Discussion Paper Series – CRC TR 224

Discussion Paper No. 632
Project B 02

From Design to Disclosure

S. Nageeb Ali[1]
Andreas Kleiner[2]
Kun Zhang[3]

January 2025

[1]Pennsylvania State University, Email: nageeb@psu.edu
[2]University of Bonn, Email: andykleiner@gmail.com
[3]University of Queensland. Email: kun@kunzhang.org

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

From Design to Disclosure*

S. Nageeb Ali† Andreas Kleiner‡ Kun Zhang§

January 6, 2025

Abstract

This paper studies games of voluntary disclosure in which a sender discloses evidence to a receiver who then offers an allocation and transfers. We characterize the set of equilibrium payoffs in this setting. Our main result establishes that any payoff profile that can be achieved through information design can also be supported by an equilibrium of the disclosure game. Hence, our analysis suggests an equivalence between disclosure and design in these settings. We apply our results to monopoly pricing, bargaining over policies, and insurance markets.

*We thank Navin Kartik, Elliot Lipnowski, Stephen Morris, Jacopo Perego, Roland Strausz, Kai Hao Yang, and various seminar and conference attendees. Kleiner acknowledges financial support from the German Research Foundation (DFG) through Germany's Excellence Strategy - EXC 2047/1 - 390685813, EXC 2126/1-390838866 and the CRC TR-224 (Project B02).

†Department of Economics, Pennsylvania State University. Email: nageeb@psu.edu.

‡Department of Economics, University of Bonn. Email: andykleiner@gmail.com.

§School of Economics, University of Queensland. Email: kun@kunzhang.org.

1 Introduction

In many strategic interactions, one player can disclose hard information to persuade others. For instance, the seller of an asset can reveal audited statements or forecasts about the asset value to prospective buyers so as to secure better price offers. An insuree can disclose her medical information or her past driving record to an insurer to obtain more favorable terms. A worker might disclose her salary to prospective employers so as to influence their offers.

An important literature, dating back to Grossman (1981) and Milgrom (1981), models these interactions as *disclosure* or *persuasion games*. These games feature two primary ingredients. First, the informed party—or the sender—shares hard information that cannot be manipulated: she can share nothing but the truth although she need not share the entire truth. Second, she cannot commit to what she discloses in that of the utterances she can make, she chooses that which results in the best outcome. Of course, the party hearing these utterances—the receiver—is no fool. He draws inferences from both what is said and what is left unsaid. Hence, what the sender communicates and how the receiver responds is inherently determined in equilibrium.

The classical analysis of these settings shows that the combination of these ingredients results in unraveling: in the unique equilibrium outcome, the sender voluntarily discloses all of her information. The logic for why there exists a fully revealing equilibrium is that were the sender to conceal information, the receiver assumes the worst and takes an action that makes concealment unprofitable. A more powerful logic pushes every equilibrium to be fully revealing: in the games studied by this prior work, for every pool of types, there exists some type of the sender that strictly prefers to separate rather than be pooled.[1] Hence, behavior with any pooling necessarily unravels and are untenable as equilibria.

While it hews to some applications that the sender may prefer to reveal her type rather than be pooled with other types, this assumption fails in many settings. Take, for instance, the interaction between a buyer and seller in which the buyer can disclose information about her value before the seller makes a take-it-or-leave-it offer. In this setting, the buyer does not profit from revealing her true value as the seller would then extract her full surplus. Analogously, in a monopolistic insurance market, disclosing all available information is guaranteed to result in a contract that leaves the insuree just indifferent between accepting and rejecting the contract. More broadly, in many principal-agent settings with flexible transfers and allocations, the agent cannot strictly gain from disclosing all information as that results in the principal

[1] Grossman (1981) and Milgrom (1981) impose a monotonicity condition that stipulates that every type strictly prefers inducing beliefs in a particular direction. More broadly, Okuno-Fujiwara, Postlewaite, and Suzumura (1990), Seidmann and Winter (1997), Mathis (2008), and Dasgupta (2023) study conditions on preferences under which every equilibrium is fully revealing.

proposing an action that makes the agent's individual rationality constraints bind. In these settings, does unraveling fail? What is the full range of equilibrium payoffs and how do they compare to those achieved through information design or cheap talk?

To answer these questions, we study a broad class of sender-receiver games in which the sender discloses evidence about her type to a receiver. We model evidence as in the classical setup of Grossman (1981) and Milgrom (1981): the sender can communicate any subset of types that contains her true type. Thus, messages can vary from being perfectly precise to completely vague. Upon receiving this disclosure, the receiver chooses an action. Our framework is agnostic about the nature of this action: consistent with monopoly pricing, it could be the price set by a seller, or a menu of insurance contracts offered by an insurer, or simply an allocation proposed by a principal.

We make three assumptions. First, the sender favors uncertainty: relative to her payoff from inducing any belief, she never strictly gains from fully revealing her type. Second, for every message, there is a "worst-case type" such that any type who could send that message would prefer its own complete-information outcome to that of the worst-case type. Third, the prior distribution is continuous and the receiver's payoff also satisfies a continuity assumption with respect to her actions and beliefs. These three assumptions are satisfied in many settings, including monopoly pricing, policy negotiations, and insurance contracting.

We characterize all equilibrium payoffs of this game. We find that the range of outcomes goes far beyond full revelation, including outcomes consistent with partial and non-disclosure. To formalize our conclusion, let us take payoffs that stem from information design as a point of comparison: suppose the sender were to commit ex ante to an information structure that reveals information to the receiver before the receiver chooses his action. We call a payoff profile (u_S^*, u_R^*) *achievable* if there exists an information structure that induces these payoffs. Clearly, every equilibrium payoff of our disclosure game must be achievable, as the sender's disclosure strategy can be replicated through an information structure. Our main result, Theorem 1, establishes that the converse also holds.

> **Main Result.** *Every achievable payoff profile can be (approximately) supported by an equilibrium of the disclosure game.*

This result identifies how, given our assumptions, there is virtually no gap between information design and voluntary disclosure: for every achievable payoff profile (u_S^*, u_R^*), and every $\varepsilon > 0$, there is an equilibrium of our disclosure game that attains payoffs within ε of (u_S^*, u_R^*). To see how we prove this result, observe that the main wedge between information design and a disclosure game is that the former allows the sender to commit to an information structure whereas in the latter, the sender does not mix between different messages unless

those messages accrue the same payoff. Therefore, in the disclosure game, a sender type cannot be "split" across segments. With this obstacle in mind, our first step shows that "partitional" segmentations—namely those in which types are partitioned into different segments, and only a measure-0 set are split—can be supported as an equilibrium of the disclosure game. Our second step evaluates the loss that the restriction to partitional segmentations entails: we show that any payoff profile achieved by an information structure can be approximated arbitrarily closely by a finite partitional segmentation. Importantly, our method of proof does not require us to know the set of achievable payoffs. Hence, the analysis applies to settings in which the implications of information design have yet to be understood. Finally, we illustrate that the three main assumptions are, in some sense, necessary in that once any assumption fails, we can find settings in which the conclusion of our main result also does not hold.

Beyond showing that disclosure games can support a rich set of outcomes, Theorem 1 offers several ancillary implications. For one, it shows that the sender does not value commitment in that she can achieve payoffs arbitrarily close to those of her optimal information structure in an equilibrium of the disclosure game. More broadly, voluntary disclosure can offer a microfoundation for information design in these settings: rather than invoking a third-party that knows the sender's type or an exogenous information structure, information flows directly from the sender to the receiver in a completely standard disclosure game. Finally, our analysis offers an indirect comparison of voluntary disclosure to cheap-talk communication; voluntary disclosure can support at least as much, and in many applications of our setting, a larger range of outcomes. Augmenting our disclosure game with cheap-talk communication would result in virtually the same set of supportable payoffs.

Our leading application is to the monopoly pricing problem studied by Bergemann, Brooks, and Morris (2015). The conclusion here is that every payoff in the "BBM-triangle" can be supported exactly or virtually as an equilibrium of the disclosure game. In this setting, we also obtain an additional conclusion: the truth-leaning refinement proposed by Hart, Kremer, and Perry (2017) refines the set of supportable payoffs to the efficiency frontier, and virtually any such payoff can be supported through a truth-leaning equilibrium.

A second application is to "veto-bargaining" models used to study policy negotiations in political and organizational contexts; see Cameron and McCarty (2004) for a survey. In these models, a proposer negotiates with a vetoer who has single-peaked preferences over policy. Romer and Rosenthal (1978)'s canonical formulation presupposes that the proposer knows the vetoer's ideal point. Our analysis speaks to a setting in which the proposer does not; instead, the privately-informed vetoer can disclose evidence about her ideal point so as to influence what the proposer offers. Although this setting lacks transferable utility, it nevertheless satisfies the assumption that a vetoer does not profit from fully revealing her

ideal point as the proposer would then follow up with an offer that leaves her just indifferent between accepting and rejecting. In this context, it is not known what information design delivers; i.e., the set of achievable payoffs has yet to be characterized.[2] Nevertheless, our result asserts that all achievable payoffs can be supported as an equilibrium of the disclosure game. Matthews (1989) studies this setting with cheap talk rather than disclosure; he shows that cheap talk supports outcomes with at most two pools. Voluntary disclosure thus supports significantly more outcomes.

Our third and final application is to insurance contracting. In this context, lawmakers debate whether firms should be allowed to offer different contracts based on an insuree's disclosure of genetic information. On one hand, offering evidence could lead to more efficient outcomes and greater market coverage. On the other hand, one might worry that giving insurees the opportunity to disclose information compels them to do so in equilibrium.[3] Against this backdrop, our result highlights rich possibilities that come from allowing disclosure. Through disclosure, the insuree could attain payoffs arbitrarily close to her (ex ante) optimal information structure. At the same time, there are equilibria that make her worse off relative to the setting in which she cannot disclose information. Thus, our analysis shows that whether regulations should allow insurers to offer contracts based on insurees' disclosures hinges on the equilibrium that is selected in the subsequent contracting game.[4]

We connect our work to prior contributions in information disclosure. We follow Grossman and Hart (1980), Grossman (1981), Milgrom (1981), and Milgrom and Roberts (1986) in that the sender can disclose any set of types that contains her true type. Our main departure is that we allow the receiver to flexibly choose allocations, transfers, or actions so that the sender does not profit from fully revealing her type. This flexibility is at the root of why the standard unraveling logic fails and other equilibrium payoffs can be supported.[5]

Several papers study how disclosure can improve the sender's payoff in settings with transferable utility. Studying various market structures, Glode, Opp, and Zhang (2018) and Ali, Lewis, and Vasserman (2023) show that the sender may benefit from disclosing evidence relative to the benchmark in which she cannot do so.[6] Closer to our study, Pram (2021) studies a

[2] Kim, Kim, and Van Weelden (2024) study the different question of what happens if the vetoer does not know her ideal point and the proposer can furnish information to her about it.

[3] With this concern in mind, the US Congress passed the *Genetic Information Nondiscrimination Act* in 2008, prohibiting health insurance companies from basing their contracts on genetic test results. See Erwin (2008) and Pram (2023) for discussion of this debate.

[4] We abstract from costs of information and evidence acquisition. Pram (2023) models this issue, offering conditions on the prior and information costs under which the insuree benefits from disclosing her riskiness.

[5] Prior work has shed light on other obstacles to unraveling such as uncertainty about whether the sender has evidence (Dye, 1985), disclosure costs (Jovanovic, 1982; Verrecchia, 1983), or the possibility that receivers are naive (e.g., Hagenbach and Koessler, 2017; Jin, Luca, and Martin, 2021).

[6] Madarasz and Pycia (2023) study how an informed buyer would choose information structures; their analysis emphasizes how information costs result in the buyer choosing a least costly information structure.

principal-agent setting with transfers to evaluate when it is possible for the sender to benefit from disclosure. With respect to these prior papers, our analysis makes several contributions. First, the main result characterizes the entire set of disclosure payoffs, showing that not only are improvements possible but also any payoff profile that can be achieved through information design. Second, our analysis distills the conditions under which this conclusion holds; one of our applications shows that utility need not be transferable.

We adopt the worst-case type assumption of Seidmann and Winter (1997) and Hagenbach, Koessler, and Perez-Richet (2014); the former studies when an unraveling equilibrium exists in a general sender-receiver game with evidence and the latter considers this question in a setting where a group of players chooses actions after sharing evidence.

We take the evidence structure from Grossman (1981) and Milgrom (1981) as a primitive. Theorem 1 implies that no other evidence structure generates a larger set of supportable payoffs, given that an equilibrium payoff under any evidence structure can be achieved via information design. In different contexts, one might be interested in how the sender or an intermediary might design evidence to influence disclosure.[7] Kamenica and Gentzkow (2011, pp. 2598-2599) show that voluntary disclosure can then support the sender-optimal information structure: they describe a different disclosure game in which an uninformed sender publicly chooses a Blackwell experiment, privately observes its realization s, and then can send any message m that contains s. In our setting, the sender can virtually obtain that value using the standard evidence structure without resorting to evidence design.

In disclosure games with finitely many actions, Titova and Zhang (2024) identify conditions under which the sender-optimal achievable payoff can be supported in an equilibrium. Their analysis is complementary in that they highlight the impact of the receiver's inflexible choices. By contrast, in the settings we study, it is the receiver's flexible action choices that make it unprofitable for the sender to reveal her type.

In our model, the receiver cannot commit to how he responds to disclosure. A related strand studies mechanism design with evidence (Hart, Kremer, and Perry, 2017; Ben-Porath, Dekel, and Lipman, 2019) and identifies how the receiver may not value commitment. In our monopoly-pricing application, we show that Hart, Kremer, and Perry's truth-leaning refinement selects efficient payoff profiles.

Our paper proceeds as follows. Section 2 describes the disclosure game. Section 3 our key assumptions and the main results. Section 4 considers our three applications. Section 5 concludes. Omitted proofs are in the Appendix.

[7]For instance, see DeMarzo, Kremer, and Skrzypacz (2019), Ali, Haghpanah, Lin, and Siegel (2022), Ben-Porath, Dekel, and Lipman (2023, 2024), Asseyer and Weksler (2024), Pollrich and Strausz (2024), and Shishkin (2024).

2 Model

We study a disclosure game in which a sender (she) shares evidence with a receiver (he) who then chooses an action. The sender privately observes her type θ drawn from the compact and convex set $\Theta \subseteq \mathbb{R}^n$ according to the probability measure F that admits a strictly positive density f with respect to the Lebesgue measure on $\mathbb{R}^n$.[8] Her type determines what she can say: the sender of type θ chooses a message in $\mathcal{M}(\theta) := \{m \in \mathcal{C} : \theta \in m\}$, where $\mathcal{C}$ denotes the collection of all non-empty closed subsets of Θ. We interpret each message as *evidence* in that the statement she makes, such as "my type is in m," must be true. Observe that $\mathcal{M}(\theta)$ includes the fully revealing message $\{\theta\}$ (which is available only to type θ), the fully concealing message Θ (which is available to every type), and a wide range of messages that reveal some but not all information about her type.[9]

After receiving the sender's message, the receiver chooses an action a from a compact metrizable space A. The sender's payoff is $u_S(a, \theta)$, which is continuous in a for each θ, and the receiver's payoff is $u_R(a, \theta)$, which is upper semicontinuous in a for each θ.

We describe strategies and equilibria for this game. The sender's strategy is a function $\rho : \Theta \to \Delta(C)$ in which the support of $\rho(\theta)$ is a subset of $\mathcal{M}(\theta)$ for every type θ. The receiver's strategy is a function $\tau : \mathcal{C} \to A$.[10] The receiver's beliefs about the sender's type are represented by the belief system $\mu : \mathcal{C} \to \Delta(\Theta)$. An assessment $((\rho, \tau), \mu)$ is a Perfect Bayesian Equilibrium (henceforth, equilibrium) if the following conditions hold:

(a) *Given her type, the sender discloses evidence optimally*: for every type θ, $\rho(\theta)$ is supported on $\arg\max_{m \in \mathcal{M}(\theta)} u_S(\tau(m), \theta)$,

(b) *Given the message, the receiver chooses actions optimally*: for every message m, $\tau(m) \in \arg\max_{a \in A} \int_\Theta u_R(a, \theta)\, \mathrm{d}\mu(\theta|m)$,

(c) *Beliefs respect evidence*: for every message m, the receiver's beliefs $\mu(m)$ have a support that is a subset of m.

(d) *Bayes' Rule*: the beliefs μ are obtained from F given ρ using Bayes' rule, i.e., μ is a regular conditional probability system.

We say that a payoff profile (u_S^*, u_R^*) is *supportable* if it corresponds to the ex ante expected payoff of some equilibrium.

This framework encompasses numerous applications, which we discuss below.

[8] In Section 3.5, we show how our results extend to a finitely-supported F.

[9] Our specification matches that of Grossman and Hart (1980), Grossman (1981), and Milgrom (1981). Equivalently, we could have formulated evidence in the space of "documents," as in Lipman and Seppi (1995), Bull and Watson (2004), Hart, Kremer, and Perry (2017), and Ben-Porath, Dekel, and Lipman (2019).

[10] We endow $\mathcal{C}$ with the Hausdorff metric. Throughout our analysis, for a compact metrizable space X, $\Delta(X)$ denotes the set of (Borel) probability measures on X endowed with the weak* topology. Our analysis allows for mixed actions; then, we interpret A as the set of probability distributions over (pure) actions.

Example 1 (Monopoly Pricing)**.** Consider a disclosure analogue of Bergemann, Brooks, and Morris (2015): the buyer of a good can disclose evidence about her value θ for a good to a monopolist who then responds with a price offer that the buyer can accept or reject. Here, the buyer's type θ is drawn from $\Theta = [\underline{\theta}, \overline{\theta}]$ and action $a \in \mathbb{R}$ represents a price. We write $u_S(a, \theta) = \max\{\theta - a, 0\}$ for the buyer's payoff and $u_R(a, \theta) = a\mathbf{1}_{a \leq \theta}$ for the monopolist's payoff, reflecting that if the buyer accepts the price a, trade happens at that price, and if she rejects, each party obtains 0.[11]

Example 2 (Bargaining Over Policy)**.** Consider a disclosure analogue of the veto bargaining model of Romer and Rosenthal (1978). A proposer and vetoer negotiate over action $a \in \mathbb{R}$. The proposer has payoffs $u(a)$ that are strictly increasing in a whereas the vetoer has payoffs $v(a, \theta)$ that are strictly single-peaked in a with a unique maximizer θ and symmetric around the maximizer. The vetoer's type θ is drawn from $\Theta = [\underline{\theta}, \overline{\theta}]$ in which $\underline{\theta} \geq 0$. Following disclosure, the proposer offers a policy that the vetoer can accept or reject; if she accepts, then the proposed policy prevails and if she rejects, then the status-quo policy $a_Q := 0$ is preserved. There are no transfers in this setting. Observe that vetoer accepts a policy $a \geq 0$ if and only if $a \leq 2\theta$. Hence, putting the vetoer in the shoes of the sender and the proposer in the shoes of the receiver, we would write $u_S(a, \theta) = \max\{v(a, \theta), v(0, \theta)\}$, and $u_R(a, \theta) = u(a)\mathbf{1}_{a \leq 2\theta} + u(0)\mathbf{1}_{a > 2\theta}$.

Example 3 (Insurance Contracting)**.** As we discuss in Section 4.3, this setup can also capture insurance contracting in which an insuree has initial wealth $w > 0$, faces a potential loss $\ell \in (0, w)$ with probability $\theta \in (0, 1)$ and discloses evidence about her riskiness to a risk-neutral insurer. The insurer then offers a menu of contracts $(x(\theta), t(\theta)) \in \mathbb{R}^2$, comprising an indemnity payment $x(\theta)$ in the event of a loss and a premium $t(\theta)$.

3 When Disclosure Attains Design

3.1 The Information Design Benchmark

We use the payoffs that stem from information design as a benchmark. In this benchmark, the receiver observes the realization of a Blackwell experiment and then chooses an action. We refer to a distribution over types $G \in \Delta(\Theta)$ as a *belief*. A *segmentation* is a distribution $\sigma \in \Delta(\Delta(\Theta))$ over beliefs that average to the prior F—that is, $\int G \, \mathrm{d}\sigma(G) = F$—and a belief in the support of a segmentation is a *segment*. A segmentation *achieves* a payoff profile (u_S^*, u_R^*) if player i's ex ante expected payoff from the segmentation, given that the

[11] Ali, Lewis, and Vasserman (2023) study a disclosure game of this form but with the additional restriction that $\mathcal{M}(\theta)$ is either $\{\{\theta\}, \Theta\}$ or all intervals that contain θ.

receiver best-responds, is u_i^*; a payoff profile is *achievable* if some segmentation achieves it. For instance, in monopoly pricing (Example 1), the set of achievable payoffs is the "BBM-triangle" characterized by Bergemann, Brooks, and Morris (2015), namely all feasible payoff profiles in which the monopolist does as well as she would from setting a uniform price.

Every equilibrium payoff profile of the disclosure game is achievable because every sender strategy induces a segmentation. Our main result pertains to the converse, namely conditions under which every achievable payoff can be approximated by an equilibrium of the disclosure game. We turn to these conditions next.

3.2 The Key Assumptions

We study the implications of three assumptions. The first stipulates that the sender does not benefit from fully disclosing her type, the second asserts that for every message, there is a "worst-case" type that no type profits from imitating, and the third is a continuity assumption that allows us to suitably perturb beliefs without significantly changing the receiver's optimal actions.

To formalize these assumptions, define $a^*(G) := \arg\max_{a\in A} \int u_R(a,\theta)\mathrm{d}G(\theta)$ to be the receiver's optimal actions given belief G. With a slight abuse of notation, $a^*(\theta)$ denotes the receiver's optimal action if the sender's type is known to be θ (with tie-breaking that results in the lowest utility for the type-θ sender).

Our first assumption states that the complete-information payoff for the type-θ sender is no more than any incomplete-information payoff.

Assumption 1. *For every type θ, belief G such that $\theta \in \operatorname{supp} G$, and action $a \in a^*(G)$,*

$$u_S(a,\theta) \geq u_S(a^*(\theta),\theta).$$

Assumption 1 asserts that the sender's payoff from inducing any belief that is consistent with her type is higher than that from fully revealing her type. In other words, the sender never strictly benefits from fully revealing her type. This assumption holds in many principal-agent settings: fully revealing the private information results in the principal making a take-it-or-leave-it offer that fully extracts the agent's surplus. For instance, in the context of monopoly pricing (Example 1), revealing the type θ results in the monopolist setting a price $a^*(\theta) = \theta$, leaving the buyer with zero payoff; Assumption 1 then holds as the buyer's payoff cannot be lower than zero regardless of the monopolist's belief.

The discussion above may make it appear that Assumption 1 is predicated on the receiver having full bargaining power. The example below shows that the assumption can be compatible with the sender having some bargaining power.

Example 4 (Monopoly Pricing Revisited)**.** Consider an adaptation of Example 1 in which following disclosure, a *random recognition* rule determines who makes the take-it-or-leave-it offer: with probability $\alpha \in [0,1]$, the buyer makes an offer and, with complementary probability, the seller makes an offer. As the buyer's disclosure affects equilibrium payoffs only when the seller makes the offer, Assumption 1 still holds.

Our second assumption identifies a "worst-case type" that no type wishes to mimic.

Assumption 2. *Every message m contains a* worst-case type $\hat{\theta}_m$ *such that for every $\theta \in m$,*

$$u_S(a^*(\theta), \theta) \geq u_S(a^*(\hat{\theta}_m), \theta).$$

Assumption 2 asserts that each message has a "worst-case type" whose complete-information outcome is worse than that of any other type that could disclose that message. We borrow this assumption from Seidmann and Winter (1997) and Hagenbach, Koessler, and Perez-Richet (2014). We use Assumption 2 in our equilibrium constructions to deter off-path messages by stipulating that for any such message, the receiver's beliefs concentrate on the worst-case type for that message. These skeptical beliefs, in combination with Assumption 1, deter off-path messages.

To see what Assumption 2 entails, let us show why it holds in monopoly pricing (Example 1). For every message m, the worst-case type is $\hat{\theta}_m = \max_{\theta \in m} \theta$, i.e., the highest type that could have sent message m.[12] Therefore, the receiver's optimal price $a^*(\hat{\theta}_m) \geq a^*(\theta)$ for any $\theta \in m$, which implies that the type-θ buyer would never profit from imitating type $\hat{\theta}_m$.

Our final assumption is a form of continuity. Let $U_R(G) := \max_{a \in A} \int u_R(a, \theta) \, \mathrm{d}G(\theta)$ be the receiver's expected payoff when he chooses his optimal action for a belief G. Of his best responses to belief G, let $\underline{a}(G)$ and $\overline{a}(G)$ be ones that minimize and maximize the sender's payoff: $\underline{a}(G) \in \arg\min_{a \in a^*(G)} \int u_S(a, \theta) \, \mathrm{d}G(\theta)$ and $\overline{a}(G) \in \arg\max_{a \in a^*(G)} \int u_S(a, \theta) \, \mathrm{d}G(\theta)$.

Assumption 3. *The following hold:*

(a) The receiver's payoff from choosing an optimal action, $U_R(G)$, is continuous and the set of optimal actions, $a^*(G)$, is upper hemicontinuous.

(b) For every belief G and strictly positive ε and δ, there are

- *a belief $\underline{H}$ such that the Radon-Nikodym derivative $\frac{\mathrm{d}\underline{H}}{\mathrm{d}G} \leq 1 + \varepsilon$, and any best response to $\underline{H}$ is in $B_\delta(\underline{a}(G))$; and*
- *a belief $\overline{H}$ such that the Radon-Nikodym derivative $\frac{\mathrm{d}\overline{H}}{\mathrm{d}G} \leq 1 + \varepsilon$, and any best response to $\overline{H}$ is in $B_\delta(\overline{a}(G))$.*

[12] Recall that every message is closed.

Moreover, the functions that send G to $\underline{H}$ and $\overline{H}$, respectively, are measurable.

Assumption 3 imposes a form of continuity on the payoffs of both the sender and receiver. To elaborate on (a), let $u_R(a, G) := \int u_R(a, \theta)\, \mathrm{d}G(\theta)$ be the receiver's expected payoff when his belief is G. If $u_R(a, G)$ is continuous in (a, G), part (a) follows from Berge's maximum theorem. However, $u_R(a, G)$ is *not* continuous in (a, G) in any of our applications. Consequently, we invoke maximum theorems with weaker assumptions to establish (a). Part (b) specifies that for every belief G, there is a "nearby" belief $\underline{H}$ such that every best response to $\underline{H}$ is close to the best response to G that minimizes the sender's payoff; analogously, there is a nearby belief $\overline{H}$ such that every best response is close to the best response to belief G that maximizes the sender's payoff.[13] In this sense, Assumption 3 allows us to perturb the receiver's beliefs without significantly changing his optimal actions.

3.3 Main Result

We characterize the entire set of equilibrium payoffs, identifying it using the information design benchmark. Our main result shows that design and disclosure result in virtually the same payoffs.

Theorem 1. *Suppose Assumptions 1, 2, and 3 are satisfied. For every achievable payoff profile (u_S^*, u_R^*) and every $\varepsilon > 0$, there is an equilibrium of the disclosure game whose payoffs are within ε of (u_S^*, u_R^*).*

Theorem 1 reveals that the disclosure game that we study supports a rich set of equilibrium outcomes, including full revelation, partial revelation of many different sorts, as well as complete concealment. These possibilities contrast with the classical unraveling force of Grossman (1981) and Milgrom (1981) in which the sender fully reveals her type in every equilibrium. In their setting, any putative equilibrium with partial revelation breaks because in every pool of types, some type strictly prefers its complete-information payoff to the incomplete-information payoff. Our analysis identifies how a starkly different conclusion emerges in settings that meet our assumptions. We view the important difference to stem from Assumption 1, namely that full revelation is not an attractive deviation relative to any incomplete-information payoff.

We offer some additional remarks about Theorem 1. An ancillary implication is that the sender does not accrue significant gains from committing to a Blackwell experiment, let alone a strategy of the disclosure game (modulo issues of equilibrium selection). The equilibria that we use to support achievable payoff profiles satisfy the natural analogue of the intuitive

[13] A sufficient condition is that for every $a \in a^*(G)$, there is a belief H such that $\frac{\mathrm{d}H}{\mathrm{d}G}$ is bounded and a is uniquely optimal.

criterion (Cho and Kreps, 1987). However, as we discuss in Section 4.1, the truth-leaning refinement of Hart, Kremer, and Perry (2017) may have some bite; in monopoly pricing, the refinement assures that all equilibrium payoff profiles are efficient.

We sketch the proof in Section 3.4. The key obstacle is that in an equilibrium of the disclosure game, the sender does not mix between different messages unless she is indifferent. By contrast, a Blackwell experiment can split a sender type across multiple segments. Lemma 1 clarifies that this is indeed the major obstacle: if a segmentation is instead "partitional"—in that only a measure-0 set of types are in more than one segment—we show that then it can be supported as an equilibrium of the disclosure game. Our second step evaluates the degree to which the restriction to partitional segmentations comes at a cost: Lemma 2 shows that for every achievable payoff profile, there exists a nearby payoff profile that is achieved by a finite partitional segmentation.[14] Although our proof uses the fact that F admits a density, Section 3.5 shows that an approximate version holds if F has a finite support so long as each type has sufficiently low probability. In Section 3.6, we clarify how the payoff equivalence of disclosure and design may not hold when our assumptions are violated.

3.4 Proof Sketch

The sender's equilibrium messaging strategy induces a segmentation: each message m defines a segment corresponding to the receiver's belief following that message. Say that a segmentation σ is *finite* if its support is a finite set. A segmentation σ is *partitional* if for every $G, H \in \operatorname{supp}(\sigma)$ with $G \neq H$, $\operatorname{supp}(G) \cap \operatorname{supp}(H)$ has F-measure zero.

Lemma 1. *If Assumptions 1 and 2 are satisfied, then every payoff profile achieved by a finite partitional segmentation can be supported as an equilibrium.*

Proof sketch. Fix a finite partitional segmentation σ and a best response $a : \operatorname{supp}\sigma \to A$ for the receiver. In the equilibrium we construct, the set of on-path messages is $\{\operatorname{supp} G\}_{G \in \operatorname{supp}\sigma}$. For any on-path message $\operatorname{supp} G$, the receiver updates her belief to G and plays $a(G)$. For any off-path message m, the receiver's belief is a point mass at the "worst case type" $\hat{\theta}_m$, which exists by Assumption 2, and the receiver plays $a^*(\hat{\theta}_m)$. For any type contained in the support of only one segment in $\operatorname{supp}\sigma$, there is only one on-path message available to her and the sender sends this message. For any type θ in the support of two or more segments in σ,

[14] We note that this payoff profile is nearby in an ex ante sense but may differ in its ex interim payoffs. Lemma 2 is distinct from the results of Zeng (2023) and Arieli, Babichenko, Smorodinsky, and Yamashita (2023), which provide conditions under which any achievable payoff can be achieved by a deterministic segmentation. Their conditions are not satisfied in our model (nor in any of our applications). Moreover, deterministic segmentations are not necessarily partitional.

the sender chooses the on-path message available to θ that results in the best possible action given $a(\cdot)$.

We establish that the above constitutes an equilibrium. Observe that the receiver's beliefs are consistent with Bayes' rule.[15] The receiver plays a best response after every message by construction. Also by construction no type of the sender has a profitable deviation to an on-path message, and Assumption 1 and Assumption 2 together imply that no type has a profitable deviation to an off-path message. ■

Lemma 2. *If Assumption 3 is satisfied, then for every achievable payoff profile (u_S^*, u_R^*) and every $\varepsilon > 0$, there is a finite partitional segmentation that achieves payoffs within ε of (u_S^*, u_R^*).*

Proof sketch. Fix an achievable payoff profile (u_S^*, u_R^*), a segmentation σ and best response achieving this payoff profile, and $\varepsilon > 0$. To illustrate the argument here, we assume that the receiver plays $\overline{a}(G)$ for all $G \in \operatorname{supp}\sigma$. Figure 1 illustrates the three steps to how we approximate σ.

FIGURE 1. The approximation process for Lemma 2.

Because optimal actions may not be lower hemicontinuous in the belief, even small perturbations of the belief can induce significant changes in the receiver's best response and therefore significantly affect the sender's payoff. We sidestep this issue by first perturbing any segment G such that in the perturbed segment, all optimal actions are close to $\overline{a}(G)$: Assumption 3(b) assures that for every $G \in \operatorname{supp}\sigma$, there is a nearby segment $\overline{H}$ such that any best response to $\overline{H}$ is arbitrarily close to $\overline{a}(G)$. We obtain a new segmentation σ_1 by replacing each $G \in \operatorname{supp}\sigma$ with its corresponding $\overline{H}$.[16] By choosing $\overline{H}$ sufficiently close to G, the receiver's payoff under σ_1 is within $\varepsilon/3$ of u_R^* because by Assumption 3(a), the receiver's payoff from choosing an optimal action is continuous in the belief. The sender's payoff under σ_1 is also within $\varepsilon/3$ of u_S^* because any best response to $\overline{H}$ is arbitrarily close to $\overline{a}(G)$ and because the sender's payoff is continuous in the action for each type.

The second step converts σ_1 to a finite segmentation σ_2 by "merging" segments in σ_1 that are sufficiently close to each other into a single segment, which is the average of the aforementioned segments. Because $\Delta(\Theta)$ is compact, the number of resulting "average segments" can

[15] Since σ is finite partitional, the set of types that are contained in multiple segments have F-measure zero. Therefore, the behavior of such types does not affect the receiver's beliefs.

[16] To maintain Bayes' plausibility, we reduce the probability of each segment slightly and create one additional segment.

be chosen to be finite. Because the receiver's best-response correspondence is upper hemi-continuous, any best response to an "average segment" in σ_2 is sufficiently close to any best response to any segment in σ_2 that is merged into it. Consequently, the payoffs under σ_2 are within $\varepsilon/3$ of those under σ_1.

Our last step identifies a finite partitional segmentation σ_3 that achieves payoffs within $\varepsilon/3$ of those under σ_2. Loosely speaking, we partition the type space into sufficiently small cubes, and approximate each of the finitely many segments in σ_2 using a collection of such cubes, which is possible because F is absolutely continuous. The statement on payoffs then follows from Assumption 3(a). Therefore, the payoffs under σ_3 are within ε of (u_S^*, u_R^*). ■

3.5 Finite Types

The proof of Theorem 1 uses the fact that the prior is atomless. The example below shows that our conclusion might not hold if, for instance, types are binary.

Example 5. Consider an adaptation of Example 1 in which the buyer's type θ is drawn from the binary set $\{\underline{\theta}, \overline{\theta}\}$ in which $\underline{\theta} > 0$ and suppose that the optimal uniform price, $\overline{p} = \overline{\theta}$. The buyer-optimal segmentation would feature two segments, one comprising $\overline{\theta}$ alone and the other featuring a pool of both types such that the monopolist prices at $\underline{\theta}$. Although this setting satisfies Assumptions 1 to 3 (shown in Section 4.1), the payoff profile from this segmentation cannot be reached in an equilibrium of the disclosure game. The issue is that, in equilibrium, type $\overline{\theta}$ mixes between the messages $\{\overline{\theta}\}$ and $\{\underline{\theta}, \overline{\theta}\}$ only if the two messages result in the same price. Consequently, the buyer's equilibrium payoff must be 0.

Example 5 shows that there can be a wedge between design and disclosure if there are types that have sufficiently high mass. Below, we prove that this is the primary issue: for any finitely supported F, so long as the mass of each type is sufficiently low, all achievable payoffs can be approximately supported through an equilibrium of the game.

Theorem 2. *Suppose Assumptions 1, 2, and 3 are satisfied. For every $\varepsilon > 0$, there is $\gamma > 0$ such that if F has finite support with $F(\{\theta\}) \leq \gamma$ for every type θ, then for every achievable payoff (u_S^*, u_R^*), there is an equilibrium of the disclosure game whose payoffs are within ε of (u_S^*, u_R^*).*

Theorem 2 offers a finite analogue of our main result. The key step shows that once types have sufficiently low mass, any finite segmentation can be approximated (in the appropriate sense) by one that is partitional.

3.6 What if the Key Assumptions Fail?

Herein, we show that our central conclusion does not hold if any assumption is dropped (while maintaining the other two assumptions).

Example 6 (Assumption 1 fails.)**.** Consider the following stylized version of Grossman's and Milgrom's model. The sender's type is uniformly distributed on $\Theta = [0, 1]$ and the receiver chooses an action a in $\mathbb{R}$. The sender's motives are transparent in that her payoffs $u_S(a, \theta)$ do not vary with θ but are strictly increasing and strictly concave in a. The receiver would like to match the action with the sender's type and his payoff is $u_R(a, \theta) = -(a - \theta)^2$. Assumption 1 fails in this game: the complete information payoff for the sender of type $\theta = 1$ is higher than her payoff if the receiver's beliefs equal the prior.

Here, the unique equilibrium outcome coincides with full revelation. However, given strict concavity, the best achievable payoff for the sender comes from the receiver obtaining no information and choosing action $1/2$.

Example 7 (Assumption 2 fails.)**.** Consider the game in which $A = \{1, 2\}$, and the sender's type θ is uniformly distributed on $\Theta = [0, 1]$. The receiver's payoffs are $u_R(1, \theta) = 1 - \theta$ and $u_R(2, \theta) = \theta$. The sender's payoffs are $u_S(1, \theta) = \mathbf{1}_{\theta \geq 1/2}$ and $u_S(2, \theta) = 1/2$. Assumption 2 fails in this game. Observe that the complete-information payoff is 0 for types strictly below $1/2$ and $1/2$ for types above $1/2$. We argue that the message $m = \Theta$ lacks a worst-case type. If the putative "worst-case" type $\hat{\theta}_m$ were assigned to be strictly below $1/2$, then types above $1/2$ accrue more than their complete-information payoff; if $\hat{\theta}_m$ were assigned to be above $1/2$, then types strictly below $1/2$ do better than their complete-information payoff.

This failure has implications for the equilibrium outcomes of the disclosure game. For instance, no equilibrium supports a payoff profile near that of the fully revealing experiment, $(1/4, 3/4)$. Were there such an equilibrium, action 2 would have to be played with high probability whenever the type is strictly above $1/2$ and action 1 would have to be played with high probability whenever the type is strictly below $1/2$. However, following the message $m = \Theta$, either action 1 is played with probability at least $1/2$ and types above $1/2$ could profitably deviate to this message, or action 2 is played with probability at least $1/2$, in which case types below $1/2$ could profitably deviate.

Example 8 (Assumption 3 fails.)**.** Suppose there are n types, $\{\theta_1, \ldots, \theta_n\}$ and $n + 1$ actions, $A = \{0, 1, \ldots, n\}$.[17] The sender's payoff is $u_S(a, \theta) = \mathbf{1}_{a=0}$ and the receiver's payoff is $u_R(a, \theta) = \mathbf{1}_{a=0} + n\mathbf{1}_{\theta=\theta_a}$. Assumption 3(b) fails here: every action is optimal under a uniform belief but whenever the belief is not uniform, action 0 is not optimal.

[17] For simplicity, we assume that the prior F is supported on a finite set, but this example can be extended to a continuum setting.

Given this failure, there are achievable payoff profiles that cannot be (approximately) supported in the disclosure game. Consider a prior that is a convex combination of a uniform belief, with weight α, and a point mass at θ_1, with weight $(1-\alpha)$. A segmentation with two segments, one of which is uniform, achieves a payoff of α to the sender. We argue that in every equilibrium, however, the sender's payoff is 0. To see why, observe that following any message $m \neq \Theta$, the receiver would not choose action 0. Thus, the only prospect for a strictly positive payoff for the sender is if the receiver played action 0 with positive probability following the message $m = \Theta$. That cannot happen in an equilibrium: were the receiver to do so, every sender-type would send this message with probability 1, which would make action 0 a sub-optimal choice for the receiver.

4 Applications

4.1 Monopoly Pricing

This section applies our results to monopoly pricing, elaborating on Example 1. A monopolist (he) sells a product to a single consumer (she), who demands a single unit. The consumer's valuation θ is drawn according to an absolutely continuous CDF F with support on $\Theta = [\underline{\theta}, \overline{\theta}]$ where $\underline{\theta} \geq 0$. The monopolist's reservation value is 0.

We augment this standard game with a disclosure stage. The timing is as follows. First, the consumer observes θ and sends a message $m \in \mathcal{M}(\theta)$ to the monopolist. The monopolist then sets a price $a \in [0, \overline{\theta}]$. A type-$\theta$ consumer's payoff is $u_S(a, \theta) = \max\{\theta - a, 0\}$, and the monopolist's payoff is $u_R(a, \theta) = a\mathbf{1}_{a \leq \theta}$. In other words, the sale happens if and only if the price a is lower than the consumer's type. It can be readily seen that $u_S(\cdot, \theta)$ is continuous and $u_R(\cdot, \theta)$ is upper semicontinuous for every type θ.

Let $\overline{u}_R$ denote the monopolist's payoff (or profit) from charging the optimal uniform price. Bergemann, Brooks, and Morris (2015) show that a payoff profile is achievable so long as (i) the consumer's payoff is nonnegative, (ii) the monopolist's payoff is no less than $\overline{u}_R$, and (iii) the total payoff is no more than the maximal aggregate surplus $\overline{w}$. This set of payoff profiles defines the "BBM triangle." We show that the BBM triangle also characterizes the set of payoffs that can be approximately supported in the disclosure game defined above.

Proposition 1. *For every payoff profile (u_S^*, u_R^*) with $u_S^* \geq 0$, $u_R^* \geq \overline{u}_R$, and $u_S^* + u_R^* \leq \overline{w}$, and every $\varepsilon > 0$, there is an equilibrium of the disclosure game that supports payoffs within ε of (u_S^*, u_R^*).*

Our argument establishes that this setting satisfies Assumptions 1 to 3, and consequently, Proposition 1 then follows from Theorem 1. Verifying the first two assumptions is straightfor-

ward. Establishing the third assumption, however, is more involved; here, we build on Yang (2023)'s analysis of continuity in the monopoly problem.[18]

Proposition 1 speaks to a challenge identified by Bergemann and Morris (2019): viewing information design as a metaphor to capture the set of achievable payoffs, they caution readers from taking a more literal interpretation.

> However, giving a literal information design interpretation of point C [the consumer-optimal segmentation] is more subtle. We would need to identify an information designer who knew consumers' valuations and committed to give partial information to the monopolist in order to maximize the sum of consumers' welfare. Importantly, even though the disclosure rule is optimal for consumers as a group, individual consumers would not have an incentive to truthfully report their valuations to the information designer.... (Bergemann and Morris, 2019, p. 67).

Against this backdrop, Proposition 1 shows that all payoff profiles in the BBM triangle can be (approximately) supported with hard information flowing directly from the consumer to the seller, without requiring an intermediary to know the consumer's value.[19]

We turn to the question of whether refinements can deliver stronger predictions for the set of supportable payoffs. We consider the "truth-leaning" refinement proposed by Hart, Kremer, and Perry (2017). Drawing on the Twainian adage, "When in doubt, tell the truth..." they study limit equilibria of perturbed games in which the sender obtains an infinitesimal bump in her utility if she shares the entire truth. While their focus is on the receiver's value for commitment, their concept turns out to have powerful implications for monopoly pricing: it selects payoff profiles on the efficiency frontier.

Formally, for a function $\boldsymbol{\varepsilon} : \Theta \to \mathbb{R}_{>0}$, consider the perturbed game $\Gamma^{\boldsymbol{\varepsilon}}$ in which the consumer's payoff increases by $\boldsymbol{\varepsilon}(\theta)$ when the type is θ and she sends message $\{\theta\}$. An equilibrium $((\rho, \tau), \mu)$ of the original game is *truth-leaning* if there exist (i) a sequence of functions $\boldsymbol{\varepsilon}^n$ that converges uniformly to $\mathbf{0}$, where $\mathbf{0}$ is a constant function that maps every θ to 0, and (ii) a sequence $((\rho^n, \tau^n), \mu^n)$ that converges uniformly to $((\rho, \tau), \mu)$ such that for each $n \in \mathbb{N}$, $((\rho^n, \tau^n), \mu^n)$ is an equilibrium of the perturbed game $\Gamma^{\boldsymbol{\varepsilon}^n}$.[20] Proposition 2 characterizes truth-leaning equilibria of this game.

[18] Although we focus on Bergemann, Brooks, and Morris (2015)'s canonical private-values formulation, our results also apply to lemon's problems (e.g., Kartik and Zhong, 2024). Consider, for instance, a procurement setting in which the sender is a seller who privately observes her quality type θ that influences both her cost and the value that the receiver, a procurer, obtains from buying the product. Given that Theorem 1 applies to this setting, disclosures about quality then can virtually support any achievable payoff profile.

[19] Crucial to this potential microfoundation is that the consumer can choose any closed set that contains her type, which points to an intermediary's role in creating this evidence. Were communication cheap talk, every equilibrium would result in a babbling outcome in which the monopolist sets his optimal uniform price.

[20] Hart, Kremer, and Perry (2017) also require that in any perturbed game, every type of the sender fully reveal her type with positive probability. However, this requirement has no bite in our setting.

Proposition 2. *The following hold:*

(a) The payoff profile of every truth-leaning equilibrium is efficient: (u_S^*, u_R^*) is supported by a truth-leaning equilibrium only if $u_S^* + u_R^* = \overline{w}$.

(b) For every efficient payoff profile, there is a nearby payoff profile supported by a truth-leaning equilibrium: for every (u_S^*, u_R^*) with $u_S^* \geq 0$, $u_R^* \geq \overline{u}_R$, and $u_S^* + u_R^* = \overline{w}$, and every $\varepsilon > 0$, there is a truth-leaning equilibrium of the disclosure game that supports payoffs within ε of (u_S^*, u_R^*).

Here is the logic. The only scope for inefficiency in monopoly pricing is that in some market segment, the monopolist's price exceeds some consumer types in that segment. Such behavior cannot emerge in an equilibrium of the perturbed game because the consumer types would then be better off revealing the whole truth to accrue the infinitesimal bump. Therefore, every truth-leaning equilibrium is efficient. A more subtle intuition underlies why all efficient payoff profiles can be approximately supported by a truth-leaning equilibrium. We show that for every finite partitional equilibrium that (exactly) supports payoff $(\tilde{u}_S, \tilde{u}_R)$, there is a truth-leaning equilibrium that (exactly) supports payoffs $(\tilde{u}_S, \overline{w} - \tilde{u}_S)$. It then follows from Proposition 1 that any efficient payoff profile can be approximated with one that is supported by a truth-leaning equilibrium.[21]

4.2 Bargaining Over Policies

Building on Example 2, this section applies our analysis to models of policy negotiations with incomplete information.[22] A policy $a \in \mathbb{R}$ is jointly chosen by the proposer and the vetoer. The proposer's payoff from policy a, $u(a)$, is strictly increasing in a. The vetoer's payoff, $v(a, \theta)$, is strictly single-peaked in a with a unique maximizer θ and symmetric around the maximizer; we call the vetoer's ideal policy θ her *type*. Her type is her private information, and is drawn according to an absolutely continuous CDF F on $\Theta = [\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} < \infty$. For simplicity, we assume $\underline{\theta} \geq 0$.

In veto bargaining, once the proposer proposes a policy a, the vetoer can accept or reject. If she accepts, the proposed policy prevails; if she rejects, then the status-quo policy $a_Q = 0$ is preserved. Given the proposer's payoffs, she never proposes any $a < 0$. Restricting attention to $a \geq 0$, the vetoer accepts if and only if $a \leq 2\theta$.

[21] Equivalently, Proposition 2(b) identifies that there is a dense set of payoff profiles on the efficiency frontier of the BBM triangle that can be supported by (truth-leaning) equilibria of the disclosure game.

[22] The literature has studied various formulations of veto bargaining with incomplete information; see, for example, Matthews (1989), McCarty (1997), Kartik, Kleiner, and Van Weelden (2021), Ali, Kartik, and Kleiner (2023), and Kim, Kim, and Van Weelden (2024).

We augment this game with a disclosure stage, with the following timing. The vetoer first observes her type θ and sends a message $m \in \mathcal{M}(\theta)$ to the proposer. The proposer then proposes a policy a. The type-θ vetoer's payoff is given by $u_S(a, \theta) = \max\{v(a, \theta), v(0, \theta)\}$, and the proposer's payoff is given by $u_R(a, \theta) = (u(a) - u(0))\mathbf{1}_{a \leq 2\theta}$. For every $\theta \in \Theta$, $u_S(\cdot, \theta)$ is continuous, and $u_R(\cdot, \theta)$ is upper semicontinuous.

In this setting, the set of achievable payoff profiles is yet to be characterized. However, our main result bears the following implication.

Proposition 3. *For every achievable payoff profile (u_S^*, u_R^*) and every $\varepsilon > 0$, there is an equilibrium of the disclosure game that supports payoffs within ε of (u_S^*, u_R^*).*

The method of proof, like Proposition 1, shows that our three key assumptions are satisfied. We note however an important distinction. This setting does not feature transferable utility and, unlike monopoly pricing, the sender and receiver may have aligned preferences that favor a higher action to a_Q. Nevertheless, the complete-information payoff is the lowest for a sender-type θ because the receiver would then propose action 2θ, which results in the same payoff as the status quo.

4.3 Insurance Markets

Our final application models disclosure in an insurance market. We consider the standard setup of an insuree purchasing an insurance contract from a single insurer, as modeled in Stiglitz (1977) and Chade and Schlee (2012). The insuree has initial wealth $w > 0$, faces a potential loss $\ell \in (0, w)$ with probability $\theta \in (0, 1)$, and has risk preferences represented by a strictly increasing, continuously differentiable, and strictly concave (Bernoulli) utility function $v : \mathbb{R}_{\geq 0} \to \mathbb{R}$. The probability of loss, θ, is the insuree's *type*, and is her private information. The insuree's outside option at every stage is to purchase zero insurance.

The insurer is risk neutral and has beliefs about the insuree's type given by the absolutely continuous CDF F with density f and support $\Theta := [\underline{\theta}, \overline{\theta}] \subseteq (0, 1)$. Without loss, the insurer chooses a menu of contracts $(x(\theta), t(\theta)) \in \mathbb{R}^2$ for each type θ comprising a premium $t(\theta)$ and an indemnity payment $x(\theta)$ in the event of a loss, subject to incentive and participation constraints. The expected profit from a contract (x, t) chosen by a type-θ insuree is $t - \theta x$.

We append a disclosure stage to this problem. After observing her type θ, the insuree sends a message $m \in \mathcal{M}(\theta)$ to the insurer. The insurer then offers a menu of contracts and the insuree selects one of the contracts or chooses no insurance.

Formulated this way, it is difficult to verify Assumption 3 directly. Therefore, we reformulate this game such that the insurer directly chooses the expected utility of every type of the insuree instead of offering a menu of contracts (see Chade and Schlee, 2012). Observe

that any incentive-compatible menu of contracts $(x(\theta), t(\theta))$ can be reformulated as a menu $(D(\theta), a(\theta))$, where

$$\begin{aligned} D(\theta) &:= v(w - t(\theta)) - v(w - \ell + x(\theta) - t(\theta)), \\ a(\theta) &:= v(w - t(\theta)) - \theta D(\theta). \end{aligned}$$

If a type-θ insuree accepts the menu, $v(w - t(\theta))$ is her utility when no loss occurs, $D(\theta)$ is the difference in utility between no loss and suffering a loss, and $a(\theta)$ is her *indirect utility*.

By standard arguments, incentive compatibility implies that a is convex; in this case, $a'(\theta) = -D(\theta)$ almost everywhere. Given the insurer's belief G, in any optimal menu the participation constraint binds for the lowest type, and $0 \leq D(\theta) \leq D_0 := v(w) - v(w - \ell)$ for G-almost every θ (Chade and Schlee, 2012, Theorem 1). The latter conditions correspond to the indemnity payment in the event of a loss being at most the loss (no overinsurance) and the utility reduction in the event of a loss being at most the utility reduction when there is no insurance. Denoting the space of continuous real-valued functions on Θ equipped with the sup-norm by $\mathscr{C}(\Theta)$, it is therefore without loss to restrict attention to the following set of indirect utilities:

$$A := \left\{ a \in \mathscr{C}(\Theta) : \begin{array}{l} a(\underline{\theta}) = \underline{\theta} v(w - \ell) + (1 - \underline{\theta}) v(w), \\ a \text{ is decreasing, convex and } D_0\text{−Lipschitz} \end{array} \right\}. \tag{1}$$

The insurer's expected profit from a menu (D, a) chosen by a type-θ insuree is

$$u(D, a, \theta) := w - \theta\ell - (1 - \theta) v^{-1}(\underbrace{a(\theta) + \theta D(\theta)}_{\text{no loss}}) - \theta v^{-1}(\underbrace{a(\theta) - (1 - \theta) D(\theta)}_{\text{loss}}),$$

where $w - \theta\ell$ is type-θ insuree's total wealth in expectation, and the remaining terms are the insurer's expected cost of providing the utilities promised to the insuree. To write the insurer's payoff as a function of a alone, we proceed with the following step. Since $a'(\theta) = -D(\theta)$ almost everywhere, the insurer's payoff is determined by a almost everywhere. At any θ where a is not differentiable, because a is convex and D_0-Lipschitz, the insurer chooses $D(\theta) \in \partial a(\theta) \cap [0, D_0]$, where $\partial a(\theta)$ is the subdifferential of a at θ. Therefore, we can interpret the indirect utility a as the insurer's action: the insuree's payoff from action $a \in A$ is $u_S(a, \theta) = a(\theta)$, and the insurer's payoff from action a is

$$u_R(a, \theta) = \max_{D \in \partial a(\theta) \cap [0, D_0]} u(D, a, \theta). \tag{2}$$

Because $\partial a(\theta)$ is closed for every θ and $u(D, a, \theta)$ is continuous, $u_R(a, \theta)$ is well-defined. In

the appendix, we verify that A is compact, $u_S(a,\theta)$ is continuous in a for each θ, and $u_R(a,\theta)$ is upper semicontinuous in (a,θ).

We characterize the equilibrium payoff set of this game.

Proposition 4. *For every achievable payoff profile (u_S^*, u_R^*) and every $\varepsilon > 0$, there is an equilibrium of the disclosure game that supports payoffs within ε of (u_S^*, u_R^*).*

Proposition 4 speaks to the debate on whether insurance companies ought to be allowed to condition their contracts based on disclosures (e.g., genetic tests). If the insuree can share partial disclosures—statements like "My test is one of these results"—then allowing disclosure could potentially lead to significant efficiency gains or even approximate the insuree-optimal information structure. Yet, the insuree could also be trapped in an equilibrium in which she has to fully disclose all evidence; then, she might be better off if disclosure were prohibited. In light of this equilibrium multiplicity, our results emphasize the role that various parties—government agencies, intermediaries, and insurance companies themselves—can play in coordinating behavior towards the public interest.

5 Conclusion

This paper studies general disclosure games in which our primary departure is that the sender does not favor the complete-information outcome to those that keep the receiver in the dark. This departure is motivated by principal-agent settings in which if the principal learns the agent's type, he would set transfers and allocations that make the agent's IR constraints bind. Combining this assumption with standard worst-case type and continuity assumptions leads to an equivalence result: the set of equilibrium payoffs in the disclosure game is virtually identical to those that obtain through information design.

This conclusion speaks to whether voluntary disclosure necessarily traps the sender in a commitment problem. Theorem 1 suggests that the sender does not benefit from commitment. At the same time, she may be worse off in some equilibria relative to the benchmark in which she cannot disclose evidence. Thus, our results highlight challenges of predicting the equilibrium effects of giving one party hard information that she can disclose to the other. One may also view our results as offering a potential microfoundation for information design: information can flow directly from the sender to the receiver in a standard disclosure game, without resorting to a metaphorical information designer or intermediary that knows the sender's type.

References

Ali, S. Nageeb, Nima Haghpanah, Xiao Lin, and Ron Siegel. 2022. "How to sell hard information." *The Quarterly Journal of Economics* 137 (1):619–678.

Ali, S. Nageeb, Navin Kartik, and Andreas Kleiner. 2023. "Sequential veto bargaining with incomplete information." *Econometrica* 91 (4):1527–1562.

Ali, S. Nageeb, Greg Lewis, and Shoshana Vasserman. 2023. "Voluntary Disclosure and Personalized Pricing." *The Review of Economic Studies* 90 (2):538–571.

Aliprantis, Charalambos D and Kim C Border. 2006. *Infinite dimensional analysis: A hitchhiker's guide*. Berlin: Springer, third ed.

Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita. 2023. "Optimal persuasion via bi-pooling." *Theoretical Economics* 18 (1):15–36.

Asseyer, Andreas and Ran Weksler. 2024. "Certification Design with Common Values." *Econometrica* 92 (3):651–686.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2019. "Mechanisms with evidence: Commitment and robustness." *Econometrica* 87 (2):529–566.

———. 2023. "Mechanism design for acquisition of/stochastic evidence." Working Paper.

———. 2024. "Sequential Mechanisms for Evidence Acquisition." Working Paper.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris. 2015. "The limits of price discrimination." *American Economic Review* 105 (3):921–957.

Bergemann, Dirk and Stephen Morris. 2019. "Information design: A unified perspective." *Journal of Economic Literature* 57 (1):44–95.

Billingsley, Patrick. 1995. *Probability and Measure*. John Wiley & Sons, third edition ed.

Bull, Jesse and Joel Watson. 2004. "Evidence disclosure and verifiability." *Journal of Economic Theory* 118 (1):1–31.

Cameron, Charles and Nolan McCarty. 2004. "Models of vetoes and veto bargaining." *Annual Review of Political Science* 7:409–435.

Chade, Hector and Edward Schlee. 2012. "Optimal insurance with adverse selection." *Theoretical Economics* 7 (3):571–607.

Cho, In-Koo and David M Kreps. 1987. "Signaling games and stable equilibria." *The Quarterly Journal of Economics* 102 (2):179–221.

Dasgupta, Sulagna. 2023. "Communication via hard and soft information." Working Paper.

DeMarzo, Peter M, Ilan Kremer, and Andrzej Skrzypacz. 2019. "Test design and minimum standards." *American Economic Review* 109 (6):2173–2207.

Dye, Ronald A. 1985. "Disclosure of nonproprietary information." *Journal of Accounting Research* :123–145.

Erwin, Cheryl. 2008. "Legal update: living with the Genetic Information Nondiscrimination Act." *Genetics in Medicine* 10 (12):869–873.

Glode, Vincent, Christian C Opp, and Xingtan Zhang. 2018. "Voluntary disclosure in bilateral transactions." *Journal of Economic Theory* 175:652–688.

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

Grossman, Sanford J and Oliver D Hart. 1980. "Disclosure laws and takeover bids." *The Journal of Finance* 35 (2):323–334.

Hagenbach, Jeanne and Frédéric Koessler. 2017. "Simple versus rich language in disclosure games." *Review of Economic Design* 21 (3):163–175.

Hagenbach, Jeanne, Frédéric Koessler, and Eduardo Perez-Richet. 2014. "Certifiable pre-play communication: Full disclosure." *Econometrica* 82 (3):1093–1131.

Hart, Sergiu, Ilan Kremer, and Motty Perry. 2017. "Evidence games: Truth and commitment." *American Economic Review* 107 (3):690–713.

Hiriart-Urruty, Jean-Baptiste and Claude Lemaréchal. 2004. *Fundamentals of convex analysis*. Springer Science & Business Media.

Jin, Ginger Zhe, Michael Luca, and Daniel Martin. 2021. "Is no news (perceived as) bad news? An experimental investigation of information disclosure." *American Economic Journal: Microeconomics* 13 (2):141–173.

Jovanovic, Boyan. 1982. "Truthful disclosure of information." *Bell Journal of Economics* 13:36–44.

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review* 101 (6):2590–2615.

Kartik, Navin, Andreas Kleiner, and Richard Van Weelden. 2021. "Delegation in veto bargaining." *American Economic Review* 111 (12):4046–4087.

Kartik, Navin and Weijie Zhong. 2024. "Lemonade from lemons: Information design and adverse selection." Working Paper.

Kim, Jenny S., Kyungmin Kim, and Richard Van Weelden. 2024. "Persuasion in veto bargaining." *American Journal of Political Science* Forthcoming.

Lipman, Barton L and Duane J Seppi. 1995. "Robust inference in communication games with partial provability." *Journal of Economic Theory* 66 (2):370–405.

Madarasz, Kristof and Marek Pycia. 2023. "Information Choice: Cost over Content." Working Paper.

Mathis, Jérôme. 2008. "Full revelation of information in sender–receiver games of persuasion." *Journal of Economic Theory* 143 (1):571–584.

Matthews, Steven A. 1989. "Veto threats: Rhetoric in a bargaining game." *The Quarterly Journal of Economics* 104 (2):347–369.

McCarty, Nolan M. 1997. "Presidential reputation and the veto." *Economics & Politics* 9 (1):1–26.

Milgrom, Paul and John Roberts. 1986. "Relying on the Information of Interested Parties." *The RAND Journal of Economics* 17 (1):18–32.

Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics* 12 (2):380–391.

Okuno-Fujiwara, Masahiro, Andrew Postlewaite, and Kotaro Suzumura. 1990. "Strategic information revelation." *The Review of Economic Studies* 57 (1):25–47.

Pollrich, Martin and Roland Strausz. 2024. "The Irrelevance of Fee Structures for Certification." *American Economic Review: Insights* 6 (1):55–72.

Pram, Kym. 2021. "Disclosure, Welfare and Adverse Selection." *Journal of Economic Theory* 197:105327.

———. 2023. "Learning and evidence in insurance markets." *International Economic Review* 64 (4):1685–1714.

Romer, Thomas and Howard Rosenthal. 1978. "Political resource allocation, controlled agendas, and the status quo." *Public Choice* :27–43.

Seidmann, Daniel J and Eyal Winter. 1997. "Strategic Information Transmission with Verifiable Messages." *Econometrica* :163–169.

Shishkin, Denis. 2024. "Evidence Acquisition and Voluntary Disclosure." Working Paper.

Stiglitz, Joseph E. 1977. "Monopoly, non-linear pricing and imperfect information: the insurance market." *The Review of Economic Studies* 44 (3):407–430.

Titova, Maria and Kun Zhang. 2024. "Persuasion with Verifiable Information." Working Paper.

Verrecchia, Robert E. 1983. "Discretionary disclosure." *Journal of Accounting and Economics* 5:179–194.

Yang, Kai Hao. 2023. "On the continuity of outcomes in a monopoly market." *Journal of Mathematical Economics* 108:102887.

Zeng, Yishu. 2023. "Derandomization of persuasion mechanisms." *Journal of Economic Theory* 212:105690.

A Proof of Theorem 1

A.1 Proof of Lemma 1

Fix any finite segmentation σ with pairwise disjoint support, let its payoff pair be (u_S^*, u_R^*), and let $a_\sigma : \operatorname{supp}\sigma \to A$ denote the receiver's best responses that yield payoffs (u_S^*, u_R^*). Note that because F has full support, it must be that $\bigcup_{G\in\operatorname{supp}(\sigma)} \operatorname{supp} G = \Theta$, as otherwise the segments cannot average back to F.

Now consider the following messaging strategy of the sender: If $\theta \in \operatorname{supp}(G)$ and $\theta \notin \operatorname{supp}(H)$ for any $H \in \operatorname{supp}(\sigma)$ such that $H \neq G$, the type-θ sender sends message $\operatorname{supp}(G)$ with probability 1. If $\theta \in \operatorname{supp}(G)$ for multiple $G \in \operatorname{supp}(\sigma)$, type θ sends message $\operatorname{supp} H$ with probability 1, where $H \in \arg\max_{\{G\in\operatorname{supp}\sigma\,:\,\theta\in\operatorname{supp} G\}} u_S(a_\sigma(G), \theta)$. Because σ is finite and has pairwise disjoint supports, such types are contained in a F-null set and hence do not affect expected payoffs.

The receiver's belief system is such that whenever she observes message $\operatorname{supp}(G)$ for some $G \in \operatorname{supp}(\sigma)$, she updates using Bayes rule and her new belief is G; following any other message m, her belief is a point mass at the worst-case type $\hat{\theta}_m$, as defined in Assumption 2. We specify the receiver's strategy as follows: if she observes message $\operatorname{supp}(G)$ for any $G \in \operatorname{supp}(\sigma)$, she chooses action $a_\sigma(G)$ (which is optimal for the receiver given belief G); for any other message m, she chooses action $a^*(\hat{\theta}_m)$.

By construction, the receiver is choosing a best response given her beliefs after any message and beliefs satisfy Bayes' rule whenever possible. Moreover, the sender never wants to deviate from her messaging strategy: by messaging according to the strategy described above, his payoff is at least $u_S(a^*(\theta), \theta)$ by Assumption 1. If he deviates to an off-path message m, his payoff is $u_S(a^*(\hat{\theta}_m), \theta)$, which is lower by Assumption 2. By construction, those types that can send multiple on-path messages choose optimally among feasible on-path messages. Hence, these strategies and beliefs form an equilibrium and this equilibrium induces the segmentation σ and payoffs (u_S^*, u_R^*). ■

A.2 A Preliminary Step for Lemma 2

Lemma 3. *Let σ be a finite segmentation with $\operatorname{supp}(\sigma) = \{F_1, \ldots, F_N\}$. There is a sequence of finite partitional segmentations $\{\sigma^m\}_{m\in\mathbb{N}}$ with $\operatorname{supp}(\sigma^m) = \{F_1^m, \ldots, F_N^m\}$ such that, for $i = 1, \ldots, N$, $\sigma^m(F_i^m) \to \sigma(F_i)$ and $F_i^m \to F_i$.*

Proof. Without loss of generality, suppose $\Theta = [0,1]^n$. For any $m, k \in \mathbb{N}$, partition Θ into at least m equal-sized cubes; denote this partition by $\mathcal{P}^m$. Partition each $P \in \mathcal{P}^m$ further into at least k equal-sized cubes, and denote the collection of all such small cubes by $\mathcal{Q}$. Choose

an assignment of all smaller cubes to $\{1,\dots,N\}$, denoted by $\ell : \mathcal{Q} \to \{1,\dots,N\}$, to minimize

$$\max_{i\in\{1,\dots,N\}} \sum_{P\in\mathcal{P}^m} \left| F\left(\bigcup_{Q\in\mathcal{Q},Q\subseteq P,\ell(Q)=i} Q\right) - \sigma(F_i)F_i(P)\right|. \tag{3}$$

Note that this expression goes to zero as $k \to \infty$. Therefore, for each m, there exists $k(m) \in \mathbb{N}$ such that we can partition each cube into $k(m)$ smaller cubes such that (3) is at most $1/m$. Denote this partition by $\mathcal{Q}^m$.

Define a finite partitional segmentation by setting, for all measurable $B \subseteq \Theta$,

$$F_i^m(B) := \frac{F(B \cap \bigcup_{Q\in\mathcal{Q}^m,\ell(Q)=i} Q)}{F(\bigcup_{Q\in\mathcal{Q}^m,\ell(Q)=i} Q)}$$

and

$$\sigma^m(F_i^m) := F\left(\bigcup_{Q\in\mathcal{Q}^m,\ell(Q)=i} Q\right).$$

It follows that $\sigma^m(F_i^m) \to \sigma(F_i)$ and $\sum_{P\in\mathcal{P}^m} |F_i^m(P) - F_i(P)| \to 0$ as $m \to \infty$.[23] Therefore, for any Lipschitz-continuous function $h : \Theta \to \mathbb{R}$ and $i \in \{1,\dots,N\}$, $\int h \,\mathrm{d}F_i^m \to \int h \,\mathrm{d}F_i$. Hence, $F_i^m \to F_i$. ■

A.3 Proof of Lemma 2

Fix arbitrary $\varepsilon > 0$ and an arbitrary segmentation σ achieving payoffs (u_S^*, u_R^*).

Step 1: We can assume that the receiver plays $\overline{a}(G)$ for each $G \in \operatorname{supp}\sigma$.

Indeed, the arguments are symmetric if the receiver plays $\underline{a}(G)$ for each $G \in \operatorname{supp}\sigma$, and the conclusion then follows for arbitrary best responses because the sender's payoff is a convex combination of the payoffs achieved from always playing $\overline{a}$ and always playing $\underline{a}$.

Step 2: We choose ε' and δ small enough and define a measurable mapping that maps any segment G to a closeby segment $\overline{H}$ such that all best responses in segment $\overline{H}$ are within δ of $\overline{a}(G)$.

Fix $\varepsilon' > 0$ and $\delta > 0$ small enough. Let $t : \Delta(\Theta) \to \Delta(\Theta)$ be a measurable function that sends any segment G to $t(G)$, where $\frac{\mathrm{d}t(G)}{\mathrm{d}G} \leq 1 + \varepsilon'$ and any best response given segment $t(G)$ is in $B_\delta(\overline{a}(G))$. Assumption 3(b) ensures that such a function exists.

[23] Indeed, $\sum_{P\in\mathcal{P}^m} |F_i^m(P) - F_i(P)| = \frac{1}{\sigma(F_i)} \sum_{\mathcal{P}} |\frac{\sigma(F_i)}{F(\bigcup_{Q\in\mathcal{Q}^m,\ell(Q)=i} Q)} F(P \cap \bigcup_{Q\in\mathcal{Q}^m,\ell(Q)=i} Q) - \sigma(F_i)F_i(P)| \to 0$ by definition of ℓ (cf. (3)) and because $\frac{\sigma(F_i)}{F(\bigcup_{Q\in\mathcal{Q}^m,\ell(Q)=i} Q)} \to 1$.

Step 3: We define σ_1 by considering a (scaled-down) version of the pushforward measure of σ under t and adding a mass point at an extra segment containing the remaining types.

For any (measurable) $Z \subseteq \Delta(\Theta)$, define $\tilde{\sigma}(Z) := \frac{1}{1+\varepsilon'}\sigma(t^{-1}(Z))$ to be a (scaled-down) pushforward of σ under t.

For any (measurable) $E \subseteq \Theta$, $(t(G))(E) = \int_E \frac{\mathrm{d}t(G)}{\mathrm{d}G}\,\mathrm{d}G \leq (1+\varepsilon')G(E)$ by the bound on the Radon-Nikodym derivative. Therefore,

$$\left(\int H\,\mathrm{d}\tilde{\sigma}(H)\right)(E) = \left(\frac{1}{1+\varepsilon'}\int t(G)\,\mathrm{d}\sigma(G)\right)(E) \leq \frac{1}{1+\varepsilon'}\int (1+\varepsilon')G(E)\,\mathrm{d}\sigma(G) = F(E)$$

where the first equality follows from the definition of $\tilde{\sigma}$ and the last equality follows since $\int G\,\mathrm{d}\sigma(G) = F$. Intuitively, $\tilde{\sigma}$ does not "exhaust" all types available under the prior (and isn't even a probability distribution over segments). Therefore, we add a segment G_{extra} that contains all the remaining types: Let $G_{extra} := \frac{F - \int H\,\mathrm{d}\tilde{\sigma}(H)}{1-\tilde{\sigma}(\Delta(\Theta))}$ and note that $\left(F - \int H\,\mathrm{d}\tilde{\sigma}(H)\right)(\Theta) = 1 - \frac{1}{1+\varepsilon'} = 1 - \tilde{\sigma}(\Delta(\Theta))$. Hence, $G_{extra} \in \Delta(\Theta)$. Now define $\sigma_1 := \tilde{\sigma} + \delta_{G_{extra}}(1 - \tilde{\sigma}(\Delta(\Theta)))$, where $\delta_{G_{extra}}$ is a Dirac measure at G_{extra}. Simple accounting shows that $\sigma_1 \in \Delta(\Delta(\Theta))$ and $\int G\,\mathrm{d}\sigma_1(G) = F$.

Step 4: We argue that the payoffs under σ_1 are within $\varepsilon/3$ of (u_S^*, u_R^*).

A fraction $\frac{1}{1+\varepsilon'}$ of types end up in a segment under σ_1 in which any best response for the receiver is within δ of the best response the receiver would have chosen under segmentation σ. Since the sender's payoff is continuous in the action for each type, we can choose $\varepsilon' > 0$ and $\delta > 0$ in Step 2 small enough such that the expected payoff for the sender is within $\varepsilon/3$ of u_S^*. Similar arguments apply to the receiver's payoff: We can choose $\varepsilon' > 0$ small enough such that, for any segment G, G and $t(G)$ are close in the Levy-Prokhorov metric (which metricizes the weak*-topology), and hence the resulting payoffs for the receiver are close by Assumption 3(a). Moreover, by choosing $\varepsilon' > 0$ small enough, the extra segment G_{extra} gets arbitrarily small probability under σ_1, and hence the receiver's payoff is within $\varepsilon/3$ of u_R^*.

Step 5: We define a new segmentation σ_2, which is an approximation of the segmentation σ_1, such that σ_2 contains finitely many segments and such that payoffs under σ_2 are within $\varepsilon/3$ of the payoffs under σ_1.

Fix $\delta_2 > 0$. For any $G \in \operatorname{supp}\tilde{\sigma}$, there is an ε_G-ball (in the Levy-Prokhorov metric) around G such that any best response to any $G' \in B_{\varepsilon_G}(G)$ is within δ_2 of any best response given G (by Assumption 3(a)). Since $\operatorname{supp}\tilde{\sigma}$ is compact, finitely many of these balls, say $\{B_1, ..., B_m\}$, cover $\operatorname{supp}\tilde{\sigma}$, where $B_i := B_{\varepsilon_{G_i}}(G_i)$. We form a new segmentation σ_2 by merging all segments that lie in a given ball. Formally, σ_2 has at most $m+1$ segments in its support; the first

segment is the barycenter of the set B_1 and, for $i > 1$, the ith segment is defined recursively as the barycenter of the set $B_i \setminus \bigcup_{j=1}^{i-1} B_j$ under $\tilde{\sigma}$,

$$H_i := \frac{\int_{B_i \setminus \bigcup_{j=1}^{i-1} B_j} G \, \mathrm{d}\tilde{\sigma}(G)}{\tilde{\sigma}(B_i \setminus \bigcup_{j=1}^{i-1} B_j)}$$

whenever $\tilde{\sigma}(B_i \setminus \bigcup_{j=1}^{i-1} B_j) > 0$ and we define $\sigma_2(\{H_i\}) := \tilde{\sigma}(B_i \setminus \bigcup_{j=1}^{i-1} B_j)$ and $\sigma_2(\{G_{extra}\}) := \sigma_1(\{G_{extra}\})$.[24] One can verify that open balls in the Levy-Prokhorov metric are convex. Therefore, $H_i \in B_i$ and any best response to the merged segment H_i is within $2\delta_2$ of any best response to any $G \in B_i$. By choosing δ_2 small enough, we obtain a segmentation with finite support such that payoffs under σ_2 are within $\varepsilon/3$ of the payoffs under σ_1.

Step 6: We define a new segmentation σ_3 which is finite partitional and approximates σ_2. We argue that the payoffs under σ_3 are within $\varepsilon/3$ of the payoffs under σ_2.

By Lemma 3, for any $\delta' > 0$ we can approximate the segmentation σ_2 by a finite partitional segmentation σ_3 such that to any segment G in σ_2 there is a unique corresponding segment H with $|\sigma_3(H) - \sigma_2(G)| < \delta'$ and $d_P(G, H) < \delta'$, where d_P denotes the Levy-Prokhorov metric. By Assumption 3(a) we can choose δ' small enough so that the receiver's payoff under segmentation σ_3 is within $\varepsilon/3$ of the payoff under segmentation σ_2. Similarly, by choosing δ' small enough, any optimal action given belief H is close to any optimal action under G and hence the sender's expected payoff given segmentation σ_3 is within $\varepsilon/3$ of the expected payoff given segmentation σ_2.

It follows that payoffs under σ_3 are within ε of (u_S^*, u_R^*). ■

Remark 1. *The proof of Lemma 2 still goes through if we replace Assumption 3(b) by the following alternative assumption:*

> *For any $G \in \operatorname{supp} \sigma$ and any $\varepsilon, \delta > 0$, there is a distribution $\underline{H}$ ($\overline{H}$) whose Radon-Nikodym derivative satisfies $\frac{\mathrm{d}\underline{H}}{\mathrm{d}G} \leq 1 + \varepsilon$ ($\frac{\mathrm{d}\overline{H}}{\mathrm{d}G} \leq 1 + \varepsilon$) and any best response a' to $\underline{H}$ ($\overline{H}$) is such that $|u_S(a', \theta) - u_S(\underline{a}(G), \theta)| < \delta$ ($|u_S(a', \theta) - u_S(\overline{a}(G), \theta)| < \delta$) for $\underline{H}(\overline{H})$-almost every θ. Moreover, the functions that send G to $\overline{H}$ and $\underline{H}$ are measurable.*

Since $\operatorname{supp} \underline{H} \subseteq \operatorname{supp} G$*, any action optimal under $\underline{H}$ yields almost the same payoff for types in* $\operatorname{supp} \underline{H}$*. Thus, by choosing sufficiently small $\varepsilon', \delta > 0$ in Step 2, Step 4 still goes through.*

[24] For simplicity, we assume $H_i \neq H_j$ for $i \neq j$ and $H_i \neq G_{extra}$. Our arguments apply without this assumption.

B Proof of Theorem 2

We first prove the following preliminary lemma. Without loss of generality, we assume that Θ is unidimensional.

Lemma 4. *Let σ be a finite segmentation with $\operatorname{supp}(\sigma) = \{F_1, \dots, F_N\}$. For every $\delta > 0$ there exists a $\gamma > 0$ such that if F has finite support with $F(\{\theta\}) < \gamma$ for all θ, then there is a finite partitional segmentation $\bar{\sigma}$ with $\operatorname{supp}(\bar{\sigma}) = \{\bar{F}_1, \dots, \bar{F}_N\}$ such that for each $i = 1, \dots, N$, $|\bar{\sigma}(\bar{F}_i) - \sigma(F_i)| < \delta$ and $d_P(\bar{F}_i, F_i) < \delta$, where d_P denotes the Lévy-Prokhorov metric.*

Proof. Fix $\delta > 0$, define $Q := \min\{\sigma(F_1), \dots, \sigma(F_N)\}$, and $\gamma := Q\delta/(N+1)$. Suppose F has finite support and $F(\{\theta\}) \leq \gamma$ for all $\theta \in \operatorname{supp} F$. Let $J := \operatorname{supp} F = \{\theta_1, \dots, \theta_M\}$, where for any $s, t \in \{1, \dots, M\}$ with $s < t$, $\theta_s < \theta_t$. Consider an assignment rule $\ell : J \to \{1, \dots, N\}$ defined as follows: $\ell(\theta_1) = 1$, and for every $s > 1$, define iteratively[25]

$$\ell(\theta_s) := \min\{i \in \{1, \dots, N\} : F(X_i^{s-1}) \leq \sigma(F_i) F_i(\theta_{s-1})\},$$

where $X_i^{s-1} = \{\theta \in \{\theta_1, \dots, \theta_{s-1}\} : \ell(\theta) = i\}$. For every $i = 1, \dots, N$, define $J_i := \{\theta \in J : \ell(\theta) = i\}$; $\{J_1, \dots, J_N\}$ is a partition of J. Define a finite partitional segmentation $\bar{\sigma}$ with $\operatorname{supp}(\bar{\sigma}) = \{\bar{F}_1, \dots, \bar{F}_N\}$ by setting, for every $B \subseteq \Theta$, $\bar{F}_i(B) := \frac{F(B \cap J_i)}{F(J_i)}$, and $\bar{\sigma}(\bar{F}_i) := F(J_i)$ for each $i = 1, \dots, N$.

Since $F(\{\theta\}) \leq \gamma$ for all $\theta \in J$, we obtain

$$\bar{\sigma}(\bar{F}_i)\bar{F}_i(\theta) \leq \sigma(F_i)F_i(\theta) + \gamma \tag{4}$$

for all $\theta \in \Theta$ and $i = 1, \dots, N$. Because both σ and $\bar{\sigma}$ are segmentations, $\sum_i \bar{\sigma}(\bar{F}_i)\bar{F}_i(\theta) = F(\theta) = \sum_i \sigma(F_i)F_i(\theta)$, and hence

$$\sum_{i=1}^{N} \sigma(F_i)F_i(\theta) = \bar{\sigma}(\bar{F}_k)\bar{F}_k(\theta) + \sum_{i \neq k} \bar{\sigma}(\bar{F}_i)\bar{F}_i(\theta) \leq \bar{\sigma}(\bar{F}_k)\bar{F}_k(\theta) + \sum_{i \neq k} (\sigma(F_i)F_i(\theta) + \gamma)$$

for all $\theta \in J$ and $k = 1, \dots, N$, where the inequality follows from (4). Consequently,

$$\bar{\sigma}(\bar{F}_i)\bar{F}_i(\theta) \geq \sigma(F_i)F_i(\theta) - (N-1)\gamma. \tag{5}$$

Taking $\theta = \theta_M$, inequalities (4) and (5) together imply $\bar{\sigma}(\bar{F}_i) - \sigma(F_i) \leq \gamma$ and $\sigma(F_i) - \bar{\sigma}(\bar{F}_i) \leq (N-1)\gamma$ for each $i = 1, \dots, N$, respectively. Furthermore, divide by $\sigma(F_i)$ on both

[25] To ease exposition, we slightly abuse notation by letting $G(\theta) := G([\underline{\theta}, \theta])$ for all $\theta \in \Theta$ for any probability measure G with support on J, where $\underline{\theta} := \inf \Theta$.

sides of (4), and rearrange, we get

$$\bar{F}_i(\theta) - F_i(\theta) \leq \frac{\gamma}{\sigma(F_i)} + \frac{\sigma(F_i) - \bar{\sigma}(\bar{F}_i)}{\sigma(F_i)}\bar{F}_i \leq N\gamma/Q$$

for all $\theta \in \Theta$. Similarly, dividing by $\sigma(F_i)$ on both sides of (5), we obtain $F_i(\theta) - \bar{F}_i(\theta) \leq N\gamma/Q$ for all $\theta \in \Theta$. Thus, $|\bar{\sigma}(\bar{F}_i) - \sigma(F_i)| \leq (N-1)\gamma < \delta$, and $\sup_{\theta\in\Theta} |\bar{F}_i(\theta) - F_i(\theta)| < \delta$ by definition of γ. By definition of the Lévy-Prokhorov metric, $d_P(\bar{F}_i, F_i) < \delta$. ■

Proof of Theorem 2. We prove the analogues of Lemma 1 and Lemma 2. Under Assumptions 1 to 3, the proof of Lemma 1 and the first five steps in the proof of Lemma 2 go through without any assumption on the prior F. Fixing $\varepsilon > 0$, it suffices to find $\gamma > 0$ such that if F has finite support with $F(\{\theta\}) \leq \gamma$, then for every finite segmentation σ, there exists a finite partitional segmentation $\bar{\sigma}$ such that the payoffs under $\bar{\sigma}$ are within $\varepsilon/3$ of those under σ.

By Lemma 4, for every $\delta > 0$ we can find a $\gamma > 0$ such that as long as F has finite support with $F(\{\theta\}) \leq \gamma$, there is a finite partitional segmentation $\bar{\sigma}$ such that to any segment G in σ there is a unique corresponding segment H with $|\bar{\sigma}(H) - \sigma(G)| < \delta$ and $d_P(G, H) < \delta$. By Assumption 3(a) we can choose δ small enough so that the receiver's payoff under segmentation $\bar{\sigma}$ is within $\varepsilon/3$ of the payoff under segmentation σ. Similarly, by choosing δ small enough, any optimal action given belief H is close to any optimal action under G and hence the sender's expected payoff given segmentation $\bar{\sigma}$ is within $\varepsilon/3$ of the expected payoff given segmentation σ. Therefore, the payoffs under $\bar{\sigma}$ are within $\varepsilon/3$ of the payoffs under σ. ■

C Proofs for Section 4

C.1 Proof of Proposition 1

Lemma 5. *Let $w : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}$ be a continuous and strictly increasing function. If A is an interval of real numbers and the receiver's payoff is given by $u_R(a, \theta) = w(a)\mathbf{1}_{a\leq\theta}$, then Assumption 3 is satisfied.*

Proof. We first verify Assumption 3(a). Recall that $u_R(a, G) = \int u_R(a, \theta)\, \mathrm{d}G(\theta)$. Let $U_R(G) := \max_{a\in A} u_R(a, G)$ be the receiver's optimal payoff in segment G. The proof of Theorem 1 in Yang (2023) can be used *mutatis mutandis* to show that the receiver's optimal payoff $U_R(G)$ is continuous under the weak* topology, and the best response correspondence $\arg\max_{a\in A} u_R(a, G)$ is upper hemicontinuous. Thus, Assumption 3(a) is satisfied.

Verifying Assumption 3(b) requires more work. Let $\underline{a} : \Delta(\Theta) \to \mathbb{R}$ and $\overline{a} : \Delta(\Theta) \to \mathbb{R}$ denote the mappings that map a segment to the lowest optimal action and the highest optimal action in that segment, respectively. Because the best response correspondence is upper

hemicontinuous, $\underline{a}$ is lower semicontinuous, and $\overline{a}$ is upper semicontinuous. Given $\varepsilon, \delta > 0$, we construct two functions, $\underline{t} : \Delta(\Theta) \to \Delta(\Theta)$ and $\overline{t} : \Delta(\Theta) \to \Delta(\Theta)$, such that both $\underline{t}$ and $\overline{t}$ are measurable, the Radon-Nikodym derivatives satisfy $\frac{\mathrm{d}\underline{t}(G)}{\mathrm{d}G} \leq 1 + \varepsilon$ and $\frac{\mathrm{d}\overline{t}(G)}{\mathrm{d}G} \leq 1 + \varepsilon$ for every $G \in \Delta(\Theta)$, and any optimal action in segment $\underline{t}(G)$ ($\overline{t}(G)$) is contained in $(\underline{a}(G) - \delta, \underline{a}(G) + \delta)$ ($(\overline{a}(G) - \delta, \overline{a}(G) + \delta)$, respectively). For notational ease, we sometimes suppress the dependence of $\underline{a}$ and $\overline{a}$ on G and simply write $\underline{a}$ and $\overline{a}$ for $\underline{a}(G)$ and $\overline{a}(G)$, respectively.

Define the function $\underline{t} : \Delta(\Theta) \to \Delta(\Theta)$ by

$$\underline{t}(G)(x) := \frac{\min\{G(x), 1 - \kappa\eta(G)\}}{1 - \kappa\eta(G)}$$

where $\eta(G)$ solves

$$\frac{\eta(G)}{1 - \eta(G)} 2\overline{\theta} = \max_{a \in [\underline{a}(G) - \delta/2, \underline{a}(G)]} w(a)[1 - G(a)] - \max_{a \leq \underline{a}(G) - \delta} w(a)[1 - G(a)], \tag{6}$$

and $\kappa \in (0, 1)$ is small enough so that $\frac{dt(G)}{dG} \leq 1 + \varepsilon$. Note that $\eta(G)$ is well-defined and takes values in $(0, 1)$ because the right-hand side is strictly positive and bounded.

We now establish that $\underline{t}$ is measurable. First, $\underline{a}$ is measurable because it is lower semicontinuous. To see that η is measurable, reformulate the first maximization problem on the RHS of (6) as choosing (a, q) with $(a, q) \in \Gamma(G)$, where

$$\Gamma(G) := \{(a, q) \in [\underline{a}(G) - \delta/2, \underline{a}(G)] \times [0, 1] : 1 - G(a) \leq q \leq 1 - G(a-)\}$$

to maximize $w(a)q$. The objective function is continuous in a and q, Γ has nonempty compact values, and by Lemma 1 in Yang (2023), Γ is continuous and hence weakly measurable. Reformulate the second maximization problem on the RHS of (6) analogously. Then we can apply the measurable maximum theorem (Theorem 18.19 in Aliprantis and Border, 2006) to conclude that the RHS of (6) is measurable in G. This implies that η is measurable and, therefore, $\underline{t}$ is measurable.

Finally, we verify that any optimal action in segment $\underline{t}(G)$ is within δ of $\underline{a}(G)$:

$$\begin{aligned}
\max_{a \in [\underline{a}(G) - \delta/2, \underline{a}(G)]} w(a)[1 - \underline{t}(G)(a)] &\geq \max_{a \in [\underline{a}(G) - \delta/2, \underline{a}(G)]} w(a) \left[1 - \frac{G(a)}{1 - \kappa\eta(G)}\right] \\
&= \max_{a \in [\underline{a}(G) - \delta/2, \underline{a}(G)]} w(a) \left[1 - G(a) - \frac{\kappa\eta(G)}{1 - \kappa\eta(G)} G(a)\right] \\
&\geq \max_{a \in [\underline{a}(G) - \delta/2, \underline{a}(G)]} w(a)[1 - G(a)] - \overline{\theta} \frac{\kappa\eta(G)}{1 - \kappa\eta(G)} \\
&> \max_{a \leq \underline{a}(G) - \delta} w(a)[1 - G(a)] \geq \max_{a \leq \underline{a}(G) - \delta} w(a)[1 - \underline{t}(G)(a)],
\end{aligned}$$

where the strict inequality follows from (6). Hence, any optimal action in segment $\underline{t}(G)$ is at least $\underline{a}(G) - \delta$. Moreover, for any $a > \underline{a}(G)$, we have

$$w\left(\underline{a}\right)\left[1 - \frac{G(\underline{a}(G))}{1-\kappa\eta(G)}\right] \geq w(a)[1-G(a)] - \frac{\kappa\eta(G)}{1-\kappa\eta(G)} w\left(\underline{a}\right) G(\underline{a}) > w(a)\left[1 - \frac{G(a)}{1-\kappa\eta(G)}\right].$$

Hence, no action strictly above $\underline{a}(G)$ is optimal in segment $\underline{t}(G)$. Thus, the function $\underline{t}$ has the desired properties.

Define the function $\overline{t} : \Delta(\Theta) \to \Delta(\Theta)$ by

$$\overline{t}(G)(x) := \begin{cases} \frac{(1-\kappa)G(x)}{1-\kappa G(\overline{a})} & \text{if } x < \overline{a}(G), \\ \frac{G(x)-\kappa G(\overline{a})}{1-\kappa G(\overline{a})} & \text{if } x \geq \overline{a}(G), \end{cases}$$

where $\kappa > 0$ is small enough such that $\frac{\mathrm{d}\overline{t}(G)}{\mathrm{d}G} < 1 + \varepsilon$. Because $\overline{a}$ is upper semicontinuous, it is measurable; then by writing $\overline{t}$ as

$$\overline{t}(G)(x) = \frac{(1-\kappa)G(x)}{1-\kappa G(\overline{a})}\mathbf{1}_{x<\overline{a}(G)} + \frac{G(x)-\kappa G(\overline{a})}{1-\kappa G(\overline{a})}\mathbf{1}_{x\geq\overline{a}(G)},$$

it is straightforward to see that $\overline{t}$ is also measurable.

Next, we show that for every segment G, $\overline{a}(G)$ is the unique optimal action in segment $\overline{t}(G)$. For any $a' > \overline{a}(G)$,

$$\begin{aligned} w\left(\overline{a}\right)\left(1 - \overline{t}(G)(\overline{a})\right) &= w\left(\overline{a}\right)\left[1 - \frac{G(\overline{a}) - \kappa G(\overline{a})}{1-\kappa G(\overline{a})}\right] = \frac{w\left(\overline{a}\right)\left(1-G(\overline{a})\right)}{1-\kappa G(\overline{a})} \\ &> \frac{w\left(a'\right)\left(1-G(a')\right)}{1-\kappa G(\overline{a})} = w\left(a'\right)\left(1-\overline{t}(G)(a')\right), \end{aligned}$$

where the strict inequality follows because $a' > \overline{a}$, w is strictly increasing, and $\overline{a}$ is the *highest* optimal action. For any $a'' < \overline{a}(G)$,

$$\begin{aligned} w\left(\overline{a}\right)\left(1 - \overline{t}(G)(\overline{a})\right) &= w\left(\overline{a}\right)\frac{1-G(\overline{a})}{1-\kappa G(\overline{a})} \\ &\geq \frac{(1-\kappa)w\left(a''\right)\left(1-G(a'')\right) + \kappa w\left(\overline{a}\right)\left(1-G(\overline{a})\right)}{1-\kappa G(\overline{a})} \\ &> \frac{(1-\kappa)w\left(a''\right)\left(1-G(a'')\right) + \kappa w\left(a''\right)\left(1-G(\overline{a})\right)}{1-\kappa G(\overline{a})} \\ &= w\left(a''\right)\left[1 - \frac{(1-\kappa)G(a'')}{1-\kappa G(\overline{a})}\right] = w\left(a''\right)\left(1-\overline{t}(G)(a'')\right), \end{aligned}$$

where the weak inequality holds because $\overline{a}$ is an optimal action, and the strict inequality

follows since $a'' < \overline{a}(G)$ and w is strictly increasing. Hence, $\bar{t}$ also has the desired properties, and Assumption 3(b) is verified. ■

Proof of Proposition 1. It suffices to verify Assumptions 1, 2, and 3. Because $a^*(\theta) = \theta$, $u_S(a^*(\theta), \theta) = 0$. Since u_S is bounded below by zero, Assumption 1 is satisfied. For Assumption 2, we claim that for every message $m \in \mathcal{C}$, one can set $\hat{\theta}_m := \max_{\theta \in m} \theta$. This is because for every $\theta \in m$, $u_S(a^*(\theta), \theta) = 0 = \max\{\theta - \hat{\theta}_m, 0\} = u_S\left(a^*\left(\hat{\theta}_m\right), \theta\right)$. Finally, Assumption 3 holds since we can appeal to Lemma 5 by letting w to be the identity function. ■

C.2 Proof of Proposition 2

To establish (a), it suffices to show that trade must happen with probability 1 in any truth-leaning equilibrium. In any equilibrium of the perturbed game $\Gamma^{\boldsymbol{\varepsilon}}$, trade must happen with probability 1: any type θ that does not trade must fully reveal her type to get payoff $\boldsymbol{\varepsilon}(\theta) > 0$, and the unique optimal action for the monopolist after receiving message $\{\theta\}$ with $\theta > 0$ is $a = \theta$, resulting in trade. By definition, a truth-leaning equilibrium is a limit point of equilibria of $\Gamma^{\boldsymbol{\varepsilon}}$, and hence trade also happens with probability 1 in any such equilibrium.

To establish (b), fix $\varepsilon > 0$ and an efficient payoff profile (u_S^*, u_R^*). By Proposition 1, there exists a payoff profile (u_S^1, u_R^1) that is within $\varepsilon/2$ of (u_S^*, u_R^*) and that is supported by an equilibrium e^*. Moreover, the proof of Theorem 1 shows that we can assume that e^* induces a finite partitional segmentation and each segment has positive probability.

We modify e^* to obtain an efficient equilibrium e^{**} of the unperturbed game that will also be an equilibrium of some perturbed games: In e^{**}, all consumer types whose payoff is strictly positive in e^* send the same message as in e^*; all types whose payoff is zero in e^* send the fully revealing messages. The monopolist's strategy is as in e^* and beliefs are derived from Bayes' rule whenever possible (with skeptical beliefs after off-path messages).

To verify that e^{**} is an equilibrium of the unperturbed game, note that each consumer type gets the same payoff in e^{**} as in e^*, any deviation to an on-path message would yield the same payoff as in e^*, and any deviation to an off-path message is not profitable. Hence, the consumer best responds. Moreover, the monopolist's actions are still optimal: Consider an on-path message m that is sent with positive probability, and denote the segment and price induced by message m in e^* by G and p_G, respectively. Under the modified strategy of the consumer in e^{**}, the segment induced by m is $G(\cdot \,|\, \theta > p_G) := (G(\theta) - G(p_G))/(1 - G(p_G))$. The monopolist's profits from charging p under $G(\cdot \,|\, \theta > p_G)$ is therefore

$$\tilde{\Pi}(p) = p(1 - G(p \,|\, \theta > p_G)) = \frac{p(1 - G(p))}{1 - G(p_G)}.$$

Since p_G is a maximizer of $p(1-G(p))$, it also maximizes $\tilde{\Pi}(p)$ among all $p \in [p_G, \overline{\theta}]$. Therefore, the monopolist best responds to m. For any other message, the monopolist holds skeptical beliefs and best responds as well.

Denote by (u_S^2, u_R^2) the payoff profile induced by e^{**}. To see that (u_S^2, u_R^2) is within ε of (u_S^*, u_R^*), note that because (u_S^*, u_R^*) and (u_S^2, u_R^2) are efficient and the efficiency frontier has slope -1,

$$|u_R^* - u_R^2| = |u_S^* - u_S^2| = |u_S^* - u_S^1| < \frac{\varepsilon}{2}.$$

It remains to argue that the equilibrium e^{**} is truth-leaning. Towards this end, for every $n \in \mathbb{N}$ and for every θ that does not send a fully revealing message in e^{**}, define $\boldsymbol{\varepsilon}^n(\theta) = (\theta - p(m))/2n$, where m is the message sent by type θ in e^{**} and $p(m)$ the resulting price, and for every type θ that sends a fully revealing message in e^{**}, define $\boldsymbol{\varepsilon}^n(\theta) = \overline{\theta}/n$. Then, for each n, $\boldsymbol{\varepsilon}^n(\theta) > 0$ for all θ and e^{**} is an equilibrium of the perturbed game $\Gamma^{\boldsymbol{\varepsilon}^n}$ since no type of the consumer has a profitable deviation, and $\boldsymbol{\varepsilon}^n$ converges uniformly to $\mathbf{0}$. Therefore, e^{**} constitutes a truth-leaning equilibrium, which completes the proof of (b). ■

C.3 Proof of Proposition 3

We verify Assumptions 1 to 3. Because $a^*(\theta) = 2\theta$ and v is symmetric around θ, $v(a^*(\theta), \theta) = v(0, \theta)$. Consequently, $u_S(a^*(\theta), \theta) = v(0, \theta)$. For every $\theta \in \Theta$, $u_S(a, \theta)$ is bounded below by $v(0, \theta)$, and hence Assumption 1 must hold. Assumption 2 also holds because for every message $m \in \mathcal{C}$, one can set $\hat{\theta}_m := \max_{\theta \in m} \theta$. Finally, by re-writing the proposer's payoff as $u_R(b, \theta) = w(b)\mathbf{1}_{b \leq \theta}$, where $b = a/2$ and $w(b) := u(2b) - u(0)$, Lemma 5 implies that Assumption 3 holds since w is continuous and strictly increasing by assumption. ■

C.4 Proof of Proposition 4

Lemma 6. *The set A defined in (1) is compact.*

Proof. Let $\{a_n\}_{n \in \mathbb{N}}$ be a sequence in A. Since a_n is uniformly bounded and D_0-Lipschitz continuous for each n, Arzela-Ascoli's theorem implies there is a subsequence, also denoted by $\{a_n\}$, that converges in the supremum-norm to some a. Clearly, $a \in A$. ■

Lemma 7. *For all θ, $u_S(a, \theta)$ is continuous in a. Also, $u_R(a, G)$ is upper semicontinuous in (a, G).*

Proof. If $a_n \to a$ then $u_S(a_n, \theta) = a_n(\theta) \to a(\theta) = u_S(a, \theta)$ and the first claim follows.

For the second claim, we show first that $u_R(a, \theta)$ is upper semicontinuous in (a, θ). Note that if $a_n \to a$, $\theta_n \to \theta$, and $\varepsilon > 0$ then $\partial a_n(\theta_n) \subseteq B_\varepsilon(\partial a(\theta))$ for all n large enough (Theorem

D.6.2.7 in Hiriart-Urruty and Lemaréchal, 2004). Hence, $(a,\theta) \mapsto \partial a(\theta)$ is upper hemicontinuous. Since $u(D,a,\theta)$ is continuous in (D,a,θ), a maximum theorem (see Lemma 17.30 in Aliprantis and Border, 2006) implies that $u_R(a,\theta)$ is upper semicontinuous in (a,θ).

Now consider $a_n \to a$ and $G_n \to G$. By Theorem 25.6 in Billingsley (1995) there are Y_n and Y on a common probability space $(\Omega, \mathcal{F}, P)$ with distributions G_n and G such that $Y_n(\omega) \to Y(\omega)$ for all ω. Then

$$\begin{aligned}&\limsup_{n\to\infty} \int u_R(a_n,\theta)\,\mathrm{d}G_n(\theta) = \limsup_{n\to\infty} \int u_R(a_n, Y_n(\omega))\,\mathrm{d}P(\omega)\\ \leq &\int u_R(a, Y(\omega))\,\mathrm{d}P(\omega) = \int u_R(a,\theta)\,\mathrm{d}G(\theta)\end{aligned}$$

where the equalities follow from a change of variables and the inequality follows from (reverse) Fatou's lemma and the fact that $u_R(a,\theta)$ is upper semicontinuous in (a,θ). We conclude that $u_R(a,G)$ is upper semicontinuous in (a,G). ■

Lemma 8. *Given $a \in A$, $G \in \Delta(\Theta)$, and $\varepsilon > 0$ there is $\tilde{a} \in A$ that is continuously differentiable such that $\|a - \tilde{a}\|_\infty < \varepsilon/2$ and $|u_R(\tilde{a},G) - u_R(a,G)| < \varepsilon/2$.*

Proof. Fix $a \in A$; let E' denote the set of θ's at which a is not differentiable. Because E' is countable, we can find a finite set $E'' = \{\theta_1, \ldots, \theta_M\} \subseteq E'$ such that $G(E' \setminus E'') < \varepsilon/(12w)$, where w is an upper bound for $|u_R(a,\theta) - u_R(b,\theta)|$ for any $a, b \in A$. Let $D^*(\theta) \in \arg\max_{D \in \partial a(\theta) \cap [0, D_0]} u(D,a,\theta)$; and for each $i = 1, \ldots, M$, let ℓ_i denote the supporting hyperplane of a at θ_i with slope $D^*(\theta_i)$.

For any $n \in \mathbb{N}$, we can find a piecewise affine function $\tilde{a}_n$ such that $\tilde{a}_n \in A$ and $\|\tilde{a}_n - a\|_\infty < 1/(9n)$. Define $\hat{a}_n := \max\{\tilde{a}_n - 1/(9n), \ell_1, \ldots, \ell_M\} + 1/(9n)$; it can be shown that $\|\hat{a}_n - \tilde{a}_n\|_\infty < 1/(9n)$, and $\hat{a}_n \in A$ for n large enough. We can approximate $\hat{a}_n$ by a differentiable function a_n such that $\|a_n - \hat{a}_n\|_\infty \leq 1/(9n)$ and $|a_n'(\theta) - D^*(\theta)| \leq 1/(9n)$ for all $\theta \in E''$.[26]

Since a_n is convex and differentiable, $a_n \to a$ implies that $a_n'(\theta) \to a'(\theta)$ for all $\theta \notin E'$. By Egoroff's theorem, there exists $E \subseteq \Theta \setminus E'$ with $G(E) < \varepsilon/(12w)$ such that $a_n' \to a$ uniformly on $\Theta \setminus (E' \cup E)$. Because $|a_n'(\theta) - D^*(\theta)| < 1/(9n)$ for all $\theta \in E''$, $|u_R(a_n,\theta) - u_R(a,\theta)| < \varepsilon/3$ for any $\theta \notin (E' \setminus E'') \cup E$ and all n large enough. Then since $G(E' \setminus E'') < \varepsilon/(12w)$ and $G(E) < \varepsilon/(12w)$, $|u_R(a_n,G) - u_R(a,G)| < \varepsilon/2$ for large enough n. ■

Lemma 9. *The receiver's optimal payoff is continuous in G and the receiver's optimal actions are upper hemicontinuous in G.*

[26] Indeed, because $\hat{a}_n$ is piecewise affine, it can be written as a positive linear combination of finite number functions taking the form of $q_c := \max\{c - \theta, 0\}$. For every function q_c, we can find a convex, decreasing, and continuously differentiable function $\tilde{q}_c^n(\theta) := \frac{1}{2}\left(\sqrt{(c-\theta)^2 + 1/(9n)} + (c-\theta)\right)$, such that $|q_c(\theta) - \tilde{q}_c^n(\theta)| \leq 1/(9n)$. Now define a_n by replacing every q_c in $\hat{a}_n$ by $\tilde{q}_c^n$. Evidently, a_n is convex, continuously differentiable, D_0-Lipschitz, and (without loss) $a_n(\underline{\theta}) = \tilde{a}_n(\underline{\theta})$, with $\|a_n - \tilde{a}_n\|_\infty \leq 1/(9n)$.

Proof. Because $u_R(a, G)$ is upper semicontinuous (Lemma 7), the optimal payoff $U_R(G) := \max_{a \in A} u_R(a, G)$ is upper semicontinuous by Lemma 17.30 in Aliprantis and Border (2006).

We argue that $U_R(G)$ is also lower semicontinuous: Consider $G_n \to G$ and suppose towards contradiction there is $\varepsilon > 0$ such that $\liminf_n U_R(G_n) + \varepsilon < U_R(G)$. By Lemma 8, there is a continuously differentiable $\tilde{a} \in A$ such that $u_R(\tilde{a}, G) + \varepsilon/2 > U_R(G)$. Then $u_R(\tilde{a}, \theta)$ is continuous in θ, and therefore $\int u_R(\tilde{a}, \theta) \, \mathrm{d}G_n(\theta) \to \int u_R(\tilde{a}, \theta) \, \mathrm{d}G(\theta)$ (since we use the weak* topology). Hence, for all n large enough, $u_R(\tilde{a}, G_n) + \varepsilon/2 \geq u_R(\tilde{a}, G)$. This implies $u_R(\tilde{a}, G_n) + \varepsilon \geq U_R(G)$, a contradiction.

Finally, we show that $a^*(G) = \arg\max_{a \in A} u_R(a, G)$ is upper hemicontinuous: Consider sequences $G_n \to G$ and $a_n \in a^*(G_n)$ such that $a_n \to a$, and suppose towards contradiction that $a \notin a^*(G)$. Because $a \in A$ this implies $U_R(G) > u_R(a, G)$. Because $u_R(a, G)$ is upper semicontinuous (Lemma 7),

$$U_R(G) > u_R(a, G) \geq \limsup_{n \to \infty} u_R(a_n, G_n) = \limsup_{n \to \infty} U_R(G_n).$$

This contradicts lower semicontinuity of U_R. ■

Proof of Proposition 4. Assumptions 1 holds because $a^*(\theta) = \theta v(w - \ell) + (1 - \theta)v(w)$, which equals the no insurance payoff. Because $a^*(\theta)$ is strictly decreasing in θ, for any $m \in \mathcal{C}$, the worst-case type is $\max_{\theta \in m} \theta$, which verifies Assumption 2. Assumption 3(a) follows from Lemma 9. Instead of verifying Assumption 3(b), we verify its alternative in Remark 1. Because $u_R(a, \theta)$ is strictly concave in a for each θ,[27] for any $G \in \Delta(\Theta)$, any $a', a'' \in a^*(G)$ satisfy $a' = a''$ G-almost everywhere. The alternative assumption holds by setting $t(G) = G$ and $\underline{H} = \overline{H} = t(G)$. ■

[27] Take $\lambda \in (0, 1)$, any $a', a'' \in A$ with $a' \neq a''$, and any $D' \in \arg\max_{D \in \partial a'(\theta) \cap [0, D_0]} u(a', D, \theta)$ and $D'' \in \arg\max_{D \in \partial a''(\theta) \cap [0, D_0]} u(a'', D, \theta)$. Then

$$\begin{aligned} \lambda u_R(a', \theta) + (1 - \lambda) u_R(a'', \theta) &< w - \theta\ell - (1 - \theta) v^{-1}((\lambda a'(\theta) + (1 - \lambda) a''(\theta)) + \theta(\lambda D' + (1 - \lambda) D'')) - \\ &\quad \theta v^{-1}((\lambda a'(\theta) + (1 - \lambda) a''(\theta)) - (1 - \theta)(\lambda D' + (1 - \lambda) D'')) \\ &\leq u_R(\lambda a' + (1 - \lambda) a'', \theta), \end{aligned}$$

where the first inequality holds because v^{-1} is strictly convex (since v is strictly concave), and the second equality follows from the fact that $\lambda D' + (1 - \lambda) D'' \in \partial(\lambda a' + (1 - \lambda) a'')(\theta) = \lambda \partial a'(\theta) + (1 - \lambda) \partial a''(\theta)$ for each θ (Theorem D.4.1.1 in Hiriart-Urruty and Lemaréchal, 2004).