a nonempty compact set. If the sender opts in, the receiver can either forgo (action $a = f$) or challenge (action $a = c$).

The utility functions have the following features. If the sender opts in, then the receiver strictly prefers to forgo in front of a high type, and to challenge a low type. That is, for every signal $s \in S_I$,

$$u_R(f, s, t_H) > u_R(c, s, t_H) \text{ and } u_R(f, s, t_L) < u_R(c, s, t_L).$$

Hence, for every $s \in S_I$, the set of optimal mixed actions as a function of the receiver's belief $p = p(t_H) \in [0, 1]$ is given by:

$$Y(s, p) = \begin{cases} \{c\} & \text{if } p < \bar{p}(s), \\ \Delta(\{c, f\}) & \text{if } p = \bar{p}(s), \\ \{f\} & \text{if } p > \bar{p}(s), \end{cases}$$

where for each $s \in S_I$, $\bar{p}(s) \in (0, 1)$ is such that the receiver is indifferent:

$$\bar{p}(s)u_R(f, s, t_H) + (1 - \bar{p}(s))u_R(f, s, t_L) = \bar{p}(s)u_R(c, s, t_H) + (1 - \bar{p}(s))u_R(c, s, t_L).$$

We normalize the sender's utility to zero when she drops out:

$$u(a, s_O, t) = 0, \quad \text{for every } a \in A \text{ and } t \in T.$$

We make the following assumptions on the sender's utility function in case she opts in. For every $s, s' \in S_I$:

$$u(c, s, t_H) := \gamma_H(s) > 0 \quad \text{and} \quad u(c, s, t_L) := \gamma_L(s) < 0, \tag{2}$$

$$u(f, s, t_H) = u(f, s', t_H) := \varphi_H > 0 \quad \text{and} \quad u(f, s, t_L) = u(f, s', t_L) := \varphi_L > 0. \tag{3}$$

Condition (2) assumes that the high type sender prefers being challenged over droping our, whereas the low type sender prefers to drop out. Condition (3) says that the sender always prefers the receiver to forgo over dropping out herself. The sender's utility does not depend on the signal when the receiver forgoes, while it may when the receiver challenges. These assumptions imply that the low type sender always prefers the receiver to forgo over being challenged: $\varphi_L > 0 > \gamma_L(s)$. However, the high type sender may prefer being challenged depending on whether $\gamma_H(s) > \varphi_H$ or $\gamma_H(s) < \varphi_H$.

Signaling ratios and benchmark values. For every signal $s \in S_I$, define the *signaling pressure ratio* by:

$$R(s) = \frac{\gamma_H(s) - \varphi_H}{\varphi_L - \gamma_L(s)}.$$

Given our assumptions, the *cost* to the low type from being challenged, $\varphi_L - \gamma_L(s)$, is always positive, while the *gain* to the high type from being challenged, $\gamma_H(s) - \varphi_H$, can be positive or negative. Hence, the signaling pressure ratio at s is positive if and only if the high type prefers being challenged over being forgone. In the former case, i.e., if $\gamma_H(s) > \varphi_H$, the high type gains from being challenged, so this ratio compares how much the high type gains from being challenged to the cost incurred by the low type when being challenged. If $\gamma_H(s) > \varphi_H$, a higher value of $R(s)$ means that the gain to the high type from being challenged is large relative to the cost to the low type from being challenged. If $\gamma_H(s) < \varphi_H$, a higher value of $R(s)$ means that the cost to the high type from being challenged is low relative to the cost to the low type from being challenged.

We call a signal $s^* \in \arg\max_{s \in S_I} R(s)$, a *maximal signaling pressure point*, that is, a signal that maximizes the signaling pressure ratio. For simplicity, we assume that s^* is unique, this assumption has very little impact on the analysis of the game. Depending on the application, the maximal signaling pressure point can be interpreted as follows. In litigation, s^* could represent filing a detailed complaint with strong supporting evidence and hiring a high-cost legal team, which significantly increases the expected trial success for a strong case. In trade policy, s^* could involve announcing tariffs on politically salient imports, where escalation benefits a hardliner sender and at the same time does not impose excessive exposure on a pragmatist. In speculative trade, the maximal signaling pressure point corresponds to a large speculative position if the sender (trader) is risk-neutral, or to a medium-sized position under risk aversion so that it is not too devastating for a poorly informed trader if challenged. In procurement, it might involve obtaining a mid-to-high tier certification and publishing some selected audit information.

Interim value correspondences. The set of interim values for the sender at p, given $s \in S_I$, is given by:

$$\mathcal{E}(s,p) = \begin{cases} \{(\gamma_H(s), \gamma_L(s))\} & \text{if } p < \bar{p}(s), \\ \text{co}\,\{(\gamma_H(s), \gamma_L(s)), (\varphi_H, \varphi_L)\} & \text{if } p = \bar{p}(s), \\ \{(\varphi_H, \varphi_L)\} & \text{if } p > \bar{p}(s), \end{cases}$$

and for $s = s_O$ by:

$$\mathcal{E}(s_O, p) = \{(0,0)\}.$$

The equation of the graph of the interim value correspondence given $s \in S_I$, projected on the sender's interim payoffs at $p = \bar{p}(s)$, is:

$$v_L = \frac{\varphi_L - \gamma_L(s)}{\varphi_H - \gamma_H(s)}(v_H - \varphi_H) + \varphi_L.$$

The slope is $\frac{\varphi_L - \gamma_L(s)}{\varphi_H - \gamma_H(s)} = -\frac{1}{R(s)}$, so it increases (the projected graph rotates anti-clockwise) with the ratio $R(s)$, as illustrated in Figures 2 and 3.

The projected graph intersects the horizontal axis at $\omega(s) := \varphi_L R(s) + \varphi_H$, which we call the *calibrated signaling value* of signal $s \in S_I$. Notice that the maximal signaling pressure point also maximizes the calibrated signaling value: $\omega(s^*) = \max_{s \in S_I} \omega(s)$.

Interim individual rationality. The interim individually rational payoffs are given by:

$$\text{INTIR}_{s_O} = \Big\{(v_H, v_L) : v_H \geq 0,\ v_L \geq 0\Big\}.$$

For $s \in S_I$, if $R(s) \leq 0$, then:

$$\text{INTIR}_s = \Big\{(v_H, v_L) : v_H \geq \gamma_H(s),\ v_L \geq \gamma_L(s)\Big\}.$$

For $s \in S_I$, if $R(s) > 0$, then:

$$\text{INTIR}_s = \Big\{(v_H, v_L) : v_H \geq \varphi_H,\ v_L \geq \gamma_L(s),\ v_L \geq -\frac{1}{R(s)}(v_H - \varphi_H) + \varphi_L\Big\}.$$

Hence, if $R(s^*) = \max_{s \in S_I} R(s) > 0$, we have $\text{INTIR}_{s_O} \cap \text{INTIR}_{s^*} \subseteq \text{INTIR}_{s_O} \cap \text{INTIR}_s$ for every $s \in S_I$, so we get

$$\begin{aligned} \text{INTIR} &= \bigcap_{s \in S} \text{INTIR}_s = \text{INTIR}_{s_O} \cap \text{INTIR}_{s^*} \\ &= \Big\{(v_H, v_L) : v_H \geq \varphi_H,\ v_L \geq 0,\ v_L \geq -\frac{1}{R(s^*)}(v_H - \varphi_H) + \varphi_L\Big\}. \end{aligned}$$

Otherwise, if $R(s^*) \leq 0$, then $R(s) \leq 0$ for every $s \in S_I$, so we get:

$$\text{INTIR} = \Big\{(v_H, v_L) : v_H \geq \max_{s \in S_I} \gamma_H(s),\ v_L \geq 0\Big\}.$$

Figure 2 illustrates the projections of the graphs of $\mathcal{E}^+(s, \cdot)$ on the sender's interim payoffs for three possible opt-in signals: s_1 (brown), s_2 (red), and s_3 (blue), where $R(s_3) > R(s_2) > 0 > R(s_1)$. The projection of the graph for $s = s_O$ is shown in green. In this figure, the maximal signaling pressure point is $s^* = s_3$, which corresponds to the projected graph that intersects the horizontal axis at the highest calibrated signaling value, $\max_s \omega(s) = \omega(s^*) > \varphi_H$, marked by a circle. The northeast region of all these curves represents interim payoffs satisfying the INTIR condition and is shaded in gray.

Figure 3 illustrates the projections of the graphs of $\mathcal{E}^+(s, \cdot)$ and the INTIR condition for three possible opt-in signals when $R(s_1) < R(s_2) < R(s_3) < 0$. Again, the maximal signaling pressure point is $s^* = s_3$, but in this case the highest calibrated signaling value satisfies $\omega(s^*) < \varphi_H$, indicated by the circle on the far right of the horizontal axis.

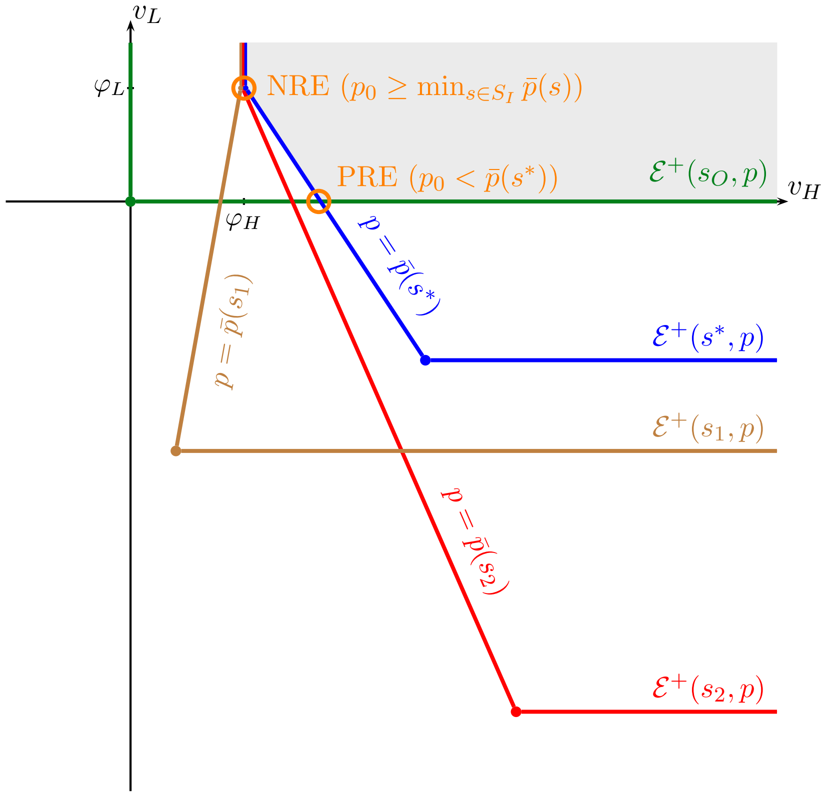

Figure 2: Projections of the graphs of $\mathcal{E}^+(s, \cdot)$ on the space of the sender's interim payoffs in a version of the intimidation game with $R(s^*) > 0$. The figure shows curves for $s = s_O$ (green), $s = s_1$ (brown), $s = s_2$ (red), and $s = s_3 = s^*$ (blue). The shaded gray area represents interim payoffs satisfying the INTIR condition.

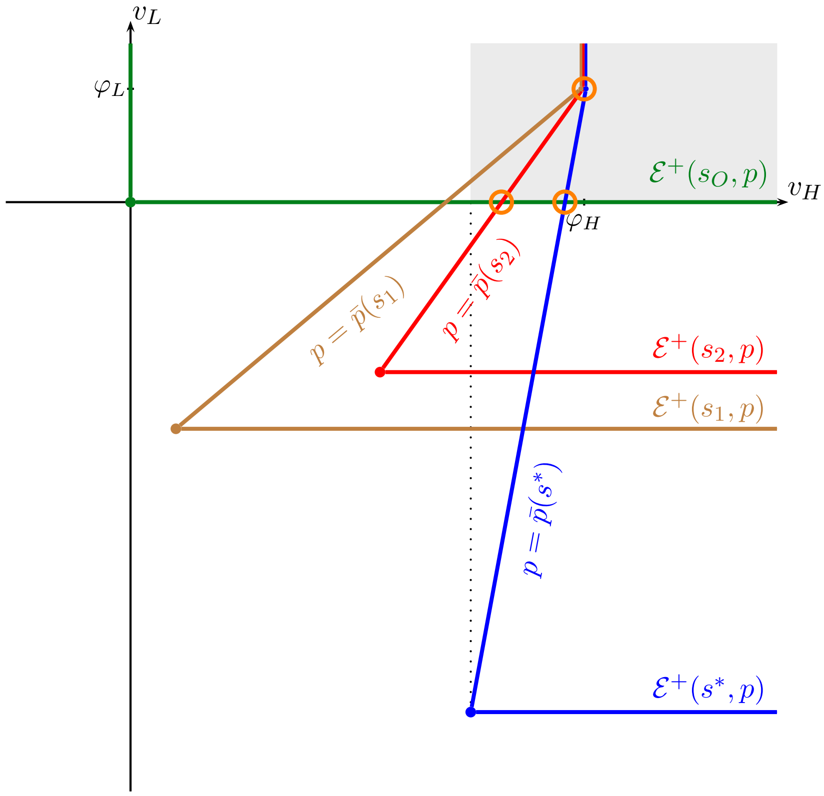

Figure 3: Projections of the graphs of $\mathcal{E}^+(s, \cdot)$ on the space of the sender's interim payoffs in a version of the intimidation game with $R(s^*) < 0$. The figure shows curves for $s = s_O$ (green), $s = s_1$ (brown), $s = s_2$ (red), and $s = s_3 = s^*$ (blue). The shaded gray area represents interim payoffs satisfying the INTIR condition.

Interim equilibrium payoffs: flat convexification. Now, our characterization theorem can be applied, using the graphs of modified interim value correspondences and the interim individually rational payoffs described above. With that, we obtain directly the set of all equilibrium outcomes of the intimidation game, for all possible priors p_0.

Consider for example the configuration depicted in Figure 2. The *non-revealing payoffs* $v \in \mathcal{E}(s, p_0)$ correspond to the points of the graph where the posterior is equal to the prior, so all the points on the projected graphs with $p = p_0$. Taking into account the individually rational conditions (the INTIR gray area), the unique non-revealing equilibrium payoff, represented by NRE in the figure, is $v = (\varphi_H, \varphi_L)$ for $p_0 > \min_{s \in S_I} \bar{p}(s)$: it corresponds to a pooling strategy on some signal s such that $p_0 \geq \bar{p}(s)$. There is no non-revealing payoff in $\mathcal{E}(s, p_0)$ which is also in INTIR for $p_0 < \min_{s \in S_I} \bar{p}(s)$. The partially revealing payoffs are those belonging to both $\mathcal{E}^+(s, p)$ and $\mathcal{E}^+(s', p')$ for some signals s and s' and posteriors p and p' with $p < p_0 < p'$. Those correspond to the points of intersections of the projected graphs. Taking into account the individually rational conditions, the unique partially revealing equilibrium payoff, represented by PRE in the figure, is $v = (\omega(s^*), 0)$ for $p_0 < \bar{p}(s^*)$. This equilibrium payoff corresponds to a partially revealing strategy using signals $s = s_O$ and $s = s^*$ and the incentive-compatible splitting on posterior $p = 0$ (with signal $s = s_O$) and $p = \bar{p}(s^*)$ (with signal $s = s^*$). For the specific prior $p = \bar{p}(s^*)$ there is also a continuum of non-revealing equilibrium payoffs in the convex hull of (φ_H, φ_L) and $(\omega(s^*), 0)$. Summing up, there are multiple equilibrium payoffs for $\min_{s \in S_I} \bar{p}(s) \leq p_0 \leq \bar{p}(s^*)$, the unique equilibrium payoff is the partially revealing equilibrium payoff $(\omega(s^*), 0)$ for $p_0 < \min_{s \in S_I} \bar{p}(s)$, and the unique equilibrium payoff is the non-revealing equilibrium payoff (φ_H, φ_L) for $p_0 > \bar{p}(s^*)$.

Consider now Figure 3 where $R(s^*) < 0$. As before, $v = (\varphi_H, \varphi_L)$ is a non-revealing equilibrium payoff for every $p_0 \geq \min_{s \in S_I} \bar{p}(s)$, $v = (\omega(s^*), 0)$ is a partially revealing equilibrium payoff for $p_0 < \bar{p}(s^*)$, and every payoff in the convex hull of (φ_H, φ_L) and $(\omega(s^*), 0)$ is a non-revealing equilibrium payoff for $p_0 = \bar{p}(s^*)$. But now, other non-revealing and partially revealing equilibria exist for some priors. There is a partially revealing equilibrium payoff $v = (\omega(s_2), 0)$ for every $p_0 < \bar{p}(s_2)$ (represented by the point of intersection of the red and green curves): it corresponds to a partially revealing strategy using signals $s = s_O$ and $s = s_2$ and the incentive-compatible splitting on posterior $p = 0$ (with signal $s = s_O$) and $p = \bar{p}(s_2)$ (with signal $s = s_2$). Finally, for $p_0 = \bar{p}(s_2)$, every point in the convex hull of (φ_H, φ_L) and $(\omega(s_2), 0)$ is a non-revealing equilibrium payoff.

More generally, our theorem directly implies the following proposition.

Proposition 1. *In the intimidation game, the set of all interim equilibrium payoffs of the sender is characterized as follows:*

(i) If $R(s^*) > 0$, then:

- *(a) If $p_0 > \min_{s \in S_I} \bar{p}(s)$, there is a non-revealing equilibrium with payoff (φ_H, φ_L);*
- *(b) If $p_0 < \bar{p}(s^*)$, there is a partially revealing equilibrium, with posterior beliefs 0 and $\bar{p}(s^*)$, and payoff $(\omega(s^*), 0)$;*
- *(c) If $p_0 = \bar{p}(s^*)$, there is a continuum of non-revealing equilibria, with payoffs in $\text{co}\,\{(\varphi_H, \varphi_L), (\omega(s^*), 0)\}$.*

(ii) If $R(s^*) \leq 0$, then:

(a) If $p_0 > \min_{s \in S_I} \bar{p}(s)$, there is a non-revealing equilibrium with payoff (φ_H, φ_L);

(b) If $p_0 < \bar{p}(\tilde{s})$ and $w(\tilde{s}) \geq \max_{s \in S_I} \gamma_H(s)$ for some $\tilde{s} \in S_I$, there is a partially revealing equilibrium, with posterior beliefs 0 and $\bar{p}(\tilde{s})$, and payoff $(w(\tilde{s}), 0)$;

(c) If $p_0 = \bar{p}(\tilde{s})$ and $w(\tilde{s}) \geq \max_{s \in S_I} \gamma_H(s)$ for some $\tilde{s} \in S_I$, there is a continuum of non-revealing equilibria, with payoffs in $\text{co}\,\{(\varphi_H, \varphi_L), (w(\tilde{s}), 0)\}$.

Notice that $R(s^*) \leq 0$ implies $\omega(s^*) \geq \max_{s \in S_I} \gamma_H(s)$. Hence, equilibrium conditions (b) and (c), identified in case (i) when $R(s^*) > 0$, also apply when $R(s^*) \leq 0$, with $\tilde{s} = s^*$. The difference is that if $R(s^*) \leq 0$, additional equilibria may exist involving signals $\tilde{s} \neq s^*$, whenever $w(\tilde{s}) \geq \max_{s \in S_I} \gamma_H(s)$.

When $R(s^*) > 0$, the best equilibrium for the high type sender is the partially revealing one —when it exists— with payoff $(\omega(s^*), 0)$, while the best equilibrium for the low type sender is the non-revealing one, with payoff (φ_H, φ_L). By contrast, when $R(s^*) < 0$, both sender types strictly prefer the non-revealing equilibrium, when it exists, since $(\varphi_H, \varphi_L) > (\omega(s), 0)$ for every $s \in S_I$.

To summarize, our equilibrium characterization shows that generically, there are two types of equilibria in the intimidation game. First, there can be a non-revealing equilibrium in which both types of the sender opt in with the same signal $s \in S_I$, provided the prior that the sender is strong is high enough, specifically, when $p_0 > \bar{p}(s)$. In this case, the receiver forgoes, and the interim payoff for the sender is (φ_H, φ_L). Second, there can be partially revealing equilibria in which the high type opts in with a signal $\tilde{s} \in S_I$, and the low type mixes between dropping out and using the same signal $\tilde{s}$. When the receiver observes signal $\tilde{s}$, which may come f rom either type, she responds with a mixture of actions. If the signaling pressure ratio $R(s)$ is positive for at least one signal s, then the partially revealing equilibrium is unique and must involve the maximal signaling pressure point $s^* = \arg\max_{s \in S_I} R(s)$. When it exists, the corresponding equilibrium with signal s^* is the most favorable outcome for the high type, and it is the unique equilibrium when the prior is low enough, that is, when $p_0 < \min_{s \in S_I} \bar{p}(s)$. If instead $R(s)$ is negative for every signal $s \in S_I$, then multiple partially revealing equilibria may exist, each involving a different signal $\tilde{s}$. These equilibria again feature the low type receiving a payoff of 0, and the high type receiving $w(\tilde{s})$, provided that $w(\tilde{s}) \geq \max_{s \in S_I} \gamma_H(s)$ and $p_0 < \bar{p}(\tilde{s})$. However, in this case, whenever the non-revealing equilibrium exists, it is preferred by both types of the sender, as it gives strictly higher payoffs than any of the partially revealing equilibria.

5 Transparent motives

In this section, we consider a signaling game with *transparent motives* wherein the sender's utility is type independent, and we simply denote it by $u : A \times S \to \mathbb{R}$. The interim equilibrium payoff of the sender is then the same regardless of her type and we denote it $v \in \mathbb{R}$. With some abuse of notation, the set G can be written as:

$$G = \{(v, p) \in \mathbb{R} \times \Delta(T) : \exists s \in S, y \in Y(s, p) \text{ s.t. } v = u(y, s)\}.$$

Then, our conditions reformulate as follows:

a) $(v, p_0) \in \mathrm{co_f}\,(G)$ if and only if for some $K \in \{1, \ldots, |T|\}$, there exist $(\lambda_k)_{k=1}^K \in \Delta(\{1, \ldots, K\})$, $(p_k)_{k=1}^K$ and $(s_k)_{k=1}^K$, with $p_k \in \Delta(T)$ and $s_k \in S$, such that $p_0 = \sum_k \lambda_k p_k$, $y_k \in Y(s_k, p_k)$ and $v = u(y_k, s_k)$ for every k;

b) $v \in$ INTIR if and only if $v \geq \max_{s \in S} \min_{p \in \Delta(T)} \min_{y \in Y(s,p)} u(y, s)$.

Hence, from Theorem 1, v is a PBE payoff of the signaling game if and only if:

$$v \geq \max_{s \in S} \min_{p \in \Delta(T)} \min_{y \in Y(s,p)} u(y, s),$$

and there exists a family $(\lambda_k, p_k, s_k, y_k)_{k=1}^K$ such that $p_0 = \sum_k \lambda_k p_k$, and for all k, $y_k \in Y(s_k, p_k)$ and $v = u(y_k, s_k)$.

5.1 Maximal equilibrium payoff

For every $s \in S$ and $p \in \Delta(T)$, the best interim value for the sender at p given signal s is denoted by:

$$w(s, p) := \max_{y \in Y(s,p)} u(y, s).$$

For every $p \in \Delta(T)$, let $w(p) := \max_{s \in S} w(s, p)$ be the best interim value at p for the sender. For every function $F : \Delta(T) \to \mathbb{R}$, we denote by $\mathrm{qcav}\, F$ the smallest (pointwise) quasi-concave function above F (the quasi-concave envelope of F).

Theorem 2. *With transparent motives, the maximal PBE payoff of the sender in the signaling game is* $\mathrm{qcav}\, w(p_0)$.

This theorem extends the results of Lipnowski and Ravid (2020) to our costly signaling environment. If we assume that signals are payoff-irrelevant, we get a cheap talk game with transparent motives as in Lipnowski and Ravid (2020) and the characterization of our Theorem 2 coincides with their Theorem 2. To get an intuition for the result, first remark that the sender can obtain the payoff $w(p_0)$ with a pooling strategy that chooses a signal s which maximizes $w(s, p_0)$ over S, for all types. In the cheap talk setting of Lipnowski and Ravid (2020), this is a babbling equilibrium. Consider now a splitting of the prior belief (a Bayes plausible distribution of posteriors with finite support) which benefits the sender in the sense that it is possible to improve the payoff over $w(p_0)$ for all posteriors. Then, the minimal optimal payoff $w(p)$ over the induced posteriors is a PBE payoff. This property is called securability in Lipnowski and Ravid (2020): their Theorem 1 shows that equilibrium payoffs satisfy an intermediate value property. We extend this logic to our signaling game. Importantly, the INTIR condition is key for showing that the set of equilibrium payoffs of the sender in an interval, even with several different signals. The proof of Theorem 2 can be found in Appendix B.

5.2 Signaling without cheap talk

A signaling game is called *without cheap talk* if $|M| = 1$.

Theorem 3. *Under transparent motives, if for each s, $w(s, \cdot)$ is quasi-concave, then every PBE payoff $v \geq w(p_0)$ of the signaling game is also a PBE payoff of the signaling game without cheap talk.*

The proof of Theorem 3 can be found in Appendix C. The following example shows that when there is a signal s such that $w(s, \cdot)$ is not quasi-concave, then cheap talk might improve payoffs.

The sender has two possible signals s_1 and s_2. The payoffs of the sender are described by Figure 4, where $w(s_1, \cdot)$ is quasi-concave, whereas $w(s_2, \cdot)$ is not. It is easy to see that, for the prior p_0, the maximal PBE payoff for the sender requires the use of cheap talk messages together with the choice of signal s_2. There is a range of priors for which the optimal strategy of the sender uses cheap-talk messages only, while for some other priors, the optimal strategy requires mixing with signals s_1 and s_2 (see the orange areas in the figure).

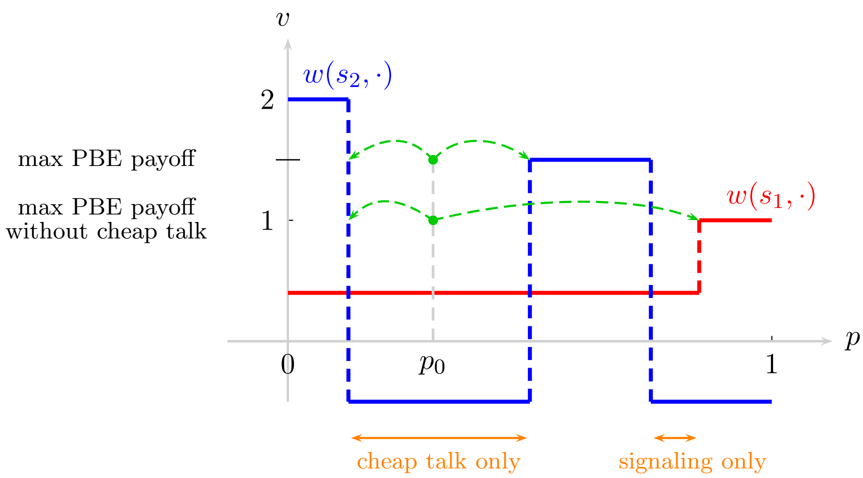

Figure 4: An example where cheap-talk might improve payoffs.

6 The value of commitment

We consider now a version of the signaling game where the sender has commitment power and chooses the messaging strategy σ at an ex-ante stage, before observing the type; this strategy is observed by the receiver. In this scenario, the sender is not able to change signals and messages after learning her type and therefore the interim incentive compatibility conditions for the sender are not relevant, neither are the interim individual rationality conditions. Thus, to study the interim payoffs that can be obtained under commitment, only the best response of the receiver matters.

An interim payoff $v = (v_t)_t$ is feasible in the signaling game with commitment if and only if there exist strategies σ and τ such that for every $t \in T$,

$$v_t = \int_{S\times M} u(\tau(s,m), s, t) d\sigma(s, m \mid t),$$

and for every $(s,m) \in S \times M$, $\tau(s,m) \in Y(s, \mu(s,m))$, and $\mu(s,m) \in \Delta(T)$ is obtained from p_0 and σ by Bayes' rule. Define the set of *adjusted* interim values for the sender at p given s as:

$$\hat{\mathcal{E}}(s,p) := \Big\{\hat{v} \in \mathbb{R}^T : \exists\, y \in Y(s,p) \text{ s.t. } \hat{v}_t = \frac{p(t)}{p_0(t)} u(y,s,t), \forall t\Big\}.$$

This set is obtained from the set of interim values $\mathcal{E}(s,p)$ by normalizing an interim payoff v_t with the likelihood-ratio $\frac{p(t)}{p_0(t)}$. The graph of the adjusted interim value correspondence of the sender given signal s is denoted by $\hat{G}_s$ and we denote by $\hat{G} = \bigcup_{s\in S} \hat{G}_s$ the union of those graphs. That is,

$$\hat{G} = \Big\{(\hat{v}, p) \in \mathbb{R}^T \times \Delta(T) : \exists s \in S, \exists y \in Y(s,p) \text{ s.t. } \hat{v}_t = \frac{p(t)}{p_0(t)} u(y,s,t), \forall t\Big\}.$$

We let $\mathrm{co}\,(\hat{G})$ be the convex hull of $\hat{G}$. Since p lies in the $(|T|-1)-$dimensional simplex and $\hat{v}$ lies in $\mathbb{R}^T$, thanks to Caratheodory's theorem it is enough to consider convex combination of pairs $(\hat{v}, p)$ of cardinality at most $2|T|$.

Theorem 4. *Assume that $|M| \geq 2|T|$. The interim payoff $v \in \mathbb{R}^T$ is a feasible interim payoff of the sender in the signaling game with commitment if and only if $(v, p_0) \in \mathrm{co}\,(\hat{G})$.*

Comparing with the result without commitment, the set of PBE interim payoffs characterized in Theorem 1 is clearly a subset of the set of feasible interim payoffs with commitment obtained in Theorem 4. In the characterization of PBE payoffs, interim values need not be normalized by the likelihood-ratios $\frac{p(t)}{p_0(t)}$ because the incentive-compatibility conditions of the sender guarantee that the (modified) interim values are the same for all posteriors. Indeed, in both characterizations, the interim payoff of type t is given by $v_t = \sum_k \lambda_k \frac{p_k(t)}{p_0(t)} v_t^k$, but in the characterization of Theorem 1, the PBE interim payoffs $(v_t^k)_t$ belong to $\mathcal{E}^+(s,p)$ and satisfy $v_t^k = v_t^{k'}$ for every k, k' and t. It follows that $v = v^k$ for every k, so $v = \sum_k \lambda_k v^k$ is a trivial convex combinaison of the modified interim values, *without* the likelihood-ratios normalization. In the setting of Theorem 4, the rationality of the receiver imposes that $(v_t^k)_t$ is in $\mathcal{E}(s,p)$ but there are no incentive-compatibility conditions for the sender so v^k is not necessarily constant in k. Yet, $v = \sum_k \lambda_k \hat{v}^k$ is a convex combinaison of the adjusted interim values $\hat{v}^k \in \hat{\mathcal{E}}(s,p)$.

Theorem 4 is a direct extension of Theorem 1 in Doval and Smolin (2024), where there is no payoff-relevant signal and the interim value correspondence $\mathcal{E}$ is single valued. This is obtained in our model by taking $|S| = 1$ and assuming that the receiver has a unique optimal action for each belief. The interim value correspondence can then be represented by an interim value function $(\mathsf{w}_t)_t$, called a welfare function in Doval and Smolin (2024), where $\mathsf{w}_t : \Delta(T) \to \mathbb{R}$ and $\mathcal{E}(s,p) = \{(\mathsf{w}_t(p))_t\}$. Defining $\hat{\mathsf{w}}_t(p) = \frac{p(t)}{p_0(t)} w_t(p)$, an interim payoff $v \in \mathbb{R}^T$ is a feasible interim payoff if and only if $(v, p_0) \in \mathrm{co}\,(\mathrm{gr}((\hat{\mathsf{w}}_t)_t))$ as in Theorem 1 in Doval and Smolin (2024).

Theorem 4 is also related to Boleslavsky and Shadmehr (2023) who consider a costly signaling setting with commitment and focus on the sender's ex-ante optimal solution. Without cheap talk messages, an interim payoff v is feasible if and only if (v, p_0) belongs to the *join* of the family of sets $\{\hat{G}_s : s \in S\}$, a subset of the convex hull which is given by convex combinations $(v, p_0) = \sum_s \lambda_s(\hat{v}_s, p_s)$ where each point $(\hat{v}_s, p_s)$ belongs to $\hat{G}_s$. As remarked in Boleslavsky and Shadmehr (2023), cheap talk messages are necessary to obtain the full convex hull. Relatedly, we can deduce the following from Theorem 4.

Corollary 5. *The maximum ex-ante expected utility of the sender is given by the concave closure* $\operatorname{cav} \max_{s \in S} w(s, p)$*, where* $w(s, p) = \max_{y \in Y(s,p)} \sum_t p(t) u(y, s, t)$ *for every* s*, and* $\operatorname{cav} F(\cdot)$ *denotes the concave closure of* $F(\cdot)$ *(the pointwise smallest concave function above* $F(\cdot)$*).*

When signals are payoff-irrelevant, or when $|S| = 1$, we obtain the characterization of Kamenica and Gentzkow (2011). Corollary 5 can be proved directly from their concave closure logic, see Proposition 4 in Boleslavsky and Shadmehr (2023).

Application: intimidation and signaling pressure with commitment. Consider again the intimidation game and assume now that the sender can commit to an arbitrary messaging strategy σ, which is observed by the receiver, who then best responds to it. The interim value function is:

$$w(s_O, p) = 0,$$

and for every $s \in S_I$,

$$w(s,p) = \begin{cases} \gamma_L(s) + p(\gamma_H(s) - \gamma_L(s)) & \text{if } p < \bar{p}(s), \\ \varphi_L + p(\varphi_H - \varphi_L) & \text{if } p > \bar{p}(s), \\ \max\left\{\gamma_L(s) + p(\gamma_H(s) - \gamma_L(s)),\, \varphi_L + p(\varphi_H - \varphi_L)\right\} & \text{if } p = \bar{p}(s). \end{cases}$$

For each $s \in S_I$, let $\psi(s)$ denote the belief about t_H that renders the sender ex-ante indifferent between the two actions of the receiver, to challenge or to forgo:

$$\psi(s) := \frac{\varphi_L - \gamma_L(s)}{\gamma_H(s) - \varphi_H + \varphi_L - \gamma_L(s)}.$$

We see that $\psi(s) = \frac{1}{1+R(s)}$, so $\psi(s)$ is minimized at the maximal signaling pressure point s^*.

If for every $s \in S_I$, we have $\bar{p}(s) = \psi(s)$, then the receiver and the (uninformed) sender are indifferent at the same belief. The game is thus strictly competitive: at each belief, the sender and receiver have opposite preferences over actions. It follows that $w(s, p)$ is continuous in p for every s. In that case, the concave closure $\operatorname{cav} \max_{s \in S} w(s, p)$ is achieved by the same optimal splitting as in the no-commitment case: using posterior 0 with signal s_O and posterior $\psi(s^*)$ with signal s^* when $p_0 < \psi(s^*)$, and pooling with any s such that $\psi(s) \geq \psi(s^*)$ when $p_0 > \psi(s^*)$. However, in the general case where $\bar{p}(s) \neq \psi(s)$, the commitment solution differs from the equilibrium outcomes without commitment, and the sender can benefit ex-ante from the ability to commit.

To illustrate the value of commitment, consider the simple case with a single opt-in signal, $S_I = \{s_I\}$, and let $\psi = \psi(s_I)$, $\bar{p} = \bar{p}(s_I)$, $\gamma_H = \gamma_H(s_I)$, and $\gamma_L = \gamma_L(s_I)$. Assume further that $\varphi_H = \varphi_L = \varphi$ and $\varphi < \gamma_H$, so that $\psi = \frac{\varphi - \gamma_L}{\gamma_H - \gamma_L} \in (0, 1)$. Denote by $\mathbb{E}v_{PBE}$ the sender's ex-ante expected payoff without commitment. From the analysis in Section 4, we have:

(i) If $p_0 \geq \bar{p}$, the unique equilibrium is non-revealing, and both sender types receive payoff φ. Hence, $\mathbb{E}v_{PBE} = \varphi$.

(ii) If $p_0 < \bar{p}$, the unique equilibrium is partially revealing: the high type always opts in ($s = s_I$), while the low type mixes between opting in ($s = s_I$) and dropping out ($s = s_O$), so that the receiver's posterior after $s = s_I$ is $\bar{p}$. The interim payoff vector is $(\omega(s_I), 0)$, where $\omega(s_I) = \frac{\gamma_H - \gamma_L}{\varphi - \gamma_L}\varphi$, and the ex-ante expected payoff is $\mathbb{E}v_{PBE} = p_0\omega(s_I)$.

First, consider the case in which $\bar{p} < \psi$, i.e., the receiver is relatively cautious—he challenges for a narrower range of beliefs than the sender would ex-ante prefer. The optimal commitment solution is illustrated in Figure 5. For $p_0 < \bar{p}$, the sender makes the receiver forgo with probability one when opting in. Compared to the no-commitment equilibrium, where the receiver challenges an opt-in signal with strictly positive probability, this commitment solution strictly benefits the ex-ante welfare of the sender ($\mathbb{E}v_{PBE} < \text{cav}\max_s w(s, p_0)$ for every $p_0 < \bar{p}$); it strictly benefits the low type, but hurts the high type. The receiver's welfare is not affected by the sender's commitment ability, as the induced splitting of beliefs and actions are the same in both scenarios. For $p_0 \geq \bar{p}$, the outcome is the same as under no commitment: both types of the sender pool on the opt-in signal and the receiver forgoes ($\mathbb{E}v_{PBE} = \text{cav}\max_s w(s, p_0)$ for every $p_0 \geq \bar{p}$).

Second, consider the case in which $\bar{p} > \psi$, i.e., the receiver is relatively aggressive—he challenges for a broader range of beliefs than the sender would ex-ante prefer. The optimal commitment solution is illustrated in Figure 6. For $p_0 < \bar{p}$, the solution is as before, except that the sender makes the receiver challenge with probability one when opting in. Compared to the no-commitment equilibrium, where the receiver forgoes with strictly positive probability, this commitment solution again strictly increases the ex-ante welfare of the sender, but now it benefits the high type and hurts the low type. For $p_0 > \bar{p}$, the solution differs substantially from the no-commitment case. Both types opt in with probability one, but the sender uses additional cheap talk messages to reveal information, as in Bayesian persuasion: the high type is revealed with certainty with positive probability (in which case the receiver forgoes), and the low type is partially pooled with the high type in such a way that the receiver always challenges the low type and sometimes challenges the high type. Such belief splittings—onto the posteriors $\bar{p}$ and 1—are not incentive-compatible without commitment, as the low type would mimic the high type to induce the receiver to forgo. Hence, with a relatively aggressive receiver, the sender strictly benefits ex-ante from commitment for all priors ($\mathbb{E}v_{PBE} < \text{cav}\max_s w(s, p_0)$ for every p_0). Commitment strictly benefits the high type but hurts the low type.

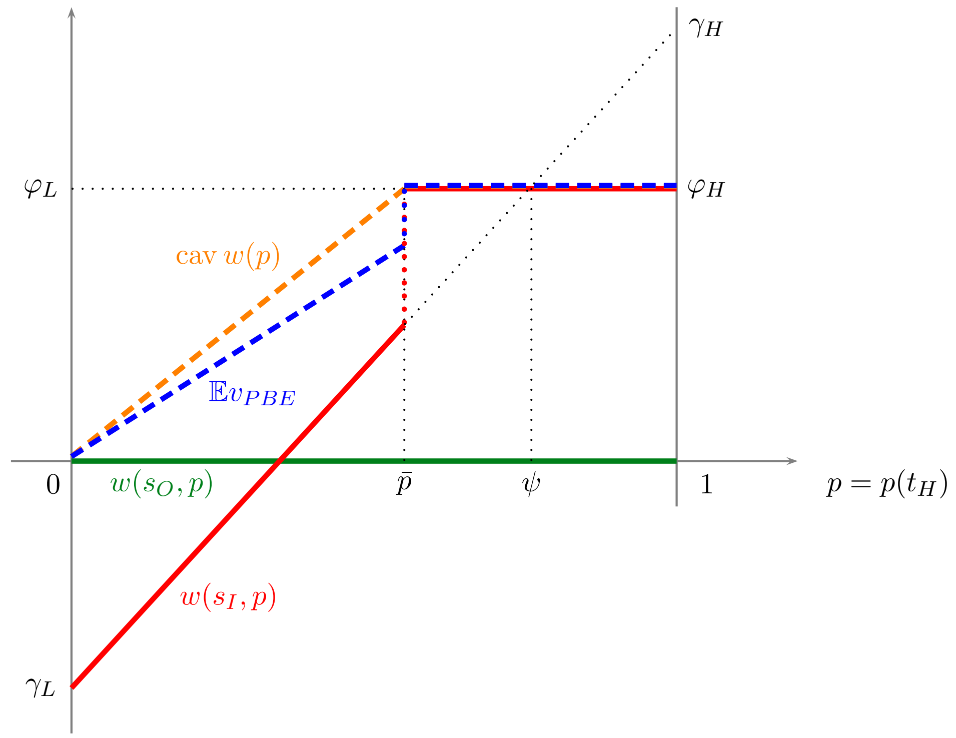

Figure 5: Intimidation game with commitment and a cautious receiver ($\psi > \bar{p}$).

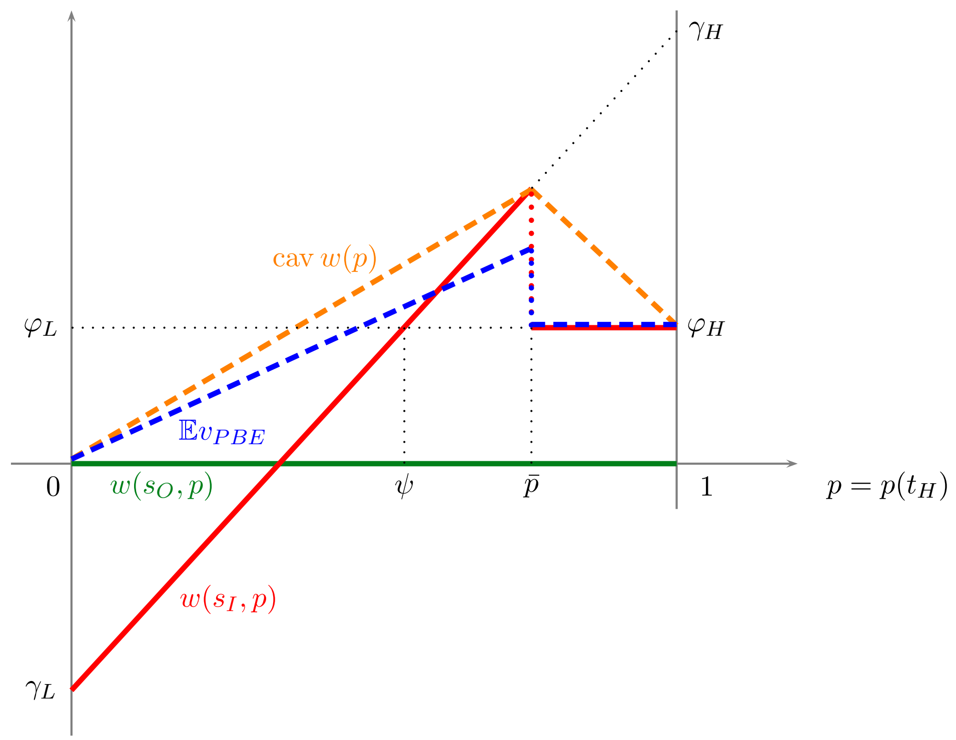

Figure 6: Intimidation game with commitment and an aggressive receiver ($\psi < \bar{p}$).

References

Aumann, R. J. and S. Hart (2003): "Long Cheap Talk," *Econometrica*, 71, 1619–1660.

Aumann, R. J. and M. Maschler (1966): "Game Theoretic Aspects of Gradual Disarmament," Report of the U.S. Arms Control and Disarmament Agency, ST-80, Chapter V, pp. 1–55.

Aumann, R. J. and M. B. Maschler (1995): *Repeated Games of Incomplete Information*, Cambridge, Massachusetts: MIT Press.

Austen-Smith, D. and J. S. Banks (2002): "Costly Signaling and Cheap Talk in Models of Political Influence," *European Journal of Political Economy*, 18, 263–280.

Boleslavsky, R. and M. Shadmehr (2023): "Signaling With Commitment," *working paper.*

Chakraborty, A. and R. Harbaugh (2010): "Persuasion by Cheap Talk," *American Economic Review*, 100, 2361–2382.

Corrao, R. and Y. Dai (2023): "The Bounds of Mediated Communication," *working paper.*

Doval, L. and A. Smolin (2024): "Persuasion and Welfare," *Journal of Political Economy.*

Forges, F. (1984): "Note on Nash equilibria in repeated games with incomplete information," *International Journal of Game Theory*, 13, 179–187.

——— (1990): "Equilibria with Communication in a Job Market Example," *Quarterly Journal of Economics*, 105, 375–398.

——— (1994): "Non-Zero Sum Repeated Games and Information Transmission," in *Essays in Game Theory: In Honor of Michael Maschler*, ed. by N. Megiddo, Springer-Verlag.

——— (2020): "Games with incomplete information: from repetition to cheap talk and persuasion," *Annals of Economics and Statistics*, 3–30.

Forges, F. and F. Koessler (2008): "Long Persuasion Games," *Journal of Economic Theory*, 143, 1–35.

Grafen, A. (1990): "Biological Signals as Handicaps," *Journal of Theoretical Biology*, 144, 517–546.

Hart, S. (1985): "Nonzero-Sum Two-Person Repeated Games with Incomplete Information," *Mathematics of Operations Research*, 10, 117–153.

Heumann, T. (2020): "On the cardinality of the message space in sender–receiver games," *Journal of Mathematical Economics*, 90, 109–118.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *The American Economic Review*, 101, 2590–2615.

Koessler, F., M. Laclau, and T. Tomala (2024): "A belief-based approach to signaling," *working paper halshs-04455227*.

Kreps, D. M. and J. Sobel (1994): "Signalling," in *Handbook of Game Theory*, ed. by R. J. Aumann and S. Hart, Elsevier Science B. V., vol. 2, chap. 25, 849–867.

Kreps, D. M. and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

Lipnowski, E. and D. Ravid (2020): "Cheap talk with transparent motives," *Econometrica*, 88, 1631–1660.

Lizzeri, A., Y. Lou, and J. Perego (2024): "Facts and Opinions: Communicating with Hard and Soft Information," *working paper*.

Manelli, A. M. (1996): "Cheap talk and sequential equilibria in signaling games," *Econometrica*, 917–942.

Pęski, M. (2014): "Repeated games with incomplete information and discounting," *Theoretical Economics*, 9, 651–694.

Ponssard, J. P. and S. Zamir (1973): "Zero-sum sequential games with incomplete information," *International Journal of Game Theory*, 2, 99–107.

Reny, P. J. (2024): "Natural Language Equilibrium," *working paper*.

Salamanca, A. (2021): "The value of mediated communication," *Journal of Economic Theory*, 192, 105191.

Sobel, J. (2020): "Signaling games," *Complex Social and Behavioral Systems: Game Theory and Agent-Based Models*, 251–268.

Spence, A. M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

Wu, W. (2022): "A Role for Cheap Talk in Disclosure," *ShanghaiTech SEM Working Paper*.

Zahavi, A. (1975): "Mate selection – A selection for a handicap," *Journal of Theoretical Biology*, 53.

A Proof of Theorem 1

(i) From constrained convexification to equilibrium. Let $v \in \text{INTIR}$ be such that $(v, p_0) \in \text{co}_{\text{f}}(G)$, we construct a PBE assessment (σ, τ, μ) with interim payoff v, where σ has finite support. Since $(v, p_0) \in \text{co}_{\text{f}}(G)$, there exists $K \in \{1, \dots, |T|\}$, $(\lambda_k)_{k=1}^K \in \Delta(\{1, \dots, K\})$, $\lambda_k > 0$, $s_k \in S$ and $y_k \in Y(s_k, p_k)$ for every k such that:

$$p_0 = \sum_{k=1}^{K} \lambda_k p_k, \tag{4}$$

and for every $k = 1, \dots K$:

$$v_t \geq u(y_k, s_k, t) \text{ and } v_t = u(y_k, s_k, t) \text{ if } p_k(t) > 0. \tag{5}$$

Consider a profile of messages $(m_k)_k$ with $m_k \in M$ and $m_k \neq m_{k'}$ for every $k \neq k'$, this is possible as $|M| \geq |T| \geq K$. Since $v \in \text{INTIR}$, for every $s \in S$, there exists $q_s \in \Delta(T)$ and

$$\bar{y}_s \in Y(q_s, s), \tag{6}$$

such that:

$$v_t \geq u(\bar{y}_s, s, t), \text{ for every } t \in T. \tag{7}$$

Define the following strategies σ for the sender and τ for the receiver:

$$\sigma(s_k, m_k \mid t) = \frac{p_k(t)\lambda_k}{p_0(t)}, \text{ for every } k, \tag{8}$$

$$\sigma(s, m \mid t) = 0, \text{ if } (s, m) \neq (s_k, m_k) \text{ for every } k, \tag{9}$$

$$\tau(s, m) = \begin{cases} y_k & \text{if } (s, m) = (s_k, m_k) \text{ for some } k, \\ \bar{y}_s & \text{otherwise.} \end{cases} \tag{10}$$

Consider the following belief system:

$$\mu(t \mid s, m) = \begin{cases} p_k & \text{if } (s, m) = (s_k, m_k) \text{ for some } k, \\ q_s & \text{otherwise.} \end{cases} \tag{11}$$

For the sender of type t, the interim payoff induced by (σ, τ) is indeed v_t because:

$$\begin{aligned} \int_{S \times M} u(\tau(s, m), s, t) d\sigma(s, m \mid t) &= \sum_{k=1}^{K} \frac{p_k(t)\lambda_k}{p_0(t)} u(y_k, s_k, t), \text{ by (8) and (10)}, \\ &= v_t, \text{ by (5) and (4)}. \end{aligned}$$

Then, sequential rationality for the sender (PBE condition (i)) follows directly from (5), (7), and (10), sequential rationality for the receiver (PBE condition (ii)) follows directly from (6), (10), and (11).

Finally, to verify belief consistency (PBE condition (iii)), it remains to show that the probability of t conditional on (s_k, m_k), given by (11), is equal to $p_k(t)$ for every k and t. From Bayes' rule and (8), this conditional probability is:

$$\frac{\sigma(s_k, m_k \mid t)p_0(t)}{\sum_{\tilde{t}\in T}\sigma(s_k, m_k \mid \tilde{t})p_0(\tilde{t})} = \frac{\frac{p_k(t)\lambda_k}{p_0(t)}p_0(t)}{\sum_{\tilde{t}\in T}\frac{p_k(\tilde{t})\lambda_k}{p_0(\tilde{t})}p_0(\tilde{t})} = \frac{p_k(t)\lambda_k}{\sum_{\tilde{t}\in T}p_k(\tilde{t})\lambda_k} = p_k(t).$$

This completes the "if" part of the theorem.

(ii) From equilibrium to constrained convexification. Consider a PBE assessment (σ, τ, μ) with interim payoff v. First, we show that $v \in \text{INTIR}$. For every $s \in S$, consider an arbitrary message $m \in M$ and let $y = \tau(s, m)$. Since τ is sequentially rational for the receiver, we have $y \in Y(s, p)$ for $p = \mu(s, m) \in \Delta(T)$. Since σ is sequentially rational for the sender, we have $v_t \geq u(y, s, t)$ for every t. We conclude that for every signal s, we have $v \in \text{INTIR}_s$, therefore $v \in \text{INTIR}$.

Second, we show that $(v, p_0) \in \text{co}_{\text{f}}(G)$. Let $\lambda(\cdot) := \sum_{t\in T} p_0(t)\sigma(\cdot|t) \in \Delta(S \times M)$ denote the marginal distribution on $S \times M$ induced by p_0 and σ. To prove that $(v, p_0) \in \text{co}_{\text{f}}(G)$, it suffices to show that there exists $\tilde{X} \subseteq S \times M$ with $\lambda(\tilde{X}) = 1$ such that (a) $v \in \mathcal{E}^+(s, \mu(s, m))$ for every $(s, m) \in \tilde{X}$, and (b) $p_0 \in \text{co}(\{\mu(s, m) : (s, m) \in \tilde{X}\})$.

The PBE condition for the sender:

$$v_t \geq u(\tau(s, m), s, t), \ \forall (s, m) \in S \times M, \tag{12}$$

implies:

$$\sum_{t\in T}(v_t - u(\tau(s, m), s, t))\,\mu(t \mid s, m) \geq 0, \forall (s, m) \in S \times M. \tag{13}$$

We also have:

$$\begin{aligned}
&\mathbb{E}_\lambda\left[\sum_t (v_t - u(\tau(s, m), s, t))\,\mu(t \mid s, m)\right] \\
&= \int_{S\times M}\sum_t (v_t - u(\tau(s, m), s, t))\,\mu(t \mid s, m)d\lambda(s, m) \\
&= \sum_t \int_{S\times M}(v_t - u(\tau(s, m), s, t))\,\mu(t \mid s, m)d\lambda(s, m) \\
&= \sum_t \int_{S\times M}(v_t - u(\tau(s, m), s, t))\,p_0(t)d\sigma(s, m \mid t) \\
&= \sum_t v_t p_0(t)\underbrace{\int_{S\times M} d\sigma(s, m \mid t)}_{=1} - \sum_t p_0(t)\underbrace{\int_{S\times M} u(\tau(s, m), s, t)d\sigma(s, m \mid t)}_{=v_t} = 0.
\end{aligned}$$

Hence, combined with (13), the previous equality implies that, λ-almost surely:

$$\sum_t (v_t - u(\tau(s, m), s, t))\,\mu(t \mid s, m) = 0.$$

Thus, from (12), there exists a Borel subset $\tilde{X} \subseteq S \times M$ with $\lambda(\tilde{X}) = 1$ such that:

$$v_t = u(\tau(s,m), s, t), \text{ for all } (s,m) \in \tilde{X} \text{ and } t \in T \text{ s.t. } \mu(t \mid s,m) > 0. \tag{14}$$

The PBE condition for the receiver implies:

$$\tau(s,m) \in Y(s, \mu(s,m)), \text{ for all } (s,m) \in \tilde{X}. \tag{15}$$

Putting (12), (14), and (15) together, we get $v \in \mathcal{E}^+(s, \mu(s,m))$ for every $(s,m) \in \tilde{X}$, which proves (a).

Now, from belief consistency with $B = \tilde{X}$ we have:

$$p_0 = \int_{\tilde{X}} \mu(s,m) d\lambda(s,m).$$

Hence, property (b) directly follows from the following lemma.

Lemma 1. *Consider $p_0 \in \Delta(T)$, $\sigma : T \to \Delta(S \times M)$, and denote by $\lambda(\cdot) := \sum_{t \in T} p_0(t)\sigma(\cdot|t) \in \Delta(S \times M)$ the marginal distribution on $S \times M$ induced by p_0 and σ. Let $\tilde{X}$ be a Borel subset of $S \times M$ such that $\lambda(\tilde{X}) = 1$, and let $\mu : S \times M \to \Delta(T)$ be a Borel measurable function satisfying:*

$$p_0(t) = \int_{\tilde{X}} \mu(t \mid x) d\lambda(x). \tag{16}$$

Then,

$$p_0 \in \text{co}\left(\{\mu(x) : x \in \tilde{X}\}\right).$$

Notice that the conclusion that p_0 belongs to the convex hull of the beliefs $\{\mu(x) : x \in \tilde{X}\}$, is stronger than requiring that it merely belongs to the *closure* of the convex hull.

Proof. Let $Z \subseteq \tilde{X}$ be a Borel subset of $\tilde{X}$ such that $\lambda(Z) = 1$ and such that the dimension $d(Z)$ of the convex hull of $\{\mu(x) : x \in Z\}$ is minimal among the $d(Z')$, for all Borel subsets $Z' \subseteq \tilde{X}$ with $\lambda(Z') = 1$. We have:

$$\forall t \in T, \; p_0(t) = \int_Z \mu(t|x) d\lambda(x).$$

Suppose that $p_0 \notin \text{co}\left(\{\mu(x) : x \in Z\}\right)$. From the separation theorem, there exists $\psi \in \mathbb{R}^T$ such that:

$$\forall x \in Z, \; \sum_t p_0(t)\psi_t \geq \sum_t \mu(t|x)\psi_t,$$

where the separating vector ψ is not 0 and belongs to the $d(Z)$-dimensional affine hull of $\{\mu(x) : x \in Z\}$. Integrating this inequality gives:

$$\sum_t p_0(t)\psi_t \geq \int_Z \sum_t \mu(t|x)\psi_t d\lambda(x) = \sum_t \int_Z \mu(t|x) d\lambda(x) \psi_t = \sum_t p_0(t)\psi_t.$$

It follows that $\sum_t \mu(t|x)\psi_t = \sum_t p_0(t)\psi_t$, λ-almost surely, that is:

$$\lambda\Big(\big\{x \in Z : \sum_t \mu(t|x)\psi_t = \sum_t p_0(t)\psi_t\big\}\Big) = 1.$$

Therefore, $Z' = \{x \in Z : \sum_t \mu(t|x)\psi_t = \sum_t p_0(t)\psi_t\}$ satisfies $\lambda(Z') = 1$ and $d(Z') < d(Z)$ which is a contradiction. We conclude that $p_0 \in \mathrm{co}\left(\{\mu(x) : x \in Z\}\right)$ and therefore $p_0 \in \mathrm{co}\left(\{\mu(x) : x \in \tilde{X}\}\right)$. ■
This completes the "only if" part of the theorem and ends the proof. □

B Proof of Theorem 2

To prove Theorem 2, we first establish the next proposition, which is an analogue of Theorem 1 in Lipnowski and Ravid (2020) for signaling games with transparent motives.

Recall that a splitting of $p_0 \in \Delta(T)$ is a finite family $(\lambda_k, p_k)_k$ with $\lambda_k \geq 0$, $p_k \in \Delta(T)$, $\sum_k \lambda_k = 1$, $\sum_k \lambda_k p_k = p_0$. One can find convex combination coefficients $(\lambda_k)_k$ such that $p_0 = \sum_k \lambda_k p_k$ if and only if $p_0 \in \mathrm{co}\{p_1, \ldots, p_K\}$.

Proposition 2. *Let $(p_k)_{k=1}^K$ be a set of posteriors with $p_k \in \Delta(T)$ for every $k = 1, \ldots, K$ such that $p_0 \in \mathrm{co}\{p_1, \ldots, p_K\}$ and $\min_k w(p_k) \geq w(p_0)$. Then, $\min_k w(p_k)$ is a PBE payoff of the signaling game with transparent motives.*

Proof of Proposition 2. Suppose that $p_0 \in \mathrm{co}\{p_1, \ldots, p_K\}$. First, we argue that, for every $j = 1, \ldots, K$ and every $\alpha \in [0, 1]$:

$$p_0 \in \mathrm{co}\{(1-\alpha)p_0 + \alpha p_j, p_k, k \neq j\}.$$

To see this, note that:

$$p_0 = \sum_k \lambda_k p_k = \lambda_j p_j + (1-\lambda_j)\sum_{k\neq j} \frac{\lambda_k}{1-\lambda_j} p_k := \lambda_j p_j + (1-\lambda_j)\bar{p}_j.$$

Consider now $q = (1-\alpha)p_0 + \alpha p_j$ and assume $0 < \alpha < 1$ (the cases $\alpha = 0$ or 1 are clear). It is easy to check that:

$$p_0 = \frac{\lambda_j}{\lambda_j + \alpha(1-\lambda_j)} q + \frac{\alpha(1-\lambda_j)}{\lambda_j + \alpha(1-\lambda_j)} \bar{p}_j.$$

Now, fix $(p_k)_k$ such that $p_0 \in \mathrm{co}\{p_1, \ldots, p_K\}$ and $w^* = \min_k w(p_k) \geq w(p_0)$. Suppose that for some j, $w(p_j) > w^*$.

Lemma 2. *There exists $\alpha_j \in [0, 1]$, $s_j \in S$ and $y_j \in Y(s_j, (1-\alpha_j)p_0 + \alpha_j p_j)$ such that*

$$u(s_j, (1-\alpha_j)p_0 + \alpha_j p_j) = w^*.$$

Proof of Lemma 2. For $q \in \Delta(T)$, denote $\underline{v}(q) = \max_{s\in S} \min_{y\in Y(s,q)} u(y, s)$ and

$$W(q) = \big\{u(y, s) : s \in S, y \in Y(s, q), u(y, s) \geq \underline{v}(q)\big\}.$$

Claim 1. *The correspondence W is Kakutani (i.e., u.h.c. with non-empty, convex, compact values).*

Proof of Claim 1. $W(q)$ is non-empty because $w(q) \in W(q)$. Verification of u.h.c. is routine: Take $q^n \to q$, $s^n \to s$, $y^n \in Y(s^n, q^n)$, $y^n \to y$ and for each s', there exists $y'^n \in Y(s', q^n)$ such that $u(y^n, s^n) \geq u(y'^n, s')$. Since $u(\cdot, \cdot)$ is continuous, from the maximum theorem, $Y(\cdot, \cdot)$ is u.h.c. and thus $y \in Y(s, q)$. For each s', take a converging subsequence to make sure that $y'^n \to y'$ and then $y' \in Y(s', q)$. So for each s', there exists $y' \in Y(s', q)$ such that $u(y, s) \geq u(y', s')$, i.e., $u(y, s) \geq \underline{v}(q)$.

We prove now that $W(q)$ is convex. Observe that $W(q) = \cup_{s \in S} W_s(q)$ with:

$$W_s(q) = \big\{u(y, s) : y \in Y(s, q), u(y, s) \geq \underline{v}(q)\big\}.$$

More precisely, $W(q) = \cup_{s \in S, W_s(q) \neq \emptyset} W_s(q)$. For each s, $Y(s, q)$ is a convex set of mixed actions and $y \mapsto u(y, s)$ is linear on this set, thus $W_s(q)$ is a compact interval of $\mathbb{R}$. To prove that the union of those intervals is convex, it is enough to show that any pair of them has a non-empty intersection. Thus, take s_1, s_2 such that $W_{s_1}(q)$ and $W_{s_2}(q)$ are non-empty and suppose by contradiction that $W_{s_1}(q) \cap W_{s_2}(q) = \emptyset$. Then, one of those intervals, say $W_{s_2}(q)$, is "above" the other, that is:

$$\max W_{s_1}(q) < \min W_{s_2}(q);$$

or equivalently:

$$\max\{u(y, s_1) : y \in Y(s_1, q), u(y, s_1) \geq \underline{v}(q)\big\} < \min\{u(y, s_2) : y \in Y(s_2, q), u(y, s_2) \geq \underline{v}(q)\big\}.$$

Consider then the interval:

$$W'_{s_2}(q) = \{u(y, s_2) : y \in Y(s_2, q)\big\}.$$

We have $W_{s_2}(q) \subseteq W'_{s_2}(q)$ and $\min W'_{s_2}(q) \leq \underline{v}(q)$. Then, either $\max W_{s_1}(q) < \min W'_{s_2}(q)$, thus $\max W_{s_1}(q) < \underline{v}(q)$ which contradicts that $W_{s_1}(q)$ is non-empty. Or $\max W_{s_1}(q) \geq \min W'_{s_2}(q)$, but then $W_{s_1}(q) \cap W'_{s_2}(q) \neq \emptyset$. This is also a contradiction because $W_{s_1}(q) \cap W'_{s_2}(q) \subseteq W_{s_1}(q) \cap W_{s_2}(q)$. This concludes the proof of Claim 1. □

Now, it follows that the correspondence $\alpha \mapsto W((1-\alpha)p_0 + \alpha p_j)$ from $[0, 1]$ to $\mathbb{R}$ is also Kakutani. From Lemma 3 in Lipnowski and Ravid (2020), its image is an interval. This proves Lemma 2 because $w(p_j) \in W(p_j)$, $w(p_0) \in W(p_0)$ and $w^* \in [w(p_0), w(p_j)]$. □

We conclude the proof of Proposition 2. For each j such that $w(p_j) > w^*$, we replace p_j by $(1-\alpha_j)p_0 + \alpha_j p_j$ given by Lemma 2. Together with the family of s_j, y_j, this defines an equilibrium with payoff w^*. □

Consider now the maximal equilibrium payoff $u^*(p_0)$ of the sender at p_0. Observe that:

$$u^*(p_0) = \max\Big\{v \in \mathbb{R} : \exists (p_k)_k, p_0 \in \text{co}\{(p_k)_k\} \text{ and } \forall k, \exists s_k \in S, \exists y_k \in Y(s_k, p_k), v = u(y_k, s_k)\Big\}.$$

The right-hand-side of this equality is greater or equal than $w(p_0)$, by considering the non-revealing splitting ($\forall k, p_k = p_0$) and $w(p_0) \geq \max_{s \in S} \min_{p, y \in Y(s,p)} u(y,s)$. Thus the INTIR condition is superfluous in defining $u^*(p_0)$. Also, $u^*(p_0) \geq w(p_0)$.

We now derive two consequences of Proposition 2.

Corollary 6. *For any family $(p_k)_k$ such that $p_0 \in \operatorname{co}\{(p_k)_k\}$, we have $u^*(p_0) \geq \min_k w(p_k)$.*

Proof of Corollary 6. If $\min_k w(p_k) \leq w(p_0)$, this follows form $u^*(p_0) \geq w(p_0)$. If $\min_k w(p_k) > w(p_0)$, from Proposition 2, $\min_k w(p_k)$ is an equilibrium payoff, thus less or equal to $u^*(p_0)$. □

Corollary 7. *Let $(p_k^*, s_k^*, y_k^*)_k$ be an optimal family in the maximisation problem defining $u^*(p_0)$. Then $u^*(p_0) = \min_k w(p_k^*)$.*

Proof of Corollary 7. For each k, $u^*(p_0) = u(y_k^*, s_k^*) \leq w(p_k^*)$, thus $u^*(p_0) \leq \min_k w(p_k^*)$. On the other hand, $\min_k w(p_k^*)$ in an equilibrium payoff from Proposition 2, thus it is less or equal to $u^*(p_0)$. □

The previous corollaries lead to the following proposition.

Proposition 3. *The function $p_0 \mapsto u^*(p_0)$ is quasi-concave.*

Proof of Proposition 3. We prove now that for any family $(p_k)_k$ such that $p_0 \in \operatorname{co}\{(p_k)_k\}$, $u^*(p_0) \geq \min_k u^*(p_k)$. For each k, let $(p_{kj}^*, s_{kj}^*, y_{kj}^*)_j$ be an optimal family in the maximisation problem defining $u^*(p_k)$. Since $p_0 \in \operatorname{co}\{(p_k)_k\}$ and $p_k \in \operatorname{co}\{(p_{kj}^*)_j\}$, then $p_0 \in \operatorname{co}\{(p_{kj}^*)_{k,j}\}$. Then from Corollary 6,

$$u^*(p_0) \geq \min_k \min_j w(p_{kj}^*)$$

and from Corollary 7, $\min_j w(p_{kj}^*) = u^*(p_k)$. Thus, u^* is quasi-concave. □

Proof of Theorem 2. We are now in position to prove the theorem. We know that the maximal equilibrium payoff $u^*(p_0)$ of the sender is a quasi-concave function of p_0 which is greater than or equal to $w(p_0)$. We argue that this is the smallest such quasi-concave function: if f is quasi-concave and $f(p_0) \geq w(p_0)$ for all p_0, then $f(p_0) \geq u^*(p_0)$ for all p_0. To see this, let $(p_k^*, s_k^*, y_k^*)_k$ be an optimal family in the maximisation problem defining $u^*(p_0)$. From Corollary 7, $u^*(p_0) = \min_k w(p_k^*)$. Then,

$$u^*(p_0) = \min_k w(p_k^*) \leq \min_k f(p_k^*) \leq f(p_0),$$

since f is quasi-concave. This concludes the proof. □

C Proof of Theorem 3

Proof. Let $v \geq w(p_0)$ be a PBE payoff of the signaling game. We construct a PBE with payoff v such that k, k', $p_k \neq p_{k'}$ and $s_k \neq s_{k'}$. In words, all the information transmitted by the equilibrium is contained in the signal: if two signals are the same, then the posterior beliefs are also the same.

Since v is a PBE payoff of the signaling game, there exists a finite family $(p_k, s_k, y_k)_{k=1}^K$ such that $p_0 \in \text{co}\{(p_k)_{k=1}^K\}$ and for all k, $y_k \in Y(s_k, p_k)$ and $v = u(y_k, s_k)$.

Since the statement is clear for $v = w(p_0)$, take $v > w(p_0)$. For each $s \in S' = \{s_k : k = 1, \ldots, K\}$, denote $K_s = \{k : s_k = s\}$. For some convex combination coefficients $\lambda_k > 0$, we have:

$$p_0 = \sum_k \lambda_k p_k = \sum_{s \in S'} \lambda_s \sum_{k \in K_s} \frac{\lambda_k}{\lambda_s} p_k = \sum_s \lambda_s p_s,$$

with $\lambda_s = \sum_{k \in K_s} \lambda_k$ and $p_s = \sum_{k \in K_s} \frac{\lambda_k}{\lambda_s} p_k$, so $p_0 \in \text{co}\{(p_s)_{s \in S'}\}$.

For each signal s, denote:

$$D_s = \{p \in \Delta(T) : \exists y \in Y(s,p), u(y,s) \geq v\} = \{p \in \Delta(T) : w_s(p) \geq v\}.$$

Since w_s is quasi-concave, D_s is convex, it is thus a non-empty convex and compact set (since w_s is u.s.c). From the equilibrium condition, for all $s \in S'$ and all $k \in K_s$, $u(y_k, s_k) = v$, thus $p_k \in D_s$. Since D_s is convex, we also have $p_s \in D_s$ for each s.

Thus for each s we have:

$$w_s(p_0) \leq w(p_0) < v \leq w_s(p_s).$$

Consider then for $\alpha \in [0, 1]$,

$$F_s(\alpha) = \{u(y, s) : y \in Y(s, (1-\alpha)p_0 + \alpha p_s)\}.$$

Similarly to the proof of Lemma 2, this correspondance is Kakutani, thus its image is an interval. It follows that there exists $\alpha_s \in [0,1]$ and $y_s \in Y(s, (1-\alpha)p_0 + \alpha p_s)$ such that $v = u(y_s, s)$. We modify then the equilibrium by replacing all the (y_k, p_k) for $k \in K_s$, by (y_s, p_s). This gives the desired conclusion. ■

D Proof of Theorem 4

Proof. The proof is a simplified version of the proof of Theorem 1 by removing the equilibrium conditions of the sender. To see how the adjustment by the likelihood ratio arises, consider a convex combination of posteriors $p_0 = \sum_k \lambda_k p_k$ and the strategy σ given by $\sigma(s_k, m_k \mid t) = \frac{p_k(t)\lambda_k}{p_0(t)}$ which induces it. The interim payoff given type t is:

$$v_t = \sum_k \sigma(s_k, m_k \mid t) v_t^k = \sum_k \lambda_k \frac{p_k(t)}{p_0(t)} v_t^k = \sum_k \lambda_k \hat{v}_t^k, \tag{17}$$

thus the interim payoff v is a convex combination of adjusted values $\hat{v}$. ■

E Proof of Corollary 5

Proof. Consider $(v, p_0) \in \operatorname{co} \hat{G}$, there is a convex combination of posteriors $p_0 = \sum_k \lambda_k p_k$, signals s_k and actions $y_k \in Y(s_k, p_k)$ such that for each t,

$$v_t = \sum_k \lambda_k \frac{p_k(t)}{p_0(t)} u(y_k, s_k, t).$$

Thus,

$$\sum_t p_0(t) v_t = \sum_k \lambda_k \sum_t p_k(t) u(y_k, s_k, t) \leq \sum_k \lambda_k w(s_k, p_k) \leq \operatorname{cav} \max_{s \in S} w(s, p_0).$$

If for each k, s_k, we choose y_k such that $w(s_k, p_k) = \sum_t p_k(t) u(y_k, s_k, t)$, then we have equality. Therefore,

$$\max \Big\{ \sum_t p_0(t) v_t : (v, p_0) \in \operatorname{co} \hat{G} \Big\} = \operatorname{cav} \max_{s \in S} w(s, p_0).$$

■