Games and Economic Behavior 158 (2026) 563–583

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

Persuasion of interacting receivers

Yishu Zeng

Department of Economics, York University, Vari Hall 1094, 4700 Keele Street, Toronto, M3J 1P3, ON, Canada

ARTICLE INFO

JEL classification:
D82
D83
D91

Keywords:
Bayesian persuasion
Information design
Revelation principle
Equilibrium selection
Psychological preferences

ABSTRACT

This paper investigates how to persuade multiple interacting receivers under arbitrary equilibrium selection and when receivers' preferences may exhibit psychological traits. Our main result characterizes the conditions under which the sender can categorize information using partitions and then disclose partition-related information without altering the equilibrium outcome. This finding broadens the traditional direct approach based on the revelation principle, which categorizes messages solely by equilibrium actions. Consequently, our results support a generalized direct approach, which we apply to study information disclosure in fostering fairness within groups where members may demonstrate altruistic behaviors.

1. Introduction

Many economic settings involve a sender disclosing information to multiple receivers before they interact strategically. In practice, these disclosures often contain more than simple action recommendations. On platforms such as Amazon or eBay, for example, sellers provide not only basic product descriptions but also expert reviews, purchase counts, and numbers of likes. Even when the implicit recommendation is unchanged (e.g., "buy the product"), such additional information can influence buyers who derive psychological utility from following trends, trusting expert opinions, or aligning with popular sentiment.

A similar pattern appears in fundraising. Rather than merely recommending investment, managers often disclose richer information, such as the identities of committed investors, early-bird opportunities, or opportunities for interaction among potential investors. These disclosures can influence beliefs about others' participation and, in turn, investment decisions. These examples intend to motivate the paper rather than serve as literal applications of the model. Our objective is to study, within a general theoretical framework, how richer information structures influence strategic outcomes beyond direct action recommendations, which remain the primary focus of much of the existing theoretical literature.

Recent studies suggest that when the sender lacks freedom in selecting the equilibrium, or when receivers' preferences reflect psychological traits, achieving optimal outcomes may require more sophisticated information structures; see, for example, Alonso and Camara (2016), Lipnowski and Mathevet (2018), Mathevet et al. (2020), Morris et al. (2024), among others. In particular, when important considerations such as psychological utility or non-sender-preferred equilibrium selection are involved, standard direct approaches that focus solely on recommending actions may be insufficient. This is because standard obedience constraints may not adequately characterize the implementable outcomes when receiver behavior is also influenced by the additional factors described above. In such cases, it is necessary to expand existing frameworks and methods to incorporate these economic considerations.

This paper takes a step toward a comprehensive understanding of persuasion among interacting receivers in general settings where the aforementioned economic factors shape their interaction. Our main result identifies the key conditions under which the sender can

E-mail address: zengyish@yorku.ca

https://doi.org/10.1016/j.geb.2026.04.009
Received 6 September 2024
Available online 8 May 2026

coarsen information using partitions, offering greater flexibility than simply categorizing information by equilibrium actions, thereby allowing for richer forms of communication. When these conditions are met in a partition, it defines a game-specific information coarsening rule, in contrast to the standard revelation principle, which seeks a universal coarsening rule based on equilibrium actions. A useful implication of our result is that it provides a constructive method for identifying the optimal information structure using these partitions, which we will illustrate through examples and applications. The conditions we characterized are also tight: if a partition violates any of these conditions, coarsening information based on that partition may lead to suboptimal designs.[1] Furthermore, if no partition satisfying these properties exists for a particular game, then this approach may not be suitable, as the primitives do not support a well-defined information coarsening rule based on partitions.

To explain our result, we first need to be specific about what we mean by "equilibrium selection". Our implicit assumption is that each receiver's equilibrium behavior is determined by some exogenous mechanism (such as environment or culture) that selects the best responses based on their *conjecture*. The concept of conjecture, borrowed from epistemic games, refers to a belief about the underlying state, others' belief hierarchies, and actions, which captures receivers' perceptions of both basic and strategic uncertainty. Selecting receivers' best response based on conjectures is thus conveniently modeled by the *selected best response correspondence*-a correspondence that maps each conjecture to the set of selected best responses for each receiver. This framework can describe several equilibrium selections in the literature.[2] Since this modeling approach allows receivers' preferences to directly depend on their higher-order beliefs, it also accounts for receivers' best responses when considering their psychological utility. Our solution concept is perfect Bayesian equilibrium with the additional constraint that the receivers' strategy profile must respect the selected best responses.

Specifically, our main result characterizes the conditions on the primitives under which the sender can construct a partition over conjectures such that, in any perfect Bayesian equilibrium, the sender can categorize information according to this partition. Only the category information is disclosed, and the receivers' equilibrium actions remain unchanged. Our characterization relies on the coarse information model called "frames" introduced by Chen et al. (2017). In our setting, a frame is a profile of a partition of conjectures-one for each receiver-such that whenever two conjectures agree on the probability concerning the events in the frame, they must be located in the same component of that receiver's partition. This imposes an implicit measurability condition: each event in receiver i's partition can be interpreted as a statement about receiver i's belief concerning the events in receiver $-i$'s partition. Building on the construction of finite-order frames in Chen et al. (2017), we define a "convex frame" as a partition constructed by recursively separating conjectures according to a convexity criterion. This partition has two key features: convexity and iterative measurability, which, as we explain in detail in the main text, are crucial for categorizing information without altering the outcome.

The selected best response correspondence naturally induces a partition of the receivers' conjectures according to their selected best responses. We refer to this as the basic partition, which is determined by the underlying primitives. Our main result shows that if this partition can be further refined by a convex frame, as defined above, then any perfect Bayesian equilibrium outcome can be implemented through a generalized direct information structure. In this framework, a generalized direct information structure is defined using an arbitrary partition of conjectures. It recommends to each receiver the specific partition component containing their conjecture, along with an incentive-compatible action. We further demonstrate that any convex frame refining the basic partition can serve as the basis for such coarsened disclosure. In this way, our result extends the standard direct approach—which, following the revelation principle, categorizes information solely by equilibrium actions—to a more flexible framework.

As mentioned, the main finding facilitates a generalized direct approach to computing optimal information structures in this setting. To illustrate its usefulness, we apply this approach to study information disclosure to psychological receivers with altruistic traits in Section 4. While altruistic traits are widely acknowledged in human behavior (see, for example, Fehr and Fischbacher, 2003 and Rand et al., 2012), there has been little work in the persuasion literature examining how this trait influences information disclosure by affecting receivers' interactions. The generalized direct approach derived from our main result, by accounting for receivers' higher-order beliefs, can be naturally applied to better understand how to leverage receivers' anticipation of others' altruistic behavior in this class of problems. This section also includes a second application, revisiting the leading example in Morris et al. (2024), along with additional examples from the literature where receiver behavior reflects psychological considerations or where equilibrium selection is arbitrary, helping to illustrate the broader relevance of our result. Section 5.2 further discusses the scope of this approach.

1.1. Literature review

This paper builds on the information design and Bayesian persuasion literature pioneered by Kamenica and Gentzkow (2011). In this literature, a sender commits to a stochastic communication device that discloses information about the underlying state to receivers whose actions depend on that information. With multiple interacting receivers, communication devices may also play a coordination role, as in Aumann (1987) and Bergemann and Morris (2016).

Our main contribution is to the literature on direct approaches in general information design settings. Under sender-preferred equilibrium selection, the standard revelation principle, introduced by Myerson (1991) and extended to information design in, among others, Kamenica and Gentzkow (2011), Bergemann and Morris (2016), and Taneva (2019), implies that, without loss of generality, the sender can communicate only through recommendations of equilibrium actions. Our setting departs from this benchmark in two ways: equilibrium selection need not favor the sender, and receivers may have psychological preferences. In this environment, richer

[1] See Section 5.1 for a detailed discussion.

[2] This includes the sender-preferred selection (e.g., Kamenica and Gentzkow, 2011), the sincere voting rule (e.g., Alonso and Camara, 2016), the sender-worse equilibrium selection in binary-action supermodular games (e.g., Morris et al., 2024), and others; More examples are provided in Section 4.

messages can affect strategic outcomes even when they induce the same recommended actions. While Mathevet et al. (2020) also study a setting with non-sender-preferred selection, their objective is different: they extend the concavification approach, whereas our focus is on generalizing the direct approach in a broader setting.

The paper is also related to the literature on persuasion with psychologically motivated receivers. Ely et al. (2015) and Lipnowski and Mathevet (2018) analyze persuasion problems with a single psychological receiver. By contrast, we study strategic interaction among multiple receivers. Our use of selected best-response correspondences provides a flexible, reduced-form approach to imposing behavioral restrictions that capture psychological motives while preserving tractability. In Section 4, we illustrate this approach in a disclosure problem with altruistic receivers, motivated by the experimental evidence in Fehr and Fischbacher (2003) and Fehr and Fischbacher (2004). Relatedly, Alonso and Camara (2016) study persuasion under the sincere voter rule in voting environments; our discussion section offers an alternative perspective and method for analyzing that class of problems.

More broadly, the paper speaks to the equilibrium-selection issue in games with strategic uncertainty. Although there is no consensus on equilibrium selection (see Samuelson, 1998), our framework provides a parsimonious way to encode classes of selection rules relevant to information design. This perspective is also useful for adversarial information design. In particular, Morris et al. (2024) characterize optimal information policies in binary-action supermodular games under sender-worst equilibrium selection. In Section 4.2, we revisit their leading example and provide an alternative solution. Although our method may generate redundant messages in that setting, it does not rely on binary actions or supermodularity and is therefore complementary to existing approaches.

The modeling also draws conceptual insight from the epistemic game theory literature. Works such as Ely and Pęski (2006), Dekel et al. (2007), and Liu (2009) develop canonical representations of both basic underlying uncertainties and the strategic uncertainty generated by higher-order reasoning. We adopt this perspective by working directly with the corresponding canonical object in our setting, i.e., the conjecture space, and by studying how information can be coarsened on it. In particular, the "finite-order frame" construction in Chen et al. (2017) motivates our partition-based approach to coarsening information while preserving strategic outcomes.

Organization: The remainder of this paper is organized as follows: Section 2 introduces the model. Section 3 presents our main result; Section 4 applies our result to applications. Section 5 discusses the scope of our result. Section 6 concludes. The Appendix collects all proofs.

2. The model

The primitives. There is a sender and finitely many receivers; denoted the set of receivers as I. Each receiver i possesses a finite action set A_i. Let $A := \prod_{i\in I} A_i$ be the product action space. The underlying state, which all players care about, takes values in a finite set Ω, and is distributed according to a probability measure μ_0. The sender has the ability to influence receivers through information structures. The message space M_i is fixed for each receiver i, with each M_i a large enough Polish space.[3] Denote their product space by $M := \prod_{i\in I} M_i$. An information structure is a conditional distribution $\pi : \Omega \to \Delta(M)$. Let Π collect all possible information structures. The sender maximizes their ex ante expected utility, with their Bernoulli utility function denoted as $u_s : \Omega \times A \to \mathbb{R}$.

Each receiver privately observes their message. We use type space to model each receiver's perception of the strategic circumstance after observing the sender's message. A type space is a tuple $(T_i, \zeta_i)_{i\in I}$ where T_i is a measurable space that indicates receiver i's type, and $\zeta_i : T_i \to \Delta(\Omega \times A_{-i} \times T_{-i})$ is a measurable mapping that associates each type t_i with its belief about the payoff-relevant parameters, including the state, opponents' actions, and the types of other players. For any type $t_i \in T_i$, we say the belief $\zeta_i(t_i)$ is *consistent with* that type. Section 2.1 details the structure of a type space and how each receiver forms their type based on the sender's information structure and the given strategy profile via Bayes' rule. Following the literature, we refer to an element in $V_i := \Delta(\Omega \times A_{-i} \times T_{-i})$ as a receiver i's *conjecture*, and use v_i to denote a generic conjecture.

We allow a general utility of each receiver i, which may directly depend on the underlying state, the actions taken by all receivers, the types of other receivers, and their own beliefs. Thus, receivers' preferences may exhibit psychological traits as in Geanakoplos et al. (1989) and depart from von Neumann—Morgenstern utilities. Instead of specifying receiver preferences in detail, we encompass each receiver's preference and the selection constraints imposed on their behavior through a selected best response correspondence $\mathcal{A}_i : V_i \to 2^{A_i}$. With conjecture v_i, receiver i would choose an action within the selected best responses characterized by $\mathcal{A}_i(v_i)$. This formulation can serve as the best response correspondence in psychological games and can also capture certain classes of equilibrium selection rules found in the literature. Section 4 provides further examples, detailing underlying preferences and selection rules, and illustrates how to transform them into the above correspondences. Section 5.2 discusses the scope of this formulation. Hence, the following parameters summarize the primitives:

$$(I \cup \{s\}, \Omega, (M_i)_{i\in I}, \Pi, u_s, (A_i)_{i\in I}, (T_i, \zeta_i)_{i\in I}, (\mathcal{A}_i)_{i\in I}).$$

To demonstrate the receiver's selected best response correspondences in the classical Judge-Prosecutor example of Kamenica and Gentzkow (2011), recall that the conjecture in a single-agent setting is the beliefs over the states Ω. Let μ^* be the threshold belief at which the Judge is indifferent between "convict" and "acquit". Under the sender-preferred selection, the selected best response correspondence for the receiver (i.e., the Judge) is $\mathcal{A}_i(\mu) = \{\text{convict}\}$ if the receiver's posterior belief μ satisfies $\mu(\{\omega = \text{gulity}\}) \geq \mu^*$ and $\mathcal{A}_i(\mu) = \{\text{acquit}\}$ otherwise.

[3] For our purposes, it suffices to consider a Polish space that includes the space of conjectures-that is, beliefs regarding the underlying state, others' types, and others' actions-as well as mixed action recommendations. We will elaborate on the concept of conjectures later in this section.

Timing. This game proceeds as follows: (i) The sender chooses an information structure $\pi \in \Pi$; (ii) Nature picks a state ω and, given ω, a message profile $m = (m_i)_{i\in I}$ according to π; (iii) each receiver i privately observes the sender's message m_i, forms their type according to Bayes' rule (details in Section 2.1); (vi) receivers simultaneously choose actions from the selected best responses $\mathcal{A}_i$ given their type-associated belief, then payoffs are realized.

Solution concept. We employ the notion of perfect Bayesian equilibrium in this game.[4] A strategy for any receiver $i \in I$ is a measurable mapping $\sigma_i : \Pi \times M_i \to \Delta(A_i)$.[5] Under a strategy profile σ, a *consistent* belief map $\beta_i : \Pi \times M_i \to V_i$ maps any information structure and any message realization to receiver i's conjecture, consistent with their type given others' strategies σ_{-i}, using Bayes' rule no matter whether they are on or off the equilibrium path. When Bayes' rule cannot be applied, the conjecture will be assigned arbitrarily. To ease the notation, for each $m_i \in M_i$, we shall often write $\sigma_i^\pi(\cdot \mid m_i)$ and $\beta_i^\pi(\cdot \mid m_i)$ instead of the more cumbersome $\sigma_i(\pi, m_i)(\cdot)$ and $\beta_i(\pi, m_i)(\cdot)$.

Definition 1. A strategy profile $(\pi, (\sigma_i, \beta_i)_{i\in I})$ is a perfect Bayesian equilibrium ("PBE" for short) if and only if (i) each receiver $i \in I$ holds a consistent belief map β_i given $(\sigma_i)_{i\in I}$; (ii) the receiver's strategy respects the selected best response correspondence in all cases: $\sigma_i^{\pi'}(\mathcal{A}_i(\beta_i^{\pi'}(\cdot \mid m_i')) \mid m_i') = 1$ for any $\pi' \in \Pi$ and $m_i' \in M_i$; and (ii) π maximizes the sender's ex-ante payoff given $(\sigma_i)_{i\in I}$.

Given any information structure π and receivers' strategy profile $(\sigma_i)_{i\in I}$, we say the induced joint distribution of the state and receivers' actions $\Delta(\Omega \times A)$ is its *outcome*.

2.1. Beliefs and types

A type space $(T_i, \zeta_i)_{i\in I}$ provides an implicit description of each receiver's beliefs about their current strategic environment, including other players' (higher-order) beliefs and so on. It is known that every type space can be uniquely mapped into the *universal type space* by a belief-preserving mapping. We will therefore identify our type space with the universal type space, the construction of which is reviewed here (see more details in Mertens and Zamir, 1985 and Brandenburger and Dekel, 1993).

Let Θ_i be a finite parameter space that captures the uncertainty receiver i faces, with the specific details to be provided below. Recall that the first-order belief of receiver i is an element of $T_i^1 := \Delta(\Theta_i)$. A k-order belief for $k \geq 2$ is an element of $T_i^k := \Delta(\Theta_i \times \prod_{l=1}^{k-1} T_{-i}^l)$. Such k can be taken to infinity. Say an infinite hierarchy of beliefs $t_i = (t_i^1, t_i^2, \ldots) \in \prod_{k=1}^\infty T_i^k$ is *coherent* if for every $k \geq 1$, the projection of t_i^k to its lower-order belief space T_i^{k-1} is compatible with the corresponding lower-order beliefs in this given belief hierarchy, i.e., $\mathrm{proj}_{T_i^{k-1}} t_i^k = t_i^{k-1}$.

Let T_i^* denote the set of receiver i's infinite belief hierarchies that satisfy coherency and common knowledge of coherency, that is, coherent belief hierarchies in which everyone knows that everyone's belief hierarchy is coherent; everyone knows that everyone knows this; and so on, ad infinitum.[6] Given the profile $(T_i^*)_{i\in I}$, there exists an isomorphism $\zeta_i^* : T_i^* \to \Delta(\Theta_i \times T_{-i}^*)$ with the belief-preserving property that each type t_i uniquely pins down the belief regarding Θ_i and conversely, each belief uniquely determines the type. The space (T_i^*, ζ_i^*) is referred to as the universal type space.

We now illustrate how each receiver forms their type after observing the sender's information under a given strategy profile. In our setting, the uncertainty space $\Theta_i := \Omega \times A_{-i}$ includes each receiver's basic uncertainty and strategy uncertainty.[7] Given the priors, the information structure, and the receivers' strategy profile, a consensus exists on how everyone should update their first-order beliefs using Bayes' rule upon receiving any specific message. Furthermore, Bayes' rule also determines the receiver's belief regarding Θ_i and others' first-order beliefs given this message, extending to higher-order beliefs. Consequently, any receiver's hierarchy of beliefs following any message is coherent, and such coherency in everyone's hierarchy of beliefs is common knowledge. As such, we henceforth equate the previous type space $(T_i, \zeta_i)_{i\in I}$ with $(T_i^*, \zeta_i^*)_{i\in I}$ by identifying types with their induced belief hierarchies that satisfy common knowledge of coherency.[8]

---

[4] For further discussion on perfect Bayesian equilibrium in the context of persuasion, see Lipnowski et al. (2026) and Wu (2023).

[5] Since the message product space M is a Polish space, the space of probability measures $\Delta(M)$ is also a Polish space with the Lévy-Prokhorov metric. The space of information structures $\Pi = \prod_{\omega\in\Omega} \Delta(M)$ is a finite product of Polish spaces and is therefore itself a Polish space. The corresponding σ-algebra is the induced Borel σ-algebra of this product space. Hence, the measurability of σ_i is with respect to the product σ-algebra on $\Pi \times M_i$.

[6] The explicit construction of T_i^* is as follows: Let $T_i^{c,1}$ denote the set of receiver i's coherent infinite belief hierarchies. By Lemma 1 in Brandenburger and Dekel (1993), there exists a homeomorphism $f_i : T_i^{c,1} \to \Delta(\Theta_i \times \prod_{l=1}^\infty T_{-i}^l)$. For any $k \geq 2$, define

$$T_i^{c,k} := \{t_i \in T_i^{c,1} \mid f_i(t_i)(\Theta_i \times T_{-i}^{c,k-1}) = 1\}.$$

Then the set T_i^* is given by the intersection $T_i^* = \cap_{k=1}^\infty T_i^{c,k}$.

[7] Including players' actions in the parameter space is not new. In fact, Mertens and Zamir (1985) allow a full list of the strategy space to be part of the space of parameters upon which the infinite hierarchy of beliefs is constructed.

[8] A related concern in Bayesian games is that passing to the universal type structure need not be without loss for equilibrium predictions. In particular, Friedenberg and Meier (2017) show that the Equilibrium Extension Property may fail: an equilibrium on a given type structure need not extend to a larger type structure posited by the analyst (which may be misspecified), including the universal type structure. Here, the universal type structure refers to the canonical terminal, non-redundant structure that represents all hierarchies of beliefs (see Definitions 8 and 9 in Friedenberg and Meier, 2017). This concern does not arise in the same way in our framework. Through the selected best-response correspondences $\mathcal{A}_i$, the conjecture space is the common strategically relevant object for both the designer and the players, and the equilibrium analysis is formulated directly on that space, rather than through an extension argument from an arbitrary type structure to the universal type structure. Likewise, players

2.2. The suboptimality of recommending actions: An example

When receivers have von Neumann-Morgenstern utilities and $\{\mathcal{A}_i\}_{i\in I}$ represents their best responses given each conjecture, Bergemann and Morris (2016) indicate that focusing on the sender recommending equilibrium actions is without loss of generality. However, such a focus may be suboptimal in general. We present a counterexample in which receivers exhibit psychological traits, and the sender cannot achieve the optimal outcome if they recommend only actions. This example will illustrate that when motivation is belief-driven, coarsening information based on equilibrium actions can alter beliefs in a way that undermines the motivation to take the desired action. Later, we will revisit it to illustrate our main result.

Example 1.

A benevolent sender seeks to promote pro-social behavior between two receivers by disclosing whether the environment is socially observable. The underlying state $\omega \in \{1,2\}$ represents the type of social context. In state $\omega = 1$, actions are publicly observable and subject to social judgment; in state $\omega = 2$, actions are private and carry no reputational consequences. The players share a common prior μ_0, with $\mu_0(\{\omega = 1\}) = \mu_0(\{\omega = 2\}) = 0.5$.

After observing the sender's chosen information structure and receiving private messages, each receiver i selects an action $a_i \in A_i = \{\text{generous act ("}g\text{"}), \text{selfish act ("}n\text{")}\}$. The sender's payoff is the number of receivers who choose the pro-social action: $v(a_I, \omega) = |\{i \in I \mid a_i = g\}|$, where $a_I = (a_1, a_2)$. In socially observable environments (i.e., $\omega = 1$), receivers are more likely to choose generous acts. We assume sender-preferred tie-breaking, that is, receivers choose g when indifferent.

Receiver 1 cares only about whether the environment is socially observable. Specifically, for any state ω, Receiver 1's utilities from actions g and n are given by:

$$u_1(\omega, g) = \mathbb{1}_{\{1\}}(\omega) - 9\mathbb{1}_{\{2\}}(\omega), \quad \text{and} \quad u_1(\omega, n) = 0.$$

Accordingly, Receiver 1's selected best response is $\mathcal{A}_1(\upsilon_1) = \{g\}$ if and only if their resulting conjecture υ_1 satisfies $\upsilon_1(\{\omega = 1\}) \geq 0.9$; otherwise, $\mathcal{A}_1(\upsilon_1) = \{n\}$.

Receiver 2 cares about the underlying state and also derives psychological utility from the other receiver's pro-social behavior, depending on their belief about the social observability of the environment. For each type t_2, the utilities from choosing g and n are given below:

$$u_2(t_2, g) = \big(\zeta_2(t_2)(\{a_1 = g\}) - 0.5\big)\big(\zeta_2(t_2)(\{\omega = 2\}) - 0.8\big), \quad \text{and} \quad u_2(t_2, n) = 0,$$

where ζ_2 is the mapping that assigns to each type its consistent conjecture.

To interpret, when Receiver 2's belief that $\{\omega = 2\}$ exceeds 0.8, their pro-social behavior is motivated by reciprocity. Specifically, they are willing to act generously if they believe the other receiver will do the same with at least 0.5 probability. In contrast, when their belief in social unobservability falls below 0.8, they derive negative psychological utility from mirroring the other's pro-social action. In this case, they view such behavior as performative, driven by a desire for social approval, and prefer to respond non-cooperatively.

Given Receiver 2's preference, their selected best response is $\mathcal{A}_2(\upsilon_2) = \{g\}$ if and only if either (i) $\upsilon_2(\{a_1 = g\}) \geq 0.5$ and $\upsilon_2(\{\omega = 2\}) \geq 0.8$, or (ii) $\upsilon_2(\{a_1 = g\}) \leq 0.5$ and $\upsilon_2(\{\omega = 2\}) \leq 0.8$; otherwise, $\mathcal{A}_2(\upsilon_2) = \{n\}$. Note that Receiver 2's psychological preference makes it insufficient to induce action g if they know only that either (i) or (ii) holds, without knowing which one specifically. This is because the incentive generated by belief-driven motivation in this case is non-convex. Coarsening information by pooling (i) and (ii) may shift the receiver's belief in a way that violates both conditions, thereby altering the motivation and leading to a different outcome.

To see this point explicitly, consider the following information structure π with six messages $\{m_1^1, m_1^2\} \times \{m_2^1, m_2^2, m_2^3\}$, where the first component is privately disclosed to Receiver 1 and the second to Receiver 2:

$$\begin{aligned} &\pi\big(m_1^1, m_2^1 | \omega = 1\big) = \frac{1}{36}, \pi(m_1^1, m_2^2 | \omega = 1) = \frac{8}{9}, \pi\big(m_1^1, m_2^3 | \omega = 1\big) = \frac{1}{12}, \\ &\pi\big(m_1^1, m_2^1 | \omega = 2\big) = \frac{1}{9}, \pi\big(m_1^2, m_2^2 | \omega = 2\big) = \frac{8}{9}. \end{aligned} \tag{1}$$

Using the approach developed from our main result, the above π is the optimal information structure for this example and yields a sender payoff of $\frac{109}{72}$. For Receiver 1, only one message is needed to induce action g. Upon receiving message m_1^1, their posterior belief is $\beta_1^\pi(\{\omega = 1\} \mid m_1^1) = 0.9$, under which g is the selected best response according to $\mathcal{A}_1$. Conversely, upon receiving message m_1^2, the belief becomes $\beta_1^\pi(\{\omega = 1\} \mid m_1^2) = 0$, leading to action n as the best response.

By contrast, Receiver 2 requires two messages to maximally induce action g. Upon receiving message m_2^1, their posterior beliefs are

$$\beta_2^\pi(\{\omega = 2\} \mid m_2^1) = 0.8, \quad \text{and} \quad \beta_2^\pi(\{a_1 = g\} \mid m_2^1) = 1,$$

under which (i) is satisfied and g is the selected best response according to $\mathcal{A}_2$. Similarly, message m_2^2 induces beliefs

$$\beta_2^\pi(\{\omega = 2\} \mid m_2^2) = 0.5, \quad \text{and} \quad \beta_2^\pi(\{a_1 = g\} \mid m_2^2) = 0.5.$$

reason about others' behavior through the same correspondences $\mathcal{A}_i$, and hence through the same conjecture space. Accordingly, the type-structure misspecification issue emphasized by Friedenberg and Meier (2017) does not arise in our setting.

This posterior belief satisfies (ii) of $\mathcal{A}_2$ and again makes g the best response. However, upon receiving message m_2^3, Receiver 2's beliefs become

$$\beta_2^{\pi}(\{\omega = 2\} \mid m_2^3) = 0, \quad \text{and} \quad \beta_2^{\pi}(\{a_1 = g\} \mid m_2^3) = 1,$$

which leads to action n under $\mathcal{A}_2$.

The outcome induced by the above information structure π cannot be replicated by a direct information structure that merely recommends equilibrium actions. To see this point more clearly, let π^d denote the "direct" information structure derived from π, in which each message is replaced by the corresponding action recommendation. For Receiver 2, this amounts to pooling messages m_2^1 and m_2^2 into a single recommendation to pro-social behaviors ("$\hat{g}$"). Note that the posterior beliefs under m_2^1 and m_2^2 satisfy conditions (i) and (ii) in $\mathcal{A}_2$, respectively. However, pooling these messages into a single recommendation alters Receiver 2's posterior to fail both conditions, since

$$\beta_2^{\pi^d}(\{\omega = 2\} \mid \hat{g}) = \frac{12}{23} < 0.8, \quad \text{and} \quad \beta_2^{\pi^d}(\{a_1 = g\} \mid \hat{g}) = \frac{37}{69} > 0.5.$$

Given these beliefs, the selected best response correspondence $\mathcal{A}_2$ indicates that Receiver 2 would not choose g. Thus, the action recommendation $\hat{g}$ under π^d fails to sustain the outcome achieved under π.

In the example above, the standard direct approach is inapplicable because Receiver 2's willingness to take pro-social actions depends crucially on belief-driven motivations. In this context, coarsening information based solely on equilibrium actions-as suggested by the direct approach-fails to satisfy the necessary regularity (such as convexity) to preserve the equilibrium outcome. Our main result identifies two key regularity conditions for preserving equilibrium outcomes under information coarsening, providing a foundation for extending the standard direct approach to more general settings in which players' behaviors may reflect psychological traits.

3. Categorizing information with partitions

Given the complexity of the conjecture space, using conjectures directly as messages may be intractable. In this section, we explore a general principle for categorizing information through partitions based on the primitives such that, for any equilibrium, the sender need only disclose partition-related information without changing the intended outcomes.

Definition 2 (Basic and Refined Partitions)**.** We take the view that the selected best response correspondence $\mathcal{A}_i$ is a single-valued function from V_i to the finite power set 2^{A_i}, rather than a multivalued function. Accordingly, its inverse $\mathcal{A}_i^{-1}$ is defined as the inverse of this function: for any $S_i \in 2^{A_i}$, we set $\mathcal{A}_i^{-1}(S_i) := \{v_i \mid \mathcal{A}_i(v_i) = S_i\}$. Under this definition, $v_i \in \mathcal{A}_i^{-1}(S_i)$ if and only if the conjecture v_i supports S_i as the set of selected best responses under $\mathcal{A}_i$.

As $\mathcal{A}_i$ maps each conjecture to a subset of A_i, its inverse $\mathcal{A}_i^{-1}$, as defined above, partitions the conjecture space by grouping conjectures that share the same set of selected best responses. We refer to this partition as the *basic partition*, denoted by $\{\mathcal{A}_i^{-1}(S_i) \mid S_i \subseteq A_i\}_{i \in I}$, where each element corresponds to a unique set of best responses specified by $\mathcal{A}_i$. We say that a partition of the conjecture space *weakly refines* the basic partition if each of its elements is contained in some element of the basic partition. Any such partition is referred to as a *refined partition*.

Under the standard direct approach, a direct obedient information structure recommends equilibrium actions. When a receiver receives such a recommendation, assuming that all others follow their recommendations, they form conjectures under which the recommended action is a best response. In this way, recommending an equilibrium action essentially identifies the component of the basic partition in which the receiver's conjecture is located.

Building on this interpretation, we propose a more general principle that categorizes information using refined partitions, allowing for the disclosure of more than just actions, while still preserving the equilibrium outcome under the selection constraint imposed by the primitives. Guided by this principle, we define a generalized direct obedient information structure as follows:

Definition 3 (Generalized direct obedient information structure)**.** Given a persuasion game, for each receiver $i \in I$, let P_i be a refined partition for receiver i. We say π is a *generalized direct obedient information structure* under $(P_i)_{i \in I}$ if:

(i) The message space is $(P_i \times \Delta(A_i))_{i \in I}$;
(ii) Given any message (p_i, α_i) with p_i a component of P_i and $\alpha_i \in \Delta(A_i)$ a mixed action, each receiver i, assuming that other receivers will obey their action recommendations, forms their posterior conjecture v_i by Bayes' rule and finds that v_i is located within the recommended component p_i;
(iii) The recommended action conforms the selected best responses given receiver i's conjecture v_i, that is, $\alpha_i(\mathcal{A}_i(v_i)) = 1$.

We say an outcome can be implemented by a generalized direct obedient information structure π under some refined partition if the given π, a profile of receivers' strategies that are (on-path) obedient, and the consistent belief maps constitute a PBE.

When receivers are von Neumann–Morgenstern utility maximizers, Bergemann and Morris (2016) show that for any fixed information structure, any Bayes Nash equilibrium (BNE) in the receivers' game induces a Bayes correlated equilibrium. This Bayes correlated equilibrium can be interpreted as a direct obedient information structure that recommends equilibrium actions to each receiver. Since Bergemann and Morris (2016) do not explicitly model a sender, to reinterpret their results within our framework,

we introduce a sender with a constant utility function. Under this interpretation, any Bayes Nash equilibrium under a fixed information structure π in their set-up can be extended to a perfect Bayesian equilibrium in our framework by including the information structure π, and by expanding each receiver's equilibrium strategy to account for the full domain $\Pi \times M_i$ of information structures and messages, along with the corresponding consistent belief mappings. This extension is subject to two additional requirements: (i) receivers update their beliefs using Bayes' rule both on and off the equilibrium path whenever possible, and (ii) they best respond to these beliefs, even off-path. Since these augmentations and constraints apply only to zero-probability events, the receivers' on-path equilibrium strategies remain unchanged.

Using our notion of basic partitions, the insight from Bergemann and Morris (2016) can be restated as follows: when $\mathcal{A}_i$ represents the best response correspondence for von Neumann–Morgenstern utility-maximizing receivers, any BNE outcome in the receivers' game—when extended to a PBE in our framework with a sender who has constant utility, as described above—can be implemented by some direct obedient information structure based on the basic partition. This information coarsening is possible because the basic partition satisfies certain regularity conditions under the above $\mathcal{A}_i$.

The main research question we address in this paper is: Under what conditions can a basic partition-or more generally, its refinement-serve as a valid information coarsening rule in a given game, allowing the resulting generalized direct obedient information structures to implement any possible PBE outcome?

3.1. A special class of partitions: The convex frames

A natural starting point for our investigation is the coarse information model known as "the frames", introduced by Chen et al. (2017). To suit our purposes, we extend this concept to the conjecture space through belief-preserving mappings.

Definition 4 (Frames). Given the type space $(T_i, \zeta_i)_{i\in I}$ in the primitive, we define a frame to be a profile $P = (P_i)_{i\in I}$, where each P_i, consisting of Borel sets in V_i, is a partition of V_i such that for each $v_i, v_i' \in V_i$, if

$$v_i\left(\omega, a_{-i}, \prod_{j\neq i}\zeta_j^{-1}(p_j)\right) = v_i'\left(\omega, a_{-i}, \prod_{j\neq i}\zeta_j^{-1}(p_j)\right)$$

for any $\omega \in \Omega$, $a_{-i} \in A_{-i}$, and $p_j \in P_j$, $j \neq i$, then v_i, v_i' must locate in the same component of the partition P_i, that is, $v_i, v_i' \in p_i$ must hold for some component $p_i \in P_i$.

In other words, a Borel partition profile is a frame if any two conjectures of player i that agree on their beliefs concerning the state, opponents' actions and the (projected) events in the partition of their opponents must belong to the same component. As interpreted in Chen et al. (2017), the frame serves as a "self-contained" coarse information model with implicitly imposed measurability conditions.

As discussed in Example 1, preserving the equilibrium outcome under information coarsening requires certain regularity conditions. Convexity, in particular, is an intuitive property to impose in this context. To see why, consider persuading a single receiver: under an information structure π, the receiver forms a posterior belief μ after receiving message m and μ' after receiving message m'. If we merge messages m and m', the receiver who receives the combined message ("either m or m'") will form a belief that is a convex combination of μ and μ'. To prevent this coarsening from altering the selected best response, each component of the frame must exhibit a certain degree of convexity.

Beyond convexity, iterative measurability is equally important for preserving equilibrium outcomes under information coarsening. This property is best described through the construction of *convex frames*, which play a central role in our main result. A convex frame is a particular type of frame, generated through a recursive partitioning process—formally defined below—that satisfies both convexity and measurability criteria. These key properties will become apparent through the construction.

3.1.1. Recursive partitioning procedure

The procedure for constructing a convex frame draws on insights from the construction of finite-order frames in Chen et al. (2017). Before presenting the procedure, we first introduce a stronger notion of convex sets that it relies on: We say a Borel set $B \subseteq \mathbb{R}$ is a *bi-convex* set if both B and its complement B^c are convex sets. The following lemma shows that a bi-convex set in $\mathbb{R}$ takes a relatively simple structure:

Lemma 1. *A Borel set $B \subseteq \mathbb{R}$ is a bi-convex set if and only if B takes one of the following forms:*

$$(-\infty, b), (-\infty, b], (b, \infty) \text{ and } [b, \infty) \text{ for some } b \in \mathbb{R}. \tag{2}$$

First-order convex frames. For each receiver $i \in I$, a *first-order characteristic pair* consists of a bi-convex set B_i^1 and a bounded function f_i^1 measurable with respect to the finite σ-algebra generated by $\Omega \times A_{-i}$. This measurability requires that f_i^1 takes the form of $f_i^1 = \sum_{l=1}^{L} c_l \cdot \mathbb{1}_{(\omega_l, a_{-i}^l)}$ for some integer L, with each $\omega_l \in \Omega$, $a_{-i}^l \in A_{-i}$, and $c_l \in \mathbb{R}$. Any such pair defines a partition $\widehat{P}_i := \{\widehat{p}_i, V_i \backslash \widehat{p}_i\}$ with at most two elements, where

$$\widehat{p}_i := \left\{ v_i \in V_i \mid \int f_i^1 \mathrm{d}v_i \in B_i^1 \right\}.$$

There may be multiple first-order characteristic pairs for each player i. A *first-order convex frame* refers to a joint partition generated by the binary partitions induced by each pair in a collection of first-order characteristic pairs. Let $\mathscr{D}_i^1$ denote the index set of characteristic pairs used to construct the frame.

For example, let us consider a two-receiver game $I = \{1, 2\}$ and the following set of first-order characteristic pairs $\{(B_i^1, f_i^1)\}_{i\in I}$, where the function $f_i^1 := 2 \cdot \mathbb{1}_{(\omega, a_{-i})}$ for a fixed ω and a_{-i} and the set $B_i^1 = (0.5, \infty)$ for each player $i \in I$. From this characteristic pair, we can construct the first-order convex frame $(P_i)_{i\in I} = (\{p_i, V_i \backslash p_i\})_{i\in I}$, where:

$$p_i := \left\{ v_i \in V_i \mid \int f_i^1 \in B_i^1 \right\} = \{v_i \mid 2v_i(\omega, a_{-i}) > 0.5\}. \tag{3}$$

Given a first-order convex frame, such as $(P_i)_{i\in I}$ above, we can recursively construct a second-order convex frame building on this lower-order structure as follows: fix another function $f_i^2 := 3 \cdot \mathbb{1}_{(\omega', \zeta_{-i}^{-1}(p_{-i}), a'_{-i})}$ with $(\omega', a'_{-i}) \in \Omega \times A_{-i}$, and $B_i^2 := (-\infty, 0.3]$ for each $i \in I$. Here, ζ_{-i} is the isomorphism that maps player $-i$'s types to their consistent conjectures. Consider the binary partition $P_i' := \{p_i', V_i \backslash p_i'\}$, where

$$p_i' := \left\{ v_i \in V_i \mid \int f_i^2 \in B_i^2 \right\} = \{v_i \mid 3v_i(\omega', \zeta_{-i}^{-1}(p_{-i}), a'_{-i}) \le 0.3\}.$$

The joint of the above partitions P_i and P_i', denoted as P_i'', partitions V_i into at most four second-order measurable events based on whether a conjecture belongs to each of the events p_i and p_i'. Thus, the profile $(P_i'')_{i\in I}$ is an example of a second-order convex frame.

k-order convex frames. In general, a *k-order convex frame* can be constructed from a similar recursive process in the above example: Given a $k-1$-order frame $(P_i^{k-1})_{i\in I}$ with $k \ge 2$, a *k-order characteristic pair* consists of a bi-convex set $B_i^k \subseteq \mathbb{R}$ and a bounded function f_i^k measurable with respect to the σ-algebra generated by $\Omega \times \zeta_{-i}^{-1}(P_{-i}^{k-1}) \times A_{-i}$. This means, f_i^k must take the form of $f_i^k = \sum_{l=1}^{L} c_l \cdot \mathbb{1}_{(\omega_l, \zeta_{-i}^{-1}(p_{-i}^l), a_{-i}^l)}$ for some integer L, with $\omega_l \in \Omega$, $p_{-i}^l \in P_{-i}^{k-1}$, $a_{-i}^l \in A_{-i}$, $c_l \in \mathbb{R}$. We refer to this requirement on f_i^k as the *iterative measurability* condition. Alongside the convexity property, it forms the second key property underlying our main result for coarsening information while preserving the equilibrium outcome.

Each k-order characteristic pair defines an (at most binary) partition on conjectures $\widehat{P}_i' := \{\hat{p}_i', V_i \backslash \hat{p}_i'\}$, where

$$\hat{p}_i' := \left\{ v_i \in V_i \mid \int f_i^k \mathrm{d}v_i \in B_i^k \right\}.$$

There may be multiple k-order characteristic pairs for each player i. A *k-order convex frame* is a joint partition formed by combining the binary partitions $\widehat{P}_i'$ induced by each member in a set of k-order characteristic pairs, constructed based on an arbitrarily given $(k-1)$-order frame $(P_i^{k-1})_{i\in I}$. Let $\mathscr{D}_i^k$ denote the index set of these characteristic pairs. This construction procedure, which generates a sequence of successively finer partitions, can be continued for an arbitrary number of finite rounds or even indefinitely to achieve an infinite-order convex frame.

Definition 5 (Convex frames)**.** A *convex frame* is a partition profile constructed from the recursive partitioning procedure described above, based on any arbitrarily given set of characteristic pairs $\{(B_{i,h}^k, f_{i,h}^k) \mid h \in \mathscr{D}_i^k, k \in \mathbb{N}\}_{i\in I}$.

By construction, the convex frame satisfies two key properties: convexity and iterative measurability. The next section will show that these properties together provide a sufficient (and tight) condition for supporting the generalized direct principle introduced at the beginning of this section.

3.2. The main result

Given any game, recall that the selected best response correspondence $\mathcal{A}_i$, which captures both the equilibrium selection and receivers' psychological preferences in the game, is a primitive element of the model. From Definition 2, the basic partition is the partition generated by $\mathcal{A}_i$, which categorizes conjectures according to the sets of best responses specified by the inverse of $\mathcal{A}_i$. Hence, the basic partition can also be regarded as a primitive of the game. Our main result states the following:

Theorem 1. *Fix an arbitrary game and recall that its basic partition is uniquely determined by the primitives of the game. Suppose there exists a convex frame $(P_i)_{i\in I}$ that weakly refines this basic partition. Then, for any perfect Bayesian equilibrium $(\pi, (\sigma_i, \beta_i)_{i\in I})$, the resulting outcome can be implemented by a generalized direct obedient information structure under $(P_i)_{i\in I}$.*
Specifically, such a generalized direct obedient information structure can be constructed by disclosing, to each receiver i, the element of P_i that contains the conjecture induced by the equilibrium message under π, along with the corresponding equilibrium action recommendation.

The above result essentially states that, in a given game, if the basic partition is itself a convex frame, then the standard direct approach applies. Since the basic partition is directly identifiable from the given primitives, this is a relatively straightforward exercise: we need only verify whether the boundaries of the partition components are characterized in the affine form by the characteristic pairs, as described in Section 3.1.1. However, if the basic partition does not form a convex frame—as the two key conditions are tight (see Section 5.1)—the standard approach may fail. Nevertheless, our main result offers a way forward: by identifying a convex frame that refines the basic partition, it becomes possible to implement the outcome of any perfect Bayesian equilibrium by categorizing information according to this convex frame and recommending both the component containing the equilibrium conjecture and the corresponding equilibrium action. This provides a generalized version of the obedience constraint, extending the standard formulation in Bergemann and Morris (2016) to settings where it may not apply.

From this result, the ability of coarsening information based on the underlying partition while preserving the equilibrium outcome hinges critically on two properties: convexity and iterative measurability, both rooted in the construction of a convex frame. As

discussed in Section 3.1, convexity provides the necessary regularity for information coarsening. Iterative measurability is equally important, as it ensures that changes in one receiver's belief, even if they affect others' higher-order beliefs, do not alter the relevant component at higher levels in all players' perceptions under the coarsened information structure. Moreover, these conditions are tight: Section 5.1 presents counterexamples showing that if either property fails, the sender's ability to coarsen information as described in the main result may break down. Appendix A provides the formal proof of this main theorem.

Our main result motivates a *generalized direct approach* to solving persuasion problems, which proceeds as follows. Starting from the game's primitives, one first identifies the basic partition determined by preferences and equilibrium selection. The next step is to check if the basic partition is a convex frame. If not, one then tries to construct a convex frame that refines this partition. Recall that the revelation principle prescribes a universal rule for coarsening information—merging messages according to equilibrium actions. In contrast, the generalized direct approach does not propose a one-size-fits-all method. Instead, it offers flexibility: if a convex frame weakly refining the basic partition can be found, it yields a valid, though possibly non-unique, information coarsening rule tailored to the specific game. In this case, our main result shows that it is without loss of generality to focus on generalized direct obedient information structures based on this refined frame. However, if no such convex frame exists, this approach may not be suitable for the game in question.

To illustrate the procedure, we begin by revisiting Example 1 and then explore two applications in Section 4. Section 5.2 further discusses the scope of the approach.

Example 2. [Example 1 continued] We will start by specifying the basic partition induced by the selected best response correspondence in this example. For Receiver 1, the basic partition induced by their preference is $\{p_1^g, V_1 \backslash p_1^g\}$, where the set $p_1^g := \{v_1 \mid v_1(\{\omega = 1\}) \geq 0.9\}$ contains conjectures that lead to action g, and all other conjectures lead to action n. This is a first-order convex frame induced by the first-order characteristic pair $([0.9, \infty), \mathbb{1}_{\{\omega=1\}})$.

For Receiver 2, the basic partition is $\{p_2^g, V_2 \backslash p_2^g\}$, where the component associated with action n under $\mathcal{A}_2$ is defined as

$$p_2^g := \left\{ v_2 \in V_2 \mid \begin{array}{l} v_2(\{\omega = 2\}) \geq 0.8 \text{ and } v_2(\{a_1 = g\}) \geq 0.5 \\ v_2(\{\omega = 2\}) \leq 0.8 \text{ and } v_2(\{a_1 = g\}) \leq 0.5 \end{array} \right\}, \tag{4}$$

and the remaining component, $V_2 \setminus p_2^g$, is associated with action n.

Consider the following first-order characteristic pairs for Receiver 2:

$$([0.8, \infty), \mathbb{1}_{\{\omega=2\}}),\ ((-\infty, 0.8], \mathbb{1}_{\{\omega=2\}}),\ ([0.5, \infty), \mathbb{1}_{\{a_1=g\}}) \text{ and } ((-\infty, 0.5], \mathbb{1}_{\{a_1=g\}}).$$

It is readily verified that the partition induced by the above characteristic pairs forms a first-order convex frame that refines Receiver 2's basic partition. Specifically, let

$$\begin{aligned} p_2^1 &:= \{v_2 \mid v_2(\{\omega = 2\}) > 0.8\}, \quad p_2^2 := \{v_2 \mid v_2(\{\omega = 2\}) = 0.8\}, \\ p_2^3 &:= \{v_2 \mid v_2(\{\omega = 2\}) < 0.8\}, \quad p_2^4 := \{v_2 \mid v_2(\{a_1 = g\}) > 0.5\}, \\ p_2^5 &:= \{v_2 \mid v_2(\{a_1 = g\}) = 0.5\}, \quad p_2^6 := \{v_2 \mid v_2(\{a_1 = g\}) < 0.5\}. \end{aligned} \tag{5}$$

The joint partition in (5) refines Receiver 2's basic partition. Guided by our main result, we focus on a generalized direct obedient information structure using this joint partition for Receiver 2 and the basic partition for Receiver 1. This leads to the computation that the following information structure is optimal:

$$\begin{aligned} &\pi^*\big((p_1^g, g), (p_2^2 \cap p_2^4, g) \mid \omega = 1\big) = \frac{1}{36}, \quad \pi^*\big((p_1^g, g), (p_2^3 \cap p_2^5, g) \mid \omega = 1\big) = \frac{8}{9}, \\ &\pi^*\big((p_1^g, g), (p_2^3 \cap p_2^4, n) \mid \omega = 1\big) = \frac{1}{12}, \quad \pi^*\big((p_1^g, g), (p_2^2 \cap p_2^4, g) \mid \omega = 2\big) = \frac{1}{9}, \\ &\pi^*\big((V_1 \backslash p_1^g, n), (p_2^3 \cap p_2^5, g) \mid \omega = 2\big) = \frac{8}{9}. \end{aligned} \tag{6}$$

Note that the above information structure is essentially the same as the one in (1), which, as shown in Example 1, cannot be implemented by a direct structure that only recommends equilibrium actions. However, our main result allows it to be implemented through a generalized direct information structure.

4. Illustrative applications

In this section, we apply the generalized direct approach to two illustrative applications and conclude with a brief discussion of additional examples in the literature where our approach is also applicable. The first application concerns information design in psychological games featuring altruistic punishment, while the second focuses on binary-action supermodular games under sender-worst selection.

4.1. Altruistic punishment

Human altruism is widely observed in practice: people frequently engage in activities such as volunteering, helping strangers, donating to charitable organizations, and joining rescue squads. However, despite being well-documented, this trait receives little

attention in the study of information design. In this section, we apply our main results to explore how to persuade interacting receivers when altruistic traits may influence their behavior.

The problem set-up is inspired by the third-party punishment experiments conducted in Fehr and Fischbacher (2004): A benevolent sender seeks to promote fairness in a community consisting of a set of receivers, denoted as I, which involves an allocator (Player 1), and $|I|-1$ bystanders (Players 2,..., $|I|$). The uncertainty that the sender can disclose is whether a social norm is fair in this community, i.e., $\omega \in \{0$ ("not fair"), 1 ("fair")$\}$. The game unfolds in two stages: In the first stage, the sender commits to an information structure that privately communicates to each receiver whether the social norm is fair. In the second stage, all receivers act simultaneously.

In the original experiment, the allocator can propose any division of a pie to a passive recipient who has no choice but to accept. Bystanders observe the proposal. If they perceive the proposal as unfair, they can choose to punish the allocator at their own expense.[9] Although purely self-interested bystanders would have no incentive to punish under this design, the experiment shows that when a fairness norm is salient, bystanders are more likely to punish unfair splits-a behavior known as "altruistic punishment".[10]

To simplify the analysis, we work with the reduced normal form of this dynamic game. In this reduced form, we further assume that if the proposal is deemed fair, the game ends and the bystanders take no action. Thus, after observing the sender's information, all players move simultaneously: the allocator chooses either a fair split ("F") or an unfair split ("U"), and each bystander submits a decision plan from the set $\{N$ ("not punish"), P ("punish")$\}$, indicating their intended response to an *unfair* proposal.

Bystanders care about both their monetary payoff and the fairness of the outcome, particularly when they hold a sufficiently strong belief that a fairness norm applies. In such cases, they may derive utility from punishing unfair behavior. To model bystanders' preferences, we adopt the reciprocity framework in psychological games (see, for example, Battigalli and Dufwenberg, 2022). Each bystander j's utility consists of two components: a monetary cost of punishment and a psychological payoff that depends on their first-order belief $\mu_j \in \Delta(\Omega \times A_{-j})$. The monetary component is denoted by $-r_j(a_j)$, where choosing not to punish yields $r_j(N)=0$, and punishing results in a cost $-r_j(P)=-1<0$.

As described in Battigalli and Dufwenberg (2022), the following functional form in the utility can model the desire to reciprocate perceived unkindness through *sign-matching*:

$$\theta_j \cdot \kappa_{1j}(a_1) \cdot \kappa_{j1}(a_j, a_1, \mu_j), \tag{7}$$

where $\theta_j > 0$ is a constant sensitivity parameter. The expression (7) is positive only when κ_{1j} and κ_{j1} share the same sign. Specifically, we set $\kappa_{1j}(U)<0$ and $\kappa_{1j}(F)\geq 0$, indicating that unkind actions are perceived negatively while fair actions are perceived positively. Since the game ends after a fair proposal, we set $\kappa_{j1}(a_j, F, \mu_j)=0$ for all a_j and μ_j. Similarly, when the proposal is unfair and the bystander chooses not to punish, we normalize $\kappa_{j1}(N, U, \mu_j)=0$ for all μ_j. Thus, whenever $a_1=F$ or $a_j=N$ with $j\neq 1$, the expression (7)$=0$.

The findings from the altruistic-punishment experiment imply that the function $\kappa_{j1}(P, U, \mu_j)$ is always strictly negative, that is, $\kappa_{j1}(P, U, \mu_j')<0$ for any μ_j, and is weakly decreasing in the belief $\mu_j(\{\omega=1\})$, that is, if $\mu_j(\{\omega=1\}) \leq \mu_j'(\{\omega=1\})$, then

$$\kappa_{j1}(P, U, \mu_j') \leq \kappa_{j1}(P, U, \mu_j) < 0.$$

This decreasing property reflects the idea that the more strongly the bystander believes the norm is fair, the greater their inclination to reciprocate unkindness. The overall utility of bystander j is given by:

$$u_j(a_I, \mu_j) = -r_j(a_j) + \theta_j \cdot \kappa_{1j}(a_1) \cdot \kappa_{j1}(a_j, a_1, \mu_j). \tag{8}$$

By this specification, $u_j(a_I, \mu_j)=0$ whenever $a_j=N$. We assume the sender-preferred equilibrium selection: when indifferent, each bystander chooses to punish.

For demonstration purposes, we set $|I|=3$ and primarily work with the following specifications, although the results extend to any primitives satisfying the stated properties:

$$\theta_2=1, \quad \kappa_{12}(U)=-\frac{1}{3}, \quad \kappa_{21}(P, U, \mu_3)=-5\cdot\mu_2(\{\omega=1\}),$$

$$\theta_3=2.5, \quad \kappa_{13}(U)=-\frac{1}{8}, \quad \kappa_{31}(P, U, \mu_3)=-5\cdot\mu_3(\{\omega=1\})^2.$$

While the allocator receives a monetary payoff from their own proposal, denoted $r_1(a_1)$, with $0\leq r_1(F)=0.5<r_1(U)=1$, they may also incur both monetary and psychological costs from unfair proposals due to altruistic punishment imposed by each bystander j. The monetary cost is given by $-r_j(a_j)$, and the psychological disutility is captured by the term $\max\{0, \mu_1(\{a_j=P\})-d_j^*\}$, where $\mu_1 \in \Delta(\Omega \times A_{-1})$ is the allocator's first-order belief, and d_j^* is a fixed constant. This disutility reflects the allocator's negative experience when their expectation of avoiding altruistic punishment after an unkind act falls short of the reference level.[11]

Thus, the allocator's utility for each state ω, action profile a_I, and first-order belief μ_1 is given by:

$$u_1(\omega, a_I, \mu_1) = r_1(a_1) - \mathbb{1}_{\{U\}}(a_1) \cdot \sum_{j\neq 1} \left(\hat{\theta}_j \cdot \max\left\{0, \mu_1(\{a_j=P\}) - d_j^*\right\}\right), \tag{9}$$

[9] Specifically, for every dollar a bystander spends on punishment, the allocator loses \$3.

[10] According to Fehr and Fischbacher (2003), if a fairness norm applies to the situation, 55% of the third parties punish the allocator for transfers below 50 (out of 100).

[11] This psychological component is inspired by the modeling of disappointment in the psychological games literature; see Battigalli and Dufwenberg (2022) for a comprehensive review.

where $\hat{\theta}_j$ denotes the allocator's sensitivity to psychological disutility from Player j's altruistic punishment. In the following analysis, we set $\hat{\theta}_2 = 5$, and $\hat{\theta}_3 = 2.5$, with $d_2^* = 0.6$ and $d_3^* = 0.3$. The benevolent sender receives a payoff of 1 whenever a fair split is proposed, and 0 otherwise. All players share a common prior, denoted μ_0; specifically, let $\mu_0(\{\omega = 1\}) = 0.3$.

Analysis. With the utilities described above, Player 2 will punish the allocator after an unfair proposal if and only if their belief that the social norm is fair exceeds 0.6, that is, if their conjecture $\upsilon_2(\{\omega = 1\}) \geq 0.6$. Similarly, Player 3 will punish an unfair split if and only if their belief in the fairness of the social norm exceeds 0.8. Thus, the selected best response correspondences of the bystanders $\mathcal{A}_j$ are given by

$$\mathcal{A}_2(\upsilon_2) := \begin{cases} \{P\} & \text{if } \upsilon_2(\{\omega = 1\}) \geq 0.6; \\ \{N\} & \text{otherwise;} \end{cases} \text{ and } \mathcal{A}_3(\upsilon_3) := \begin{cases} \{P\} & \text{if } \upsilon_3(\{\omega = 1\}) \geq 0.8; \\ \{N\} & \text{otherwise.} \end{cases}$$

Based on the allocator's incentive, they will propose a fair split, $\mathcal{A}_1(\upsilon_1) = \{F\}$, if and only if one of the following conditions holds:

$$(i)\ \upsilon_1(\{a_2 = P\}) \geq 0.7 \,\&\, \upsilon_1(\{a_3 = P\}) < 0.3; \quad (ii)\ \upsilon_1(\{a_3 = P\}) \geq 0.5 \,\&\, \upsilon_1(\{a_2 = P\}) < 0.6;$$
$$(iii)\ 5\upsilon_1(\{a_2 = P\}) + 2.5\upsilon_1(\{a_3 = P\}) \geq 4.25 \,\&\, \upsilon_1(\{a_2 = P\}) \geq 0.6 \,\&\, \upsilon_1(\{a_3 = P\}) \geq 0.3.$$

For any allocator's conjecture υ_1 that does not satisfy the above conditions, $\mathcal{A}_1(\upsilon_1) = \{U\}$.

If the sender reveals no information, given the prior that the probability of $\{\omega = 1\}$ is 0.3, the allocator will propose an unfair split, and the sender's payoff will be 0. If the sender fully reveals the state, the fair split will be proposed with a probability of 0.3, resulting in an expected payoff of 0.3 for the sender.

As a first step of the generalized direct approach, we write down the basic partition for the bystanders (that is, the partition induced by $\mathcal{A}_i$ for $i = 2, 3$) as follows:

$$P_2 := \left\{ p_2^1 := \{\upsilon_2 \,|\upsilon_2(\{\omega = 1\}) < 0.6\};\ p_2^2 := \{\upsilon_2 \mid \upsilon_2(\{\omega = 1\}) \geq 0.6\} \right\};$$
$$P_3 := \left\{ p_3^1 := \{\upsilon_3 \,|\upsilon_3(\{\omega = 1\}) < 0.8\};\ p_3^2 := \{\upsilon_3 \mid \upsilon_3(\{\omega = 1\}) \geq 0.8\} \right\}.$$

In the above partition P_i for $i = 2, 3$, the component p_1^i is associated with the action N while p_2^i is associated with P. By definition, these partitions are first-order convex frames as they can be constructed from first-order characteristic pairs.[12]

As for the allocator, they are essentially concerned about how likely proposing an unfair split will incur altruistic punishment. Their concern therefore can be captured by the second-order convex frame recursively constructed from the above first-order frames $\{P_i\}_{i\in\{2,3\}}$ as follows: Recall that ζ_i is the isomorphism that links Player i's types with their consistent conjectures. Then:

$$\begin{aligned} P_1 := \Bigg\{ & p_1^1 := \{\upsilon_1 \mid \upsilon_1(\zeta_2^{-1}(p_2^2)) \geq 0.7 \,\&\, \upsilon_1(\zeta_3^{-1}(p_3^2)) < 0.3\}; \\ & p_1^2 := \{\upsilon_1 \mid \upsilon_1(\zeta_2^{-1}(p_2^2)) < 0.6 \,\&\, \upsilon_1(\zeta_3^{-1}(p_3^2)) \geq 0.5\}; \\ & p_1^3 := \left\{ \upsilon_1 \,\middle|\, \begin{array}{l} \upsilon_1(\zeta_2^{-1}(p_2^2)) \geq 0.6 \,\&\, \upsilon_1(\zeta_3^{-1}(p_3^2)) \geq 0.3 \\ \,\&\, 5\upsilon_1(\zeta_2^{-1}(p_2^2)) + 2.5\upsilon_1(\zeta_3^{-1}(p_3^2)) \geq 4.25 \end{array} \right\}; \\ & p_1^4 := V_1 \backslash \left(\cup_{j=1}^3 p_1^j \right) \Bigg\}. \end{aligned}$$

In this partition, components p_1^1, p_1^2, and p_1^3 are associated with the allocator's action F, and p_1^4 is associated with U. It is readily verified that $\{P_i\}_{i\in\{1,2,3\}}$ is a second-order convex frame.[13] Given that $\{P_i\}_{i\in\{1,2,3\}}$ refines the basic partition induced by $\{\mathcal{A}_i\}_{i\in\{1,2,3\}}$, we can apply Theorem 1, which allows us to focus on the class of direct information structures based on the frame $\{P_i\}_{i\in\{1,2,3\}}$. Specifically, the message set for this class is

$$\left\{(p_1^1, F), (p_1^2, F), (p_1^3, F), (p_1^4, U)\right\} \times \left\{(p_2^1, N), (p_2^2, P)\right\} \times \left\{(p_3^1, N), (p_3^2, P)\right\}.$$

By our computation, the optimal information structure in this problem is the following:

$$\pi((p_1^2, F), (p_2^2, P), (p_3^2, P) \mid \omega = 1) = 1; \quad \pi((p_1^1, F), (p_2^2, P), (p_3^1, N) \mid \omega = 0) = \frac{2}{7};$$
$$\pi((p_1^2, F), (p_2^1, N), (p_3^2, P) \mid \omega = 0) = \frac{3}{28}; \quad \pi((p_1^1, F), (p_2^1, N), (p_3^1, N) \mid \omega = 0) = \frac{1}{14};$$
$$\pi((p_1^2, F), (p_2^1, N), (p_3^1, N) \mid \omega = 0) = \frac{15}{28}.$$

[12] Specifically, the characteristic pair is $([0.6, \infty), \mathbb{1}_{\{\omega=1\}})$ for Player 2 and $([0.8, \infty), \mathbb{1}_{\{\omega=1\}})$ for Player 3.

[13] The second-order characteristic pairs that construct P_1 are $([0.7, \infty), \mathbb{1}_{\zeta_2^{-1}(p_2^2)})$, $([0.5, \infty), \mathbb{1}_{\zeta_3^{-1}(p_2^3)})$, $([0.6, \infty), \mathbb{1}_{\zeta_2^{-1}(p_2^2)})$, $([0.3, \infty), \mathbb{1}_{\zeta_3^{-1}(p_2^3)})$, and $([4.25, \infty), 5 \cdot \mathbb{1}_{\zeta_2^{-1}(p_2^2)} + 2.5 \cdot \mathbb{1}_{\zeta_3^{-1}(p_2^3)})$. To reduce message redundancy, we merge several components in the refined partition into p_3^1. This merging does not affect the solution, as the merged component p_3^1 is still characterizable by lower-order characteristic pairs in the affine manner described in Section 3.1.1.

Table 1
Payoff matrix for the investment game.

b	Not	Invest	g	Not	Invest
Not	0, 0	0, −8	Not	0, 0	0, 1
Invest	−7, 0	−4, −5	Invest	2, 0	5, 4

Under this information structure, the sender achieves a maximum payoff of 1. The receivers' beliefs and actions can be interpreted as follows: When the allocator receives the message (p_1^1, F), they understand that, although the social norm is not fair and Player 3 will not punish an unfair proposal, Player 2 will be led to believe that the norm is sufficiently fair to punish an unfair proposal with probability 0.8. Consequently, it is optimal to propose a fair split. Similarly, when the allocator receives the message (p_1^2, F), they recognize that altruistic punishment may come from both bystanders; in particular, the punishment from Player 3 is greater than 0.5, while the punishment from Player 2 is strictly less than 0.6. This type of punishment again makes a fair split the proposer's best response.

For the other receivers, when Player 2 receives the message (p_2^2, P), their belief that $\{\omega = 1\}$ is 0.6, making it optimal to punish an unfair proposal. For the alternative message (p_2^1, N), their belief that $\{\omega = 1\}$ is 0, so not punishing becomes optimal. Similarly, when Player 3 receives the message (p_3^2, P), their belief that $\{\omega = 1\}$ is 0.8, leading to altruistic punishment of the unfair proposal. For the other message (p_3^1, N), their belief that $\{\omega = 1\}$ is again 0, so not punishing is optimal.

In conclusion, persuasion through private messages allows the sender to maximally misalign the bystanders' altruistic punishment in the state $\omega = 0$. In this example, such misalignment completely eliminates any possibility that the allocator believes they can get away with an unfair proposal, thereby ensuring that a fair proposal is always made.

4.2. *On sender-worst equilibrium selection in binary-action supermodular games*

Morris et al. (2024) provides important insights into adversarial information design in a general class of supermodular games with binary receiver actions. We find that our approach is also applicable to their framework. To illustrate this, we revisit their leading example through the lens of our main result, which provides an alternative perspective on how strategic information disclosure can maximize desirable outcomes, even under the pessimistic assumption of sender-worst equilibrium selection in the receivers' interactions.

A sender aims to persuade two receivers, $I = \{1, 2\}$, to encourage investment. There are two equally likely states: b ("bad") and g ("good"). Let action 1 denote "invest" and action 0 denote "not invest". The sender's objective is to maximize the expected number of receivers who choose to invest, regardless of the state. The sender's utility is given by $v(a_I, \omega) = |\{i \in I | a_i = 1\}|$ for each action profile a_I.

The receivers' game is supermodular: Receiver 1 faces an investment cost of 7, while Receiver 2 faces a cost of 8. Each receiver earns a return of 3 from investing if the other receiver also invests. In the good state, both players receive an additional return of 9 from investing. The payoff from not investing is always 0. Thus, in the good state, both have a dominant strategy to invest, while in the bad state, their dominant strategy is not to invest. Table 1 summarises the payoffs, where Player 1 is the row player and Player 2 is the column player:

Let us now consider the selected best response correspondences of this game under the sender-worst equilibrium selection. Note that each receiver can be persuaded to invest if the state is sufficiently good. In particular, Player 1 will choose to invest, even knowing with certainty that the other player will not invest, if their belief in the good state is high enough.

$$p_1^1 := \left\{ v_1 \in V_1 \mid -7v_1(\{\omega = b\}) + 2v_1(\{\omega = g\}) > 0 \Longrightarrow v_1(\{\omega = g\}) > \frac{7}{9} \right\}. \tag{10}$$

In other words, if player1ı́s conjecture falls into p_1^1, then investing is the unique best response. Similarly, for Player 2, the region in which investment becomes a unique best response strategy is given by

$$p_2^1 := \left\{ v_2 \in V_2 \mid -8v_2(\{\omega = b\}) + v_2(\{\omega = g\}) > 0 \Longrightarrow v_2(\{\omega = g\}) > \frac{8}{9} \right\}. \tag{11}$$

The sender can further persuade players to invest by leveraging the possibility of the other player's investment. However, under the sender-worst equilibrium selection, if a player's incentive to invest relies on the optimistic expectation that the other player will invest even when not investing is equally optimal, and this incentive vanishes once that expectation is removed, then the sender-worst equilibrium must exclude such an investment action.

That said, since each player will choose to invest when their first-order belief is sufficiently high regardless of the other's behavior, such beliefs provide a robust foundation for encouraging further investment. When a player's second-order belief assigns sufficiently high probability to the other player's first-order belief under which investment is uniquely optimal, such a belief can motivate further investment. In this way, second-order beliefs expand the set of beliefs under which action 1 is the unique best response and can further promote investment. This recursive reasoning extends to belief hierarchies of arbitrary order: when a player's belief assigns sufficiently high probability to lower-order beliefs under which investment is the unique best response, this (higher-order) belief may also support investment as the unique best response.

To formalize this reasoning, let p_i^1 be as defined in (10) and (11). By construction, the partition induced by p_i^1 and its complement form a first-order convex frame. Inductively, for any fixed integer $n \geq 2$, suppose the following two conditions hold:

(i) for each $l \leq n-1$, the set p_i^l identifies all l-order beliefs of receiver i for which action 1 is the unique best response.
(ii) The joint partition formed by $p_i^1, \ldots, p_i^{n-1}$ and their complements constitutes an $(n-1)$-order convex frame.

Let $F_i^{n-1} := \bigcup_{l=1}^{n-1} p_i^l$. Recall that ζ_i is the isomorphism mapping Player i's types to their consistent conjectures. Then, let us define

$$\begin{aligned} f_1^n &:= 5 \cdot \mathbb{1}_{(g,\zeta_2^{-1}(F_2^{n-1}))} + 2 \cdot \mathbb{1}_{(g,\zeta_2^{-1}(F_2^{n-1})^c)} + (-4) \cdot \mathbb{1}_{(b,\zeta_2^{-1}(F_2^{n-1}))} + (-7) \cdot \mathbb{1}_{(b,\zeta_2^{-1}(F_2^{n-1})^c)}; \\ f_2^n &:= 4 \cdot \mathbb{1}_{(g,\zeta_1^{-1}(F_1^{n-1}))} + \mathbb{1}_{(g,\zeta_1^{-1}(F_1^{n-1})^c)} + (-5) \cdot \mathbb{1}_{(b,\zeta_1^{-1}(F_1^{n-1}))} + (-8) \cdot \mathbb{1}_{(b,\zeta_1^{-1}(F_1^{n-1})^c)}. \end{aligned} \tag{12}$$

Note that each f_i^n captures Player i's utility from choosing action 1 under the pessimistic expectation that the other player will invest if and only if action 1 is their unique best response, given beliefs of order lower than n. Thus, the set of conjectures under which investment is the unique best response, grounded in n-order belief reasoning, is characterized as follows:

$$p_i^n := \left\{ v_i \in V_i \mid \int f_i^n \mathrm{d}v_i > 0 \right\}.$$

Note that, by construction of (12), each f_i^n is measurable with respect to the σ-algebra generated by $\{g,b\} \times \bigcup_{l=1}^{n-1} \{p_{-i}^l, V_{-i} \setminus p_{-i}^l\}$. Thus, iterative measurability is satisfied, and pairing f_i^n with the bi-convex set $(0,\infty)$ defines a well-defined n-order characteristic pair. It follows that the joint partition from characteristic pairs $\left\{\left((0,\infty), f_i^l\right)\right\}_{l=1,\ldots,n}$ constitutes an n-order convex frame. Continuing this procedure as $n \to \infty$, the joint partition induced by $\{p_i^l\}_{l=1}^{\infty}$ and their complements defines an infinite-order convex frame. Thus, the sender-worst equilibrium selection in this game can be characterized by the following selected best response correspondence:

$$\mathcal{A}_i(v_i) = \begin{cases} \{1\}, & \text{if } v_i \in \bigcup_{l=1}^{\infty} p_i^l; \\ \{0\}, & \text{otherwise.} \end{cases}$$

From the above reasoning, the associated basic partition can be refined by the convex frame induced by $\{p_i^l\}_{l=1}^{\infty}$ and their complements. Therefore, our main result implies that, without loss of generality, we can focus on the generalized direct information structure defined over this convex frame. Moreover, it is straightforward to verify that $p_i^{l''} \subseteq p_i^l$ for any $l' < l$, thus the generalized direct information structure reveals whether each receiver i's conjecture falls into one of the following components: (a) p_i^1; (b) $p_i^{l+1} \setminus p_i^l$ for $l = 1, \ldots$; or (c) $\mathcal{A}_i^{-1}(\{0\})$.

Our computation shows that achieving the sender's maximum payoff in this game requires at most three messages per receiver. In particular, the following generalized direct information structure, which involves a small positive parameter ϵ, achieves the sender's maximum payoff in the limit as $\epsilon \searrow 0$:

$$\begin{aligned} &\pi^*((p_1^1,1),(p_2^2 \setminus p_2^1,1) \mid \omega = g) = \frac{7}{9}, \quad \pi^*((p_1^3 \setminus p_1^2,1),(p_2^1,1) \mid \omega = g) = \frac{2}{9}; \\ &\pi^*((p_1^1,1),(p_2^2 \setminus p_2^1,1) \mid \omega = b) = \frac{2}{9} - \epsilon, \quad \pi^*((p_1^3 \setminus p_1^2,1),(p_2^1,1) \mid \omega = b) = \frac{1}{36} - \epsilon; \\ &\pi^*((p_1^3 \setminus p_1^2,1),(p_2^2 \setminus p_2^1,1) \mid \omega = b) = \frac{1}{4}, \quad \pi^*((\mathcal{A}_1^{-1}(\{0\}),0),(\mathcal{A}_2^{-1}(\{0\}),0) \mid \omega = b) = \frac{1}{2}. \end{aligned}$$

In particular, when receiving the message $(p_1^1,1)$, Player 1 holds a strong enough belief that the state is good and will invest even if the other player refrains from investing with certainty. The sender then associates this robust investment with the private recommendation for Player 2 in the message $(p_2^1,1)$, persuading this player to invest based both on the belief in a good state and on Player 1's committed investment. Thus, investing becomes the unique best response for Player 2. The sender can then leverage this investment to further persuade Player 1 through the message $(p_1^3 \setminus p_1^2,1)$, drawing on both the belief in the state and the other player's investment.

Unlike the (approximately) optimal information structure in Morris et al. (2024), which relies entirely on public disclosure to both receivers, the information structure presented above uses private disclosure. Despite this difference, both structures lead to the same conclusion: the sender can approximately attain the maximal payoff of $\frac{3}{2}$. Mathevet et al. (2020) study a similar application with both private and public disclosure. In this example, the intuition underlying our construction is closely related to their private-disclosure step (the "maximization within" step in their terminology), which determines the optimal minimal distribution through private messages to maximize investment under adversarial selection.

4.3. More examples in the literature

We provide two additional examples from the literature that consider either equilibrium selections differing from the sender-preferred one or the incorporation of psychological traits into receivers' behavior, where our main result is also applicable.

Example 3. (The Sincere Voting Rule) Empirical evidence demonstrates that voters derive utility from expressing support for their favorite candidates in large elections. Consequently, the literature on persuading voters is particularly interested in the equilibrium selection known as the "sincere voting" rule. Specifically, this rule states that for any two policies A and B, if policy A yields a weakly higher expected payoff to a voter than policy B, then the voter will vote for policy A regardless of whether their vote can make a difference to the outcome (see, for example, Alonso and Camara, 2016). We now apply the insight from our main result to understand how this rule influences voter interactions in the context of persuasion.

For simplicity, consider a binary underlying state $\omega \in \{L, R\}$. A sender seeks to persuade a set I of voters to pass a proposal. Each voter has a binary action set $\{Y, N\}$, and their joint action, via a social choice function, determines one alternative from the binary policy set $X := \{x_0, x_1\}$, where x_0 represents the status quo and x_1 the proposal. Let $u_i : X \times \{L, R\} \to \mathbb{R}$ denote the utility function of each voter i. Under the sincere voting rule, a voter selects Y if and only if, given the sender's message, the policy x_1 yields a strictly higher expected payoff than the alternative x_0. In cases of indifference, the voter chooses the sender's preferred action Y. This equilibrium selection can be expressed through the following selected best response correspondence:

$$\mathcal{A}_i(v_i) = \begin{cases} \{Y\} & \text{if } \int u_i(x_1, \omega)\mathrm{d}v_i \geq \int u_i(x_0, \omega)\mathrm{d}v_i; \\ \{N\} & \text{otherwise.} \end{cases}$$

Note that although observing other voters' choices could provide information that might influence an individual's beliefs, this voting rule requires each voter to base their decision solely on their own information, ignoring any signals from others' choices. This captures the psychological trait underlying "sincerity" in voting-a sincere voter may not be fully rational in making use of all available information.[14]

Given this feature of the selection rule, it is straightforward to verify that the basic partition induced by the selection correspondence forms a first-order convex frame. Therefore, our main result implies that, without loss of generality, we can focus on the generalized direct information structure that reveals whether each receiver i's belief falls into $\mathcal{A}_i^{-1}(Y)$ or not.

Example 4. (The Skeptical Posture) In the context of selling an object with unknown quality to a buyer, Milgrom and Roberts (1986) introduce the concept of a "skeptical posture" to capture sophisticated buyer behavior. Specifically, a buyer exhibits a skeptical posture when, regardless of their beliefs about the quality states, they always assume that the true state is the one that minimizes the purchased quantity among all possibilities. The perspective from our main result can also be applied to understand such a skeptical buyer's purchasing behavior.

Let us start by describing the setting of Milgrom and Roberts (1986): There is one seller and one buyer. The seller owns a product with unknown quality. It is known that the set of all possible qualities Ω is finite, i.e., $\Omega := \{\omega_1, \omega_2, \dots, \omega_N\}$ with $\omega_1 < \omega_2 < \dots < \omega_N$. Let $A := \{q_n\}_{n=1,\dots,N}$ represent all possible purchase quantities, where each q_n corresponds to the optimal purchase quantity when the buyer knows for sure that the quality is ω_n. These quantities follow an increasing order, such that $q_1 < q_2 < \dots < q_N$. In this single-receiver case, higher-order beliefs play no role. Thus, the buyer's conjecture space coincides with their beliefs regarding the state $\Delta(\Omega)$. The skeptical posture requires that, for any posterior belief $\beta \in \Delta(\Omega)$ derived from the seller's information, the buyer will purchase the minimum quantity $q_{I(\beta)}$ among all the possible states supported by this belief, i.e., $I(\beta) := \arg\min\{n \mid \omega_n \in \operatorname{supp}\beta\}$. Thus, the selected best response correspondence is that for any $\mu \in \Delta(\Omega)$, $\mathcal{A}(\mu) := \{q_{I(\mu)}\}$.

It is readily verified that the basic partition is a first-order convex frame. For instance, $\mathcal{A}^{-1}(q_1)$ can be characterized by the collection of beliefs that satisfies $\mu(\{\omega_1\}) > 0$. Similarly, $\mathcal{A}^{-1}(q_2)$ can be characterized by the collection of beliefs that satisfies $\mu(\{\omega_1\}) = 0$ and $\mu(\{\omega_2\}) > 0$ simultaneously; and so forth for the other possible components. Our main result is again applicable, allowing us to focus on the class of generalized direct information structures that disclose which component of the basic partition the receiver's belief falls into.

5. Discussion

This section discusses related aspects of our main result. Section 5.1 demonstrates that our condition is tight: if a refined partition fails to satisfy either convexity or iterative measurability, then coarsening information based on this partition may alter the outcome in some equilibria. Section 5.2 then returns to the fundamental game primitives—specifically, the receivers' preferences and arbitrary equilibrium selection—to further discuss the applicability of our approach beyond the case of selected best response correspondences.

5.1. On the tightness of the key regularity properties

Recall that the key element of our generalized direct approach is whether a refined partition can be constructed as a convex frame. Such a frame possesses two essential properties—iterative measurability and convexity—which together form a valid information coarsening rule. We now show that these properties are tight: if a refined partition fails to satisfy either one, the resulting coarsening rule may not preserve some equilibrium outcomes in the game.

As discussed in Section 3.1, convexity provides the regularity required for information coarsening. The following example illustrates this point.

Example 5. Consider a game with a binary state space $\Omega := \{0, 1\}$, one sender (Player s), and one receiver (Player 1). The receiver has binary actions $A_1 := \{0, 1\}$. Since there is only one receiver, the receiver's conjecture reduces to a first-order belief $v_1 \in \Delta(\Omega)$. The receiver's preferences are represented by the following selected best response correspondence:

$$\mathcal{A}_1(v_1) = \begin{cases} 1, & \text{if } v_1(\{\omega = 0\}) \in \mathbb{R} \setminus \mathbb{Q}; \\ 0, & \text{if } v_1(\{\omega = 0\}) \in \mathbb{Q}. \end{cases}$$

[14] See "Equilibrium Selection" and "Electoral Outcome" on pp. 3593–3594 of Alonso and Camara (2016).

Both players share a common prior under which the two states are equally likely. The sender receives a payoff of 1 if the receiver chooses action 1, and 0 otherwise.

The basic partition in this game is $\mathcal{A}_1^{-1}(\{1\}) := \{v_1 \mid v_1(\{\omega = 0\}) \in \mathbb{R} \setminus \mathbb{Q}\}$ and its complement. This partition is generated by the first-order characteristic pair $(B_1, \mathbb{1}_{\{\omega=0\}})$, where $B_1 = \mathbb{R} \setminus \mathbb{Q}$—a set that is not bi-convex. We show below that the non-convexity of B_1 implies that restricting attention to generalized direct information structures based on this partition can lead to a loss of generality, as coarsening information according to this partition may alter the outcome of some PBEs. Consider a PBE in which the sender achieves the maximum payoff of 1 by the following information structure:

$$\pi(m_1 \mid \omega = 0) = \frac{\sqrt{2}}{4},\ \pi(m_1' \mid \omega = 0) = 1 - \frac{\sqrt{2}}{4},\ \pi(m_1 \mid \omega = 1) = 1 - \frac{\sqrt{2}}{4},\ \pi(m_1' \mid \omega = 1) = \frac{\sqrt{2}}{4}.$$

This information structure induces two posterior beliefs, v_1 and v_1', with $v_1(\{\omega = 0\}) = \frac{\sqrt{2}}{4}$ and $v_1'(\{\omega = 0\}) = 1 - \frac{\sqrt{2}}{4}$. Under both beliefs, the receiver strictly prefers action 1. Moreover, since $\int \mathbb{1}_{\{\omega=0\}} \mathrm{d}v_1 \in B_1$ and $\int \mathbb{1}_{\{\omega=0\}} \mathrm{d}v_1' \in B_1$, these beliefs fall within the same component of the basic partition defined by $(B_1, \mathbb{1}_{\{\omega=0\}})$.

However, if we adopt the generalized direct version of this information structure under the basic partition, both messages m_1 and m_1' would be mapped to the same recommendation, i.e., $(\mathcal{A}_1^{-1}(\{1\}), 1)$. The resulting information structure sends the pooled message with probability one, yielding the posterior belief $\hat{v}_1(\{\omega = 0\}) = 0.5$, under which the receiver optimally chooses action 0. Hence, coarsening information based on the basic partitions can change the equilibrium outcome, since it violates the bi-convexity property. In particular, while the original information structure π induces the receiver to choose action 1 with certainty, the corresponding generalized direct information structure under the basic partition leads the receiver instead to choose action 0 with certainty.

We now turn to the next example to demonstrate the importance of the other key property, iterative measurability.

Example 6. Consider a game with one sender, two receivers $I := \{1, 2\}$, and two states $\Omega := \{0, 1\}$. All players share a common prior under which the states are equally likely. Each receiver i chooses an action $a_i \in \{0, 1\}$.

Receiver 1's selected best response correspondence is defined as follows: $\mathcal{A}_1(v_1) = \{1\}$ if and only if $v_1(\{\omega = 0\}) \geq 0.6$, and $\mathcal{A}_1(v_1) = \{0\}$ otherwise. Recall that ξ_i is the mapping that maps receiver i's type to their consistent conjectures. Receiver 2's selected best response correspondence is defined as: $\mathcal{A}_2(v_2) = 1$ if and only if

$$v_2\big(\{\omega = 1\} \times \xi_1^{-1}\big(\{v_1 \mid v_1(\{\omega = 0\}) \geq 0.8\}\big)\big) \geq \frac{3}{32},$$

and $\mathcal{A}_2(v_2) = \{0\}$ otherwise. The sender receives a positive payoff from each receiver's action 1, with greater weight placed on Receiver 1's action. Specifically, for each action profile $a_I := (a_1, a_2)$, the sender receives a payoff of $v(a_I) = 4.5 \cdot \mathbb{1}_{\{1\}}(a_1) + \mathbb{1}_{\{1\}}(a_2)$.

Note that Receiver 1's basic partition is a first-order convex frame generated by the characteristic pair $(B_1, \mathbb{1}_{\{\omega=0\}})$ with $B_1 = [0.6, \infty)$. Receiver 2's basic partition is generated by the second-order characteristic pair $(B_2, \mathbb{1}_{\{\omega=1\}\times\xi_1^{-1}(S')})$, where $B_2 = [\frac{3}{32}, \infty)$ and $S' := \{v_1 \mid v_1(\{\omega = 0\}) \geq 0.8\}$. Since the above S' is not a partition component of Receiver 1's first-order convex frame, the second-order characteristic pair for Receiver 2 fails the iterative measurability condition under the given Receiver 1 basic partition. As a result, the basic partition, consisting of these two partitions, is *not* a convex frame.

There exists a PBE in which the following information structure constitutes part of the equilibrium:

$$\pi((m_1^1, m_2^1) \mid \omega = 0) = \frac{1}{3}, \quad \pi((m_1^2, m_2^2) \mid \omega = 0) = \frac{2}{3},$$
$$\pi((m_1^1, m_2^1) \mid \omega = 1) = \frac{1}{12}, \quad \pi((m_1^2, m_2^2) \mid \omega = 1) = \frac{4}{9}, \quad \pi((m_1^3, m_2^1) \mid \omega = 1) = \frac{17}{36}.$$

In equilibrium, Receiver 1 chooses action 1 whenever they receive message m_1^1 or m_1^2, and action 0 otherwise. Receiver 2 chooses action 1 whenever they receive message m_2^1, and action 0 otherwise. Under this equilibrium, the sender achieves a maximum payoff of approximately 3.88.

However, if we construct the corresponding generalized direct information structure based on the basic partition which, as discussed earlier, fails to be a convex frame due to violating the iterative measurability condition, the equilibrium outcome changes. Specifically, consider merging messages m_1^1 and m_1^2, since their induced conjectures fall within the same component of Receiver 1's basic partition. While this merge preserves Receiver 1's optimal response-because their first-order belief remains within the same component of their basic partition-the merge changes how Receiver 2 perceives Receiver 1's beliefs.

Under the original information structure, Receiver 2 perceives that Receiver 1's posterior on $\{\omega = 0\}$ was 0.8 upon observing m_1^1, and 0.6 upon observing m_1^2. This belief incentivizes Receiver 2 to take action 1 at m_2^1 and action 0 at m_2^2. However, after pooling m_1^1 and m_1^2, Receiver 2 updates their belief and infers that Receiver 1's posterior on $\{\omega = 0\}$ is now strictly below 0.8 upon observing the merged message $m_1^1 \vee m_1^2$. This shift places Receiver 2's conjecture outside the belief component that incentivizes action 1, leading them to choose action 0 instead. Thus, even though the merge does not alter Receiver 1's behavior, it *does* change Receiver 2's best response, thereby altering the equilibrium outcome.

This example illustrates the role of the iterative measurability condition, which helps to ensure consistency throughout the belief hierarchy in the selected best-response correspondences under coarsened information. Together, iterative measurability and convexity form the minimal conditions needed for a refined partition to allow the sender to merge messages without affecting any PBE outcomes.

5.2. Scope of the generalized direct approach

In this section, we revisit the fundamental primitives-namely, the receiver's preferences and the equilibrium selection rule-to examine when these elements produce a sufficiently regular selected best response to support the general direct approach. We first consider receivers with psychological preferences under sender-preferred equilibrium selection, and then address the case of arbitrary equilibrium selection.

Psychological preferences. Psychological games provide a valuable framework for modeling belief-dependent motivations in decision making. Under the sender-preferred equilibrium selection, the key condition in our main result appears to be broadly applicable to this class of games—if not exactly, then in an approximate sense, as clarified below. For illustration, we examine the leading example from the "Reciprocity" section in Battigalli and Dufwenberg (2007), pp. 837–838. A comprehensive analysis of which functional forms of psychological preferences exactly satisfy the key conditions in our main result would be an interesting direction for future research.

Two players, 1 and 2, move sequentially in a game. Despite the original dynamic formulation, our insight applies to its normal form. Player 1 selects $a_1 \in A_1 = \{\text{Stay, Reach}\}$, and Player 2 selects $a_2 \in A_2 = \{\text{Take, Give}\}$. Let $r_i(a_1, a_2)$ denote the material payoff for player i. If Player 1 chooses Stay, the game ends immediately and both players receive a material payoff of $r_i(\text{Stay}, a_2) = 5$. If Player 1 chooses Reach, Player 2 then makes a choice. The material payoffs for both players for the remaining cases are: $r_1(\text{Reach, Give}) = 9 = r_2(\text{Reach, Take})$, and $r_2(\text{Reach, Give}) = 1 = r_1(\text{Reach, Take})$.

This example involves only strategic uncertainty. Let $\mu_i \in \Delta(A_j)$ denote Player i's first-order belief about Player j's actions. For $i = 1, 2$, let $\theta_i \in (0, 1]$ be a sensitivity parameter, commonly known to both players. Each player i's overall utility, including a reciprocity component, is:

$$u_i(a_1, a_2, \mu_1) = r_i(a_1, a_2) + \theta_i \cdot \kappa_{21}(a_2) \cdot \kappa_{12}(a_1, \mu_1), \tag{13}$$

where $\kappa_{12}(\text{Stay}, \mu_1) = 2 - 4 \cdot \mu_1(\{a_2 = \text{Take}\}) = -\kappa_{12}(\text{Reach}, \mu_1)$, $\kappa_{21}(\text{Give}) = 4$, and $\kappa_{21}(\text{Take}) = -4$. Note that Player 1 will choose Stay under belief $\mu_1 \in \Delta(A_2)$ only if the expected utility from Stay is at least as good as that from Reach. This condition can be written as:

$$\int u_1(\text{Stay}, \cdot, \mu_1)\, d\mu_1 \geq \int u_1(\text{Reach}, \cdot, \mu_1)\, d\mu_1 \iff \mu_1(\{a_2 = \text{Take}\}) \in (-\infty, 0.5 - \frac{1}{8\theta_1}) \cup [0.5, \infty).$$

Thus, the basic partition for Player 1 is a first-order convex frame, generated by the characteristic pair $([0.5, \infty), \mathbb{1}_{\{a_2=\text{Take}\}})$ and, whenever $0.5 - \frac{1}{8\theta_1} \geq 0$, also by the pair $((-\infty, 0.5 - \frac{1}{8\theta_1}], \mathbb{1}_{\{a_2=\text{Take}\}})$. As for Player 2, they can move only after Player 1 chooses Reach. Let $\nu_2 \in \Delta(\Delta(A_2))$ denote Player 2's belief about Player 1's belief. Given Player 1's choice of Reach, Player 2 will choose Give only if

$$\int u_2(\text{Reach, Give, } \mu_1) d\nu_2 \geq \int u_2(\text{Reach, Take, } \mu_1) d\nu_2 \iff \int \mu_1(\{a_2 = \text{Take}\}) d\nu_2 \geq \frac{1}{4\theta_2} + 0.5.$$

Note that this incentive constraint identifies a second-order characteristic pair $([\frac{1}{4\theta_2} + 0.5, \infty), f_2(\mu_1))$ where $f_2(\mu_1) := \mu_1(\{a_2 = \text{Take}\})$ may not satisfy the iterative measurability property defined earlier. However, $f_2(\mu_1)$ can be approximated by the simple functions $\sum_{j=1}^{L} c_j \mathbb{1}_{\{c_j \leq \mu_1(\{a_2=\text{Take}\}) \leq c_{j+1}\}}$ to arbitrary precision, with $\{c_j\}_{j=1}^{L}$ an increasing sequence of numbers partitioning $[0, 1]$ and L a finite nonnegative integer. Consider now an approximate version of the game, identical in all respects except that the psychological term κ_{12} is replaced by its approximation:

$$\hat{\kappa}_{12}(\text{Stay}, \mu_1) = 2 - 4 \cdot \sum_{j=1}^{L} c_j \mathbb{1}_{\{c_j \leq \mu_1(\{a_2=\text{Take}\}) \leq c_{j+1}\}} = -\hat{\kappa}_{12}(\text{Reach}, \mu_1).$$

In this setting, Receiver 1's basic partition is refined by the first-order frame $\{([c_j, \infty), \mathbb{1}_{\{a_2=\text{Take}\}})\}_{j=1}^{L}$. Based on this first-order frame, the function $\sum_{j=1}^{L} c_j \mathbb{1}_{\{c_j \leq \mu_1(\{a_2=\text{Take}\}) \leq c_{j+1}\}}$ satisfies the iterative measurability condition. The joint partition generated by these characteristic pairs, together with $([\frac{1}{4\theta_2} + 0.5, \infty), \sum_{j=1}^{L} c_j \mathbb{1}_{\{c_j \leq \mu_1(\{a_2=\text{Take}\}) \leq c_{j+1}\}})$ forms a second-order convex frame.

It is well known that any bounded Lebesgue measurable function can be uniformly approximated arbitrarily closely by simple functions.[15] For any general psychological utility $u_i(\omega, a_I, \mu_I)$, where μ_I denotes the profile of first-order beliefs, one can apply simple function approximation to the receivers' psychological preferences over beliefs (holding ω and a_I fixed) as in the example above, to obtain the corresponding approximate games to which our approach extends.[16]

Arbitrary equilibrium selection. We now turn to the case of arbitrary equilibrium selection. As noted in the introduction, equilibrium selection is a complex issue that often depends on broader contextual factors, and a comprehensive treatment is beyond the scope of this paper. In contrast to psychological games, the usefulness of our approach here may be limited to situations where equilibrium actions are selected primarily based on (higher-order) beliefs, such as those discussed in Section 4. For certain well-known equilibrium selection rules, some of which we discuss below, the selected best response correspondence is only defined after the final equilibrium outcome is determined. In such cases, our approach is not applicable.

[15] See, for example, the simple function approximation lemma, Section 18.1, p. 363, Royden (1988).

[16] The utility function $u_i(\omega, a_I, \mu_I)$ follows the framework of Battigalli and Dufwenberg (2022), who define psychological utility largely in terms of first-order beliefs; see pp. 862–863 for more details therein.

For example, equilibrium selection based on payoff dominance (also known as Pareto dominance) or risk dominance (see Harsanyi and Selten, 1988) requires first identifying the full set of equilibria and then selecting the "best" one according to a specific criterion. In these cases, although the selected best response correspondence can be defined, it is only determined *after* all equilibria have been found and the relevant selection rule has been applied. In the context of persuasion, this means that before defining the selected best response correspondence, one must first consider every possible information structure and examine all the equilibria that could result from each. Such equilibrium selection rules will not be compatible with our approach. Similarly, the intuitive criterion (Cho and Kreps, 1987) eliminates an equilibrium if there exists a sender type and a deviating signal that guarantees this type a payoff higher than their equilibrium payoff, provided the receiver holds a "reasonable" belief as specified by the criterion. Since this criterion must be applied after candidate equilibria have been identified, our approach is likewise unsuitable for persuasion problems that involve equilibrium selection criteria of this kind.

6. Conclusion

Persuasion schemes in practice are often complex and may involve conveying richer forms of information than simply recommending actions. This paper examines a setting in which multiple receivers interact under an equilibrium selection that may differ from the sender-preferred selection, and where receivers' preferences may directly depend on their own beliefs, including higher-order beliefs. Our main result shows that if a convex frame can be constructed to refine the basic partition in this game, then information can be categorized according to this partition without altering the equilibrium outcome.

As an important implication, our result provides a generalized direct approach for examining persuasion schemes where the traditional direct approach might not be applicable. To demonstrate its usefulness, we apply our approach to examples from the existing literature as well as to a new persuasion problem involving a psychological game inspired by the experiment of Fehr and Fischbacher (2004), in which receivers' behavior may exhibit altruistic traits. Despite the widespread occurrence of such behavior, there has been little study on persuading interacting altruistic receivers. Our approach helps to derive the optimal information structure in this setting. Given the modeling framework, we conjecture that our approach may exhibit the most utility in information design problems where equilibrium actions are selected based on higher-order beliefs or where receivers' preferences possess psychological traits.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yishu Zeng reports financial support was provided by Social Sciences and Humanities Research Council of Canada. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

I am grateful to the editors and anonymous referees for their thoughtful comments and suggestions, which greatly improved the paper. I am deeply indebted to David Miller and Uday Rajan for their continued support. I also benefited from valuable feedback from Tilman Börgers, Fei Hu, Emil Kamenica, Ming Li, Elliot Lipnowski, Andrey Malenko, Vasiliki Skreta, Siyang Xiong, and the audiences at the Michigan Theory Lunch Seminar, York University Seminar, GAMES 2020, and the 2021 Midwest Economic Theory Conference (at Michigan State University). I gratefully acknowledge the generous financial support of the Social Sciences and Humanities Research Council of Canada (Insight Development Grant 430-2023-00907). The first draft was written while I was at the University of Michigan.

Appendix A. Proofs

Proof of Lemma 1. Note that any bi-convex set B in $\mathbb{R}$ is a connected set, as by the definition of convexity, B cannot be divided into two disjoint open sets. Recall that an open set in $\mathbb{R}$ must be a union of open intervals. Therefore, B must take the form of either (b', b), $[b', b)$, $(b', b]$, or $[b', b]$ for some $b', b \in \mathbb{R}$. The fact that B is a bi-convex set in $\mathbb{R}$ implies that B^c is also connected. Thus B^c must take one of these four specific forms as well. Combining these two expressions, B must take one of the following forms: $(-\infty, b)$, (b, ∞), $(-\infty, b]$, or $[b, \infty)$ for some $b \in \mathbb{R}$. The converse can be readily verified. □

Proof of Theorem 1. Let $(P_i)_{i\in I}$ be a convex frame that refines the basic partition induced by $(\mathcal{A}_i)_{i\in I}$. For each $i \in I$, define $\xi_i : V_i \to P_i$ as the mapping that maps each $v_i \in V_i$ to the partition component $\xi_i(v_i) \in P_i$ that satisfies $v_i \in \xi_i(v_i)$. Denote by $\xi = (\xi_i)_{i\in I}$.

Fix an arbitrary PBE $(\pi, \sigma = (\sigma_i)_{i\in I}, \beta = (\beta_i)_{i\in I})$. Consider the following information structure $\hat{\pi} : \Omega \to \Delta((P_i \times \Delta(A_i))_{i\in I})$ such that for each $\omega \in \Omega$ and each $(p_i, \alpha_i) \in P_i \times \Delta(A_i)$,

$$\hat{\pi}((p_i, \alpha_i)_{i\in I} \mid \omega) := \pi\left(\left\{m_i \in M_i \mid \xi_i(\beta_i^{\pi}(\cdot \mid m_i)) = p_i, \sigma_i^{\pi}(\cdot \mid m_i) = \alpha_i\right\}_{i\in I} \mid \omega\right). \tag{A.1}$$

Given that we assume the message space M_i is a large enough Polish space, without loss of generality, we assume that $P_i \times \Delta(A_i)$ is a subset of M_i for each $i \in I$. Define receiver i's strategy $\hat{\sigma}_i$ as follows: for any $(\pi', m_i') \in \Pi \times M_i$,

$$\hat{\sigma}_i^{\pi'}(\cdot \mid m_i') := \begin{cases} \alpha_i & \text{if } \pi' = \hat{\pi}, m_i' = (p_i, \alpha_i) \in P_i \times \Delta(A_i); \\ \sigma_i^{\pi'}(\cdot \mid m_i') & \text{otherwise.} \end{cases} \tag{A.2}$$

Thus $\hat{\sigma}_i$ specifies an obedient receiver strategy on the path when the above $\hat{\pi}$ is chosen and the message $m_i' \in P_i \times \Delta(A_i)$ is realized. Let $\hat{\beta}_i : \Pi \times M_i \to V_i$ be the consistent belief map under the strategy profile $\hat{\sigma}$ in (A.2). In particular, for any (π', m_i'), whenever either $\pi' \neq \hat{\pi}$ or $m_i' \notin P_i \times \Delta(A_i)$, we set

$$\hat{\beta}_i^{\pi'}(\cdot \mid m_i') = \beta_i^{\pi'}(\cdot \mid m_i'). \tag{A.3}$$

The remaining of this proof will show that $(\hat{\pi}, (\hat{\sigma}_i)_{i\in I}, (\hat{\beta}_i)_{i\in I})$ is a PBE. If this holds, then $\hat{\pi}$ is a generalized direct obedient information structure, where the obedience outcome under $\hat{\pi}$ is essentially the same as the outcome under the given PBE. Thus our conclusion follows naturally.

By the construction and (A.3), the consistency of the belief map $(\hat{\beta}_i)_{i\in I}$ is automatically satisfied. To show that $(\hat{\pi}, (\hat{\sigma}_i)_{i\in I}, (\hat{\beta}_i)_{i\in I})$ is a PBE, we only need to verify that

(a) Each $\hat{\sigma}_i^{\pi'}(\mathcal{A}_i(\hat{\beta}_i^{\pi'}(\cdot \mid m_i')) \mid m_i') = 1$ holds for each $\pi' \in \Pi$ and $m_i' \in M_i$;
(b) The above $\hat{\pi}$ maximizes the sender's ex ante payoff given $\hat{\sigma}$.

To proceed, we first present an important conclusion, which relies on a crucial claim whose proof we will postpone to the end. Recall that $(P_i)_{i\in I}$ is a convex frame that refines the basic partition induced by $(\mathcal{A}_i)_{i\in I}$. By definition, for each $i \in I$, there exists a set of characteristic pairs $\{(B_{i,h}^k, f_{i,h}^k) \mid k \in \mathbb{N} \cup \{\infty\}, h \in \mathscr{D}_i^k\}_{i\in I}$ that recursively constructs P_i. Suppose that the maximum belief order in such construction is $K \in \mathbb{N} \cup \{\infty\}$. For any belief order $k \leq K$, denote by P_i^k the k-order frame that emerges in the recursive construction of P_i. Note that the frame becomes finer with each recursive round. Thus, the final frame P_i refines any P_i^k with $k \leq K$ and $i \in I$. Therefore, for each $p_i \in P_i$ and each $k \leq K$, there exists a unique element, denoted as $\phi_i^k(p_i) \in P_i^k$, such that $p_i \subseteq \phi_i^k(p_i)$. In particular, when $k = K$, then $p_i = \phi_i^k(p_i)$ holds. We claim that

Claim 1. *Given the above $(\hat{\pi}, (\hat{\sigma}_i)_{i\in I}, (\hat{\beta}_i)_{i\in I})$ constructed in (A.1)–(A.3), for any receiver i and any message $(p_i, \alpha_i) \in P_i \times \Delta(A_i)$ under $\hat{\pi}$, for any finite $k \leq K$, we have $\hat{\beta}_i^{\hat{\pi}}(p_i, \alpha_i) \in \phi_i^k(p_i)$.*

If the above claim holds, then for any receiver i and any message $(p_i, \alpha_i) \in P_i \times \Delta(A_i)$ under $\hat{\pi}$, we must have

$$\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i) \in p_i. \tag{A.4}$$

This is because, by the recursive construction, a violation of $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i) \in p_i$ must imply that there exists some finite k such that $\hat{\beta}_i^{\hat{\pi}}(p_i, \alpha_i) \in \phi_i^k(p_i)$ is violated at that level, which would contradict Claim 1.

Building upon the above conclusion, we now show the above (a) and (b). Let us start with (a). Note that when either $\pi' \neq \hat{\pi}$ or $m_i' \notin P_i \times \Delta(A_i)$, the statement (a) holds because the strategy profile and belief maps, in this case, coincide with the given PBE. For the case when $\pi' = \hat{\pi}$ and any message $m_i' = (p_i, \alpha_i) \in P_i \times \Delta(A_i)$, recall that in the given PBE, $\beta_i^{\pi}(\cdot \mid m_i) \in p_i$ holds for any $m_i \in (\beta_i^{\pi})^{-1}(p_i) \cap (\sigma_i^{\pi})^{-1}(\alpha_i)$, where β_i^{π} is a mapping from each message m_i' to a consistent conjecture and σ_i^{π} a mapping from each message to the corresponding strategy. By the above (A.4), $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i) \in p_i$. Since P_i is a refined partition of the basic partition, the conjectures in the same $p_i \in P_i$ result in the same set of selected best responses. As such, for the given $m_i' = (p_i, \alpha_i)$,

$$\mathcal{A}_i(\beta_i^{\pi}(\cdot \mid m_i)) = \mathcal{A}_i(p_i) = \mathcal{A}_i(\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i)), \forall\, m_i \in (\beta_i^{\pi})^{-1}(p_i) \cap (\sigma_i^{\pi})^{-1}(\alpha_i). \tag{A.5}$$

In the given PBE, for any $m_i \in (\beta_i^{\pi})^{-1}(p_i) \cap (\sigma_i^{\pi})^{-1}(\alpha_i)$, $\sigma_i^{\pi}(\mathcal{A}_i(\beta_i^{\pi}(\cdot \mid m_i)) \mid m_i) = 1$. Thus, for any such m_i, we have the following expression:

$$\begin{aligned} 1 &= \sigma_i^{\pi}(\mathcal{A}_i(\beta_i^{\pi}(\cdot \mid m_i)) \mid m_i) = \alpha_i(\mathcal{A}_i(\beta_i^{\pi}(\cdot \mid m_i))) \\ &\overset{\text{(A.5)}}{=} \alpha_i(\mathcal{A}_i(\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i))) \overset{\text{(A.2)}}{=} \hat{\sigma}_i^{\hat{\pi}}(\mathcal{A}_i(\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i)) \mid (p_i, \alpha_i)). \end{aligned}$$

Hence, we verify (a) holds for the remaining case when $\pi' = \hat{\pi}$ and $m_i' = (p_i, \alpha_i) \in P_i \times \Delta(A_i)$ for each $i \in I$.

We now turn to (b). By construction, the new receivers' strategy profile $\hat{\sigma}$ is the same as σ whenever the sender's choice is other than the given $\hat{\pi}$ constructed in (A.1), that is, for any $\pi' \neq \hat{\pi}$,

$$\sigma^{\pi'}(\cdot \mid m_i') = \sigma^{\pi'}(\cdot \mid m_i'), \forall\, m_i' \in M_i. \tag{A.6}$$

Hence, under the new strategy profile $\hat{\sigma}$, the sender's expected payoff of any choice other than $\hat{\pi}$ remains the same as that in the given PBE. The given PBE implies that the sender achieves the maximum expected payoff by choosing π as compared to other choices under σ. Given that (a) is satisfied, by construction, the receivers' behavior is essentially the same under $\hat{\pi}$ as those of σ under π. Hence the sender's payoff when she chooses $\hat{\pi}$ under $\hat{\sigma}$ is the same as her choice of π under σ. Therefore, $\hat{\pi}$ is an optimal choice for the sender under $\hat{\sigma}$.

The proof of Claim 1 below concludes our proof. □

Proof of Claim 1. We will prove this claim by induction. The notations we use will follow those in the main proof. For notational convenience, for an arbitrarily fixed message (p_i, α_i), let

$$\widehat{M}_i^{\pi}(p_i, \alpha_i) := (\beta_i^{\pi})^{-1}(p_i) \cap (\sigma_i^{\pi})^{-1}(\alpha_i), \tag{A.7}$$

where β_i^{π} is a mapping from each message m_i' to a consistent conjecture and σ_i^{π} is a mapping from each message to the corresponding strategy.

We first establish Claim 1 for $k = 1$. Recall from the construction that (i) $\hat{\pi}$ essentially pools the set of messages $\widehat{M}_i^{\pi}(p_i, \alpha_i)$ of π into a single message (p_i, α_i); and that (ii) P_i^1 is constructed from the set of first-order characteristic pairs $\{(B_{i,h}^1, f_{i,h}^1) \mid k \in \mathbb{N}, h \in \mathscr{D}_i^1\}_{i \in I}$. To show $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i) \in \phi_i^1(p_i)$, it is suffice to show that for any binary partition $(F_{i,h}^1, V_i \backslash F_{i,h}^1)$ induced by any first-order characteristic pair $(B_{i,h}^1, f_{i,h}^1)$, the conjecture $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i)$ under $\hat{\pi}$ remains in the same component as $\beta^{\pi}(\cdot \mid m_i)$ does for all $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ under the given π.

Whether $\beta_i^{\pi}(\cdot \mid m_i) \in F_{i,h}^1$ for all $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ or $\beta_i^{\pi}(\cdot \mid m_i) \in V_i \backslash F_{i,h}^1$ for all $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$, given that $B_{i,h}^1$ is bi-convex, by Lemma 1, there must exist a set B' that takes one of the following forms: $(-\infty, b)$, (b, ∞), $(-\infty, b]$, or $[b, \infty)$ for some $b \in \mathbb{R}$ such that

$$\beta_i^{\pi}(\cdot \mid m_i) \in \left\{ v_i \in V_i \mid \int f_{i,h}^1 \mathrm{d}v_i \in B' \right\}, \forall\, m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i). \tag{A.8}$$

For the given message (p_i, α_i) under $\hat{\pi}$ and the obedient strategy $\hat{\sigma}$ in (A.2), by the pooling construction of $\hat{\pi}$, receiver i's consistent conjecture $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i)$ when projected to $\Delta(\Omega \times A_{-i})$ is a convex combination amongst elements in $\left\{ \mathrm{proj}_{\Delta(\Omega \times A_{-i})} \beta_i^{\pi}(\cdot \mid m_i) \mid m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i) \right\}$. Hence, $\int f_{i,h}^1 \mathrm{d}\hat{\beta}_i^{\hat{\pi}}(p_i, \alpha_i) \in B'$ follows immediately from the convex combination relation and the fact that each component $\beta_i^{\pi}(\cdot \mid m_i)$ in this combination satisfies (A.8).

Now, suppose that for an arbitrarily chosen finite positive integer $n \leq K$, Claim 1 holds for any k with $1 \leq k \leq n-1$. We will show that for any fixed message (p_i, α_i) and $i \in I$, $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i) \in \phi_i^n(p_i)$ as well.

Given that P_i^n is the partition that joins all the binary partitions induced by n-order characteristic pairs and P_i^{n-1}, to show $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i) \in \phi_i^n(p_i)$, it suffices to show that for any binary partition $(F_{i,h}^n, V_i \backslash F_{i,h}^n)$ induced by any n-order characteristic pair $(B_{i,h}^n, f_{i,h}^n)$, the conjecture $\hat{\beta}_i^{\hat{\pi}}(\cdot \mid p_i, \alpha_i)$ is located in the same component of the binary partition as $\beta_i^{\pi}(\cdot \mid m_i)$ for any $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ under the given π. By the bi-convex of $B_{i,h}^n$ and Lemma 1, regardless of whether $\beta_i^{\pi}(\cdot \mid m_i) \in F_{i,h}^n$ or $\beta_i^{\pi}(\cdot \mid m_i) \in V_i \backslash F_{i,h}^n$, there exists a Borel set $\hat{B}$ that takes one of the following forms: $(-\infty, b)$, (b, ∞), $(-\infty, b]$, or $[b, \infty)$ for some $b \in \mathbb{R}$ such that

$$\beta_i^{\pi}(\cdot \mid m_i) \in \left\{ v_i \in V_i \mid \int f_{i,h}^n \mathrm{d}v_i \in \hat{B} \right\}, \forall\, m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i). \tag{A.9}$$

Given the type space $(T_i, \zeta_i)_{i \in I}$, where ζ_i is the isomorphism mapping each receiver's type to their consistent conjecture, recall that we require the function $f_{i,h}^n$ in the n-order characteristic pair to satisfy the iterative measurability condition, which is

$$f_{i,h}^n = \sum_{l=1}^{L} c_l \cdot 1_{(\omega_l, \zeta_{-i}^{-1}(F_{-i}^l), a_{-i}^l)} \tag{A.10}$$

with $\omega_l \in \Omega$, $F_{-i}^l \in P_{-i}^{n-1}$, $a_{-i}^l \in A_{-i}$, and $c_l \in \mathbb{R}$ for some finite integer L. This condition will play an important role in our proof later on. For notational convenience, for each index l with $1 \leq l \leq L$, denote

$$S_l := \omega_l \times \zeta_{-i}^{-1}(F_{-i}^l) \times a_{-i}^l = \omega_l \times \prod_{j \neq i} \zeta_j^{-1}(F_j^l) \times \prod_{j \neq i} a_j^l. \tag{A.11}$$

Thus, the probability that the belief map $\beta_i^{\pi}(\cdot \mid m_i)$ assigns to S_l given message $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ under π and σ is as follows:

$$\begin{aligned} \beta_i^{\pi}(S_l \mid m_i) &= \int_{(\beta_{-i}^{\pi})^{-1}(F_{-i}^l)} \sigma_{-i}^{\pi}(a_{-i}^l \mid m_{-i}) \mathrm{d}\Pr(\omega_l \times m_{-i} | m_i, \pi) \\ &= \sum_{\alpha_{-i}' \in (\Delta(A_j))_{j \neq i}} \alpha_{-i}'(a_{-i}^l) \mathrm{d}\Pr(\omega_l \times (\beta_{-i}^{\pi})^{-1}(F_{-i}^l) \cap (\sigma_{-i}^{\pi})^{-1}(\alpha_{-i}') \mid m_i, \pi). \end{aligned} \tag{A.12}$$

To proceed further, it is essential to establish several important dichotomy relations between $\beta_i^{\pi}(S_l \mid m_i)$ with $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ and $\hat{\beta}_i^{\hat{\pi}}(S_l \mid p_i, \alpha_i)$. By the induction hypothesis, for any fixed message $(\hat{p}_j, \hat{\alpha}_j) \in P_j \times \Delta(A_j)$ under the $\hat{\pi}$ constructed in (A.1) and any $j \in I$, we have:

$$\hat{\beta}_j^{\hat{\pi}}(\cdot \mid \hat{p}_j, \hat{\alpha}_j) \in \phi_j^{n-1}(\hat{p}_j). \tag{A.13}$$

For any such $\hat{p}_j \in P_j$ with $j \neq i$, recall that we denote $\phi_j^{n-1}(\hat{p}_j)$ as the unique component in P_j^{n-1} such that $\hat{p}_j \subseteq \phi_j^{n-1}(\hat{p}_j)$. Given that $F_{-i}^l = (F_j^l)_{j \neq i} \in P_{-i}^{n-1}$ in (A.11), we have:

$$\text{either } \phi_j^{n-1}(\hat{p}_j) \subseteq F_j^l \text{ or } \phi_j^{n-1}(\hat{p}_j) \subseteq V_j \backslash F_j^l \text{ must hold.} \tag{A.14}$$

Given that P_j is a refine partition of P_j^{n-1} for any $j \in I$, it follows that for any $\hat{p}_j \in P_j$ with $j \neq i$, we also have:

$$\text{either } \hat{p}_j \subseteq F_j^l \text{ or } \hat{p}_j \subseteq V_j \backslash F_j^l \text{ must hold.} \tag{A.15}$$

Therefore, for any such $\hat{p}_j \in P_j$ with $j \neq i$, the above dichotomies, along with the inclusion that $\hat{p}_j \in \phi_j^{n-1}(\hat{p}_j)$, together imply

$$\begin{aligned}&\hat{p}_j \subseteq F_j^l \text{ if and only if } \phi_j^{n-1}(\hat{p}_j) \subseteq F_j^l\\&\hat{p}_j \subseteq V_j \backslash F_j^l \text{ if and only if } \phi_j^{n-1}(\hat{p}_j) \subseteq V_j \backslash F_j^l.\end{aligned} \tag{A.16}$$

For (ω_l, F_{-i}^l) appearing in (A.11), and any arbitrarily chosen $\hat{\alpha}_{-i} \in \prod_{j\neq i} \Delta(A_j)$, define:

$$\widehat{M}_{-i}^{\pi}(F_{-i}^l, \hat{\alpha}_{-i}) := \{m'_{-i} \mid \beta_{-i}^{\pi}(\cdot \mid m'_{-i}) \in F_{-i}^l \& \sigma_{-i}^{\pi}(\cdot \mid m'_{-i}) = \hat{\alpha}_{-i}\}. \tag{A.17}$$

Recall from the construction that for any $j \in I$, $\hat{\pi}$ essentially pools the set of messages $m_j \in \widehat{M}_j^{\pi}(p_j, \alpha_j)$ under π into a single message (p_j, α_j) according to the conjecture location and equilibrium actions. For the arbitrarily fixed (p_i, α_i) for receiver i, any F_{-i}^l in (A.11), and $\hat{\alpha}_{-i} \in (\Delta(A_j))_{j\neq i}$,

$$\begin{aligned}&\Pr(\omega_l \times (\hat{\beta}_{-i}^{\hat{\pi}})^{-1}(F_{-i}^l) \cap (\hat{\sigma}_{-i}^{\hat{\pi}})^{-1}(\hat{\alpha}_{-i}) \mid (p_i, \alpha_i), \hat{\pi})\\ \overset{(A.13)\&(A.14)}{=}\ &\Pr(\omega_l \times \{(p'_{-i}, \alpha'_{-i}) \mid \phi_{-i}^{n-1}(p'_{-i}) \subseteq F_{-i}^l \,\&\, \hat{\sigma}_{-i}^{\hat{\pi}}(p'_{-i}, \alpha'_{-i}) = \hat{\alpha}_{-i}\} \mid (p_i, \alpha_i), \hat{\pi})\\ \overset{(A.15)\&(A.16)}{=}\ &\Pr(\omega_l \times \widehat{M}_{-i}^{\pi}(F_{-i}^l, \hat{\alpha}_{-i}) \mid \widehat{M}_i^{\pi}(p_i, \alpha_i), \pi) \quad \text{(recall the definitions in(A.7)\&(A.17))}\\ =\ &\Pr(\omega_l \times (\beta_{-i}^{\pi})^{-1}(F_{-i}^l) \cap (\sigma_{-i}^{\pi})^{-1}(\hat{\alpha}_{-i}) \mid \widehat{M}_i^{\pi}(p_i, \alpha_i), \pi).\end{aligned} \tag{A.18}$$

The above equation establishes an important connection between $\beta_i^{\pi}(S_l \mid m_i)$ for all $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ and $\hat{\beta}_i^{\hat{\pi}}(S_l \mid p_i, \alpha_i)$. With this equation, we can now compute receiver i's conjecture $\hat{\beta}_i^{\hat{\pi}}(S_l \mid p_i, \alpha_i)$ for the set S_l defined in (A.11) as follows:

$$\begin{aligned}\hat{\beta}_i^{\hat{\pi}}(S_l \mid p_i, \alpha_i) &= \int_{(\hat{\beta}_{-i}^{\hat{\pi}})^{-1}(F_{-i}^l)} \hat{\sigma}_{-i}^{\hat{\pi}}(a_{-i}^l \mid m'_{-i}) \mathrm{d}\Pr(\omega_l \times m'_{-i} \mid (p_i, \alpha_i), \hat{\pi})\\ &= \sum_{\alpha'_{-i} \in (\Delta(A_j))_{j\neq i}} \alpha'_{-i}(a_{-i}^l) \mathrm{d}\Pr(\omega_l \times (\hat{\beta}_{-i}^{\hat{\pi}})^{-1}(F_{-i}^l) \cap (\hat{\sigma}_{-i}^{\hat{\pi}})^{-1}(\alpha'_{-i}) \mid (p_i, \alpha_i), \hat{\pi})\\ &\overset{(A.18)}{=} \sum_{\alpha'_{-i} \in (\Delta(A_j))_{j\neq i}} \alpha'_{-i}(a_{-i}^l) \mathrm{d}\Pr(\omega_l \times (\beta_{-i}^{\pi})^{-1}(F_{-i}^l) \cap (\sigma_{-i}^{\pi})^{-1}(\alpha'_{-i}) \mid \widehat{M}_i^{\pi}(p_i, \alpha_i), \pi)\\ &= \int_{\widehat{M}_i^{\pi}(p_i, \alpha_i)} \underbrace{\sum_{\alpha'_{-i} \in (\Delta(A_j))_{j\neq i}} \alpha'_{-i}(a_{-i}^l) \mathrm{d}\Pr(\omega_l \times (\beta_{-i}^{\pi})^{-1}(F_{-i}^l) \cap (\sigma_{-i}^{\pi})^{-1}(\alpha'_{-i}) \mid m_i, \pi)}_{\beta_i^{\pi}(S_l \mid m_i)} \mathrm{d}\Pr(m_i \mid \widehat{M}_i^{\pi}(p_i, \alpha_i), \pi)\\ &= \int_{\widehat{M}_i^{\pi}(p_i, \alpha_i)} \beta_i^{\pi}(S_l \mid m_i) \mathrm{d}\Pr(m_i \mid \widehat{M}_i^{\pi}(p_i, \alpha_i), \pi).\end{aligned}$$

Given that the above relation between $\hat{\beta}_i^{\hat{\pi}}(S_l \mid p_i, \alpha_i)$ and $\beta_i^{\pi}(S_l \mid m_i)$ for any message $m_i \in \widehat{M}_i^{\pi}(p_i, \alpha_i)$ and recalling the specifications in (A.10) for the characteristic function $f_{i,h}^n$, the linearity of the integration allows us to conclude the following:

$$\int f_{i,h}^n \mathrm{d}\hat{\beta}_i^{\hat{\pi}}(p_i, \alpha_i) = \int_{\widehat{M}_i^{\pi}(p_i, \alpha_i)} \int f_{i,h}^n \mathrm{d}\beta_i^{\pi}(m_i) \mathrm{d}\Pr(m_i \mid \widehat{M}_i^{\pi}(p_i, \alpha_i), \pi). \tag{A.19}$$

Therefore, for the above fixed $\widehat{B}$ in (A.9) that takes one of the following forms: $(-\infty, b)$, (b, ∞), $(-\infty, b]$, or $[b, \infty)$ for some $b \in \mathbb{R}$, it follows immediately from (A.19) and the condition (A.9) that $\int f_{i,h}^n \mathrm{d}\hat{\beta}_i^{\hat{\pi}}(p_i, \alpha_i) \in \widehat{B}$. Note that this conclusion hinges on $f_{i,h}^n$ satisfying the iterative measurability condition. Thus, Claim 1 holds for any $1 \leq k \leq n$, and we conclude our proof. □

References

Alonso, R., Camara, O., 2016. Persuading voters. Am. Econ. Rev. 106 (11), 3590–3605. https://doi.org/10.1257/aer.20140737

Aumann, R.J., 1987. Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55 (1), 1–18. https://doi.org/10.2307/1911154

Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Am. Econ. Rev. 97 (2), 170–176. https://doi.org/10.1257/aer.97.2.170

Battigalli, P., Dufwenberg, M., 2022. Belief-dependent motivations and psychological game theory. J. Econ. Lit. 60 (3), 833–882. https://doi.org/10.1257/jel.20201378

Bergemann, D., Morris, S., 2016. Bayes correlated equilibrium and the comparison of information structures in games. Theor. Econ. 11, 487–522. https://doi.org/10.3982/TE1808

Brandenburger, A., Dekel, E., 1993. Hierarchies of beliefs and common knowledge J. Econ. Theory 59 (1), 189–198. https://doi.org/10.1006/jeth.1993.1012

Chen, Y.-C., Di Tillio, A., Faingold, E., Xiong, S., 2017. Characterizing the strategic impact of misspecified beliefs. Rev. Econ. Stud. 84, 1424–1471. https://doi.org/10.1093/restud/rdw061

Cho, I.-K., Kreps, D.M., 1987. Signaling games and stable equilibria. Q. J. Econ. 102 (2), 179–221. https://doi.org/10.2307/1885060

Dekel, E., Fudenberg, D., Morris, S., 2007. Interim correlated rationalizability. Theor. Econ. 2 (1), 15–40. https://econtheory.org/ojs/index.php/te/article/view/20070015.

Ely, J., Frankel, A., Kamenica, E., 2015. Suspense and surprise. J. Polit. Econ. 123 (1), 215–260. https://doi.org/10.1086/677350

Ely, J., Peski, M., 2006. Hierarchies of belief and interim rationalizability. Theor. Econ. 1 (1), 19–65. https://www.econtheory.org/ojs/index.php/te/article/view/20060019/0.

Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425, 785–791. https://doi.org/10.1038/nature02043

Fehr, E., Fischbacher, U., 2004. Third-party punishment and social norms. Evol. Hum. Behav. 25 (2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Friedenberg, A., Meier, M., 2017. The context of the game. Econ. Theory 63, 347–386. https://doi.org/10.1007/s00199-015-0938-z

Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1 (1), 60–79. https://doi.org/10.1016/0899-8256(89)90005-5
Harsanyi, J.C., Selten, R., 1988. A General Theory of Equilibrium Selection in Games. MIT Press, Cambridge, MA.
Kamenica, E., Gentzkow, M., 2011. Bayesian persuasion. Am. Econ. Rev. 101 (6), 2590–2615. https://doi.org/10.1257/aer.101.6.2590
Lipnowski, E., Mathevet, L., 2018. Disclosure to a psychological audience. Am. Econ. J. Microecon. 10 (4), 67–93. https://doi.org/10.1257/mic.20160247
Lipnowski, E., Ravid, D., Shishkin, D., 2026. Perfect Bayesian persuasion. J. Polit. Econ. Microecon. In Press. https://doi.org/10.1086/740148
Liu, Q., 2009. On redundant types and Bayesian formulation of incomplete information. J. Econ. Theory 144 (5), 2115–2145. https://doi.org/10.1016/j.jet.2009.02.002
Mathevet, L., Perego, J., Taneva, I., 2020. On information design in games. J. Polit. Econ. 128 (4), 1370–1404. https://doi.org/10.1086/705332
Mertens, J.-F., Zamir, S., 1985. Formulation of Bayesian analysis for games with incomplete information. Int. J. Game Theory 14 (1), 1–29. https://doi.org/10.1007/BF01770224
Milgrom, P., Roberts, J., 1986. Relying on the information of interested parties. RAND J. Econ. 17 (1), 18–32. https://doi.org/10.2307/2555625
Morris, S., Oyama, D., Takahashi, S., 2024. Implementation via information design in binary-action supermodular games. Econometrica 92 (3), 775–813. https://doi.org/10.3982/ECTA19149
Myerson, R.B., 1991. Game Theory: Analysis of Conflict. Harvard University Press.
Rand, D.G., Greene, J.D., Nowak, M.A., 2012. Spontaneous giving and calculated greed. Nature 489 (7416), 427–430. https://doi.org/10.1038/nature11467
Royden, H.L., 1988. Real analysis. Macmillan Publishing Company, New York. third edition edition.
Samuelson, L., 1998. Evolutionary Games and Equilibrium Selection. Economic Learning and Social Evolution, The MIT Press. second edition edition.
Taneva, I., 2019. Information design. Am. Econ. J. Microecon. 11 (4), 151–185. https://doi.org/10.1257/mic.20170351
Wu, W., 2023. Sequential bayesian persuasion. J. Econ. Theory 214, 105763. https://doi.org/10.1016/j.jet.2023.105763