Coalition-Proof Disclosure[*]

Germán Gieczewski[†] Maria Titova[‡]

March 2025

Abstract

We analyze the equilibrium set of a general game of verifiable disclosure with type-independent sender preferences and propose coalition proofness among sender types as an equilibrium selection criterion. 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 sender 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.

Keywords: verifiable disclosure, evidence, coalitions, neologism proofness, cheap talk

[*] We thank Nageeb Ali, Steve Callander, Jeff Ely, Elliot Lipnowski, Navin Kartik, Jacopo Perego and the audience members at Penn State, QMUL-City Theory Workshop, 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 the existing notion of neologism proofness (Farrell, 1993) for cheap talk games. The gist of our equilibrium selection argument is that credible coalitional deviations by groups of senders should be correctly interpreted by the receiver.

Due to the generality of the setting, coalition-proof equilibria may fail to exist, for similar reasons that neologism-proof equilibria fail to exist in the cheap talk literature. However, 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 uses cheap talk in addition to verifiable messages. If the sender has access to cheap talk, then a coalition-proof PBE exists when the message mapping is complete or the sender's payoff from fully revealing her type is sufficiently low.

We characterize coalition-proof PBEs and provide tools for finding them. We show that *all* PBE strategies—including coalition-proof ones—are *partition strategies*; they partition the type space into *coalitions*. A coalition consists of a set of sender types, a set of messages that only they 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. The coalition-proof PBE algorithm is simply a greedy version of the PBE algorithm—it removes a coalition that reaches the highest payoff at each step. Therefore, 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, which we call the *tent*, of the sender's ex-ante utility in a coalition-proof PBE of a disclosure game with a rich message space. The tent is comparable to the concave closure in information design (Kamenica and Gentzkow, 2011) and the quasiconcave envelope in cheap talk (Lipnowski and Ravid, 2020).

Related Literature

Our paper relates to three strands of literature. First, a number of applied papers on disclosure, motivated by similar concerns to ours, employ notions close to coalition proofness to select "reasonable" equilibria (Callander, Lambert, and Matouschek, 2021; Aybas and Callander, 2024; Farina et al., 2024). We contribute by providing a general characterization of and existence conditions for this solution concept.

Second, our paper relates to the literature on belief-based refinements for cheap talk games that includes announcement proofness (Matthews, Okuno-Fujiwara, and Postlewaite, 1991), undefeated equilibria (Mailath, Okuno-Fujiwara, and Postlewaite, 1993), with the closest analogue being neologism proofness (Farrell, 1993). These papers, along with Bertomeu and Cianciaruso (2018) who adopt neologism proofness to disclosure games, rule out deviations to a single off-path message. We rule out coalitional deviations to mixed strategies and messages used on path, and that allows us obtain stronger results. Our findings are also related to Koessler and Skreta (2023) who study informed information design and apply a related selection criterion that they term interim sender optimality. We discuss this literature, along with Hart, Kremer, and Perry (2017), in more detail in Section 6.

Third, there are two papers that provide algorithmic equilibrium characterizations for disclosure games, Rappoport (2022) for receiver-optimal PBEs and Wu (2022) for all PBEs but in a special case where and cheap talk is available and there is a total order on evidence messages. Our algorithm outputs all PBEs for any game of verifiable disclosure with type-independent preferences of the sender.

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 theorems for 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

Example 1 (Implausible revelation in receiver-optimal equilibria)**.** *A centrist incumbent (S) learns the state of the world $\theta \in \{-1, 1\}$ drawn from a uniform prior; she can reveal it to the voter (R) or not by choosing message $m \in \{\theta, \varnothing\}$. R sees m and elects a left-wing challenger ($a = -1$), a right-wing one ($a = 1$), or the incumbent ($a = S$). S is office-motivated and gets $\mathbb{1}_{a=S}$. R gets* 1 *if he matches the state ($a = \theta$) and 0 if not ($a = -\theta$), but reelecting the incumbent is a safe alternative that pays* 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, $\Pr(\theta = 1 | m = \varnothing)$, 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.

For S, 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] □

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 ex ante sender-optimal equilibria)**.** *Consider a Grossman (1981) game with a "nuisance dimension." Let* $\Theta = \{(L, A), (L, B), (H, A), (H, B)\}$*; all 4 types are equally likely. Type* $\theta = (\theta(1), \theta(2))$ *can send any message* $m \subseteq 2^{\Theta}$ *such that* $\theta \in m$*. Assume R's best response to m is* $a^*(m) = \sqrt{\Pr(\theta(1) = H \mid m)} - 8\left[\Pr(\theta(2) = A \mid m) - 0.5\right]^2$ *and S's payoff is* $u_S = a$*.*

Notice that revealing $\theta(1) = H$ is good for S but revealing any information about the second dimension hurts her. The receiver-optimal PBE outcome features full revelation: S reveals $(\theta(1), \theta(2))$ and R interprets off-path messages adversarially. In contrast, in the ex ante sender-optimal PBE, all types pool on $m = \Theta$ and get $u_S = \sqrt{0.5}$. Off the path, R thinks that message $m \in \{(H, A), (H, B)\}$ is disproportionally likely to be sent by a single type, much like in Example 1.

Next, consider a simpler game with the second dimension $\theta(2)$ removed.[2] Now, there is a unique PBE in which H separates from L, and the types get 1 and 0, respectively. S's ex ante utility is now 0.5, lower than the pooling payoff of $\sqrt{0.5}$. We argue that a natural equilibrium in Example 2 should mirror this outcome: types (H, A), (H, B) should "form a coalition" (send message $m = \{(H, A), (H, B)\}$ thus separating themselves from other types) to get the highest payoff in the game. More generally, when S has incentives to reveal some dimensions of the state but not others, then that should be the equilibrium outcome. □

[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] That is, now $\Theta = \{L, H\}$ (equally likely) and $a^*(m) = \sqrt{\Pr(\theta = H \mid m)}$.

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 $\theta \in \Theta := \{\theta_1, \dots, \theta_n\}$, which is drawn from a common prior distribution $\mu^0 := (\mu^0_1, \dots, \mu^0_n) \in \Delta\Theta$. 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. Upon observing m, R forms a posterior belief $\mu \in \Delta\Theta$ and chooses action $a \in A$. Finally, payoffs $u_S(a)$ and $u_R(a, \theta)$ are realized. Note that the sender's preferences are type-independent and only depend on R's action.

We employ the belief-based approach that is common in the information design literature. Specifically, we let $\mu \in \Delta\Theta$ be R's posterior belief, $a^*(\mu)$ be R's best response, and $v(\mu) := u_S(a^*(\mu))$ be S's payoff when R has that belief. For much of the paper, we assume without loss that R breaks ties in favor of S when indifferent, which leads to v being upper semicontinuous under mild assumptions.[3] We thus forget about R as a player and work with an upper semicontinuous function v. A more careful tie-breaking is required for some of our results; we make it clear then and 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 $\mu^0_C \in \Delta\Theta$ 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 a 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

[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 $u_S(a)$, $u_R(a, \theta_i)$ are continuous in a for each i (see Lemma 5). In particular, this condition holds if A is a compact subset of $\mathbb{R}^k$; it also holds automatically for any finite A as we can take the discrete metric on A.

[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 $\theta \in \widetilde{\Theta}$.

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 $\bigcup\limits_{\theta \in C} \operatorname{supp} \sigma(\cdot \mid \theta) = X$.

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.

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 $m \in X \cap M(\theta)$.

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 $m \in \bigcup\limits_{\theta \in \widetilde{\Theta}} M(\theta)$ 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, $(C, \{m\}, \sigma, v(\mu^0_C))$ 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

- $C_1, \ldots, C_T$ 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 $(C_t, X_t, \sigma_t, w_t) \in \mathcal{C}(\Theta_t)$.

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
```

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 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 on equilibrium path. For m off path, $\mu(\cdot \mid m)$ can be any feasible belief.

The set of feasible beliefs $\mu(\cdot|m)$ for a message m is 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 $M(\theta_i) = \{m\}$.

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 <u>individually rational</u> (IR) if

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

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 modify Algorithm 1 to return 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]

[5] The IR requirement on $\theta \in C_t$ 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 latter requirement yields the same set of partitions because if $\max_{\theta \in C_t} \leq w_t < \max_{\theta \in \Theta_t} \underline{v}(\theta)$, then a failure of IR is inevitable: there is $\widetilde{\theta} \in \Theta_t \smallsetminus C_t$ 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 $\underline{v}(\theta) > w_t \geq w_\tau$.

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*. Here we define this concept, characterize coalition-proof PBE and provide a modification of Algorithm 2 that returns 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 receiver believed this announcement and updated his 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. Specifically, R should expect all types using m on path to participate in the announced deviation

[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 $\widetilde{w}$.

because their equilibrium payoff is strictly less than $\widetilde{w}$.[7]

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.

Now, we introduce a modification of the PBE Partition Algorithm that selects a payoff-maximizing coalition 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.

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 $\Theta_t$;
    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.[8]

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.*

[7] A related issue that motivates the concept of *announcement proofness* (Matthews, Okuno-Fujiwara, and Postlewaite, 1991), is that if multiple blocking coalitions exist, a potential deviator may consider joining only the best blocking coalition(s) he has access to. This is not a concern in our setting as senders with type-independent preferences share a common ranking over blocking coalitions.

[8] 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.

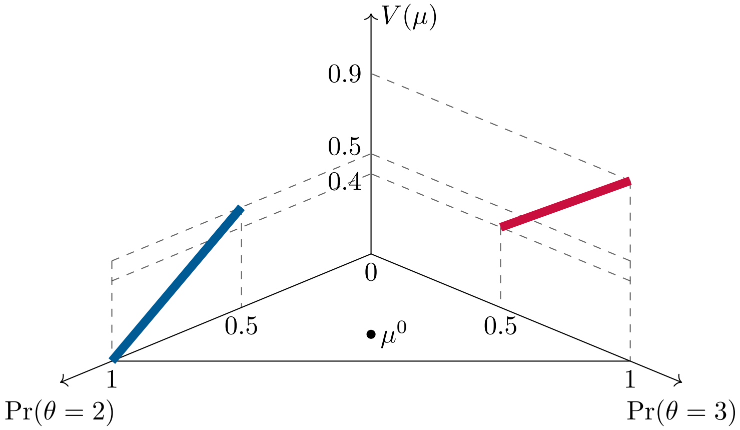


Figure 1. *Feasible sender payoffs in Example 3. Type 2 must send message a (blue), while type 3 must send message b (red). Type 1 has access to both messages.*

Example 3 has a unique PBE, which is not coalition-proof. To see why, follow Algorithm 2. The possible coalitions at step 1 are (i) $X_1 = \{b\}$, $C_1 = \{1,3\}$, $\sigma_1(b \mid 1,3) = 1$, $w_1 = v(0.5, 0, 0.5) = 0.4$ and (ii) $X_1 = \{a\}$, $C_1 = \{1,2\}$, $\sigma_1(a \mid 1,2) = 1$, $w_1 = v(0.5, 0.5, 0) = 0.5$. There is no coalition with $X_1 = \{a, b\}$ because no mixed strategy for type 1 equalizes the payoffs of messages a and b. Hence, $W_1 = \{0.4, 0.5\}$.

Now, if the coalition $(\{1,3\}, \{b\}, \cdot, 0.4)$ is chosen at step 1, then the only possible coalition at step 2 is $(\{2\}, \{a\}, \cdot, 0)$. The associated strategy is a PBE strategy but not coalition-proof since $0.4 < \max W_1$; the blocking coalition is $(\{1,2\}, \{a\}, \cdot, 0.5)$. However, choosing $(\{1,2\}, \{a\}, \cdot, 0.5)$ at step 1 leaves only one possible coalition at step 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. We also establish additional conditions under which Algorithm 3 always terminates, and 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 $\widetilde{w} = v(\mu^0_C) \geq w$.*

Proof. Consider a coalition $(C, \{m\}, \widetilde{\sigma}, \widetilde{w})$. By M-C, there exists a message m such that $M^{-1}(\{m\}) = M^{-1}(X) = C$ and a coalition $(C, \{m\}, \cdot, v(\mu^0_C))$. Since μ^0_C is a linear combination of the posteriors $\mu(\cdot \mid m)$ generated by $m \in X$ under σ, by QC,

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

□

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 $v(\mu^0_C) = v(\mu^0_{C'})$ only if $C = C'$), then all coalition-proof PBE are payoff-equivalent.

To prove this result, we show that $\max W_t \leq w_{t-1}$ at each step of Algorithm 3, so that picking a 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 at step 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}. 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 (i.e., no two "pooling" coalitions pay the same), there is at most one choice of C_t at each step that maximizes 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.[9]

(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 if v satisfies betweenness, then 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 $\mu(\cdot \mid m)$

[9] 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.

for $m \in X$. From B,

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

□

In analogous fashion to Theorem 1, betweenness of v guarantees existence of a coalition-proof PBE; strict betweenness guarantees that every way of choosing coalitions in Algorithm 3 yields a coalition-proof PBE; and 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 the proof of Theorem 1: we show, by contradiction, that $\max W_t \leq w_{t-1}$ at each step of Algorithm 3. 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 $(C_{t-1}, X_{t-1}, \sigma_{t-1}, w_{t-1})$ and $(C_t, X_t, \sigma_t, w_t)$ satisfy $w_{t-1} < w_t$, then there must exist *some* coalition at step $t-1$ that pays at least $v(\mu^0_{C_{t-1} \cup C_t}) \in (w_{t-1}, w_t)$.

new stuff

Lemma 3. *Suppose B holds. Take $X^* \in \arg\max_{X \subseteq \mathcal{M}} v(\mu_{M^{-1}(X)}))$ that is minimal in this set, i.e., such that there is no $X' \subsetneq X^*$ that is also in the argmax. Take any partition $X^* = Y \cup Z$ with $Y \cap Z = \emptyset$. Then $M^{-1}(Y) \cap M^{-1}(Z) \neq \emptyset$.*

Proof. Suppose that, for some Y, Z as described, we have $M^{-1}(Y) \cap M^{-1}(Z) = \emptyset$. Then, by B, at least one of $v(\mu_{M^{-1}(Y)})$, $v(\mu_{M^{-1}(Z)})$ is weakly greater than $v(\mu_{M^{-1}(X^*)}))$, which violates the minimality of X^*. □

Lemma 4. *Suppose B holds. Take $X^* \in \arg\max_{X \subseteq \mathcal{M}} v(\mu_{M^{-1}(X)}))$ that is minimal in this set, i.e., such that there is no $X' \subsetneq X^*$ that is also in the argmax. Then there is a coalition $(M^{-1}(X^*), X^*, \cdot, v(\mu_{M^{-1}(X^*)}))$.*

Proof. We know what C, X and w we want; the question is if there is a σ that equalizes the payoffs of all messages in X^*.

Suppose not. Recall Lemma 7. Denote $v_0 = v(\mu_{M^{-1}(X^*)})$ and let $L_{v_0}(v)$ be the level set containing $\mu_{M^{-1}(X^*)}$. Using the notation in the proof of Theorem 2, take $\lambda \in [0,1]$ such that H_λ is the hyperplane within $L_{v_0}(v)$ containing $\mu_{M^{-1}(X^*)}$. We can then partition $\Delta\Theta - H_\lambda$ into two half-planes, one containing all μ with $v(\mu) > v_0$ and the other containing all μ with $v(\mu) < v_0$. Denote these sets by H^+, H^-.

We will find σ such that all messages in X generate posteriors in H_λ. Begin with an arbitrary σ with full support. (?) If it does not induce only posteriors on H_λ, it must induce some posteriors in H^+ and also some on H^-. Let X_+ be the (nonempty) set of messages in X that, given σ, generate posteriors in H^+. Note that $v(\mu_{M^{-1}(X^+)}) < v_0$: indeed, $>$ would violate the maximality of v_0 and $=$ would violate the minimality of X. Then there must be a message $m \in X^+$ and a type θ such that $\mu_{\{\theta\}} \in H^-$; and θ can send m but is instead sending messages $m' \notin X^+$ with positive probability under σ.

Then, increase $\sigma(m|\theta)$ while lowering some $\sigma(m'|\theta)$ to compensate. If m' generates an equilibrium posterior in H^-, we can do this until either the posterior from m hits H_λ or the one from m' does. If m' generates an equilibrium posterior in H_λ, we can do a little bit of this so that both m and m' then generate posteriors in H^+; if necessary, we can do this until no messages generate posteriors in H_λ; then, once done, the same argument yields that there must be a type sending a H^- message with positive prob, which can be shifted to one of these H^+ messages. Can be better explained, but the argument is that so long as there are messages yielding posteriors in H^+, we can use at least one of these perturbations, and they always ultimately increase the total probability that a message with $H^+ \cup H_\lambda$ posteriors is sent. This probability can't hit one, so... etc. □

Proposition 3. *If B holds, then, for set of messages $X \subseteq \mathcal{M}$ such that $v(\mu_{M^{-1}(X)})$ is maximized, there is a coalition $(M^{-1}(X), X, \cdot, v(\mu_{M^{-1}(X)}))$, and it is chosen at the first step. The analogous claim also holds at later steps. Thus, the solution is "as if" the message space were complete.*

Proof. follows from the Lemma? The lemma gives us that, at each step, the highest attainable payoff is just $\max_{X \subseteq \mathcal{M}} v(\mu_{M^{-1}(X)}))$, so... □

why is this useful? In general it makes it easier to describe what CPPBE look like in the betweenness case.

——

It is easy to produce examples where B holds but receiver optimality and CPPBE do not coincide. For example, take 3 types 1, 2, 3 and 2 messages a, b such that 2 can send both messages, 1 only a, 2 only b, and v satisfying B such that revealing yourself as 2¿revealing 1, revealing 3. Generally this game will have 3 equilibria: 2 pools with 1, 2 pools with 3, or 2 mixes so that both messages give the same payoff. CPPBE picks the equilibrium that's optimal for type 2. Receiver optimality picks depending on R's payoff, which we are mostly agnostic to in our paper.

In fact, I believe that the following holds:

Claim 1. *There are games with the same set of types and prior belief; same message space and mapping; same set of receiver actions A; same sender payoff $u_S(a)$; and same receiver best response function $a^*(\mu)$ (i.e., where BR is a singleton), where receiver optimality picks different equilibria. In other words, there are ways to perturb u_R without changing anything else, and in ways that don't change R's behavior given any posterior belief, that nevertheless change what receiver optimality selects.*

Here is another conjecture: take a game and perturb it by adding an additional message m with preimage $C \subseteq \Theta$. Then, for any CPPBE of the original game, there is a CPPBE of the new game, in which the payoffs of all members of C are weakly higher. In other words, adding a message cannot hurt the sender types with access to it; it just gives them more options. I think this is true. But it is also easy to find examples showing that receiver optimality does not have this property.

end new stuff

Adding Cheap Talk

The last two existence conditions apply when the sender has access to cheap talk as well as evidence. 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 exist 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 because it allows us to rewrite the message space as follows. Let two messages m, m' be equivalent if $M^{-1}(\{m\}) = M^{-1}(\{m'\})$

and denote the equivalence class of m by $\widetilde{m}$. Then, each message m maps to a pair $(\widetilde{m}, j)$, where $\widetilde{m}$ is the verifiable content of m and $j = 1, 2, \ldots, J_{\widetilde{m}}$, the index denoting which "copy" of $\widetilde{m}$ was sent, is cheap talk. Conversely, any message mapping M can be augmented to allow S to send cheap talk in addition to evidence.

When cheap talk is available, useful equilibria may be lost if R always breaks ties in favor of the sender (the same issue shows up in Lipnowski and Ravid, 2020). Hence, in this section, we treat v as an upper hemicontinuous, compact and convex-valued correspondence that returns all possible sender payoffs when R best-responds to μ.[10] Condition 4 in Definition 1 of a coalition then becomes $w \in v(\mu(\cdot \mid m))$ for each $m \in X$. We let $\overline{v}(\mu) := \max(v(\mu))$ 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 restrictions 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. Using Theorem 1, we first find a coalition-proof PBE of a modified game with the same message mapping and the sender's payoff being the quasiconcave closure of v rather than v itself. We then show that cheap talk can be used to reconstruct the same equilibrium in the original game. The result is related to Lipnowski and Ravid (2020)'s insight that cheap talk effectively quasiconcavifies the sender's payoff function. In particular, when M allows *only* cheap talk (i.e., $M^{-1}(\{m\}) = \Theta$ for all $m \in \mathcal{M}$), the unique coalition-proof PBE is simply the sender-optimal PBE found by Lipnowski and Ravid (2020).

Our fourth existence condition is that M satisfies the cheap talk property and giving full information to the receiver is (potentially) bad for the sender:

Theorem 4. *Suppose that*

- *v is an upper hemicontinuous, compact and convex-valued correspondence;*

[10] Again, these properties of v follow under mild assumptions on u_S, u_R and A (Lemma 5).

- *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 ever halts at step t when the set $W_t \cap [\max_{\theta\in\Theta_t} \underline{v}(\theta), w_{t-1}]$ is empty, which can only happen if $\max W_t > w_{t-1}$. Under the conditions of Theorem 4 we can continuously "degrade" a coalition that pays $\max W_t$ by adding cheap talk messages that bring R's posterior closer to complete information. 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.[11]

We finish this section with three 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, our results are relatively "tight": 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.[12] Third, any coalition-proof PBE partition $\{(C_t, X_t, \sigma_t, w_t)\}_{t=1}^T$ must be *lexicographically maximal* in the following sense: w_1 is the maximal payoff obtained across all PBEs; w_2 is the maximal payoff obtained in any PBE with $(C_1, X_1, \sigma_1, w_1)$ as its first coalition; and so on. Then, even when coalition-proof PBE fails to exist, it may be reasonable to focus on lexicographically maximal PBEs, which of course always exist.[13]

5 Rich Message Spaces

This section provides an explicit characterization of coalition-proof PBE when the message space is maximally "rich". In particular, we not only require there to be a

[11] 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.

[12] 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 still no coalition-proof PBE.

[13] Mailath, Okuno-Fujiwara, and Postlewaite (1993) propose a related selection criterion for signaling games, *undefeated equilibrium*, and show that the lexicographically maximal PBE is undefeated. However, their assumptions do not map well into ours: they require that messages affect payoffs in a particular way and that v is continuous and FOSD-monotonic in μ.

message that allows any set of types to pool together and separate from others (as in Grossman (1981) and Milgrom (1981), where any subset $m \subseteq \Theta$ such that $\theta \in m$ is a valid message) but also that *fractions* of each type can pool, while excluding otherwise identical copies of themselves.[14]

To accommodate this formally, we begin with an upper semicontinuous sender payoff function $v(\mu)$ and a finite "payoff-relevant" type space Θ, as in Section 2. We then take the actual type space of our game to be $\Omega := \Theta \times [0, 1]$, and assume that, for each nonzero vector $(p_1, \dots, p_n) \in [0, 1]^n$, 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 j is payoff-irrelevant for both S and R. 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 do not directly apply to this game (because the type space Ω is infinite), a coalition-proof PBE does exist, and its payoffs can be tightly characterized under a genericity assumption.

Proposition 4. *In the model with a rich message space,*

(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 $C = \operatorname{supp}(\mu^0_{\Omega_t}) \subseteq \Theta$;

(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 existence results. Part (iii) describes the structure of coalition-proof PBEs. It establishes that every sender type (θ, j) who is in a coalition

[14] 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*; but it is useful for the sender when v is general.

$(C_t, X_t, \sigma_t, w_t)$ must receive the highest payoff attainable by the types left in stage t, i.e., by the set C of $\theta \in \Theta$ that are represented in Ω_t. Indeed, if positive masses of each $\theta \in C$ are still left, some fractions of them can pool appropriately to induce a constrained-optimal posterior $\mu \in \mu^{*C}$. And, in a coalition-proof PBE, they must do so. Part (iv) gives us uniqueness: there is effectively a unique coalition-proof PBE if there is a unique maximizer of $v(\mu)|_{\Delta C}$ for each C, which holds for "almost all" v.

Proposition 4.(iii) and its proof also provide a recipe for constructing coalition-proof PBEs, which we sketch here. Figure 2 provides an example with three types.

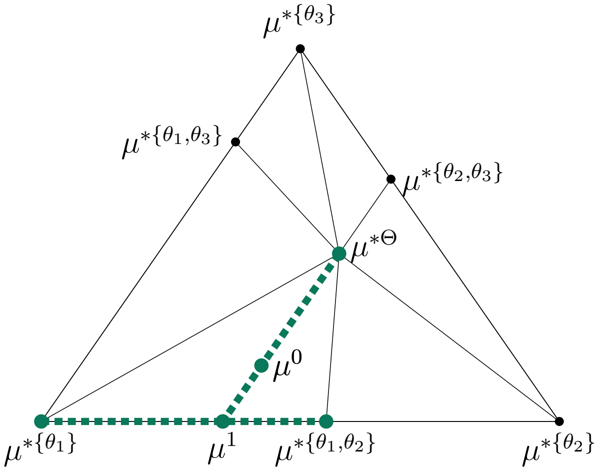


Figure 2. *Posterior beliefs in a coalition-proof PBE of a game with a rich message space.*

First, find the largest sender payoff in the game. We claim that the first coalition reaches this payoff and, without loss, fully removes some $\theta \in \Theta$ from the game. Indeed, by (iii), the first coalition must get $w_1 = v(\mu^{*\Theta})$. To do so, it must induce a belief $\mu^{*1} \in \mu^{*\Theta}$. This can be done with a message 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 $\lambda > 0$. If choosing coalitions of maximal size [(ii)], we must pick the highest feasible λ such that $p_i \leq 1$ for all i, i.e., $\lambda = \min_j \frac{\mu^0_j}{\mu^{*1}_j}$. The first coalition is then $(M^{-1}(\{m_1\}), \{m_1\}, \cdot, v(\mu^{*1}))$, and $\Omega_2 = \Omega \setminus (M^{-1}(\{m_1\})$. Since λ is maximal, we have $p_i = 1$ for some i, so (θ_i, j) participates in this coalition for all j, i.e., θ_i is fully removed in step 1.[15] To simplify exposition, suppose that $p_i = 1$ for a single i and without loss assume that $i = n$. Let $\Theta_2 = \{\theta_1, \ldots, \theta_{n-1}\}$. In

[15] If not choosing coalitions of maximal size, all coalitions formed until type θ_i is fully removed from the game must get the payoff $v(\mu^{*\Theta})$; all coalitions attaining this payoff can then be merged ex post.

Figure 2, $\mu^{*1} = \mu^{*\Theta}$ is the posterior induced by the first coalition; θ_3 is removed from the game at step 1, and the residual posterior μ^1 is supported on θ_1 and θ_2.

Let $\overline{\mu}^2 = \overline{\mu}^0_{\Omega_2}$ be R's posterior (over types $(\theta, j) \in \Omega$) conditional on not receiving message m_1. By construction, $\mu^2 := \mu^0_{\Omega_2} \in \Delta\Theta_2$ is proportional to $\mu^0 - \lambda\mu^{*1}$. We repeat the same construction, with $\mu^{*2} \in \arg\max_{\mu \in \Delta\Theta_2} v(\mu)$: since all θ_n-senders were removed in stage 1, the best the remaining types can do is induce the posterior μ^{*2}.

We repeat until all types are assigned to a coalition. Thus, if θ_{n-1} senders are 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. The equilibrium is effectively unique if the arg max at each step is unique.

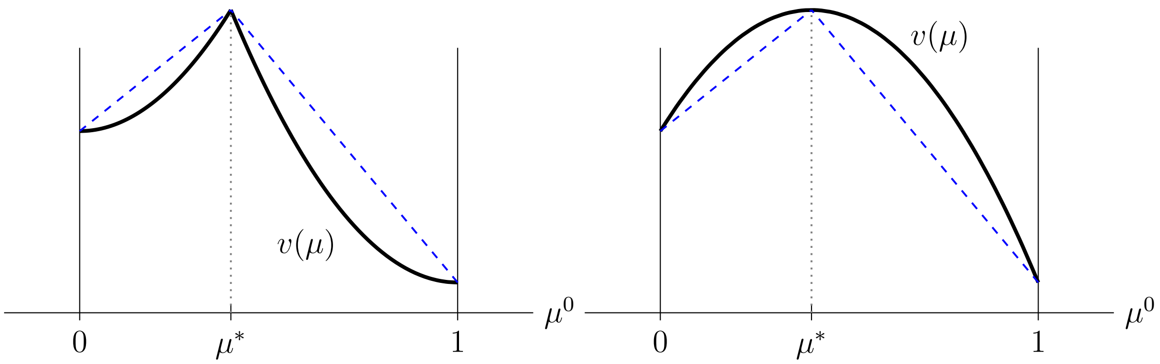
Existing literature has characterized the ex ante sender payoffs attainable under Bayesian persuasion (Kamenica and Gentzkow, 2011) and cheap talk (Lipnowski and Ravid, 2020) with general state-independent sender payoff $v(\mu)$ and an arbitrary prior belief μ^0. In Bayesian persuasion, the sender attains $v^C(\mu^0)$, the value of the *concave closure* of v evaluated at the prior, while under cheap talk she attains (at most) $v^{QC}(\mu^0)$, the value of the quasiconcave envelope. An analogous characterization for games of disclosure with general sender payoff v is not available in the literature.

Assuming as in Proposition 4.(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 in a coalition-proof PBE, $v^{\text{tent}}(\mu^0)$, as a function of the prior μ^0. The function $v^{\text{tent}}(\cdot)$, which we call the *tent* of v, admits a geometric characterization.

Proposition 5. *If every μ^{*C} is a singleton, then the sender's ex ante payoff given prior μ^0 is $v^{tent}(\mu^0)$. v^{tent} is the unique function such that:*

- *$v^{tent}(\mu^{*C}) = v(\mu^{*C})$ for all $C \subseteq \Theta$;*
- *for each permutation $(\theta_{i1}, \ldots, \theta_{in})$ of Θ, v^{tent} is linear when restricted to conv$(\mu^{*\Theta}, \mu^{*\{\theta_{i1},\ldots,\theta_{i(n-1)}\}}, \ldots, \mu^{*\{\theta_{i1}\}})$.*

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^*. In a coalition-proof PBE, 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 $v(\mu^*)$.



(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.*

If $\mu^0 = \mu^*$, then *both* types obtain this payoff with probability 1, so $v^{\text{tent}}(\mu^*) = v(\mu^*)$. If $\mu^0 > \mu^*$, so that θ_2 is more numerous than needed to produce the optimal posterior μ^*, *all* θ_1-senders will pool with *some* 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 θ_2-senders are unable to pool with θ_1 senders, so $v^{\text{tent}}(\mu^0)$ is close to $v(1)$. As μ^0 increases from μ^* to 1, $v^{\text{tent}}(\mu^0)$ decreases linearly from $v(\mu^*)$ to $v(1)$, as in fact $v^{\text{tent}}(\mu^0) = v(\mu^*)\left(1 - \mu^0 + \frac{\mu^*}{1-\mu^*}(1-\mu^0)\right) + v(1)\left(\mu^0 - \frac{\mu^*}{1-\mu^*}(1-\mu^0)\right)$, which is linear in μ^0. By similar logic, for $\mu^0 < \mu^*$, there are "excess" θ_1 senders, so all θ_2 senders get to pool with some θ_1 senders and obtain $v(\mu^*)$, while the leftover θ_1 senders get $v(0)$. $v^{\text{tent}}(\mu^0)$ varies linearly between $v(0)$ and $v(\mu^*)$, creating the "tent" shape seen in Figure 3.

Clearly, payoffs in a coalition-proof PBE depend *only* on the "peaks" of the payoff function v and the prior μ^0. Indeed, as seen on Figure 3.(b), 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 $\mu^* = \arg\max v$) are unchanged. Note that when v is convex on either side of the peak, as in Figure 3.(a), v^{tent} coincides with the concave closure of v, so S does as well in a coalition-proof PBE as when he has commitment power, and better than cheap talk (which would be uninformative). On the other hand, as seen in Figure 3.(b), the sender's ex ante payoff in a coalition-proof PBE can be also be lower than with cheap talk, Bayesian persuasion, or no communication at all. The reason is that there is no "loyalty" across sender types: types who can get the highest payoff $v(\mu^*)$ will do so even if it hurts other types and even ex ante payoffs.

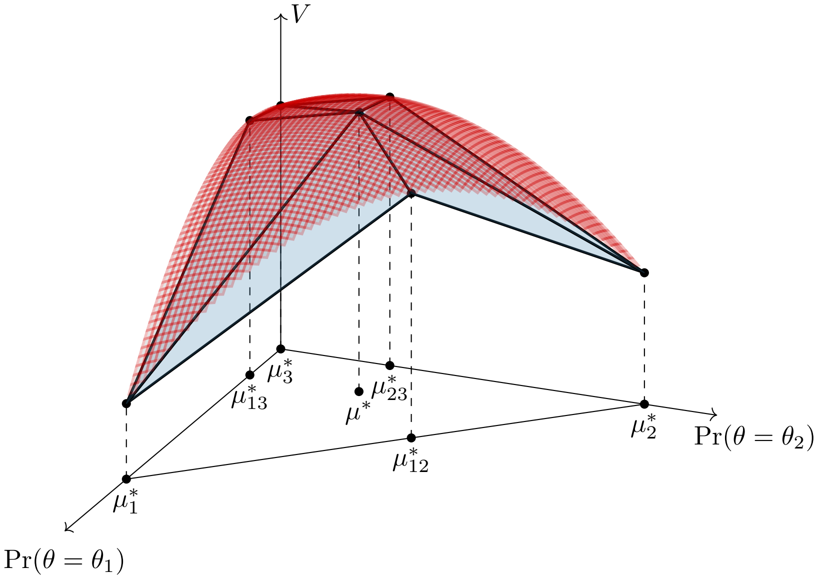


Figure 4. *Tent of v with three types. We denote $\mu^{*\{\theta_i,\theta_j\}} = \mu^*_{ij}$ and $\mu^{*\{\theta_i\}} = \mu^*_i$.*

Figure 4 further illustrates how the tent of v is constructed in the three-type example from Figure 2. 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 $\operatorname{supp}\mu \subseteq \{\theta_1,\theta_2\}$, $\{\theta_1,\theta_3\}$, $\{\theta_2,\theta_3\}$ respectively; and the corner beliefs $\mu^{*\{\theta_1\}}, \mu^{*\{\theta_2\}}, \mu^{*\{\theta_3\}}$. These 7 beliefs ($2^n - 1$ in general) 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 triangle, v^{tent} is linear and equal to v on the vertices of the triangle. Thus the graph of v^{tent} is made up of 6 triangles joined 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 truth-leaning equilibrium is also coalition-proof in HKP's setting, and in fact it is essentially the *only* output of Algorithm 3 in most cases—namely, whenever v satisfies *strict* betweenness.[16]

Proposition 6. *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, the conceptual connection between both concepts breaks down if betweenness is not satisfied. The intuition is as follows: receiver-optimal equilibria involve as much revelation (separation of sender types) as possible. When v satisfies betweenness, high types prefer to separate from others, and coalitional deviations ensure that any profitable opportunities to separate do not go unused. If v does not satisfy betweenness, then higher payoffs may instead be achievable with pooling.

Bertomeu and Cianciaruso (2018)

Our notion of coalition-proof PBE generalizes Bertomeu and Cianciaruso (2018)'s Grossman-Perry-Farrell equilibrium (GPFE). The authors also provide an algorithmic characterization of GPFE analogous to our Algorithm 3.[17] The main advantage of

[16] 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.

[17] 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.

our solution concept over theirs is that we allow for mixed strategies, while GPFE only allows coalitions of the form $(M^{-1}(\{m\}), \{m\}, \cdot, \mu^0_{M^{-1}(\{m\})})$. Their definition of a coalition cannot accommodate cheap talk, nor the kind of mixing that often occurs in HKP's truth-leaning equilibria. As a result, they require stronger conditions than ours to guarantee existence (for instance, a GPFE does not exist in their Example 7 which does have a coalition-proof PBE). Their Proposition 1 shows that a GPFE exists under B + M-C; in contrast, a coalition-proof PBE exists under B (Theorem 2), under QC + M-C (Theorem 1), as well as other conditions (Theorems 3 and 4) that do not guarantee existence of GPFE. Our characterization results for maximally flexible message spaces (Propositions 4 and 5) are also novel.

Other Selection Criteria Based on Coalitional Deviations

Other selection criteria for communication games in the spirit of neologism proofness include announcement proofness (Matthews, Okuno-Fujiwara, and Postlewaite, 1991) and undefeated equilibrium (Mailath, Okuno-Fujiwara, and Postlewaite, 1993). Beyond the differences discussed in footnotes 7 and 13, they (along with the aforementioned paper on neologism-proofness) only allow (blocking) "coalitions" to form using a single message.

Another selection criterion that can be viewed as coalition-based was proposed by Koessler and Skreta (2023) in a model of mechanism design with an informed designer.[18] Their interim optimality (IO) criterion effectively rules out credible deviations by "coalitions" of types in any proportion, assuming a message is always available to include these types and exclude others. Their analysis is thus related to our Section 5 (rich message spaces), where the existence of both coalition-proof PBE and IO mechanisms is guaranteed. However, IO imposes looser constraints than coalition proofness. In particular, the set of IO mechanisms in their Section V is larger than our set of coalition-proof PBEs in Section 5, and the ex ante preferred mechanism IO* is generally not coalition-proof in our setting. The discrepancy arises because Koessler and Skreta (2023)'s mechanisms allow agents to commit to mixing over messages that are *not* payoff-equivalent, and agents know their type but *not* the realization of their mixed message when deviating (compare with Section 5, where

[18] Their sender commits to a disclosure mechanism after learning his type, which makes that setup related to models of verifiable disclosure.

types know if they will be "excluded" from the equlibrium's top coalition).

7 Conclusion

We show that all PBE strategies in games of verifiable disclosure are partitional in nature. As such, it makes sense to view equilibrium refinements as ruling out coalitions that pay more than the equilibrium payoff, provided that the receiver correctly interprets the coalitional deviation. Unlike the existing coalition-based refinements such as neologism proofness, announcement proofness, undefeatedness and interim sender optimality, we allow our blocking coalitions to form using multiple messages (so that mixing is involved) and messages that are already on the equilibrium path.

Our flexible framework allows us to state existence results for a general class of disclosure games, ranging from the seminal models of disclosure (Milgrom, 1981; Grossman, 1981) to cheap talk (Lipnowski and Ravid, 2020), while also clarifying clarifying the relationship with receiver-optimal equilibria when evidence is structured (Hart, Kremer, and Perry, 2017). Our geometric characterization of the sender's ex-ante utility (the tent of the sender's value function) is, to our knowledge, the first analysis of a disclosure game with general state-independent sender preferences that is comparable to the concave closure in information design (Kamenica and Gentzkow, 2011). Finally, one takeaway from Theorems 3 and 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 PBEs under various conditions, similarly to what we do for the case of rich message spaces in Propositions 4, 5 or 6.

References

Aybas, Yunus C and Steven Callander (2024), “Cheap Talk in Complex Environments”. (p. 3.)

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

Bertomeu, Jeremy and Davide Cianciaruso (2018), “Verifiable Disclosure”, *Economic Theory*, 65, 4, pp. 1011-1044. (pp. 2, 3, 25.)

Callander, Steven, Nicolas S Lambert, and Niko Matouschek (2021), “The Power of Referential Advice”, *Journal of Political Economy*, 129, 11, pp. 3073-3140. (p. 3.)

Farina, Agata, Guillaume Fréchette, Alessandro Lizzeri, and Jacopo Perego (2024), “The Selective Disclosure of Evidence: an Experiment”. (p. 3.)

Farrell, Joseph (1993), “Meaning and Credibility in Cheap-Talk Games”, *Games and Economic Behavior*, 5, 4 (Oct. 1993), pp. 514-531. (pp. 2, 3.)

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, 5, 20, 27.)

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, 14, 19, 24, 27.)

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. 34, 43.)

Kamenica, Emir and Matthew Gentzkow (2011), “Bayesian Persuasion”, *American Economic Review*, 101, 6, pp. 2590-2615. (pp. 3, 22, 27.)

Koessler, Frédéric and Vasiliki Skreta (2023), “Informed Information Design”, *Journal of Political Economy*, 131, 11, pp. 3186-3232. (pp. 3, 26.)

Lipnowski, Elliot and Doron Ravid (2020), “Cheap Talk with Transparent Motives”, *Econometrica*, 88, 4, pp. 1631-1660. (pp. 3, 18, 22, 27, 39.)

Mailath, George J, Masahiro Okuno-Fujiwara, and Andrew Postlewaite (1993), “Belief-Based Refinements in Signalling Games”, *Journal of Economic Theory*, 60, 2, pp. 241-276. (pp. 3, 19, 26.)

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green, et al. (1995), *Microeconomic Theory*, Oxford University Press, New York, vol. 1. (p. 43.)

Matthews, Steven A, Masahiro Okuno-Fujiwara, and Andrew Postlewaite (1991), “Refining Cheap-talk Equilibria”, *Journal of Economic Theory*, 55, 2, pp. 247-273. (pp. 3, 11, 26.)

Milgrom, Paul R (1981), “Good News and Bad News: Representation Theorems and Applications”, *The Bell Journal of Economics*, pp. 380-391. (pp. 2, 20, 27.)

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 5. *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 $\Theta_2 := \Theta \smallsetminus C_1$. 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 IR 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 $\forall m \in \mathcal{M} \setminus \operatorname{supp} \sigma$, $\mu(\cdot \mid m) = \arg \min_{\mu(\cdot|m) \text{ feasible}} v(\mu(\cdot|m))$. 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. □

Proof of Remark 2

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

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, 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 $\Theta_t$;
    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
```

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 happens to satisfy $w_t \leq w_{t-1}$. (As shown next, imposing IR is unnecessary because the output of Algorithm 4 automatically satisfies IR.) We now prove that COEs are indeed coalition-proof PBEs:

Lemma 6. *Every COE is also a coalition-proof PBE. 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 $\min\limits_{\mu(\cdot|m) \text{ feasible}} v(\mu(\cdot|m))$. Therefore, for each t, t-th coalition of Algorithm 4 satisfies $w_t = \max W_t$ and $w_t \in [\max\limits_{\theta\in\Theta_t} \underline{v}(\theta), w_{t-1}]$. 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 $\max W_t \leq w_{t-1}$ for all t, then at each step $\max(W_t \cap [\max\limits_{\theta\in\Theta_t} \underline{v}(\theta), w_{t-1}]) = \max W_t$, 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 $\max W_t \leq w_{t-1}$ 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 $w_t = \max W_t \leq w_{t-1}$ for all $t \leq s$, but $\Theta_{s+1} \neq \varnothing$ and $\max W_{s+1} > w_s$. Let $(C_{s+1}, X_{s+1}, \sigma_{s+1}, w_{s+1})$ 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 $M^{-1}(\{m'\}) = M^{-1}(X_s) \cup M^{-1}(\{m\}) = C_s \cup C_{s+1} =: \widetilde{C}$.[19] 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 $w_s = \max W_s$.

[19] 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., $M^{-1}(X_t) \cap \Theta_t$.

(i) Suppose that QC and M-C hold. We will show that a COE exists, which implies the result by Lemma 6. To do this, we will construct a *maximal* coalition-optimal 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 4, 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 $x \neq x'$.[20] 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. □

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.[21] 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 6 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

[20] 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.

[21] 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.

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.

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 7. *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.

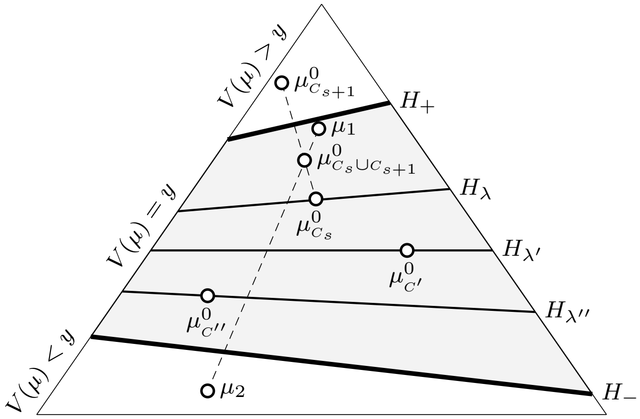


Figure 5. *Choice of λ-maximizing coalition*

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 $H_+ \cap H_- \cap \Delta\Theta = \varnothing$.[22] 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

[22] 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 $\widetilde{H}_1$.

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 $\max W_{t+1} \leq w_t$ must hold. Suppose that B holds, but our modified algorithm yields an incomplete partition $(C_t, X_t, \sigma_t, w_t)^s_{t=1}$ 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.[23] Label the posteriors generated by each coalition as $\mu_1, \ldots, \mu_k$; their convex hull must contain $\mu^0_{C_s \cup C_{s+1}}$.[24] 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 $x \neq x'$.[25] Then the set of v with $C \mapsto v(\mu^0_C)$

[23] 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.

[24] 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 $\mu^0_{\widetilde{C}_z}$.

[25] 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 $\{x, x'\}$.

injective is a finite intersection of open and dense sets, hence open and dense. □

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.,

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

Lemma 8. *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 $\mu \in \operatorname{Conv}(\mu^{1t}, \ldots, \mu^{nt})$ 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 $\mu \in \operatorname{Conv}(\{\mu^{1\infty}, \ldots, \mu^{n\infty}\})$, 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. □

Now, as in Theorem 1, we aim to show the existence of a COE, which must be a coalition-proof PBE by Lemma 6. 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 8, 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 $\widetilde{\mu}_t^1, \ldots, \widetilde{\mu}_t^n$ whose convex hull contains $\mu^0_{C_t}$ and such that

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

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

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 $\widetilde{\mu}_t^i$. 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 $w_t = v^{QC}(\mu^0_{C_t}) \in v(\widetilde{\mu}_t^i)$ in the original game G. A partition thus constructed attains in G the same payoffs that $(C_t, X_t, \sigma_t, w_t)_{t=1}^T$ 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 $(C_t, \widetilde{X}_t, \widetilde{\sigma}_t, w_t)_{t=1}^T$ is a COE of G. □

Proof of Theorem 4

We will use the following lemma.

Lemma 9. *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 $\mu^i \in \operatorname{Conv}(\mu^i_1, \ldots, \mu^i_n)$ 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 $\mu^i \in \operatorname{Conv}(\mu^i_1, \ldots, \mu^i_n)$. 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 8, $\max W_t$ 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 4

(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, $v(\mu^{*\text{supp }(\mu^0_{\Omega_t})})$ is weakly decreasing in t, as 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 $w_1 = \max W_1$, $\max W_2 \leq w_1 \implies w_2 = \max W_2$, $\max W_3 \leq w_2 \implies w_3 = \max W_3$, and so on. Thus $w_t = v(\mu^{*\text{supp }(\mu^0_{\Omega_t})})$, as desired.

(ii) Lemma 6 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 supp $(\mu^0_{\Omega_{t+1}})$ = 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 $w \in V$, $\nu(S + w) = 0$. That is, 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 $h_K(a) = \sup_{k \in K} \langle a, k\rangle$. 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 Lipschitz: a supremum of L-Lipschitz functions is L-Lipschitz, and $\langle a, k\rangle$ is $||k||$-Lipschitz as a function of a, so $h_K(a)$ 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 5

The construction in Proposition 4.(iv) gives an essentially unique[26] way to write any belief $\mu \in \Delta\Theta$ as a convex combination $\sum_{j=1}^{n} \lambda_j \mu^{*C_j}$ for some permutation $(\theta_{i1}, \ldots, \theta_{in})$ of Θ and $C_j := \{\theta_{i1}, \ldots, \theta_{i(n-j+1)}\}$ for all j. Moreover, it yields that $v^{\text{tent}}(\mu) = \sum_{j=1}^{n} \lambda_j v(\mu^{*C_j})$. The identity $\mu = \sum_{j=1}^{n} \lambda_j(\mu) v(\mu^{*C_j})$ for $\mu \in \text{conv}(\mu^{*C_1}, \ldots, \mu^{*C_n})$ yields the linearity because the λ_j are linear functions of μ over this set. If $\mu = \mu^{*C}$ for some $C \subseteq \Theta$, then such a decomposition is given by: $\mu^{*C_j} = \mu^{*C}$ for some j and $\lambda_j = 1$, yielding $v^{\text{tent}}(\mu^{*C}) = v(\mu^{*C})$.

Conversely, any function $\tilde{v}$ satisfying the given properties must equal v^{tent}. Indeed, any $\mu \in \Delta\Theta$ can be written as $\mu = \sum_{j=1}^{n} \lambda_j \mu^{*C_j}$ as above, whence

$$\tilde{v}(\mu) = \sum_{j=1}^{n} \lambda_j \tilde{v}(\mu^{*C_j}) = \sum_{j=1}^{n} \lambda_j v(\mu^{*C_j}) = v^{\text{ea}}(\mu).$$

□

Proof of Proposition 6

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 $\theta' \leq \theta$.[27] 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

[26] The only indeterminacies are that there may be multiple choices of C_j in steps where $\lambda_j = 0$ anyway, and that there may be multiple (redundant) valid choices of λ_j when $\mu^{*C_j} = \mu^{*C_{j+1}}$. These do not affect our argument.

[27] 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.[28] 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 10. *$\widetilde{\mathcal{C}}_1$ 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 $\alpha \mu^0_{C_1} + (1-\alpha)\mu^0_{C_1'} = \beta \mu^0_{C_1 \cup C_1'} + (1-\beta)\mu^0_{C_1 \cap C_1'}$, if we take $\alpha = \frac{\mu^0(C_1)}{\mu^0(C_1) + \mu^0(C_1')}$ and

[28] 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 $\widetilde{C} \subseteq \Theta_s$.

$\beta = \frac{\mu^0(C_1 \cup C_1')}{\mu^0(C_1)+\mu^0(C_1')}$. 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 $\max W_1$.[29] 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.

□

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