Coalition-Proof Disclosure*

Germán Gieczewski[†] Maria Titova[‡]

May 2024

Abstract

We analyze the equilibrium set of a general game of verifiable disclosure with type-independent sender preferences and propose an equilibrium selection criterion of coalition proofness among sender types. We provide recursive algorithms that output all equilibrium strategies and all coalition-proof equilibrium strategies. We provide four sets of conditions on the sender's payoff function and the mapping from types to available messages that guarantee existence of a coalition-proof equilibrium. We show when coalition proofness coincides with existing equilibrium selection methods such as receiver optimality and truth-leaning. We geometrically characterize the sender's ex-ante utility in the coalition-proof equilibrium of a disclosure game with a rich message space and compare it to its counterparts in cheap talk and Bayesian persuasion.

*We thank Nageeb Ali, Steve Callander, Jeff Ely, Elliot Lipnowski, Jacopo Perego and the audience members at Penn State, the Berlin Verifiability in Mechanism and Information Design Workshop and MPSA. All errors are our own.

[†]Department of Politics, Princeton University.

[‡]Department of Economics, Vanderbilt University. E-mail: motitova@gmail.com.

1 Introduction

Games of verifiable information are used to model many important economic situations. The canonical models of verifiable disclosure (Grossman, 1981, Milgrom, 1981) gave us the classic unraveling result that predicts full revelation. More recently, various papers (Glazer and Rubinstein (2004), Hagenbach, Koessler, and Perez-Richet (2014), Hart, Kremer, and Perry (2017), Rappoport (2022), and Sher (2011, 2014)) have studied verifiable disclosure games with various properties, such as uncertainty about how much information the sender has, or limited ability of the sender to reveal her type. In these models partial revelation in rich patterns is possible but typically accompanied by severe multiplicity of equilibria. As a result, most of the literature has focused on receiver-optimal equilibria, which often coincide with the receiver commitment solution (Ben-Porath, Dekel, and Lipman (2019), Glazer and Rubinstein (2004), and Hart, Kremer, and Perry (2017)).

In this paper we take a different approach. We study a game of disclosure with one substantive assumption: the sender's preferences are type-independent. Our setup is general: the sender's payoff is some function of the receiver's posterior and there is some mapping from sender types to available messages. Rather than focusing on receiver-optimal equilibria, we introduce a notion of *coalition proofness*, which is closely related to existing notions of neologism proofness (Farrell, 1993) and announcement proofness (Matthews, Okuno-Fujiwara, and Postlewaite, 1991) for cheap talk games. The gist of our equilibrium selection argument is that the receiver should correctly interpret coalitional deviations by groups of senders when these deviations are credible.

Due to the generality of the considered model, coalition-proof equilibria may fail to exist, for similar reasons that neologism- and announcement-proof equilibria fail to exist in the cheap talk literature. We provide four sets of conditions on the sender's payoff function and the message mapping that guarantee existence of a coalition-proof perfect Bayesian equilibrium (PBE). First, if the message mapping is *complete* (the set of messages available to each type is sufficiently rich as in Bertomeu and Cianciaruso, 2018), then it is sufficient for the sender's payoff to be quasiconcave. Second, if the sender's payoff satisfies *betweenness* (Hart, Kremer, and Perry, 2017), then no condition on the message mapping is required. The last two existence conditions require that the sender have access to cheap talk in addition to verifiable messages. Under this assumption, a coalition-proof PBE exists as long as the message mapping is complete or the sender's payoff from fully revealing her type is sufficiently low.

We also provide a characterization of coalition-proof PBE and tools for finding them. In fact, we show that *all* PBE strategies—including coalition-proof ones—belong to a class of strategies that we term *partition strategies*. These strategies partition the type space into *coalitions*. A coalition consists of a set of sender types, a set of messages that only these types could send, a (possibly mixed) strategy assigning types to messages, and a common payoff for all types in the coalition.

We provide algorithms that return (i) the set of all partition strategies; (ii) the set of all PBE strategies; and (iii) the set of all coalition-proof PBE strategies. These algorithms are recursive and remove one coalition from the game at each step. We show that the coalition-proof PBE algorithm is simply a greedy version of the PBE algorithm; that is, it follows the same steps but with the added requirement that a coalition that reaches the highest payoff is removed at each step. This logic yields an additional result: in a number of cases, if the sender's payoff function is "generic" (such that no ties occur across coalition payoffs), then the coalition-proof PBE is unique.

Finally, we provide a geometric characterization of the sender's ex-ante utility for disclosure games with a rich message mapping. In so doing, we provide a benchmark result for disclosure games that can be seen as analogous to Kamenica and Gentzkow (2011)'s in information design and Lipnowski and Ravid (2020)'s in cheap talk.

There are two strands of literature, besides the aforementioned, that our paper connects to. First, a number of applied papers on disclosure, motivated by similar concerns to ours, employ notions of equilibrium selection close to coalition proofness to select "reasonable" equilibria. Among these are Aybas and Callander (2024), Callander, Lambert, and Matouschek (2021), and Farina et al. (2024). We contribute to this literature by giving a general characterization of and existence conditions for this solution concept. Second, two existing papers, Bertomeu and Cianciaruso (2018) and Koessler and Skreta (2023), perform closely related theoretical exercises. Bertomeu and Cianciaruso (2018) is very much in the same spirit as our approach; our notion of coalition proofness generalizes their equilibrium selection concept, and allows us to extend their results in several directions. Koessler and Skreta (2023) studies a problem of informed information design which shares some similarities with models of disclosure, and yields related but different results to ours in a case related to our benchmark exercise. We discuss both of these, along with Hart, Kremer, and Perry (2017), in more detail in Section 6. Finally, Rappoport (2022) is noteworthy for presenting an algorithmic characterization of equilibria in disclosure games, though his algorithm yields receiver-optimal equilibria rather than coalition-proof ones. Wu (2022) contains some related results—namely, a characterization of PBE, though under a "strong normality" assumption on evidence and the presence of cheap talk, and an algorithmic characterization in a further special case where there is a total order on messages. It contains no analogous notion of coalition proofness, and its results do not imply ours.

The rest of the paper proceeds as follows. We begin with two examples that illustrate the shortcomings of receiver optimality and (ex ante) sender optimality as equilibrium selection approaches. Section 2 introduces the model and notation. Section 3 introduces coalition-proof PBE as a solution concept and presents a recursive characterization alongside other basic results. Section 4 presents four sets of conditions that guarantee existence of a coalition-proof PBE. Section 5 solves the benchmark case with a maximally rich message mapping. Section 6 discusses related papers in more detail. Section 7 concludes. All proofs are in Appendix A.

Motivating Examples

Two examples serve to illustrate the motivation for our exercise.

Example 1 (Implausible revelation in receiver-optimal equilibria)**.** *A centrist incumbent (Sender) knows the state of the world* $\theta \in \{-1, 1\}$ *(with prior* $P(\theta = 1) = 0.5$*) and can reveal it to a voter (Receiver) or not, by choosing a message* $m \in \{\theta, \varnothing\}$*.* R *sees* m *and elects a left-wing challenger (*$a = -1$*), a right-wing one (*$a = 1$*), or reelects* S *(*$a = S$*).* S *wants to be reelected (*$u_S = \mathbb{1}_{a=S}$*);* R *wants to match the state (*$u_R(1) = \mathbb{1}_{\theta=1}$*,* $u_R(-1) = \mathbb{1}_{\theta=-1}$*) but reelecting* S *is a safe alternative (*$u_R(S) = 0.9$*).*

This game has two types of (perfect Bayesian) equilibria. First, there is an equilibrium in which all senders send $m = \varnothing$ and get reelected. Second, there are equilibria in which the sender is reelected with probability 0. In these equilibria, the empty message is either never sent (but interpreted by R as coming disproportionately from one type off-path) or sent disproportionately by one sender type.

In any receiver-optimal equilibrium, S reveals θ w.p. 1 and R's posterior after seeing the empty message, $P(\theta = 1|m = \varnothing)$, is any element of $[0, 0.1] \cup [0.9, 1]$. R extracts full revelation from S by threatening to interpret the off-path message asymmetrically, despite both sender types having identical incentives to go off-path.

Revealing the state is weakly dominated by saying nothing: if S sends θ, she is never reelected and gets 0; if she sends $\varnothing$, she gets 0 at worst. Simply speaking, anything the sender might say can be used against her. Thus, the sender could argue: "it is in my interest to reveal no information, and I would benefit from making this announcement regardless of the true state. Hence, you should retain your prior belief when I reveal nothing." In other words, the sender could announce a deviation to

$m \equiv \varnothing$, which both types would participate in.[1] $\square$

One may think that limiting the receiver's ability to adversarially interpret deviations ought to increase the sender's payoff, so perhaps we should focus on equilibria that are ex ante optimal for the sender. Our next example provides a counterpoint.

Example 2 (Implausible equilibria that are ex ante sender-optimal)**.** *Take a game of disclosure in the spirit of Grossman (1981) but with a "nuisance dimension". Assume* $\Theta = \{(L, A), (L, B), (H, A), (H, B)\}$*, with all 4 types equally likely. Denote a generic type by* $\theta = (\theta(1), \theta(2))$*. A type* θ *can send any message* $m \subseteq 2^{\Theta}$ *such that* $\theta \in m$*.* $u_S \equiv a$*, and R's best response is* $a^*(m) \equiv \sqrt{\Pr(\theta(1) = H|m)} - 8\left[\Pr(\theta(2) = A|m) - 0.5\right]^2$.

Here, revealing a high $\theta(1)$ is good for S, but revealing anything on the second dimension $\theta(2)$ is worse than the prior. Yet, any receiver-optimal equilibrium features full revelation: $m \equiv \{\theta\}$ and the receiver interprets off-path messages adversarially.[2]

In contrast, in the ex ante sender-optimal PBE, all types pool on $m = \Theta$ and get $u_S = \sqrt{0.5}$. To deter deviations, R must think that the message $m = \{(H, A), (H, B)\}$ is disproportionately likely to come from a single type, much as in Example 1.

Compare this with a simpler game in which (L, A), (L, B) are merged into a single type L, and (H, A), (H, B) are merged into a single type H. Now, unraveling happens: H separates from L. Note that unraveling *lowers* ex ante sender payoffs: H gets 1 and L gets 0, compared to $\sqrt{0.5}$ for both if they were to pool.

Our intuition is that a natural equilibrium in Example 2 should mirror this outcome: types (H, A), (H, B) should pool on $m = \{(H, A), (H, B)\}$, leaving (L, A),

[1] Refinements such as the intuitive criterion or D1 are not helpful here, as they generally give conditions to rule out certain types as potential deviators. Our argument is that multiple sender types are all equally plausible deviators.

[2] Assume R thinks that m comes from (L, A) w.p. 1 if $(L, A) \in m$; if not, but $(L, B) \in m$, then R thinks m came from (L, B); if not, but $(H, A) \in m$, R thinks m came from (H, A). Since the receiver does better with more information, a PBE with full revelation is receiver-optimal if it exists.

(L, B) to pool on their own. More generally, when S has incentives to reveal some dimensions of the state but not others, this ought to be the equilibrium outcome. However, receiver optimality can predict full revelation even if S has incentives to conceal some (or all) dimensions of the state. Ex ante sender optimality can deter sender types from revealing information they want to reveal if this would lower ex ante payoffs, and does so by leveraging the same threat of (implausibly) adversarial beliefs on dimensions that S would rather conceal. $\square$

2 Model

There are two players, a sender (S, she/her) and a receiver (R, he/him). The game proceeds as follows. First, S observes her type $\theta \in \Theta := \{\theta_1, \ldots, \theta_n\}$, which is drawn from a common prior distribution $\mu^0 := (\mu^0_1, \ldots, \mu^0_n) \in \Delta\Theta$. Then, S chooses message $m \in M(\theta)$, where $M : \Theta \to 2^{\mathcal{M}} \setminus \varnothing$ is a mapping that determines the set of messages available to type θ and $\mathcal{M}$ is the "grand" message space. Having observed m, R forms posterior belief $\mu \in \Delta\Theta$, and chooses action $a \in A$. After that, payoffs $u_S(a)$ and $u_R(a, \theta)$ are realized. Note that the sender's preferences are state-independent, and his payoff is affected only by R's action.

We employ the belief-based approach that is common in the information design literature. Specifically, we focus on R's posterior belief $\mu \in \Delta\Theta$ and let $a^*(\mu)$ be the receiver's best response and $v(\mu) := u_S(a^*(\mu))$ be the sender's payoff when R has that belief. For much of the paper, it is without loss to assume that R breaks ties in favor of S if indifferent, which leads to v being upper semicontinuous under mild assumptions.[3] We thus forget about R as a player and simply work with an upper

[3] v is upper semicontinuous if R breaks ties in favor of S, and there is a metric on A under which A is compact and $u_S(a)$, $u_R(a, \theta_i)$ for each i are continuous in a (see Lemma 3). In particular, this condition can hold if A is a compact subset of $\mathbb{R}^k$, and holds automatically for any finite A, as we can take the discrete metric on A.

semicontinuous function v. In a few places that will be made clear, more careful tie-breaking is needed, and we will treat v as a correspondence returning all possible values of u_S when R best-responds.

Next, we introduce some helpful notation. Given a set of messages $X \subseteq \mathcal{M}$, we let $M^{-1}(X) := \{\theta \in \Theta \mid M(\theta) \cap X \neq \varnothing\}$ be the set of types with access to at least one message in X. Also, given a non-empty set of types $C \subseteq \Theta$, we let $\mu^0(C) := \sum_{i \in C} \mu^0(\theta_i)$ be the prior measure of C and $\mu^0_C \in \Delta\Theta$ be the prior distribution conditional on C defined as $\mu^0_C(\theta) := \frac{\mu^0(\theta) \cdot \mathbb{1}(\theta \in C)}{\mu^0(C)}$ for all $\theta \in \Theta$. In our analysis, we often consider a *restricted game* with non-empty type space $\widetilde{\Theta} \subseteq \Theta$, prior distribution $\mu^0_{\widetilde{\Theta}}$ and message mapping $M|_{\widetilde{\Theta}}$, which is simply M restricted to the domain $\widetilde{\Theta}$.[4]

Coalitions and Partitions

We introduce a new object that we term a *partition* and focus on a particular class of strategies that we term *partition strategies*. These strategies partition the type space Θ into *coalitions* of sender types that get the same payoff and send messages that are only available to them.

Definition 1. A coalition is a quadruple (C, X, σ, w), where

1. $C \subseteq \Theta$ is a non-empty set of types.
2. $X \subseteq \mathcal{M}$ is a set of messages such that $M^{-1}(X) = C$.
3. $\sigma : C \to \Delta\mathcal{M}$ is a sender strategy for types $\theta \in C$ such that $\operatorname{supp} \sigma(\cdot \mid \theta) \subseteq X \cap M(\theta)$ for all $\theta \in C$ and $\bigcup_{\theta \in C} \operatorname{supp} \sigma(\cdot \mid \theta) = X$.
4. $w := v(\mu(\cdot \mid m))$ for each $m \in X$, where $\mu(\cdot \mid m)$ is calculated from μ^0, given σ, using Bayes' rule.

---

[4] That is, $M|_{\widetilde{\Theta}} : \widetilde{\Theta} \to 2^{\mathcal{M}} \setminus \varnothing$ is given by $M|_{\widetilde{\Theta}}(\theta) = M(\theta)$ for all $\theta \in \widetilde{\Theta}$.

Coalitions have two important features. First, a coalition strategy σ only specifies what types in C do. In particular, every type $\theta \in C$ uses messages from $X \cap M(\theta)$; all messages in X are "on path." While σ does not specify what types in $\Theta \setminus C$ do, we know that they do not have access to messages in X. Second, although different messages in X may induce different posteriors, the sender's payoff is the same for all of them. In particular, types in the coalition receive the same payoff and are indifferent between all the messages $m \in X \cap M(\theta)$ that they use.

Remark 1. Let $\mathcal{C}(\widetilde{\Theta})$ be the set of coalitions of the restricted game with a non-empty type space $\widetilde{\Theta} \subseteq \Theta$. Then, $\mathcal{C}(\widetilde{\Theta})$ is non-empty.

Proof. Let $m \in \bigcup_{\theta \in \widetilde{\Theta}} M(\theta)$ be a message; $C = (M|_{\widetilde{\Theta}})^{-1}(\{m\})$ be the non-empty set of types with access to it; $\sigma(m \mid \theta) = 1$ for $\theta \in C$ be the strategy prescribing that everyone in $\widetilde{\Theta}$ who can send m does so. Then, $(C, \{m\}, \sigma, v(\mu_C^0))$ is a coalition. □

While the existence of coalitions that pool on a single message is clear, there may also exist coalitions that pool on a larger set of messages. Next, we recursively define a partition and provide an algorithm that outputs them.

Definition 2. A collection $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ is a partition if

- C_1, …, C_T are disjoint and $\Theta = \bigcup_{t=1}^T C_t$;
- for each $t \in \{1, \ldots, T\}$, $(C_t, X_t, \sigma_t, w_t)$ is a coalition of the restricted game with type space $\Theta_t := C_t \cup \ldots \cup C_T$, or $(C_t, X_t, \sigma_t, w_t) \in \mathcal{C}(\Theta_t)$.

We will refer to each $(C_t, X_t, \sigma_t, w_t)$ simply as a coalition when there is no possibility of confusion, although generally (for $t > 1$) it is not a coalition of the original game. Since Θ is finite and each C_t contains at least one type, Algorithm 1 terminates in at most $|\Theta|$ steps. Furthermore, the set of partitions is non-empty since Remark 1 ensures existence of a coalition at each step of the algorithm. When the algorithm

Algorithm 1: Partition Algorithm

Let $t := 1$ and $\Theta_1 := \Theta$;

while $\Theta_t \neq \varnothing$

  let $(C_t, X_t, \sigma_t, w_t) \in \mathcal{C}(\Theta_t)$;

  let $\Theta_{t+1} := \Theta_t \setminus C_t$ and $t := t + 1$;

end

terminates, $\{\sigma_t\}_{t=1}^T$ specifies the strategy for all sender types $\theta \in \Theta$ and R's posterior beliefs for all on-path messages.

Definition 3. $\sigma : \Theta \to \Delta\mathcal{M}$ is a partition strategy if, for some partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$, we have $\sigma|_{C_t} = \sigma_t$ for all $t \in \{1, \ldots, T\}$. We say that σ is associated with $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$.

3 Analysis

PBE Partitions and Individual Rationality

The standard solution concept for communication games is perfect Bayesian equilibrium (PBE). We say that a Sender's strategy $\sigma : \Theta \to \Delta\mathcal{M}$ is a PBE strategy if there exists a Receiver's belief system $\mu : \mathcal{M} \to \Delta\Theta$ such that

(PBE-1) $\forall\theta \in \Theta$, $\sigma(\cdot \mid \theta)$ is supported on $\arg\max_{m \in M(\theta)} v(\mu(\cdot \mid m))$;

(PBE-2) μ is obtained from μ^0, given m, using Bayes' rule, for all m used with positive probability on equilibrium path.

We begin by characterizing the set of PBE strategies in terms of partitions. Consider a partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ and the associated strategy σ. For σ to be a PBE strategy, S must not have profitable deviations to on-path or off-path messages.

At the very least, the sender's payoff must exceed his best deviation to any off-path message assuming that R is maximally skeptical.

Definition 4. A partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ is individually rational (IR) if

$$w_t \geq \underline{v}(\theta) := \max_{m \in M(\theta)} \min_{\mu(\cdot|m) \text{ feasible}} v(\mu(\cdot|m)) \quad \text{for all } t \in \{1, \ldots, T\} \text{ and } \theta \in C_t.$$

The set of feasible beliefs $\mu(\cdot|m)$ for a message m is simply the set $\Delta M^{-1}(\{m\})$ if all types in $M^{-1}(\{m\})$ have access to more than one message. More generally, it is the set of all beliefs proportional to $(\mu_i^0 \sigma_i)_{\theta_i \in M^{-1}(\{m\})}$ for any $\sigma_i \in [0,1]$ if $|M(\theta_i)| > 1$ and $\sigma_i = 1$ if $M(\theta_i) = \{m\}$.

Our first result is an equilibrium characterization in terms of partition strategies. We show that all PBE strategies are partition strategies. Given a strategy associated with partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$, sender deviations on path are ruled out as long as w_t is decreasing; sender deviations off path are ruled out by IR.

Proposition 1. *σ is a PBE strategy $\iff$ σ is associated with an individually rational partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ such that $w_1 \geq \ldots \geq w_T$.*

Using Proposition 1, we propose a modification of Algorithm 1 that returns all *PBE partitions* (those satisfying IR and decreasing payoffs) and hence all PBE strategies.

Algorithm 2: PBE Partition Algorithm

Let $t := 1$, $\Theta_1 := \Theta$, and $w_0 := \infty$;

while $\Theta_t \neq \varnothing$

  let $(C_t, X_t, \sigma_t, w_t) \in \mathcal{C}(\Theta_t)$ be a coalition such that $w_t \in [\max_{\theta \in \Theta_t} \underline{v}(\theta), w_{t-1}]$;

  let $\Theta_{t+1} := \Theta_t \smallsetminus C_t$ and $t := t+1$;

end

In words, at each step of the algorithm, we select coalitions such that payoffs are non-increasing ($w_{t-1} \geq w_t$) and that satisfy individual rationality ($w_t \geq \max_{\theta \in \Theta_t} \underline{v}(\theta)$).[5] Unlike Algorithm 1, Algorithm 2 may not terminate: depending on coalitions selected at earlier steps, there may not exist a coalition at step t with payoff $w_t \in [\max_{\theta \in \Theta_t} \underline{v}(\theta), w_{t-1}]$. When that happens, we say that the algorithm *halts* with no output. Of course, since a PBE must exist (by Fan-Glicksberg), there is at least one way to select coalitions at each step so that the algorithm terminates.

Coalition Proofness

Our proposed criterion to select among strategies is *coalition proofness*. We first provide a definition of coalition proofness, which is closely related to neologism proofness (Farrell, 1993) and announcement proofness (Matthews, Okuno-Fujiwara, and Postlewaite, 1991) in cheap talk games, and Grossman-Perry-Farrell equilibrium in disclosure games (Bertomeu and Cianciaruso, 2018). We then provide a simple modification of Algorithm 2 that yields all coalition-proof PBE strategies.

Definition 5. Let σ be a strategy associated with partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^{T}$.

(i) $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ is a blocking coalition of σ if it is a coalition of the restricted game with type space $\bigcup_{t: w_t < \widetilde{w}} C_t$.

(ii) σ is coalition-proof if there are no blocking coalitions.

Intuitively, a strategy σ is coalition-proof if it rules out coalitional deviations. A coalitional deviation involves a set of types $\widetilde{C}$ announcing that they would like to switch to a message strategy $\widetilde{\sigma}$ with domain $\widetilde{C}$ and codomain $\widetilde{X}$, such that if the

[5] The IR requirement on types $\theta \in C_t$ is $w_t \geq \max_{\theta \in C_t} \underline{v}(\theta)$ rather than $w_t \geq \max_{\theta \in \Theta_t} \underline{v}(\theta)$. However, the stronger requirement $w_t \geq \max_{\theta \in \Theta_t} \underline{v}(\theta)$ yields the same set of partitions. The reason is that, if $\max_{\theta \in C_t} \leq w_t < \max_{\theta \in \Theta_t} \underline{v}(\theta)$, a failure of IR is inevitable: there is $\widetilde{\theta} \in \Theta_t \setminus C_t$ with $\underline{v}(\theta) > w_t$, and this type must be included in a coalition $(C_\tau, X_\tau, \sigma_\tau, w_\tau)$ with $\tau > t$, so $\underline{v}(\theta) > w_t \geq w_\tau$.

receiver believed this announcement and updated her beliefs accordingly in response to messages in $\widetilde{X}$, the types in $\widetilde{C}$ would obtain payoff $\widetilde{w}$ from the deviation. For this announcement to be credible, the participating types in $\widetilde{C}$ must be exactly those who have access to at least one message in $\widetilde{X}$ and benefit from the deviation.[6]

An issue that needs careful consideration is that the messages that the deviators mean to use may already be used on path. In such cases, it is unclear how the receiver should interpret the announcement $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ followed by a message $m \in \widetilde{X}$: does it come from a deviator, or from an on-path user? The following remark confirms that announced deviations to on-path messages do not cause any ambiguity.

Remark 2. Let σ be a PBE strategy associated with partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^{T}$ and let $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ be a blocking coalition. Then $M^{-1}(\widetilde{X} \cap X_t) \cap C_t \subseteq \widetilde{C}$ for all t.

Proof. Let $\tau \in \{1, \ldots, T\}$ be minimal such that $\widetilde{w} > w_\tau$. By definition, $\tilde{C} = M^{-1}(\widetilde{X}) \cap \left(\bigcup_{t:w_t<\widetilde{w}} C_t\right) = M^{-1}(\widetilde{X}) \cap \Theta_\tau$.

For $t < \tau$, note that if $m \in X_t$ then $M^{-1}(\{m\}) \subseteq C_1 \cup \ldots \cup C_t \subseteq C_1 \cup \ldots \cup C_{\tau-1} = \Theta - \Theta_\tau$, which implies $m \notin \widetilde{X}$. Hence $\widetilde{X} \cap X_t = \varnothing$, so $M^{-1}(\widetilde{X} \cap X_t) \cap C_t = \varnothing$. For $t \geq \tau$, it is obvious that $M^{-1}(\widetilde{X} \cap X_t) \cap C_t \subseteq M^{-1}(\widetilde{X}) \cap \Theta_\tau = \widetilde{C}$. □

Any sender type θ who in equilibrium uses a message $m \in X_t$ also used by the blocking coalition ($m \in \widetilde{X}$) must be in the set $M^{-1}(\widetilde{X} \cap X_t) \cap C_t$. The proof shows that all such types must be in coalitions with equilibrium payoff below $\widetilde{w}$. Therefore, R should expect *all* of them to participate in the announced deviation.

Now, we introduce a modification of the PBE Partition Algorithm that selects a coalition attaining the highest possible payoff at each stage. We refer to outputs of this algorithm as *greedy partitions*. Our next result, Proposition 2, establishes that Algorithm 3 returns all PBE strategies that are coalition-proof.

[6] The notion of coalition proofness can be extended to non-partitional strategies. In the general case, types are presumed to participate in a deviation iff their expected equilibrium payoff is less than $\widetilde{w}$.

Algorithm 3: Greedy Partition Algorithm

Let $t := 1$, $\Theta_1 := \Theta$, and $w_0 := \infty$;

while $\Theta_t \neq \varnothing$

let $W_t = \{w \in \mathbb{R} \mid \exists (C, X, \sigma, w) \in \mathcal{C}(\Theta_t)\}$ be the set of payoffs attainable by coalitions of the restricted game with type space Θ_t;

let $(C_t, X_t, \sigma_t, w_t) \in \mathcal{C}(\Theta_t)$ be such that $w_t = \max(W_t \cap [\max_{\theta \in \Theta_t} \underline{v}(\theta), w_{t-1}])$;

let $\Theta_{t+1} := \Theta_t \setminus C_t$ and $t := t + 1$;

end

Proposition 2. *Consider a partition strategy σ associated with $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$. Then, σ is a PBE strategy and coalition-proof if and only if $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ is a greedy partition.*

As might be expected from the literature on neologism proofness, coalition-proof PBEs do not always exist, meaning that Algorithm 3 may halt no matter what choices are made at each step. We now provide a minimal example of non-existence.[7]

Example 3 (Non-existence of coalition-proof PBE)**.** *Let $\Theta = \{1, 2, 3\}$. Let $M(1) = \{a, b\}$, $M(2) = \{a\}$, $M(3) = \{b\}$, $\mu^0 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Let $v : \Delta\Theta \to \mathbb{R}$ be any continuous function such that $v(x, 1-x, 0) \equiv x$, $v(x, 0, 1-x) \equiv 0.9 - x$ for all $x \in [0, 0.5]$, as illustrated in Figure 1.*

Example 3 has a unique PBE, which is not coalition-proof. To see why, follow Algorithm 2. The possible coalitions in stage 1 are:

$$X_1 = \{a\},\ C_1 = \{1, 2\},\ \sigma_1(a \mid 1, 2) = 1,\ w_1 = v(0.5, 0.5, 0) = 0.5;$$

$$X_1 = \{b\},\ C_1 = \{1, 3\},\ \sigma_1(b \mid 1, 3) = 1,\ w_1 = v(0.5, 0, 0.5) = 0.4,$$

[7] Example 3 uses three sender types. It can be shown that a coalition-proof PBE always exists with two types under only technical conditions on v.

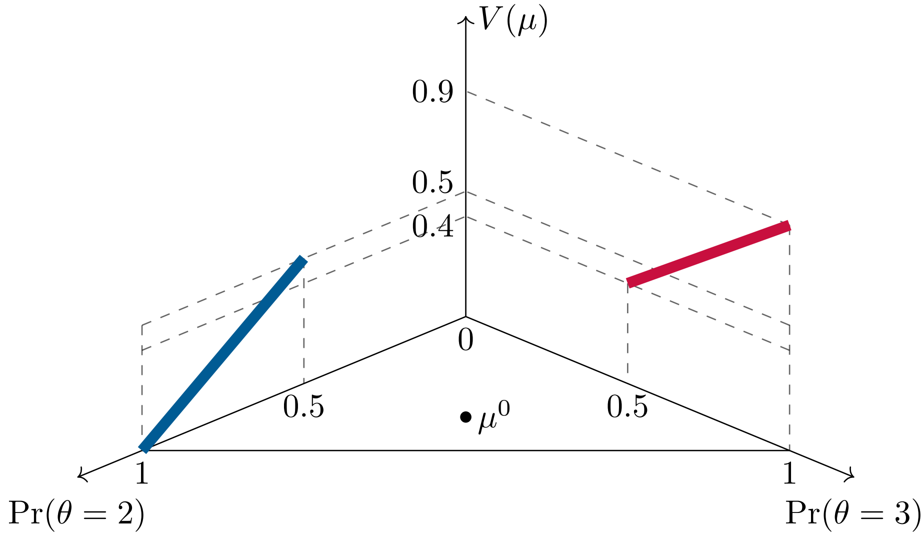


Figure 1. *Feasible sender payoffs in Example 3. Type 2 must send message a (blue), while type 3 must send message b (red). Type 1 has access to both messages.*

and there is no coalition with $X_1 = \{a, b\}$, as there is no mixed strategy for type 1 that equalizes the payoffs of messages a and b.

Thus $W_1 = \{0.4, 0.5\}$. If the coalition $(\{1,3\}, \{b\}, \cdot, 0.4)$ is chosen, then the only possible coalition in stage 2 is $(\{2\}, \{a\}, \cdot, 0)$. The associated strategy is a PBE strategy that is not coalition-proof, since $0.4 < \max W_1$: the blocking coalition is $(\{1,2\}, \{a\}, \cdot, 0.5)$. However, choosing $(\{1,2\}, \{a\}, \cdot, 0.5)$ in stage 1 leaves only one possible coalition in stage 2: $(\{3\}, \{b\}, \cdot, 0.9)$; the associated strategy is not a PBE strategy as type 1 has a profitable deviation to message b. □

Although coalition-proof PBEs can fail to exist, we show next that they exist under relatively weak conditions, covering many settings studied previously in the disclosure literature.

4 Theorems for Existence of Coalition-Proof PBE

We provide four sets of conditions on the sender's payoff function v and the message mapping M that guarantee existence of a coalition-proof PBE. Under certain additional conditions, we also show that Algorithm 3 always terminates, and that the coalition-proof PBE is unique.

Quasiconcavity of v and Completeness of M

The first existence condition is that v is quasiconcave (QC; we denote strict quasiconcavity by QC*) and M is complete (M-C).

(QC) For all $\alpha \in (0,1)$ and $\mu, \mu' \in \Delta\Theta$, $\min\{v(\mu), v(\mu')\} \leq v(\alpha\mu + (1-\alpha)\mu')$.

(QC*) For all $\alpha \in (0,1)$ and $\mu, \mu' \in \Delta\Theta$, $\min\{v(\mu), v(\mu')\} < v(\alpha\mu + (1-\alpha)\mu')$.

(M-C) For any two messages $m, m' \in \mathcal{M}$, there exists $m'' \in \mathcal{M}$ such that $M^{-1}(\{m''\}) = M^{-1}(\{m\}) \cup M^{-1}(\{m'\})$.

Completeness of the message mapping requires that the collection of types that can pool together on a single message—that is, the collection $\{M^{-1}(\{m\}) : m \in \mathcal{M}\}$—is closed under unions. In other words, if message m means "my type is in A", and message m' means "my type is in B" (where $A = M^{-1}(\{m\})$ and $B = M^{-1}(\{m'\})$), then there is a way to say "my type is in A or B".

Under these conditions, the following lemma allows us to effectively restrict attention to simple "pooling" coalitions.

Lemma 1. *If QC and M-C hold, then for any coalition (C, X, σ, w), there is a coalition $(C, \{m\}, \widetilde{\sigma}, \widetilde{w})$ with $M^{-1}(\{m\}) = C$, $\widetilde{\sigma}(m \mid C) = 1$ and $\widetilde{w} = v(\mu_C^0) \geq w$.*

Proof. M-C implies that there is a message m with $M^{-1}(\{m\}) = M^{-1}(X) = C$, and $(C, \{m\}, \widetilde{\sigma}, \widetilde{w})$ is a valid coalition that pays $\widetilde{w} = v(\mu_C^0)$. Since μ_C^0 is a linear combination of the posteriors $\mu(\cdot \mid m)$ generated by $m \in X$ under σ, we have that, by QC,

$$v(\mu_C^0) \geq \underbrace{\min_{m \in X} v(\mu(\cdot \mid m))}_{=w \text{ for all } m \in X} = w.$$

□

QC and M-C guarantee existence of a coalition-proof PBE, and adding QC* ensures that **every** way of choosing coalitions in Algorithm 3 terminates.

Theorem 1.

(i) If QC and M-C hold, then there exists a coalition-proof PBE.

(ii) If QC and M-C hold, then Algorithm 3 always terminates.*

(iii) If QC and M-C hold and v is generic (such that $v(\mu_C^0) = v(\mu_{C'}^0)$ only if $C = C'$), then all coalition-proof PBE are payoff-equivalent.

The gist of the proof is as follows. We show that at each step of Algorithm 3, $\max W_t \leq w_{t-1}$, which implies that simply picking the payoff-maximizing coalition (one with $w_t = \max W_t$) guarantees coalition proofness and ensures that the resulting partition has a non-increasing sequence of payoffs, as required for PBE. We prove that by contradiction: if a coalition paying more than w_{t-1} exists in stage t, then we can "merge" it with the coalition obtained at step $t-1$ to obtain a feasible coalition for step $t-1$ that pays more than w_{t-1}. Crucially, M-C ensures existence of messages that pool types in C_{t-1} and C_t together, while QC guarantees that those types receive a higher payoff from the merged coalition. Finally, when v is generic (in the sense that no two "pooling" coalitions pay the same), there is at most one choice of C_t at each step that can maximize w_t, yielding the uniqueness result.

Betweenness of v

The second existence condition is "betweenness" of v (v is quasiconcave and quasiconvex) and involves no restriction on M.[8]

(B) for all $\alpha \in (0,1)$ and $\mu, \mu' \in \Delta\Theta$,

$$\min\{v(\mu), v(\mu')\} \leq v(\alpha\mu + (1-\alpha)\mu') \leq \max\{v(\mu), v(\mu')\}$$

(B^*) B holds and for all $\alpha \in (0,1)$ and $\mu, \mu' \in \Delta\Theta$ such that $v(\mu) \neq v(\mu')$,

$$\min\{v(\mu), v(\mu')\} < v(\alpha\mu + (1-\alpha)\mu') < \max\{v(\mu), v(\mu')\}$$

A key observation is that, when v satisfies betweenness, all types in a coalition *must* receive their pooling payoff even if they pool by mixing across multiple messages, and even if there does not exist a single message available to all of them.

Lemma 2. *If B holds, then $w = v(\mu_C^0)$ for any coalition (C, X, σ, w).*

Proof. Let (C, X, σ, w) be a coalition. Then μ_C^0 is a linear combination of $\mu(\cdot \mid m)$ for $m \in X$. From B,

$$\underbrace{\min_{m\in X} v(\mu(\cdot \mid m))}_{=w \text{ for all } m\in X} \leq v(\mu_C^0) \leq \underbrace{\max_{m\in X} v(\mu(\cdot \mid m))}_{=w \text{ for all } m\in X} \implies v(\mu_C^0) = w.$$

□

In analogous fashion to Theorem 1, betweenness of v guarantees existence of a coalition-proof PBE; adding its strict version guarantees that **every** way of choosing

[8] Hart, Kremer, and Perry (2017) (HKP henceforth) provide a compelling microfoundation of betweenness: it holds if the receiver's expected utility $E_\mu(u_R(a,\theta))$ is single-peaked in her action a, for any belief $\mu \in \Delta\Theta$. When Θ is binary, B (B^*) is equivalent to $v(\mu)$ being (strictly) monotone.

coalitions in Algorithm 3 yields a coalition-proof PBE. Also, the coalition-proof PBE is essentially unique when v is generic.

Theorem 2.

(i) If B holds, then there exists a coalition-proof PBE.

(ii) If B holds, then Algorithm 3 always terminates.*

(iii) If B holds and v is generic (that is, $v(\mu_C^0) = v(\mu_{C'}^0)$ only if $C = C'$), then all coalition-proof PBE are payoff-equivalent.

The proof is similar to Theorem 1: we show, by contradiction, that at each step of Algorithm 3, we have $\max W_t \leq w_{t-1}$. We can no longer "merge" coalitions to arrive at a contradiction because no condition on M is assumed. However, condition B gives us enough structure on v to prove that, if two successive coalitions $(C_{t-1}, X_{t-1}, \sigma_{t-1}, w_{t-1})$, $(C_t, X_t, \sigma_t, w_t)$ satisfy $w_{t-1} < w_t$, then there must exist *some* coalition at step $t-1$ that pays at least $v(\mu^0_{C_{t-1} \cup C_t}) \in (w_{t-1}, w_t)$.

Adding Cheap Talk

The last two existence conditions apply when the sender has access to cheap talk in addition to verifiable messages. Formally, the mapping $M : \Theta \to 2^{\mathcal{M}} \setminus \varnothing$ satisfies the *cheap talk property* if

(M-CT) for each message $m \in \mathcal{M}$, there are at least n messages $m' \in \mathcal{M}$ (including m) such that $M^{-1}(\{m'\}) = M^{-1}(\{m\})$.

We call M-CT the cheap talk property since it allows the message space to be rewritten as follows. Let two messages m, m' be equivalent ($m \sim m'$) if $M^{-1}(\{m\}) = M^{-1}(\{m'\})$. Then, each m can be identified with a pair $(\widetilde{m}, j)$, where $\widetilde{m} \in \widetilde{\mathcal{M}} =$

$\mathcal{M}/\sim$, the set of equivalence classes of the relation $\sim$, and $j = 1, 2, \ldots, k_{\tilde{m}}$ ($k_{\tilde{m}} \geq n$) is an index denoting which element of the equivalence class was sent, and which we can interpret as a cheap talk message. Conversely, any message space can be augmented with a second dimension to obtain a message space satisfying M-CT, which represents that the sender has access to hard information as described by M and can simultaneously send cheap talk.

Lipnowski and Ravid (2020) showed that useful equilibria may be lost if the receiver always breaks ties in favor of the sender in cheap talk games. Hence, in this section, we treat v as an upper hemicontinuous correspondence that is compact and convex-valued, and which returns all possible sender payoffs when R best-responds to a belief μ.[9] Condition 4 in Definition 1 of a coalition then becomes that $w \in v(\mu(\cdot \mid m))$ for each $m \in X$. We denote $\overline{v}(\mu) = \max(v(\mu))$ and $\check{v}(\mu) = \min(v(\mu))$ for each μ.

Our third existence condition requires that M is complete and satisfies the cheap talk property, with *no further conditions on* v.

Theorem 3. *Suppose that*

- v *is an upper hemicontinuous, compact and convex-valued correspondence;*
- M *satisfies M-C and M-CT.*

Then, there exists a coalition-proof PBE.

The gist of the proof is simple: for any posterior μ that may result from the use of verifiable messages, using cheap talk allows the sender to obtain not just $v(\mu)$ but the value of the quasiconcave closure of v at μ (Lipnowski and Ravid, 2020). Since the quasiconcave closure of v is itself quasiconcave, Theorem 1 applies.

[9]Again, these properties of v follow under mild assumptions on u_S, u_R and A (Lemma 3).

Our fourth existence condition is that M satisfies the cheap talk property and full revelation is (potentially) bad enough for the sender:

Theorem 4. *Suppose that*

- v *is an upper hemicontinuous, compact and convex-valued correspondence;*
- M *satisfies M-CT;*
- $\check{v}(\mu^0_{\{\theta\}}) = \min_{\mu \in \Delta\Theta} \check{v}(\mu)$ *for all* $\theta \in \Theta$.

Then, there exists a coalition-proof PBE.

The idea behind Theorem 4 is as follows. Recall that Algorithm 3 only halts at step t when the set $W_t \cap [\max_{\theta \in \Theta_t} \underline{v}(\theta), w_{t-1}]$ is empty, in particular, if $\max W_t > w_{t-1}$ for some t. Under the conditions of Theorem 4, there is a way to continuously "degrade" a coalition that pays $\max W_t$ by adding cheap talk messages that bring the receiver's posterior closer to full revelation. We can calibrate this leakage of information to produce a coalition that pays exactly w_{t-1}, which must be a valid choice in Algorithm 3.[10]

We finish this section with two observations. First, note that Theorem 2 provides sufficient conditions for existence that *only* constrain the payoff function v, while Theorem 3 provides sufficient conditions that *only* constrain the message mapping M. Second is a comment about the "tightness" of our results: continuity and quasiconcavity of v, evidence structure on M (Hart, Kremer, and Perry, 2017) and availability of cheap talk (which makes no difference if v is quasiconcave) are not jointly sufficient for existence of a coalition-proof PBE. Example 3 illustrates this point.[11]

---

[10] The same logic would apply if some messages were costlier than others, and S could voluntarily increase the cost of a message—a form of burning money.

[11] We can modify Example 3 to add a revealing message for type 1 and a payoff $v(1, 0, 0)$ so that M has evidence structure and v is quasiconcave, and there are no coalition-proof PBE.

5 Rich Message Spaces

This section provides an explicit characterization of coalition-proof PBE in a special case where the message space is maximally "rich". In particular, we not only require there to be a message that allows any set of types to pool together and separate from others (as in Grossman (1981) and Milgrom (1981), where any subset $\theta \in C \subseteq \Theta$ is a valid message) but also that *fractions* of each type can pool, while excluding otherwise identical copies of themselves.[12]

Formally, we begin with a finite type space Θ and an upper semicontinuous sender payoff function $v(\mu)$, and consider an augmented space $\Omega := \Theta \times [0, 1]$. We assume that, for each nonzero vector $(p_1, \ldots, p_n) \in [0, 1]^n$, there is a message m available precisely to all types (θ_i, j) with $j \leq p_i$ (hence, "a fraction p_i of senders of type θ_i").

Note that messages with $(p_1, \ldots, p_n) \in \{0, 1\}^n$ correspond to traditional Grossman-Milgrom style messages that allow any subset of Θ to pool together and exclude the rest; however, our setup allows even richer messages. The second component j of the sender's type is payoff-irrelevant for both the sender and the receiver. However, types (θ_i, j) with lower j may attain higher payoffs because they have access to more messages.

Since only the first dimension of the sender's type matters to the receiver, it is useful to denote a distribution over sender types by $\overline{\mu} \in \Delta\Omega$, and let $\mu \in \Delta\Theta$ be the marginal distribution of θ given $\overline{\mu}$. For any subset $C \subseteq \Theta$, we denote by μ^{*C} the argmax of $v(\mu)$ subject to the constraint supp $\mu \subseteq C$.

We proceed to characterize the coalition-proof PBE of this game. While our previous existence results (Theorem 3) do not apply to this game (because the type space Ω is infinite), a coalition-proof PBE does exist, and its payoffs can be tightly

[12]This type of flexibility would add nothing if $v(\mu)$ depended only on $E_\mu(\theta)$, as in Milgrom (1981), or more generally if v satisfied B*.

characterized under a genericity assumption.

Proposition 3.

(i) There exists a coalition-proof PBE.

(ii) Algorithm 3 never halts. Moreover, if restricted to choosing coalitions of maximal size, it ***always*** *terminates in at most n steps.*

(iii) In any coalition-proof PBE written as a finite partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$, every coalition obtains payoff $w_t = v(\mu^{*C})$ for $C = \text{supp}\ (\mu^0_{\Omega_t}) \subseteq \Theta$.

(iv) If v is generic (μ^{*C} is a singleton for every $C \subseteq \Theta$), then all coalition-proof PBEs are payoff-equivalent.

Parts (i) and (ii) are simply existence results. Part (iii) describes the structure of coalition-proof PBEs. It establishes that every sender type (θ, j) must receive a payoff of the form $v(\mu)$ for $\mu \in \bigcup_{C \subseteq \Theta} \mu^{*C}$, and furthermore pins down the relevant C for each type: a type (θ, j) who is part of a coalition $(C_t, X_t, \sigma_t, w_t)$ must have a payoff $w_t = v(\mu^{*C})$, where C is the set of $\theta \in \Theta$ that are represented in Ω_t. The intuitive reason is that, if there are still sender types left for each $\theta \in C$, there will be fractions of them able to pool in the appropriate way to induce an optimal posterior $\mu \in \mu^{*C}$, and they must do so in equilibrium, as otherwise a coalitional deviation would be available. Finally, part (iv) gives us uniqueness: there is effectively a unique coalition-proof PBE if there is a unique way to maximize $v(\mu)|_{\mu:\text{supp}\ \mu \subseteq C}$ for each C, and this happens for "almost all" v.

Moreover, Proposition 3.(iii) and its proof effectively provide a recipe for constructing coalition-proof PBEs, which we sketch here. An illustration with three types is given in Figure 2.

First, find $\mu^{*1} \in \mu^{*\Theta} = \arg\max_{\mu \in \Delta\Theta} v(\mu)$ and construct m_1 such that $\mu^0_{M^{-1}(\{m\})} = \mu^{*1}$, i.e., by setting $p_i = \lambda \frac{\mu^{*1}_i}{\mu^0_i}$ for all i, for some $\lambda > 0$. Choose the highest fea-

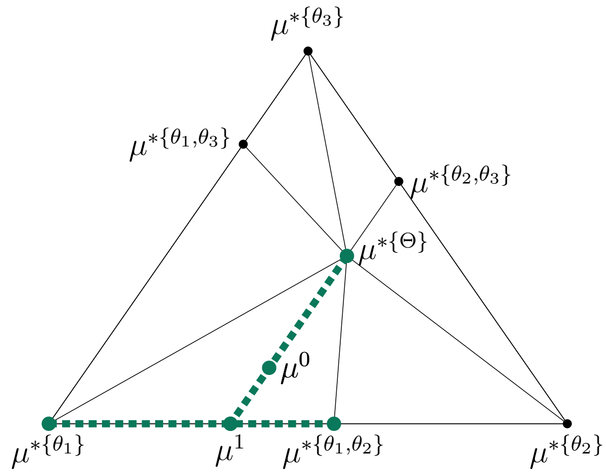


Figure 2. *Posterior beliefs in a coalition-proof PBE of a game with a rich message space.*

sible λ so that $p_i \leq 1$ for all i, i.e., $\lambda = \min_j \frac{\mu_j^0}{\mu_j^{*1}}$. The first coalition is then $(M^{-1}(\{m_1\}), \{m_1\}, \cdot, v(\mu^{*1}))$, and $\Omega_1 = \Omega - (M^{-1}(\{m_1\})$. Note that, since λ was chosen to be maximal, we must have $p_i = 1$ for some i, so there is at least one $\theta \in \Theta$ for whom **all** types (θ, j) are in coalition 1 and get $v(\mu^{*1})$. To simplify exposition, suppose this holds for a single θ_{i_1}, and without loss suppose $i_1 = n$. Let $\Theta_2 = \{\theta_1, \ldots, \theta_{n-1}\}$.

Let $\overline{\mu}^1 = \overline{\mu}^0_{\Omega_1}$ be the receiver's posterior (over types $(\theta, j) \in \Omega$) conditional on not receiving message m_1. Note that, by construction, $\mu^1 = \mu^0_{\Omega_1} \propto \mu^0 - \lambda\mu^{*1} \in \Delta\Theta_2$. We repeat the same construction in this game, with $\mu^{*2} \in \arg\max_{\mu \in \Delta\Theta_2} v(\mu)$: since type θ_n was fully removed from the game in stage 1, the best the sender types left in stage 2 can do is induce the posterior μ^{*2}.

We repeat until all types are assigned to a coalition. Thus, if type $n-1$ is fully removed in stage 2, then $\Theta_3 = \{\theta_1, \ldots, \theta_{n-2}\}$ and $\mu^{*3} \in \arg\max_{\mu \in \Delta\Theta_3} v(\mu)$, and so on. The algorithm terminates in at most $|\Theta|$ steps. Proposition 3 shows that the equilibrium thus obtained is effectively unique if the arg max at each step is unique, i.e., if v has a unique maximizer when restricted to each sub-domain ΔC for $C \subseteq \Theta$.

Existing literature has characterized the (ex ante) sender payoffs attainable under Bayesian persuasion (Kamenica and Gentzkow, 2011) and cheap talk (Lipnowski and Ravid, 2020) with state-independent sender payoffs and an arbitrary prior belief μ^0. Under Bayesian persuasion, the sender attains $v^C(\mu^0)$, the value of the *concave closure* of v at the prior, while under cheap talk she attains $v^{QC}(\mu^0)$, as discussed in Theorem 3. An analogous characterization for games of disclosure is not available in the literature, as most models of disclosure impose additional restrictions on the sender's payoff function that preclude any systematic results for general v.

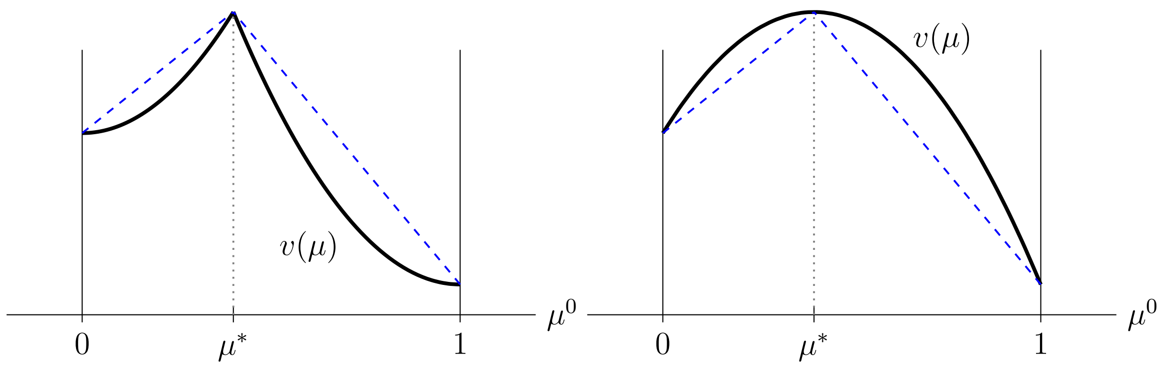
Assuming as in Proposition 3.(iv) that the argmax of $v(\mu)$ subject to supp $\mu \subseteq C$ is unique for all $C \subseteq \Theta$, we now characterize the (unique) ex ante payoff of the sender under any coalition-proof PBE, as a function of the prior μ^0.[13] We show that this expected ex ante payoff, $v^{\text{ea}}(\mu^0)$, admits a geometric characterization, which we call the *tent* of v.

Proposition 4. *Suppose every μ^{*C} is a singleton. For every $\mu \in \Delta\Theta$, there is a unique way to write μ as a convex combination $\mu = \sum_{i=1}^{n} \lambda_i \mu^{*C_i}$ for a collection of sets $C_i \subseteq \Theta$ such that $C_{i+1} \subsetneq C_i$ for each i. Then $v^{\text{ea}}(\mu) = \sum_{i=1}^{n} \lambda_i v(\mu^{*C_i})$.*

The result is illustrated in Figure 3 for the case of two types, θ_1 and θ_2. By an abuse of notation, we denote $\Pr(\theta = \theta_2)$ by μ. On both sides of the figure, v is single-peaked with a peak at μ^*. The nature of coalition-proof PBE requires that at least one of these types must attain the optimal payoff $v(\mu^*)$ with probability 1. Indeed, were this not the case, a group of θ_1-senders and θ_2-senders in the correct proportion could deviate to a message giving them payoff $v(\mu^*)$.

If $\mu^0 = \mu^*$, then in the coalition-proof equilibrium, *both* types obtain the optimal payoff with probability 1, so $v^{\text{ea}}(\mu^*) = v(\mu^*)$. If $\mu^0 > \mu^*$, so that type θ_2 is more

[13] While the sender's ex ante payoff is less relevant in this setting, as decisions over messages are made ex post and sender types are not all on the same footing, it serves as a useful benchmark relative to the literature on Bayesian persuasion and cheap talk.



(a) *Sender does as well in CP-PBE as in BP; better than cheap talk.* ***(b)*** *Sender does worse in CP-PBE than in BP and cheap talk.*

Figure 3. *Coalition-proof PBE with two types of Sender and rich message space.*

numerous than needed to produce the optimal posterior belief μ^*, the coalition-proof outcome will have *all* type θ_1-senders pooling with *some* type θ_2-senders, in the correct proportion to induce the posterior μ^*. The leftover θ_2 types are forced to receive the payoff $v(1)$, as their type is effectively revealed. If μ^0 is close to 1, then almost all type θ_2-senders are unable to pool with θ_1 types, so $v^{\text{ea}}(\mu^0)$ is close to $v(1)$. As μ^0 increases from μ^* to 1, $v^{\text{ea}}(\mu^0)$ decreases linearly from $v(\mu^*)$ to $v(1)$, as in fact $v^{\text{ea}}(\mu^0) = v(\mu^*)\left(1 - \mu^0 + \frac{\mu^*}{1-\mu^*}(1-\mu^0)\right) + v(1)\left(\mu^0 - \frac{\mu^*}{1-\mu^*}(1-\mu^0)\right)$, which is linear in μ^0. By similar logic, for $\mu^0 < \mu^*$, there are "excess" θ_1 types, so all θ_2 types get to pool with some θ_1 types and obtain the optimal payoff, while the leftover θ_1 types get $v(0)$. $v^{\text{ea}}(\mu^0)$ varies linearly between $v(0)$ and $v(\mu^*)$, creating the "tent" shape seen in Figure 3.

It is clear from this argument that payoffs in a coalition-proof PBE depend *only* on the "peaks" of the payoff function v, and the prior μ^0; hence, as seen on the right side of Figure 3, changing the payoff function to make it concave leaves payoffs unaffected if the peaks (i.e., $v(0)$, $v(\mu^*)$, $v(1)$, and the fact that $\max v = v(\mu^*)$) are unchanged. Note that, when v is convex on either side of the peak, as in the left side of Figure 3, v^{ea} coincides with the concave closure of v, so verifiable disclosure allows the Sender to do (ex ante) as well as when she has full commitment power, and better than cheap talk (as cheap talk would be uninformative). On the other hand, as seen on the right hand of the figure, the sender's payoff in a coalition-proof PBE can be ex ante *lower* than cheap talk or no communication at all. The reason is that, because there is no "loyalty" across sender types (as the sender seeks an optimal payoff conditional on knowing her type), a sender strategy may be played that allows some additional sender types to attain the optimal payoff $v(\mu^*)$, but at the cost of relegating other sender types to a very low payoff $v(1)$, which drags down the average.

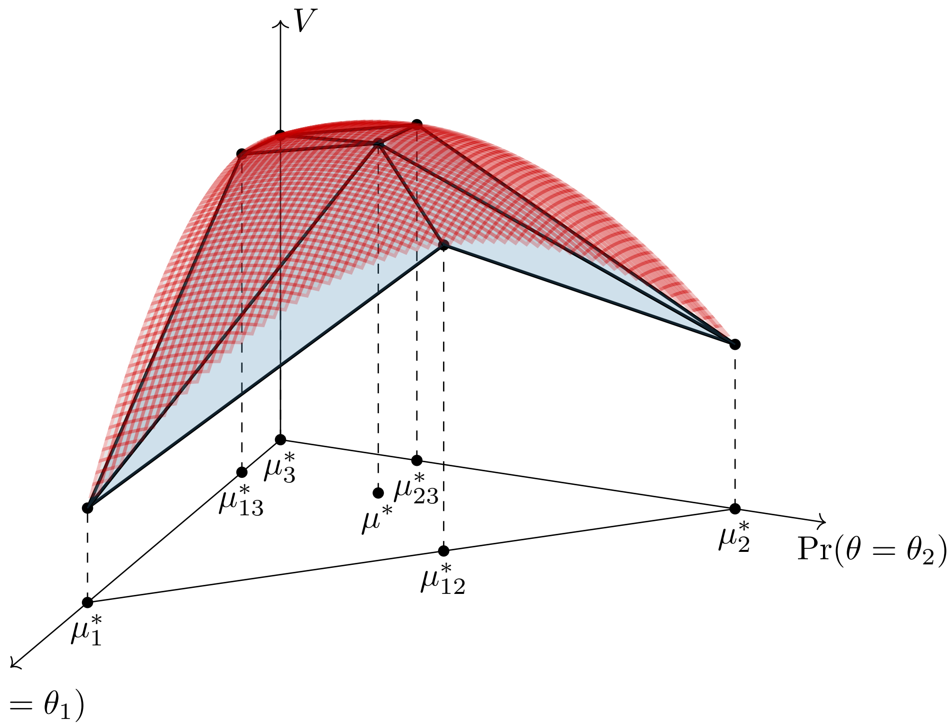


Figure 4. *Tent of v with three types. We denote $\mu^{*\{\theta_i,\theta_j\}} = \mu_{ij}^*$ and $\mu^{*\{\theta_i\}} = \mu_i^*$.*

Figure 4 further illustrates how the tent of v is constructed in the same three-type example from Figure 2. In the example, there is an (interior) optimal belief μ^*; three beliefs $\mu^{*\{\theta_1,\theta_2\}}$, $\mu^{*\{\theta_1,\theta_3\}}$, $\mu^{*\{\theta_2,\theta_3\}}$ that are optimal constrained to supp $\mu \subseteq \{\theta_1, \theta_2\}$, $\{\theta_1, \theta_3\}$, $\{\theta_2, \theta_3\}$ respectively; and the corner beliefs $\mu^{*\{\theta_1\}}, \mu^{*\{\theta_2\}}, \mu^{*\{\theta_3\}}$. These 7 beliefs ($2^n - 1$ in the general case) partition the simplex into 6 ($n!$) triangles (sub-simplices). Each triangle has as vertices μ^*; one of $\mu^{*\{\theta_1,\theta_2\}}$, $\mu^{*\{\theta_1,\theta_3\}}$, $\mu^{*\{\theta_2,\theta_3\}}$; and one compatible corner belief. On each of these triangles, v^{ea} is affine and equal to v

on the vertices of the triangle. Thus the graph of v^{ea} is made up of 6 triangles joined to each other along line segments that grow from μ^*.

6 Comparison to Existing Literature

Hart, Kremer, and Perry (2017)

Hart, Kremer, and Perry (2017) (HKP) study truth-leaning equilibria for games of disclosure with evidence structure. Their evidence structure translates into the following assumptions on the message mapping.

Definition 6. The message mapping $M : \Theta \to \Theta$ has evidence structure if it satisfies

- $\theta \in M(\theta)$ (**reflexivity**);
- if $\theta_j \in M(\theta_i)$ and $\theta_k \in M(\theta_j)$, then $\theta_k \in M(\theta_i)$ (**transitivity**).

HKP's truth-leaning is an equilibrium refinement which requires that (A0) type θ sends message θ with probability 1 if it is weakly optimal to do so, and (P0) when the receiver hears an off-path message θ, he believes it came from type θ.

Definition 7. (σ, μ) is a truth-leaning equilibrium if it is a PBE and

(A0) $\forall \theta \in \Theta$, if $v(\mu(\cdot \mid \theta)) = \max_{m \in M(\theta)} v(\mu(\cdot \mid m))$, then $\sigma(\theta \mid \theta) = 1$;

(P0) $\forall m \in \Theta$, if $\sum_{\theta \in \Theta} \mu^0(\theta)\sigma(m \mid \theta) = 0$, then $\mu(\cdot \mid m) = \mu^0_{\{m\}}$.

HKP find that if v satisfies betweenness and M satisfies $A0$ and $P0$, then there is a unique truth-leaning equilibrium outcome, which coincides with the unique receiver commitment outcome (see their Theorem 1). In particular, the truth-leaning equilibrium is receiver-optimal. Although coalition proofness has little to do *a priori* with receiver optimality, we find that the (receiver-optimal) truth-leaning equilibrium is also coalition-proof in HKP's setting, and in fact it is essentially the *only* thing coalition proofness can select in most cases that HKP's model is concerned with—namely, whenever v satisfies *strict* betweenness.[14]

Proposition 5. *If v satisfies B and M has evidence structure, then the truth-leaning equilibrium is coalition-proof. Moreover, if B^* is also satisfied, then every coalition-proof PBE is payoff-equivalent to the truth-leaning equilibrium.*

However, note that the conceptual connection between both concepts breaks down if betweenness is not satisfied. The intuition is as follows: receiver-optimal equilibria generally involve as much revelation as possible, and thus high separation between sender types. When v satisfies betweenness, coalition proofness also leads to high separation, because "high" types want to separate from "low" types if they can, and coalitional deviations enable them to do so. However, for more general v, coalitional deviations may instead be used to move towards greater pooling.

Bertomeu and Cianciaruso (2018)

Our notion of coalition-proof PBE generalizes Bertomeu and Cianciaruso (2018)'s Grossman-Perry-Farrell equilibrium (GPFE). Bertomeu and Cianciaruso (2018) also provide an algorithmic characterization of GPFE that is analogous to our Algorithm 3.[15] The main advantage of our solution concept over theirs is that we allow for mixed strategies. In contrast, GPFE only allows coalitions of the form $(M^{-1}(\{m\}), \{m\}, \cdot, \mu^0_{M^{-1}(\{m\})})$. Such a definition of coalitions cannot accommodate useful cheap talk, nor the kind of mixing that often occurs in HKP's truth-leaning

[14] If v satisfies B but not B*, coalition-proof PBEs that are not equivalent to the truth-leaning one may exist. To see why, suppose R's action is binary; R takes the high action if she believes S's type is "high" enough; and all types can reveal themselves. Then truth-leaning leads to full revelation, but there may be coalition-proof PBEs where some "low" sender types pool with high ones.

[15] Note however that, while their algorithm yields a GPFE when there is one, it can terminate and yield a non-GPFE strategy when no GPFE exists, in contrast to our Proposition 2.

equilibria. As a result, they require stronger conditions than ours to guarantee existence. (Note for example that, in their Example 7, a GPFE does not exist, but there is a coalition-proof PBE.) Their Proposition 1 shows that a GPFE exists under B + M-C; in contrast, coalition-proof PBE exists under B (Theorem 2) or under QC + M-C (Theorem 1), as well as other conditions. GPFE is also not guaranteed to exist under the conditions of our Theorem 3 or Theorem 4. Although the concept of GPFE would be sufficient for our example with maximally flexible messages (Proposition 3 and Proposition 4), Bertomeu and Cianciaruso (2018) do not perform this exercise.

Koessler and Skreta (2023)

Koessler and Skreta (2023) model information design by an informed designer, in a way that resembles a disclosure problem despite using a different modeling technology. Their setting is more general in allowing state-dependent sender preferences, but more restrictive in that there is no analog of an arbitrary message space M; their information technology is effectively analogous to having a maximally flexible M under disclosure. The closest overlap between their paper and ours is thus between their Section V (state-independent preferences) and our Section 5 (flexible M).

Even here, our models yield related but distinct solutions. To see why, some detail is needed. In Koessler and Skreta (2023), an equilibrium is summarized by a mechanism $\gamma : \Theta \to \Delta A$ that all sender types commit to in equilibrium *after learning their type*. A mechanism is "interim optimal" (IO) if it is incentive-compatible for the receiver given the induced beliefs, and there are no credible interim deviations by the sender. With state-independent preferences, a credible deviation boils down to a belief $\mu \in \Delta\Theta$ such that $v(\mu)$ is strictly greater than θ's equilibrium payoff for all

$\theta \in \text{supp}\ \mu$.[16]

Assuming $\mu^* = \arg\max_\mu v(\mu)$ is unique, it can be shown that all IO mechanisms coincide with the coalition-proof PBE of Proposition 3 "up to the first coalition". But the set of IO mechanisms is generally large, as weaker constraints are imposed on the behavior of "leftover types". Koessler and Skreta (2023) further define an IO* mechanism as one that is ex ante optimal among IO mechanisms; this is generally unique but also not equivalent to coalition-proof PBE.

This divergence stems from the nature of interim knowledge in the two settings. In our model, senders who are excluded on-path from the optimal message (e.g., the "leftover" types forced to reveal their type in Figure 2) are aware of their predicament when they choose whether to deviate. In contrast, in Koessler and Skreta (2023), all senders commit to the same γ on path; a type θ thus commits to a random action $\gamma(\theta) \in \Delta A$. She chooses whether to deviate after learning her type θ, but *before* learning the realization of $\gamma(\theta)$ she would get on-path. Because $\gamma(\theta)$ may give her a chance of sending the optimal message, deviations are less tempting. Indeed, it can be shown that, if μ^* is interior and μ^0 is close enough to it, then IO imposes no constraints whatsoever on the behavior of "leftover" types. That is, in our notation, a strategy σ yields an IO outcome iff it induces the belief μ^* with probability λ (i.e., the probability of message m_1 in Proposition 3.(iii)); the behavior of later coalitions is unrestricted. In particular, IO* selects Bayesian-persuasion behavior for all "leftover" types in this case, rather than coalition-proof behavior.

[16] This is in the spirit of coalition proofness: the set of types willing to deviate, supp μ, is akin to a coalition. But it allows many more deviations: effectively, it allows arbitrary fractions of types who would gain from a deviation to participate. With anything less than a fully flexible M, this notion of credible deviations would generally yield equilibrium non-existence even under the conditions of Theorem 1, Theorem 2, Theorem 3, or Theorem 4. In contrast, our notion of coalitions pins down the belief μ by presuming that *all* types who would benefit *will* participate.

7 Conclusion

We provide an approach to equilibrium selection in games of disclosure. While the general principle behind coalition proofness is well understood, our definition is flexible enough to obtain existence results in fairly general classes of disclosure games, thus generalizing existing results on games of disclosure (Bertomeu and Cianciaruso, 2018), cheap talk (Lipnowski and Ravid, 2020) and also clarifying the relationship with receiver-optimal equilibria in disclosure (Hart, Kremer, and Perry, 2017). Our application in Section 5 is, to our knowledge, the first analysis of a disclosure game with general state-independent sender preferences, comparable with the canonical setting in information design (Kamenica and Gentzkow, 2011). Finally, one takeaway from Theorem 3 and Theorem 4 is that adding cheap talk to a disclosure game—a substantively innocuous assumption—can simplify and discipline the analysis rather than complicate it.

While Example 3 shows that some obvious candidates for stronger existence results are false, other useful sets of sufficient conditions for existence may have escaped our attention—for example, ones involving message mappings with evidence structure. Furthermore, there is room for more work in not just showing existence but also further characterizing the structure of coalition-proof PBE under various conditions, in the spirit of Proposition 3, Proposition 4 or Proposition 5.

In our view, communication problems, even ones with some hard information available, are not modeled as problems of disclosure as often as they ought to, due simply to the poorer theoretical properties of the disclosure framework relative to cheap talk and Bayesian persuasion. We hope to reduce this gap.

References

Aybas, Yunus C and Steven Callander (2024), "Cheap Talk in Complex Environments". (p. 4.)

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019), "Mechanisms With Evidence: Commitment and Robustness", en, *Econometrica*, 87, 2, pp. 529-566. (p. 2.)

Bertomeu, Jeremy and Davide Cianciaruso (2018), "Verifiable Disclosure", *Economic Theory*, 65, 4, pp. 1011-1044. (pp. 3, 4, 12, 29, 30, 32.)

Callander, Steven, Nicolas S Lambert, and Niko Matouschek (2021), "The Power of Referential Advice", *Journal of Political Economy*, 129, 11, pp. 3073-3140. (p. 4.)

Farina, Agata, Guillaume Fr'echette, Alessandro Lizzeri, and Jacopo Perego (2024), "The Selective Disclosure of Evidence: an Experiment". (p. 4.)

Farrell, Joseph (1993), "Meaning and Credibility in Cheap-Talk Games", *Games and Economic Behavior*, 5, 4 (Oct. 1993), pp. 514-531. (pp. 2, 12.)

Glazer, Jacob and Ariel Rubinstein (2004), "On Optimal Rules of Persuasion", *Econometrica*, 72, 6, pp. 1715-1736. (p. 2.)

Grossman, Sanford J (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality", *The Journal of Law and Economics*, 24, 3, pp. 461-483. (pp. 2, 6, 22.)

Hagenbach, Jeanne, Frédéric Koessler, and Eduardo Perez-Richet (2014), "Certifiable Pre-Play Communication: Full Disclosure", *Econometrica*, 82, 3, pp. 1093-1131. (p. 2.)

Hart, Sergiu, Ilan Kremer, and Motty Perry (2017), "Evidence Games: Truth and Commitment", *American Economic Review*, 107, 3, pp. 690-713. (pp. 2-4, 18, 21, 28, 32.)

Hunt, Brian R, Tim Sauer, and James A Yorke (1992), "Prevalence: a Translation-Invariant "Almost Every" on Infinite-Dimensional Spaces", *Bulletin of the American Mathematical Society*, 27, 2, pp. 217-238. (pp. 39, 48.)

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian Persuasion", *American Economic Review*, 101, 6, pp. 2590-2615. (pp. 3, 25, 32.)

Koessler, Frédéric and Vasiliki Skreta (2023), "Informed Information Design", *Journal of Political Economy*, 131, 11, pp. 3186-3232. (pp. 4, 30, 31.)

Lipnowski, Elliot and Doron Ravid (2020), "Cheap Talk with Transparent Motives", *Econometrica*, 88, 4, pp. 1631-1660. (pp. 3, 20, 25, 32, 44.)

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green, et al. (1995), *Microeconomic Theory*, Oxford University Press, New York, vol. 1. (p. 48.)

Matthews, Steven A, Masahiro Okuno-Fujiwara, and Andrew Postlewaite (1991), "Refining Cheap-talk Equilibria", *Journal of Economic Theory*, 55, 2, pp. 247-273. (pp. 2, 12.)

Milgrom, Paul R (1981), "Good News and Bad News: Representation Theorems and Applications", *The Bell Journal of Economics*, pp. 380-391. (pp. 2, 22.)

Rappoport, Daniel (2022), "Evidence and Skepticism in Verifiable Disclosure Games", *Working paper*. (pp. 2, 4.)

Sher, Itai (2011), "Credibility and Determinism in a Game of Persuasion", *Games and Economic Behavior*, 71, 2, pp. 409-419. (p. 2.)

— (2014), "Persuasion and Dynamic Communication", *Theoretical Economics*, 9, 1, pp. 99-136. (p. 2.)

Wu, Wenhao (2022), "A Role for Cheap Talk in Disclosure". (p. 4.)

A Proofs

Lemma 3. *Suppose there is a metric on A under which A is compact and $u_S(a)$, $u_R(a,\theta_i)$ for each i are continuous in a. Let $a^*(\mu) = \arg\max_{a\in A} \sum_{i=1}^{n} \mu_i u_R(a,\theta_i)$ and $v(\mu) = \{E_\rho(u_S(a)) : \rho \in \Delta a^*(\mu)\}$. Then, v is upper hemicontinuous, compact and convex-valued. Furthermore, $\overline{v}(\mu)$ is upper semicontinuous.*

Proof. By Berge's maximum theorem, $\mu \mapsto a^*(\mu)$ is upper hemicontinuous, nonempty and compact-valued. It follows that $\mu \mapsto u_S(a^*(\mu))$ inherits these properties as well since u_S is continuous. Since $v(\mu) = \mathrm{Conv}(u_S(a^*(\mu)))$ for each μ (as R is always willing to mix over best responses), v is also upper hemicontinuous, nonempty and compact-valued, and also convex-valued.

For the semicontinuity of $\overline{v}$, suppose for the sake of contradiction that $\overline{v}(\mu) < \limsup_{n\to\infty} \overline{v}(\mu^n)$ for some sequence $\mu^n \to \mu$. By taking a subsequence, we can assume without loss that $\overline{v}(\mu^n)$ converges to $\limsup_{n\to\infty} \overline{v}(\mu^n)$. Next, let $\hat{a}(\mu) = \arg\max_{a\in a^*(\mu)} u_S(a)$. Since A is a compact metric space, there is a subsequence $(\mu^{n_k})_k$ along which $\hat{a}(\mu^{n_k})$ converges to some $a^* \in A$. By construction, $a^* \in a^*(\mu)$, and $u_S(a^*) = \limsup_{n\to\infty} \overline{v}(\mu^n) > \overline{v}(\mu)$, a contradiction. □

Proof of Proposition 1

($\Longrightarrow$): let (σ,μ) be a PBE. Given the PBE (σ,μ), let $w_1 = \max_{\theta\in\Theta} \max_{m\in M(\theta)} v(\mu(\cdot \mid m))$ be the highest equilibrium payoff across all sender types. This maximum exists by (PBE-1) and because Θ is finite. Let $X_1 := \{m \in \mathcal{M} \mid v(\mu(\cdot \mid m)) = w_1 \text{ and } \sum_{\theta\in\Theta} \sigma(m \mid \theta) > 0\}$ be the set of on-path messages that obtain that payoff and $C_1 := \{\theta \in \Theta \mid \sigma(m \mid \theta) > 0 \text{ for some } m \in X_1\}$ be the set of types getting that payoff. Now, for each $\theta \in C_1$, $\operatorname{supp} \sigma(\cdot \mid \theta) \subseteq X_1 \cap M(\theta)$ by (PBE-1). Also, $M^{-1}(X_1) = C_1$ and $\sum_{\theta\in C_1} \operatorname{supp} \sigma(\cdot \mid \theta) = X_1$ as every type with access to messages in X_1 (and payoff w_1) would be sending these messages in equilibrium. Thus, $(C_1, X_1, \sigma|_{C_1}, w_1)$ is a coalition.

Next, consider the restricted game with type space $\Theta_2 := \Theta \setminus C_1$. Let $w_2 = \max_{\theta\in\Theta_2} \max_{m\in M(\theta)} v(\mu(\cdot \mid m))$ be the highest equilibrium payoff across all types in Θ_2 and X_2 (C_2) be the set of messages (types) getting that payoff. Then, $(C_2, X_2, \sigma|_{C_2}, w_2)$ is a coalition of the restricted game with type space Θ_2. Proceed in a similar fashion to obtain a partition $\{(C_t, X_t, \sigma|_{C_t}, w_t)\}$, where w_t is strictly decreasing by construction. The partition is individually rational or else there exists a type with a profitable deviation to an off-path message.

($\Longleftarrow$): let σ be the strategy associated with an individually rational partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^{T}$ such that $w_1 \geq \ldots \geq w_T$. Let R's off-path beliefs be skeptical for all off-path messages, meaning that $\forall m \in \mathcal{M} \setminus \text{supp}\sigma$, $\mu(\cdot \mid m) = \arg \min_{\mu(\cdot|m) \text{ feasible}} v(\mu(\cdot|m))$. Now, $\forall t$, every type $\theta \in C_t$ does not have profitable deviations to any on-path messages (she does not have access to messages in coalitions prior to t that obtain a higher payoff and coalitions after t receive a lower payoff) or to off-path messages (by individual rationality). Therefore, σ is a PBE strategy. $\square$

Proof of Proposition 2

($\Longleftarrow$) Let $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^{T}$ be a greedy partition. Because any greedy partition is also a feasible output of Algorithm 2, σ is a PBE strategy by Proposition 1.

It remains to show the coalition proofness. Suppose that there is a blocking coalition $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ of σ. Since $w_1 \geq \ldots \geq w_T$, there exists τ such that $w_{\tau-1} \geq \widetilde{w} > w_\tau \geq w_{\tau+1} \geq \ldots \geq w_T$. Then, $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ is a coalition of the restricted game with type space $\bigcup_{t: w_t < \widetilde{w}} C_t = \bigcup_{t \geq \tau} C_t = \Theta_\tau$. Furthermore, $\widetilde{w} \in W_t \cap (w_\tau, w_{\tau-1}]$. That, combined with $\widetilde{w} > w_\tau$, contradicts the choice of coalition at step τ.

($\Longrightarrow$) Suppose that σ is a PBE strategy and coalition-proof. By Proposition 1, the former implies that σ is associated with a partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^{T}$ with $w_1 \geq \ldots \geq w_T$ that satisfies IR.

We prove by induction that, at each step t of Algorithm 3, $(C_t, X_t, \sigma_t, w_t)$ is a feasible choice of coalition. For $t = 1$, note that $(C_1, X_1, \sigma_1, w_1)$ must attain the payoff $w_1 = \sup W_1$ (hence $= \max W_1$), or else a coalition (C, X, σ, w) with $w > w_1$ would be a blocking coalition of σ. At step 1, a coalition can be chosen by Algorithm 3 iff it attains the payoff $\max W_1$, so $(C_1, X_1, \sigma_1, w_1)$ is a feasible choice.

Consider now an arbitrary step τ, and suppose that the algorithm has chosen $(C_t, X_t, \sigma_t, w_t)$ for $t = 1, \ldots, \tau - 1$. Since $(C_\tau, X_\tau, \sigma_\tau, w_\tau)$ is a coalition of the restricted game with type space Θ_τ and $w_\tau \in [\max_{\theta \in \Theta_\tau} \underline{v}(\theta), w_{\tau-1}]$, this coalition can fail to be a feasible choice at step τ only if $w_\tau < \max(W_t \cap [\max_{\theta \in \Theta_\tau} \underline{v}(\theta), w_{\tau-1}])$, i.e., if there is another coalition (C, X, σ, w) of the restricted game with type space Θ_τ that pays

$w = \max(W_t \cap (-\infty, w_{\tau-1}]) \in (w_\tau, w_{\tau-1}]$. In that case, (C, X, σ, w) is a blocking coalition of σ (since $\Theta_\tau = \bigcup_{t \geq \tau} C_t = \bigcup_{t: w_t < w} C_t$), a contradiction. ☐

Coalition-Optimal Partitions

To prove our existence results, it is useful to define a strengthening of coalition-proof PBE, which we call coalition-optimal equilibrium (COE). A strategy σ is a COE if it is associated with a COE partition obtainable from the following algorithm:

Algorithm 4: COE Partition Algorithm

Let $t := 1$ and $\Theta_1 := \Theta$, $w_0 = \infty$;

while $\Theta_t \neq \varnothing$

  let $W_t = \{w \in \mathbb{R} \mid \exists (C, X, \sigma, w) \in \mathcal{C}(\Theta_t)\}$ be the set of payoffs attainable by coalitions of the restricted game with type space Θ_t;

  let $(C_t, X_t, \sigma_t, w_t) \in \mathcal{C}(\Theta_t)$ be such that $w_t = \max W_t$ and $w_t \leq w_{t-1}$;

  let $\Theta_{t+1} := \Theta_t \setminus C_t$ and $t := t + 1$;

end

In other words, rather than requiring a choice of coalition $(C_t, X_t, \sigma_t, w_t)$ that maximizes w_t *subject to* $w_t \in [\max_{\theta \in \Theta_t} \underline{v}(\theta), w_{t-1}]$, Algorithm 4 requires w_t to be maximized with no constraints, and only admits the choice as valid if it still happens to satisfy $w_t \leq w_{t-1}$. (Note that, as shown next, imposing IR is unnecessary because the output of Algorithm 4 as is automatically satisfies IR.) We now prove that COE is indeed a strengthening of coalition-proof PBE:

Lemma 4. *Every COE is also a coalition-proof PBE. Moreover, a coalition-proof PBE partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ is a COE partition if and only if, for each t, $\max W_t \leq w_{t-1}$.*

Proof. For the first part, we show that any output of Algorithm 4 must satisfy IR, i.e., $w_t \geq \max_{\theta \in \Theta_t} \underline{v}(\theta)$ for all t. Indeed, if at any step t we have that $\max W_t < \max_{\theta \in \Theta_t} \underline{v}(\theta)$, then $\max W_t < \underline{v}(\theta)$ for some θ, so $\max W_t < \min_{\mu(\cdot|m) \text{ feasible}} v(\mu(\cdot|m))$ for some message $m \in M(\theta)$. This is impossible as the trivial coalition with $\widetilde{X} = \{m\}$, $\widetilde{C} = M^{-1}(\{m\})$ is in $\mathcal{C}(\Theta_t)$ and must pay at least $\min_{\mu(\cdot|m) \text{ feasible}} v(\mu(\cdot|m))$. Therefore, for each t, t-th coalition of Algorithm 4 satisfies $w_t = \max W_t$ and $w_t \in [\max_{\theta\in\Theta_t} \underline{v}(\theta), w_{t-1}]$. Hence, that coalition is a feasible choice of coalition at step t of Algorithm 3.

For the second part, if $\max W_t > w_{t-1}$ for some (minimal) t, then an attempt to obtain the partition $\{(C_t, X_t, \sigma_t, w_t\})_{t=1}^T$ as an output of Algorithm 4 would fail at step t, since when $\max W_t > w_{t-1}$, Algorithm 4 and Algorithm 3 require w_t to take different values. (In fact, Algorithm 4 would halt at this step.) On the other hand, if $\max W_t \le w_{t-1}$ for all t, then at each step $\max(W_t \cap [\max_{\theta\in\Theta_t} \underline{v}(\theta), w_{t-1}]) = \max W_t$, so $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ would also be a valid output of Algorithm 4. □

While there are generally fewer COEs than coalition-proof PBEs, it can be easier to show the existence of COE.

Proof of Theorem 1

(ii) We show that, if QC* and M-C hold, then $\max W_t \le w_{t-1}$ at every step of either Algorithm 3 or Algorithm 4, so COE and coalition-proof PBE are equivalent, as are Algorithm 3 and Algorithm 4, and both algorithms always terminate.

Take an incomplete partition $(C_t, X_t, \sigma_t, w_t)_{t=1}^s$ generated by Algorithm 4 with $w_t = \max W_t \le w_{t-1}$ for all $t \le s$, but $\Theta_{s+1} \ne \varnothing$ and $\max W_{s+1} > w_s$. Let $(C_{s+1}, X_{s+1}, \sigma_{s+1}, w_{s+1})$ be an element of $\mathcal{C}_{s+1}$ with $w_{s+1} = \max W_{s+1}$, so $w_s < w_{s+1}$. By Lemma 1, without loss of generality, $X_{s+1} = \{m\}$ for some $m \in \mathcal{M}$, so $w_{s+1} = v(\mu^0_{C_{s+1}})$. Similarly, since $w_s = \max W_s$, without loss of generality, $X_s = \{m''\}$ and $w_s = v(\mu^0_{C_s})$.

By M-C, there is a message m' such that $M^{-1}(\{m'\}) = M^{-1}(X_s) \cup M^{-1}(\{m\}) = C_s \cup C_{s+1} =: \widetilde{C}$.[17] Therefore, $(\widetilde{C}, \{m'\}, \cdot, v(\mu^0_{\widetilde{C}}))$ is a coalition (in which all types in $\widetilde{C}$ send m' with probability one) of the restricted game with type space Θ_s. By QC*,

$$v(\mu^0_{\widetilde{C}}) > \min\{v(\mu^0_{C_s}), v(\mu^0_{C_{s+1}})\} = \min\{w_s, w_{s+1}\} = w_s,$$

which contradicts that $w_s = \max W_s$.

(i) Suppose that QC and M-C hold. We will show that a COE exists, which implies the result by Lemma 4. To do this, we will construct a *maximal* coalition-optimal

[17] By an abuse of notation, we use $M^{-1}(X_t)$ to denote the preimage of X_t through M in the restricted game at stage t rather than in the original game, i.e., $M^{-1}(X_t) \cap \Theta_t$.

partition by executing a modified version of Algorithm 4. At a given stage t, denote by $\overline{\mathcal{C}}_t$ the collection of all coalitions yielding the payoff $\max W_t$. Algorithm 4 simply picks *any* element of $\overline{\mathcal{C}}_t$, as long as $\max W_t \leq w_{t-1}$. In the modified algorithm, we pick a coalition $(C_t, X_t, \sigma_t, \max W_t) \in \overline{\mathcal{C}}_t$ that is also maximal in the sense of set inclusion, i.e., such that, for all $(C'_t, X'_t, \sigma'_t, \max W_t) \in \overline{\mathcal{C}}_t$, either $C_t = C'_t$ or $C_t - C'_t \neq \varnothing$. If the algorithm terminates, we refer to its output as a *maximal COE*, which is of course a COE.

We argue this algorithm always terminates and returns a COE. Suppose for the sake of contradiction that there is an incomplete partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^{s}$ generated by this modified algorithm, with $w_t = \max W_t \leq w_{t-1}$ for all $t \leq s$, but $\Theta_{s+1} \neq \varnothing$ and $\max W_{s+1} > w_s$. Let $(C_{s+1}, \{m\}, \cdot, w_{s+1})$ be an element of $\mathcal{C}(\Theta_{s+1})$ with $w_{s+1} = \max W_{s+1}$, so $w_s < w_{s+1} = v(\mu^0_{C_{s+1}})$. As before, without loss, $X_s = \{m''\}$ and $w_s = v(\mu^0_{C_s})$ as well. Repeat the same argument as in the case (QC*+MC) to construct the coalition $(\widetilde{C}, \{m'\}, \cdot, v(\mu^0_{\widetilde{C}}))$ with $\widetilde{C} = C_s \cup C_{s+1}$. Now, because v is only weakly quasiconcave, we have that

$$v(\mu^0_{\widetilde{C}}) \geq \min\{v(\mu^0_{C_s}), v(\mu^0_{C_{s+1}})\} = \min\{w_s, w_{s+1}\} = w_s.$$

Again, if the inequality holds strictly, the property $w_s = \max W_s$ is violated. But if it holds at equality, then the maximality of $(C_s, \{m''\}, \cdot, w_s)$ is violated, since $(\widetilde{C}, \{m\}, \cdot, v(\mu^0_{\widetilde{C}}))$ pays the same as $(C_s, \{m''\}, \cdot, w_s)$ and strictly contains it, i.e., $\widetilde{C} = C_s \cup C_{s+1} \supsetneq C_s$.

(iii) If $C \mapsto v(\mu^0_C)$ is injective, then there is at most one feasible choice of C_t at each step in the proof of (i). Indeed, the mapping $C \mapsto v(\mu^0_C)$ must have a unique maximum $C^* = M^{-1}(X^*)$ among feasible C at that step, and the only valid coalitions must then be $(C^*, X^*, \cdot, v(\mu^0_{C^*}))$, as well as others with type set C^* if they happen to yield the same payoff (by Lemma 1, higher payoffs are not possible). Since all valid coalitions have type set C^*, pay $v(\mu^0_{C^*})$, and leave the game in the same state in step $t+1$, and this is true at every step, all coalitions resulting from Algorithm 3 are payoff-equivalent.

As for the genericity, there is to our knowledge no measure-theoretic notion of genericity for subsets of the space of quasiconcave functions. (Hunt, Sauer, and Yorke (1992)'s notion of prevalence, used in Proposition 3, is only defined for subsets of vector spaces; the space of quasiconcave functions is not a vector space.)

However, if we endow the space $\mathcal{Q}$ of quasiconcave functions from Δ^{n-1} to $\mathbb{R}$ with the metric induced by $||\cdot||_\infty$, then the set of functions $v \in \mathcal{Q}$ such that $v(x) \neq v(x')$ is clearly open and dense, for any $x \neq x'$.[18] Then the set of functions with $C \mapsto v(\mu^0_C)$ injective is a finite intersection of open and dense sets, hence open and dense. $\square$

Proof of Theorem 2

(ii) We will show that, if B* holds, then when following either Algorithm 3 or Algorithm 4, the property $\max W_t \leq w_{t-1}$ is always satisfied. For the sake of contradiction, suppose that there is an (incomplete) output $(C_t, X_t, \sigma_t, w_t)_{t=1}^s$ of Algorithm 4 such that $\Theta_{s+1} \neq \varnothing$ and $\max W_{s+1} > w_s$. Let $(C_{s+1}, X_{s+1}, \sigma_{s+1}, w_{s+1})$ be an element of $\mathcal{C}_{s+1}$ with $w_{s+1} = \max W_{s+1}$. Then, $w_s < w_{s+1}$. By Lemma 2, this is equivalent to $v(\mu^0_{C_s}) < v(\mu^0_{C_{s+1}})$.

Consider an auxiliary game in which the type space is $C_s \cup C_{s+1}$, the prior is $\mu^0_{C_s \cup C_{s+1}}$ and the message space is $X_s \cup X_{s+1}$ (i.e., the message mapping is as in the original game, except that messages outside of $X_s \cup X_{s+1}$ are unavailable). Suppose that $X_s \cup X_{s+1}$ is finite.[19] Then this auxiliary game has a PBE by standard existence theorems (Fan-Glicksberg). Using Proposition 1, take the first coalition of a PBE partition strategy $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$. By B* and Proposition 1, $\widetilde{w} \geq v(\mu^0_{C_s \cup C_{s+1}})$. Moreover, by B*, $w_s < v(\mu^0_{C_s \cup C_{s+1}}) < w_{s+1}$. Then, $\widetilde{w} > w_s$. But $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ is in fact a feasible coalition in stage s, contradicting that $w_s = \max W_s$.

The same argument applies to outputs of Algorithm 3. In fact, because at each step the property $\max W_t \leq w_{t-1}$ is guaranteed, Lemma 4 implies that the set of COEs and coalition-proof PBEs coincide. Moreover, Algorithm 4 halts only when $W_t > w_{t-1}$ for some t. Thus, Algorithm 4 (and equivalently Algorithm 3) always terminates, as desired.

(i) We assume that v satisfies B, and show that a COE exists, which implies existence of a coalition-proof PBE. The general strategy of the proof will be to choose coalitions at each step of the algorithm in a careful way that ensures $\{w_t\}$ is weakly decreasing.

[18]The openness is obvious. For the density, given $v \in \mathcal{Q}$, we want $v' \in B(v, \epsilon)$ such that $v(x) \neq v(x')$ for arbitrarily small $\epsilon > 0$. If $v(x) \neq v(x')$, take $v' = v$. If $v(x) = v(x') = y$, let $K^+ = \{x \in \Delta^{n-1} : v(x) \geq y\}$ and $K^{++} = \{x \in \Delta^{n-1} : v(x) > y\}$. Find a set K' convex such that $K^{++} \subseteq K' \subseteq K^+$ and $x \in K'$, $x' \notin K'$ or $x \notin K'$, $x' \in K'$, then set $v' = v + \frac{\epsilon}{2}\mathbb{1}_{K'}$. For a choice of K', either $K' = \text{Conv}(K^{++} \cup \{x\})$ or $K' = \text{Conv}(K^{++} \cup \{x'\})$ must work.

[19]If $X_s \cup X_{s+1}$ is infinite, a similar argument goes through: since the type space is finite, we can take at most n messages with each possible preimage in the type space and discard the rest.

A few observations are in order. First, note that the set of payoffs that can be possibly attained by a coalition in the game (and in any restricted game) is finite. The reason is that, given a type set C, any coalition supported on that type set must receive the payoff $v(\mu_C^0)$, and there are at most 2^n type sets that a coalition can be supported on.Label these possible payoffs $y_1 < y_2 < \ldots < y_m$.

Next, we will provide a characterization of the level sets of v that these payoffs exist in. Let $L_y^+(v) = \{\mu : v(\mu) \geq y\}$, $L_y^{++}(v) = \{\mu : v(\mu) > y\}$, $L_y^-(v) = \{\mu : v(\mu) \leq y\}$, $L_y^{--}(v) = \{\mu : v(\mu) < y\}$, and $L_y(v) = \{\mu : v(\mu) = y\}$ be the upper level set, strict upper level set, lower level set, strict lower level set, and level set of v at y, respectively.

Remark 3. For any v satisfying B and any y, $L_y^+(v)$, $L_y^{++}(v)$, $L_y^-(v)$, $L_y^{--}(v)$ $L_y(v)$ are convex sets.

Lemma 5. *The boundaries of $L_y(v)$, that is, the sets $\overline{L_y(v)} \cap \overline{L_y^{++}(v)}$ and $\overline{L_y(v)} \cap \overline{L_y^{--}(v)}$, can be written as $\Delta\Theta \cap H_+$, $\Delta\Theta \cap H_-$ respectively, for some hyperplanes H_+, $H_- \subseteq \mathbb{R}^n$. Moreover, H_+ and H_- do not intersect on the interior of $\Delta\Theta$, unless they coincide.*

Proof. Because $L_y^-(v)$ $L_y^{++}(v)$ are convex and disjoint, and $L_y^{++}(y)$ is open, it follows from the separating hyperplane theorem that there is a hyperplane $H_+ = \{x \in \mathbb{R}^n : \langle x, v_+ \rangle = c_+\}$ such that $\langle x, v_+ \rangle \leq c_+$ for all $x \in L_y^-(v)$, and $\langle x, v_+ \rangle > c_+$ for all $x \in L_y^{++}(v)$. We can define H_- analogously so that $\langle x, v_- \rangle < c_-$ for all $x \in L_y^{--}(y)$ and $\langle x, v_- \rangle \geq c_+$ for all $x \in L_y^+(y)$.

But because $L_y^-(v) \cup L_y^{++}(v) = \Delta\Theta$, we must have $L_y^-(y) = \{x \in \Delta\Theta : \langle x, v_+ \rangle \leq c_+\}$, $L_y^{++}(y) = \{x \in \Delta\Theta : \langle x, v_+ \rangle > c_+\}$. Indeed, any $x \in \Delta\Theta$ such that $\langle x, v_+ \rangle \leq c_+$ must be in $L_y^-(y)$ because if it were not, it would have to be in $L_y^{++}(y)$, implying $\langle x, v_+ \rangle > c_+$, a contradiction. The same argument applies to c_-. It follows that the boundaries of $L_y(v)$ are $\Delta\Theta \cap H_+$ (where it meets $L_y^{++}(v)$) and $\Delta\Theta \cap H_-$ (where it meets $L_y^{--}(v)$).

Finally, we prove that H_+ and H_- either do not intersect in the interior of $\Delta\Theta$, or they coincide. Suppose that there is $x \in \text{int}(\Delta\Theta)$ such that $\langle x, v_+ \rangle = c_+$ and $\langle x, v_- \rangle = c_-$. If H_+ and H_- do not coincide (i.e., v_+ and v_- are not parallel), there is a vector u such that $\langle u, v_+ \rangle > 0 > \langle u, v_- \rangle$ (indeed, if v_+, v_- are not orthogonal, either v_+ or $-v_+$ works; if they are, $v_+ - v_-$ works). Then, for $\epsilon > 0$ small enough, $x + \epsilon u \in \Delta\Theta$ and $\langle x + \epsilon u, v_+ \rangle > c_+$, $c_- > \langle x + \epsilon u, v_- \rangle$, so $x + \epsilon u \in L_y^{++}(v) \cap L_y^{--}(v) = \varnothing$, a contradiction. □

Thus, for each y, $L_y(v)$ is either a hyperplane (when H_+, H_- coincide) or the space between two hyperplanes (if not) intersected with the simplex $\Delta\Theta$. The first case needs no special care.

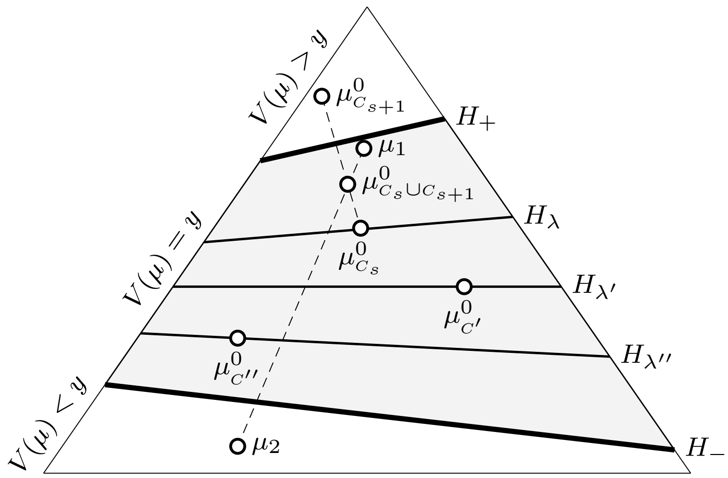


Figure 5. *Choice of λ-maximizing coalition*

For the second case, we provide a transitive and complete preference relation on $L_y(v)$, as follows. The construction is illustrated in Figure 5. For simplicity, suppose $H_+ \cap H_- \cap \Delta\Theta = \varnothing$.[20] Define $H_\lambda = \{x : \langle x, v_\lambda \rangle = c_\lambda\}$, where $v_\lambda = \lambda v_+ + (1-\lambda)v_-$, and $c_\lambda = \lambda c_+ + (1-\lambda)c_-$, for $\lambda \in [0,1]$. Define $\widetilde{H}_\lambda = H_\lambda \cap \Theta$. It is easy to show that $(\widetilde{H}_\lambda)_{\lambda\in[0,1]}$ partitions $L_y(v)$. We then say that, for any $\mu \in \widetilde{H}_\lambda$, $\mu' \in \widetilde{H}_{\lambda'}$, $\mu \succeq \mu'$ iff $\lambda \geq \lambda'$.

Armed with this ordering on each $L_{y_i}(v)$ with nonempty interior, we tweak Algorithm 4 as follows: if the optimal payoff feasible at stage t of the algorithm, $\max W_t$, satisfies $\max W_t \leq w_{t-1}$, and belongs to a level set $L_y(v)$ with nonempty interior, then we pick a coalition $(C_t, X_t, \sigma_t, w_t)$ with $w_t = \max W_t$ such that $\mu^0_{C_t}$ is top-ranked with respect to the preference relation on $L_y(v)$, relative to $\mu^0_{C'_t}$ for all other C'_t that can support a coalition yielding w_t at that stage. (Thus, in Figure 5, $\lambda > \lambda' > \lambda''$, and we pick C_t in stage t rather than C' or C''. Intuitively, in this way we ensure that $\mu^0_{C_t}$ is as close as possible to $L^{++}_y(v)$.) If $L_y(v)$ is just a hyperplane, we simply pick any coalition with $w_t = \max W_t$.

We now retread the argument used in the first case (where B* holds) for why

[20] If H_+ and H_- intersect at the boundary of $\Delta\Theta$, the same argument works if we take all points in $H_+ \cap H_-$ to be in $\widetilde{H}_1$.

$\max W_{t+1} \leq w_t$ must hold. Suppose that B holds, but our modified algorithm yields an incomplete partition $(C_t, X_t, \sigma_t, w_t)_{t=1}^{s}$ such that $\max W_{s+1} > w_s$. Again, define a coalition $(C_{s+1}, X_{s+1}, \sigma_{s+1}, w_{s+1})$ with $w_{s+1} = \max W_{s+1}$ regardless. Again, $w_s < w_{s+1}$, so $v(\mu^0_{C_s}) < v(\mu^0_{C_{s+1}})$. Consider the same auxiliary game with type space $C_s \cup C_{s+1}$, prior $\mu^0_{C_s \cup C_{s+1}}$ and message space $X_s \cup X_{s+1}$. This game has a PBE. Label it as a partition strategy with strictly decreasing payoffs, as in Proposition 1.[21] Label the posteriors generated by each coalition as $\mu_1, \ldots, \mu_k$; their convex hull must contain $\mu^0_{C_s \cup C_{s+1}}$.[22] Now, if $L_{w_s}(v)$ is only a hyperplane, then $\mu^0_{C_s \cup C_{s+1}} \in L^{++}_{w_s}(v)$, because $\mu^0_{C_s \cup C_{s+1}}$ is a convex combination of $\mu^0_{C_s} \in L_{w_s}(v)$ and $\mu^0_{C_{s+1}} \in L^{++}_{w_s}(v)$. Then at least one $\mu_i \in L^{++}_{w_s}(v)$, so the PBE's top coalition (which is a feasible coalition at stage s of the algorithm) pays more than w_s, contradicting that $w_s = \max W_s$. If $L_{w_s}(v)$ has nonempty interior, then either $\mu^0_{C_s \cup C_{s+1}} \in L^{++}_{w_s}(v)$ (in which case the same argument applies) or $\mu^0_{C_s \cup C_{s+1}} \in L_{w_s}(v)$, but $\mu^0_{C_s \cup C_{s+1}} \succ \mu^0_{C_s}$. To see why, suppose that $\mu^0_{C_s} \in H_\lambda$, so $\langle \mu^0_{C_s}, v_\lambda \rangle = c_\lambda$. Note that this implies $\langle \mu^0_{C_s}, v_- \rangle > c_-$ and $\langle \mu^0_{C_s}, v_+ \rangle < c_+$. Since $\mu^0_{C_s \cup C_{s+1}}$ is a convex combination of $\mu^0_{C_s}$ and $\mu^0_{C_{s+1}}$, and $\mu^0_{C_{s+1}}$ is above even H_+ (i.e., $\langle \mu^0_{C_{s+1}}, v_+ \rangle > c_+$), we have $\langle \mu^0_{C_s \cup C_{s+1}}, v_\lambda \rangle > c_\lambda$. But then there must be μ_i which is also above H_λ. And it must be μ_1, the posterior generated by the PBE's top coalition, because all other coalitions pay less than it, hence less than y. The PBE's top coalition thus yields payoff $w_s = \max W_s$ *and* should have been chosen over $(C_s, X_s, \sigma_s, w_s)$ at step s of Algorithm 3 due to being higher-ranked with respect to $\succeq_{L_{w_s}(v)}$, a contradiction.

(iii) The argument is analogous to Theorem 1.(iii). Again, there is no suitable measure-theoretic notion of genericity, but within the set $\mathcal{B}$ of $v : \Delta^{n-1} \to \mathbb{R}$ satisfying B, endowed with the metric induced by $|| \cdot ||_\infty$, the set of all v such that $v(x) \neq v(x')$ is open and dense, for any $x \neq x'$.[23] Then the set of v with $C \mapsto v(\mu^0_C)$ injective is a finite intersection of open and dense sets, hence open and dense. □

[21] Proposition 1 only states that payoffs are weakly decreasing, but we can merge coalitions that yield the same payoff into one, to obtain a partition with strictly decreasing payoffs.

[22] Note that the μ_i are not necessarily posteriors generated by any message; instead, each coalition $(\widetilde{C}_z, \widetilde{X}_z, \widetilde{\sigma}_z, \widetilde{w}_z)$ is mapped to the posterior $\mu^0_{\widetilde{C}_z}$.

[23] Again, the openness is trivial. For the density, it suffices to construct $v' = v + \mathbb{1}_K$, where K can now be a half-plane intersected with Δ^{n-1} that is nested between K^+ and K^{++} and contains exactly one of $\{x, x'\}$.

Proof of Theorem 3

We prove a helpful lemma first. Let $v^{QC}(\cdot)$ be the quasiconcave closure of $\overline{v}$ (Lipnowski and Ravid, 2020), i.e.,

$$v^{QC}(\mu) = \sup_{\mu \in \mathrm{Conv}(\mu^1,\ldots,\mu^n)} \min_{i=1,\ldots,n} \{\overline{v}(\mu^i)\}. \tag{1}$$

Lemma 6. *If $\overline{v}$ is upper semicontinuous, so is v^{QC}. Moreover, the maximum is attained for all μ in (1).*

Proof. We first prove the second claim. Suppose that, for some μ, the maximum is not attained in (1). Then there is a sequence $(\mu^{1t}, \ldots, \mu^{nt})_t$ such that $\mu \in \mathrm{Conv}(\mu^{1t}, \ldots, \mu^{nt})$ for all t and $\min\{\overline{v}(\mu^{1t}), \ldots, \overline{v}(\mu^{nt})\} \to v^{QC}(\mu) =: y$ as $t \to \infty$. Take a subsequence t_m along which $(\mu^{1t_m}, \ldots, \mu^{nt_m})_m \to (\mu^{1\infty}, \ldots, \mu^{n\infty})$. Then $\mu \in \mathrm{Conv}(\{\mu^{1\infty}, \ldots, \mu^{n\infty}\})$, and the upper semicontinuity of $\overline{v}$ implies $\min\{\overline{v}(\mu^{1\infty}), \ldots, \overline{v}(\mu^{n\infty})\} \geq y$. But then $v^{QC}(\mu) \leq \min\{\overline{v}(\mu^{1\infty}), \ldots, \overline{v}(\mu^{n\infty})\}$, a contradiction.

As for the first claim, $\overline{v}$ is upper semicontinuous iff its level sets $\{\mu : \overline{v}(\mu) \geq y\}$ are closed. Because $v^{QC}(\mu)$ can always be written as $\min\{\overline{v}(\mu^1), \ldots, \overline{v}(\mu^n)\}$ for some $\mu^1, \ldots, \mu^n$ that contain μ in its convex hull, the level set $\{\mu : v^{QC}(\mu) \geq y\}$ is simply the convex hull of $\{\mu : \overline{v}(\mu) \geq y\}$, hence also closed. $\square$

Now, as in Theorem 1, we aim to show the existence of a COE, which must be a coalition-proof PBE by Lemma 4. Denote the game by G. Denote by G^{QC} a disclosure game with the same message mapping M as G, but with payoff function v^{QC} instead of payoff correspondence v. We know that G^{QC} has a COE $(C_t, X_t, \sigma_t, w_t)_{t=1}^T$ because M satisfies M-C and v^{QC} is quasiconcave (i.e., satisfies QC), so Theorem 1 applies. We will use this to construct a COE $(C_t, \widetilde{X}_t, \widetilde{\sigma}_t, w_t)_{t=1}^T$ of G that is payoff-equivalent to $(C_t, X_t, \sigma_t, w_t)_{t=1}^T$.

Without loss of generality, we can assume that each coalition $(C_t, X_t, \sigma_t, w_t)$ uses a single message m_t and so can be written as $(C_t, \{m_t\}, \cdot, v^{QC}(\mu^0_{C_t}))$ (Lemma 1). By Lemma 6, take for each t a collection $\mu^1_t, \ldots, \mu^n_t$ whose convex hull contains $\mu^0_{C_t}$ and such that $v^{QC}(\mu^0_{C_t}) = \min(\overline{v}(\mu^1_t), \ldots, \overline{v}(\mu^n_t))$.

We will construct $\widetilde{\mu}^1_t, \ldots, \widetilde{\mu}^n_t$ whose convex hull contains $\mu^0_{C_t}$ and such that

$$v^{QC}(\mu^0_{C_t}) \in v(\widetilde{\mu}^i_t) \text{ for all } i = 1, \ldots, n.$$

If $v^{QC}(\mu^0_{C_t}) \in v(\mu^0_{C_t})$, we are done (take $\check{\mu}^i_t = \mu^0_{C_t}$). Clearly $v^{QC}(\mu^0_{C_t}) < \widetilde{v}(\mu^0_{C_t})$ is impossible as in fact $v^{QC} \geq \overline{v}$. If instead $v^{QC}(\mu^0_{C_t}) > \overline{v}(\mu^0_{C_t})$, then, for each i s.t. $\overline{v}(\mu^i_t) > v^{QC}(\mu^0_{C_t})$, we can choose $\widetilde{\mu}^i_t$ to be a belief on the line segment $[\mu^i_t, \mu^0_{C_t}]$. Any such choice preserves the property that $\mu^0_{C_t}$ is in the convex hull of $\widetilde{\mu}^1_t, \ldots, \widetilde{\mu}^n_t$. And, because v is upper hemicontinuous, and goes from $\overline{v}(\mu^i_t) > v^{QC}(\mu^0_{C_t})$ to $\check{v}(\mu^0_{C_t}) < v^{QC}(\mu^0_{C_t})$ over this line segment, there is an intermediate point $\widetilde{\mu}^i_t$ where $v^{QC}(\mu^0_{C_t}) \in v(\widetilde{\mu}^i_t)$.

Returning to the main proof, we can take $\widetilde{X}_t = \{(m_t, 1), \ldots, (m_t, n)\}$, and $\widetilde{\sigma}_t$ such that message (m_t, i) induces belief $\widetilde{\mu}^i_t$.(Such a message strategy exists by the "fundamental lemma of information design", i.e., it is possible to construct a message strategy to produce posteriors that are any mean-preserving spread of $\mu^0_{C_t}$.)

By construction, all messages (m_t, i) can then induce the payoff $w_t = v^{QC}(\mu^0_{C_t}) \in v(\widetilde{\mu}^i_t)$ in the original game G. A partition thus constructed attains in G the same payoffs that $(C_t, X_t, \sigma_t, w_t)^T_{t=1}$ attains in G^{QC}. Because $v \leq \overline{v} \leq v^{QC}$, any message strategy yields weakly lower payoffs in G than it does in G^{QC}. Hence, if $w_t = \max W_t$ is the maximal payoff attainable by coalitions in $\mathcal{C}(\Theta_t)$ at stage t in G^{QC}, then $\max \widetilde{W}_t$, the analogous maximum in G, must satisfy $\max \widetilde{W}_t \leq \max W_t$. But since $(C_t, \widetilde{X}_t, \widetilde{\sigma}_t, w_t)$ is a coalition at stage t in G, $w_t \in \widetilde{W}_t$, so $w_t = \max \widetilde{W}_t = \max W_t$. Thus $w_t = \max \widetilde{W}_t$ as required by Algorithm 4. Moreover, the partition we constructed inherits from the original the property that w_t is weakly decreasing in t, so $w_t = \max \widetilde{W}_t \leq w_{t-1}$ for all t. Thus $(C_t, \widetilde{X}_t, \widetilde{\sigma}_t, w_t)^T_{t=1}$ is a COE of G. □

Proof of Theorem 4

We will use the following lemma.

Lemma 7. *Consider a coalition $(C, X, \sigma, w) \in \mathcal{C}(\widetilde{\Theta})$ of a restricted game with non-empty type space $\widetilde{\Theta} \subseteq \Theta$. Then, there exists a coalition $(C, X', \sigma', w') \in \mathcal{C}(\widetilde{\Theta})$ for all $w' \in [\min \check{v}, w]$.*

Proof. Label the posteriors induced by the coalition (C, X, σ, w) as $\mu^1, \ldots, \mu^k$. By construction, we have $w \in v(\mu^i)$ for all $i \in \{1, \ldots, k\}$.

We will argue that, for each i and $w' \in [\min \check{v}, w]$, there are beliefs $\mu^i_1, \ldots, \mu^i_n$ such that $\mu^i \in \text{Conv}(\mu^i_1, \ldots, \mu^i_n)$ and $w' \in v(\mu^i_j)$ for all ij. Indeed, any choice of the form $\mu^i_j = \alpha_j \mu^i + (1 - \alpha_j)\mu^0_{\{\theta_j\}}$ for $\alpha_j \in [0, 1]$ $(j = 1, \ldots, n)$ satisfies $\mu^i \in \text{Conv}(\mu^i_1, \ldots, \mu^i_n)$. And, because v is upper hemicontinuous and nonempty, compact and convex valued, and goes from $w \in v(\mu^i)$ to $\min \underline{v} \in v(\mu^0_{\{\theta_j\}})$ over the line segment $[\mu^i, \mu^0_{\{\theta_j\}}]$, there must be α_j such that $w' \in v(\mu^i_j)$. □

Following the argument in the text, there are two cases when executing step t of Algorithm 3. If $\max W_t \leq w_{t-1}$, then there is always a viable coalition, as we can simply choose $w_t = \max W_t$, which is always IR. Moreover, by Lemma 6, $\max W_t$ is always attainable. If $\max W_t > w_{t-1}$, we have that $W_t = [\min \check{v}, \max W_t]$ by the lemma, so $\max(W_t \cap [\min \check{v}, w_{t-1}]) = w_{t-1}$, and of course $w_{t-1} \geq \min \check{v}$. Hence Algorithm 3 can never halt. □

Proof of Proposition 3

(iii) Take a restricted game with type space Ω_t, and write $\Omega_t = \bigcup_{i=1}^n \{\theta_i\} \times A_i$ with $A_i \subseteq [0,1]$. Since $C_s = M^{-1}(X_s)$ for $s < t$ and $M^{-1}(\{m\})$ is always of the form $\bigcup_{i=1}^n \{\theta_i\} \times [0, p_i]$, A_i must be of the form $[0,1]$, $(q_i, 1]$ $(0 \leq q_i < 1)$, or $\varnothing$. Note that $\mu^0_{\Omega_t}(\theta_i) = 0$ iff $A_i = \varnothing$.

Consider the set of payoff-relevant beliefs $\mu \in \Delta\Theta$ that can be induced by a coalition given the type space Ω_t. Clearly $\operatorname{supp} \mu$ must be a subset of $\operatorname{supp}(\mu^0_{\Omega_t})$: if $A_i = \varnothing$ then $\mu(\theta_i \mid m) = 0$ for any on-path message $m \notin X_1 \cup \ldots \cup X_{t-1}$.

But in fact all beliefs with $\operatorname{supp} \mu \subseteq \operatorname{supp}(\mu^0_{\Omega_t})$ are attainable. Indeed, to attain a posterior $\mu = (\mu_1, \ldots, \mu_n)$, we can use a message m accessible to types (θ_i, j) with $j \leq z_i$, where

$$z_i = q_i + \lambda \frac{\mu_i}{\mu^0_i} \tag{2}$$

for each i. We can take $\lambda > 0$ to be any value small enough that $z_i \leq 1$ for all i.

Since all posteriors μ with $\operatorname{supp} \mu \subseteq \operatorname{supp}(\mu^0_{\Omega_t})$ are attainable, and for each of them a coalition can be built using a single message, we have that

$$W_t = \{v(\mu) : \mu \in \Delta \operatorname{supp}(\mu^0_{\Omega_t})\}.$$

Then $\max W_t = v(\mu^{*\operatorname{supp}(\mu^0_{\Omega_t})})$. Since this argument applies to all t, it also yields by induction that $\max W_t \leq w_{t-1}$ for all t. Indeed, $v(\mu^{*\operatorname{supp}(\mu^0_{\Omega_t})})$ is weakly decreasing in t, as $\operatorname{supp}(\mu^0_{\Omega_t})$ must weakly shrink as t increases, $v(\mu^{*C})$ weakly decreases if C shrinks. Then $\max W_t$ is weakly decreasing in t. So $w_1 = \max W_1$, $\max W_2 \leq w_1 \Longrightarrow w_2 = \max W_2$, $\max W_3 \leq w_2 \Longrightarrow w_3 = \max W_3$, and so on. Thus $w_t = v(\mu^{*\operatorname{supp}(\mu^0_{\Omega_t})})$, as desired.

(ii) Lemma 4 and the same argument used in Theorem 2 and Theorem 1 applies here: since we have shown that $\max W_t \leq w_{t-1}$ always holds, Algorithm 3 and Algorithm 4 are equivalent and neither ever halts.

As for the second part, note that, if the t-th coalition $(C_t, X_t, \sigma_t, w_t)$ is such that $\text{supp}\,(\mu^0_{\Omega_{t+1}}) = \text{supp}\,(\mu^0_{\Omega_t})$, then $w_{t+1} = w_t$ by (iii). But then $(C_t, X_t, \sigma_t, w_t)$ and $(C_{t+1}, X_{t+1}, \sigma_{t+1}, w_{t+1})$ can be joined into a single coalition $(C'_t, X'_t, \sigma'_t, w_t)$ with $C_t \subsetneq C'_t$, contradicting the maximality of $(C_t, X_t, \sigma_t, w_t)$. Then $|\text{supp}\,(\mu^0_{\Omega_{t+1}})| \leq |\text{supp}\,(\mu^0_{\Omega_t})| - 1$ for all t; the result follows as $|\text{supp}\,(\mu^0)| = n$.

(i) Follows from (ii): any way of picking (maximal) coalitions through Algorithm 3 terminates and yields a coalition-proof PBE.

(iv) First restrict attention to partitions made with coalitions of maximal size and using a single message if possible. We claim that there is a unique coalition-proof PBE under these restrictions. Indeed, if μ^{*C} is a singleton for every $C \subseteq \Theta$, then the t-th coalition $(C_t, X_t, \sigma_t, w_t)$ in a coalition-proof PBE partition must induce the single belief in $\mu^{*\text{supp}\,(\mu^0_{\Omega_t})}$ with probability 1, by (iii). Then, only messages as constructed in (iii) (Equation (2)) can be used. Denote these messages by $m_t(\lambda)$, indexed by the λ used in Equation (2). To obtain a coalition of maximal size, we must use the maximal λ s.t. $q_i + \lambda \frac{\mu_i}{\mu_i^0} \leq 1\, \forall i$, that is, we must have $X_t = \{m_t(\lambda^*)\}$, with $\lambda^* = \min_{i=1}^n (1-q_i)\frac{\mu_i^0}{\mu_i}$. We can iterate on t to construct a coalition-proof PBE $(C_t, X_t, \sigma_t, w_t)_{t=1}^T$ with $T \leq n$.

Next, we argue by induction that any coalition-proof PBE σ' must be payoff-equivalent to this one. By construction, w_1 is the global maximum of v, so no higher payoff can be obtained under σ'. Let C'_1 be the set of types (θ, j) obtaining payoff w_1 under σ. Because the only way to obtain this payoff is with the unique posterior $\mu^{*\Theta}$, we must have $\mu^0_{C'_1} = \mu^{*\Theta}$. Because all types with access to a message attaining this payoff must use it, and lower types have access to more messages, $C'_1 \cap \{\theta_i\} \times [0,1]$ must be of the form $[0, p_i]$ for all i. Then $p_i = \lambda \frac{\mu_i^{*\Theta}}{\mu_i^0}$ for all i and some fixed λ (Equation (2)). Clearly $\lambda > \min_{i=1}^n \frac{\mu_i^0}{\mu_i^{*\Theta}}$ is impossible as it would imply $p_i > 1$ for some i. If $\lambda < \min_{i=1}^n \frac{\mu_i^0}{\mu_i^{*\Theta}}$, then $(D, \{m_1(\lambda^*)\}, \cdot, w_1)$ with $D = \bigcup_{i=1}^n \{\theta_i\} \times (\,p_i, \lambda^* \frac{\mu_i}{\mu_i^0}\,]$ is a blocking coalition—intuitively, we can pack the remaining types who "should" have been in the coalition C_1 into a new coalition using message $m_1(\lambda^*)$, so σ' is not coalition-proof. If $\lambda = \min_{i=1}^n \frac{\mu_i^0}{\mu_i^{*\Theta}}$ then $C_1 = C'_1$ and σ and σ' are payoff-equivalent up to the first coalition. We can iterate the same argument for all $t \leq T$.

This argument proves the first claim. Finally, we argue that for generic v, μ^{*C}

is indeed a singleton for all $C \subseteq \Theta$. The notion of genericity we use is that of prevalence (Hunt, Sauer, and Yorke (1992)), and we consider the space of functions $V = \{v : \Delta\Theta \to \mathbb{R} \text{ continuous}\}$. (Effectively the same proof works if we instead consider all upper semicontinuous functions with this domain and codomain.) We denote by $V' \subseteq V$ the set of functions v with unique μ^{*C} for all C.

Equivalently, we aim to show that the set of v such that at least one μ^{*C} is not a singleton is shy. By Fact 3' in Hunt, Sauer, and Yorke (1992), it is enough to show that, for *each* $C \subseteq \Theta$, the set of v such that μ^{*C} is not a singleton is shy.

Fix C. Denote by $S \subseteq V$ the set of functions such that $|\mu^{*C}| > 1$. We aim to use as a probe the subspace Z of linear functions, i.e.,

$$Z = \{f_a : a \in \mathbb{R}^n\},$$

where $f_a : \Delta^{n-1} \to \mathbb{R}$ is defined by $f_a(x) \equiv \langle a, x\rangle$. We then define a measure ν as follows if $W \subseteq V$ and $W \cap Z = \{f_a : a \in A\}$, then $\nu(W) = \mathcal{L}(A)$, where $\mathcal{L}$ is the Lebesgue measure in $\mathbb{R}^n$. Then we need to show that, for any $w \in V$, $\nu(S + w) = 0$. In other words, we need to show that the set $a \in \mathbb{R}^n : f_a \in S + w$ has measure zero for all $w \in V$.

For $v \in S$, $v + w \in Z$ iff $v(x) + w(x) \equiv f_a(x)$ for some $a \in \mathbb{R}^n$. Equivalently, $f_a \in (S + w) \cap Z$ iff $v(x) := \langle a, x\rangle - w(x)$ has multiple maxima. In turn, v has multiple maxima if and only if its concave closure v^C does, i.e., iff $\langle a, x\rangle + (-w)^C(x)$ does. Thus, without loss, we can restrict attention to convex w. Moreover, for convex w, $\langle a, x\rangle - w(x)$ has multiple maxima iff the supporting hyperplane H_a of the graph of $-w$ with normal vector a meets the graph of $-w$ at multiple points. Thus, it is enough to show that, for any compact convex set $K \subseteq \mathbb{R}^n$, its supporting hyperplanes H_a satisfy that $K \cap H_a$ is a singleton for almost all $a \in \mathbb{R}^n$.

Let $h_K : \mathbb{R}^n \to \mathbb{R}$ be the *support function* of K, defined as $h_K(a) = \sup_{k\in K}\langle a, k\rangle$. It is well-known that, for $a \neq 0$, h_K is differentiable at a iff the supporting hyperplane H_a meets K at a single point (see Mas-Colell, Whinston, Green, et al. (1995), Proposition 3.F.1). In addition, h_K is always Lipschitz: since an arbitrary supremum of L-Lipschitz functions is L-Lipschitz, and $\langle a, k\rangle$ is $||k||$-Lipschitz as a function of a, $h_K(a)$ is $\sup_{k\in K} ||k||$-Lipschitz.

Hence, it is enough to prove that a Lipschitz function from $\mathbb{R}^n$ to $\mathbb{R}$ is differentiable almost everywhere, which is a special case of Rademacher's theorem.

□

Proof of Proposition 4

First we note that the algorithm in Proposition 3 yields a way to write any prior μ^0 precisely as a convex combination of beliefs with nested support. Then $v^{\text{ea}}(\mu^0) = \sum_{i=1}^{n} \lambda_i v(\mu^{*C_i})$ follows from the fact that, by construction, this is the ex ante payoff the sender obtains in the equilibrium given by Proposition 3.(iii), and that is in fact the only possible payoff (Proposition 3.iv).

It remains to show that there is no other way to write μ as a convex combination of beliefs with nested support. Suppose there is, and denote the belief supports by $\widetilde{C}_1, \widetilde{C}_2, \ldots$. Note that, because $\widetilde{C}_1 \supsetneq \widetilde{C}_2 \supsetneq \ldots$, we must have $\widetilde{C}_1 = \Theta$, and there must be a type θ_i that is *only* in $\widetilde{C}_1$. But then $\mu^{*\widetilde{C}_1}$ is simply μ^{*1}, and there is a unique feasible λ_1 for which type i is eliminated: $\lambda_1 \mu_i^{*1} = \mu_i^0 \implies \lambda_1 = \frac{\mu_i^0}{\mu_i^{*1}}$. Furthermore, this λ is only feasible if i is a type for which $\frac{\mu_j^{*1}}{\mu_j^0}$ is maximized. Thus λ_1 is uniquely pinned down, and hence so is the remaining distribution of types $\mu^0 - \lambda_1 \mu^{*1}$. Then $\widetilde{C}_2 = \text{supp}\,(\mu^0 - \lambda_1 \mu^{*1})$ is also pinned down and coincides with what we obtained in Proposition 3.(iii). Iterating yields the result.

Proof of Proposition 5

For the first part, take a truth-leaning equilibrium σ. Since truth-leaning equilibria are PBE, by Algorithm 2, σ is associated with an IR partition $(C_t, X_t, \sigma_t, w_t)_{t=1}^{T}$ with w_t weakly decreasing.

We argue that any type θ with separating payoff $v(\mu^0_{\{\theta\}}) \geq w_1$ must be truth-telling in a truth-leaning equilibrium, i.e., $\sigma(\theta \mid \theta) = 1$. To see why, suppose $v(\mu(\cdot \mid \theta')) > v(\mu(\cdot \mid \theta))$ for some $\theta' \leq \theta$.[24] Then any type who can send θ can (by evidence structure) and would rather send θ', so θ is off-path. But then R interprets θ as coming from θ (P0), so $v(\mu(\cdot \mid \theta')) > v(\mu(\cdot \mid \theta)) = v(\mu^0_{\{\theta\}}) \geq w_1$, a contradiction, as w_1 is the highest payoff in this equilibrium. Then, since truth-telling is weakly optimal for θ, θ must be truth-telling (A0).

Now suppose σ is not coalition-proof, so there is a blocking coalition $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$. Suppose first that $\widetilde{w} > w_1$, so $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ is simply a coalition of the original game.

Let μ be R's posterior if S is playing σ and R observes the following hypothetical

[24] We write $\theta' \leq \theta$ if $\theta' \in M(\theta)$.

information: R sees that $m \in \widetilde{C}$ but not its realization. Our argument implies that μ has full weight on all types $\theta \in \widetilde{C}$ such that $v(\mu^0_{\{\theta\}}) \geq w_1$. In addition, no types outside of $\widetilde{C}$ can send a message in $\widetilde{C}$: indeed, if $\theta \in \Theta$ can send $m \in \widetilde{C}$, then $\theta \geq m$. Since $\widetilde{C} = M^{-1}(\widetilde{X})$, $m \in \widetilde{C}$ implies that type m can send some $m' \in \widetilde{X}$, so $m \geq m'$. Then $\theta \geq m'$, so $\theta \in \widetilde{C}$.

But then, since μ has full weight on all types $\theta \in \widetilde{C}$ such that $v(\mu^0_{\{\theta\}}) \geq w_1$, *at most* full weight on other members of $\widetilde{C}$, and no weight on any other types, we must have $v(\mu) \geq \widetilde{w} = v(\mu^0_{\widetilde{C}})$ by B. But then, since μ is a linear combination of $\mu(\cdot \mid m)$ for $m \in \widetilde{C}$, we must have $v(\mu(\cdot \mid m)) \geq \widetilde{w} > w_1$ for some $m \in \widetilde{C}$, a contradiction, as w_1 is the highest sender payoff attained under σ.

Next we consider the case where $w_{s-1} \geq \widetilde{w} > w_s$ for $s \geq 2$, so $(\widetilde{C}, \widetilde{X}, \widetilde{\sigma}, \widetilde{w})$ is a coalition of the restricted game with type space Θ_s. The same argument applies: all types θ in Θ_s with $v(\mu^0_{\{\theta\}}) \geq w_s$ must be truth-telling. If R knows $m \in \widetilde{C}$ but not the exact value of m, his posterior μ must have full weight on types $\theta \in \widetilde{C}$ such that $v(\mu^0_{\{\theta\}}) \geq w_s$, at most full weight on other types in $\widetilde{C}$, and no weight on other types.[25] Then $v(\mu) \geq \widetilde{w}$, so some message in $\widetilde{C}$ pays at least $\widetilde{w}$, a contradiction.

For the second part, it is enough to show that, if B* holds and M has evidence structure, then all coalition-proof PBEs are payoff-equivalent (as, by the above argument, the truth-leaning equilibrium is one of them). Since B* implies that coalition-proof PBE is equivalent to COE (Theorem 2), we need to show that all COEs are payoff-equivalent.

Let $\widetilde{\mathcal{C}}_1 = \{C_1 \subseteq \Theta : \exists (C_1, X_1, \sigma_1, w_1) \in \mathcal{C}_1 \text{ with } w_1 = \max W_1\}$ be the collection of all type sets that the first coalition in a COE can be supported on. We first show the following lemma:

Lemma 8. *$\widetilde{\mathcal{C}}_1$ is closed under union and intersection.*

Proof. Let $(C_1, X_1, \sigma_1, w_1)$, $(C_1', X_1', \sigma_1', w_1') \in \widetilde{\mathcal{C}}_1$, so $v(\mu^0_{C_1}) = v(\mu^0_{C_1'}) = \max W_1$. The claim is trivial if $C_1 \subseteq C_1'$ or vice versa, so suppose not. By B*, $v\left(\alpha \mu^0_{C_1} + (1-\alpha)\mu^0_{C_1'}\right) = \max W_1$ for any $\alpha \in (0,1)$. But note that

[25] Here is the only difference from the previous case: we need to check not just that types in $\Theta_s \setminus \widetilde{C}$ don't have access to these messages, but also that no types in $\Theta - \Theta_s$ send them, i.e., that $(X_1 \cup \ldots \cup X_{s-1}) \cap \widetilde{C} = \varnothing$. But if the message $m \in \widetilde{C}$ were in X_i $(i \leq s-1)$, then type m would be in C_i, contradicting $\widetilde{C} \subseteq \Theta_s$.

$\alpha\mu^0_{C_1} + (1-\alpha)\mu^0_{C_1'} = \beta\mu^0_{C_1\cup C_1'} + (1-\beta)\mu^0_{C_1\cap C_1'}$, if we take $\alpha = \frac{\mu^0(C_1)}{\mu^0(C_1)+\mu^0(C_1')}$ and $\beta = \frac{\mu^0(C_1\cup C_1')}{\mu^0(C_1)+\mu^0(C_1')}$. So $v\left(\beta\mu^0_{C_1\cup C_1'} + (1-\beta)\mu^0_{C_1\cap C_1'}\right) = \max W_1$ for some $\beta \in (0,1)$.

Because v satisfies B*, this implies that either $v(\mu^0_{C_1\cup C_1'}) > \max W_1 > v(\mu^0_{C_1\cap C_1'})$, $v(\mu^0_{C_1\cup C_1'}) < \max W_1 < v(\mu^0_{C_1\cap C_1'})$, or $v(\mu^0_{C_1\cup C_1'}) = v(\mu^0_{C_1\cap C_1'}) = \max W_1$. The first two cases lead to a contradiction by a similar argument as in Theorem 2: if $v(\mu^0_{C_1\cup C_1'}) > \max W_1$, then the game with type space $C_1 \cup C_1'$ **and** message space $C_1 \cup C_1'$ has a PBE with a top coalition that receives at least $v(\mu^0_{C_1\cup C_1'})$, and this is also a coalition of the original game, a contradiction. Similarly, if $v(\mu^0_{C_1\cap C_1'}) > \max W_1$, then the game with type space $C_1 \cap C_1'$ **and** message space $C_1 \cap C_1'$ has a PBE with a top coalition that receives at least $v(\mu^0_{C_1\cap C_1'})$, and this is also a coalition of the original game, a contradiction. (Importantly, the set of types θ with access to messages $m \in C_1 \cup C_1'$ is exactly $C_1 \cup C_1'$, and the set of types with access to $m \in C_1 \cap C_1'$ is exactly $C_1 \cap C_1'$.)

Then $v(\mu^0_{C_1\cup C_1'}) = v(\mu^0_{C_1\cap C_1'}) = \max W_1$. To show that there is a coalition with type set $C_1 \cup C_1'$, consider again a PBE $\widetilde{\sigma}$ of the restricted game with type space and message space both equal to $C_1 \cup C_1'$. By B*, either the top coalition's payoff under $\widetilde{\sigma}$ is strictly greater than $\max W_1$ (leading to a contradiction), or *all* coalitions receive exactly $\max W_1$, in which case $(C_1 \cup C_1', \operatorname{supp}\widetilde{\sigma}, \widetilde{\sigma}, \max W_1)$ is a coalition with type set $C_1 \cup C_1'$. The same argument applies for $C_1 \cap C_1'$. □

Let $\overline{C}_1 = \bigcup_{C_1 \in \widetilde{\mathcal{C}}_1} C_1$ be the largest coalition yielding $\max W_1$. Take any COE and relabel it if necessary so that only the first coalition pays $\max W_1$.[26] Clearly, the set of types receiving payoff $\max W_1$, C_1, is a subset of $\overline{C}_1$. We will now show that, in fact, $C_1 = \overline{C}_1$.

Suppose for the sake of contradiction that $C_1 \subsetneq \overline{C}_1$, and let $D = \overline{C}_1 \setminus C_1$. By B*, $v(\mu^0_D) = \max W_1$. Then the game with restricted type space D and message space D has a PBE whose top coalition receives at least $\max W_1$. This coalition is a valid coalition of the game with type space $\Theta_2 = \Theta - C_1$, because types in $\Theta - C_1 - D = \Theta - \overline{C}_1$ have no access to messages in D. But this contradicts $w_1 > w_2 = \max W_2$ (Algorithm 4).

This argument shows that, if we relabel partitions so that payoffs are strictly decreasing, then all COEs are payoff-equivalent up to the first coalition (i.e., $C_1 = C_1' = \overline{C}_1$)). We can iterate to show the result for all coalitions.

[26]That is, if $w_1 = w_2$, join the first two coalitions into a single one, and so on.

□