Persuasion with Verifiable Information*

Maria Titova† Kun Zhang‡

March 2025

Abstract

This paper studies a game in which an informed sender with state-independent preferences uses verifiable messages to convince a receiver to choose an action from a finite set. We characterize the equilibrium outcomes of the game and compare them with commitment outcomes in information design. We provide conditions for a commitment outcome to be an equilibrium outcome and identify environments in which the sender does not benefit from commitment power. Our findings offer insights into the interchangeability of verifiability and commitment in applied settings.

*We thank the editor, the associate editor, and two anonymous referees for their excellent comments. We are indebted to Renee Bowen, Andreas Kleiner, and Joel Sobel for their guidance and support. We are grateful to Nageeb Ali, Arjada Bardhi, Simone Galperti, Germán Gieczewski, Navin Kartik, Jan Knoepfle, Frédéric Koessler, Remy Levin, Aleksandr Levkun, Elliot Lipnowski, Claudio Mezzetti, Denis Shishkin, Joel Watson, Mark Whitmeyer, and various seminar and conference audiences for their helpful comments and suggestions. Any remaining errors are our own.

†Department of Economics, Vanderbilt University. Email: motitova@gmail.com.

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

1. Introduction

Persuasion with verifiable information plays an essential role in many economic settings, including courtrooms, electoral campaigns, product advertising, financial disclosure, and job market signaling. For example, in a courtroom, a prosecutor tries to persuade a judge to convict a defendant by selectively presenting inculpatory evidence. In an electoral campaign, a politician carefully chooses which campaign promises he can credibly make to win over voters. In advertising, a firm convinces consumers to purchase its product by highlighting only specific product characteristics. In finance, a CEO discloses only certain financial statements and indicators to board members to obtain higher compensation. In a labor market, a job candidate lists specific certifications to make her application more attractive to an employer.

We consider the following model of persuasion with verifiable information. First, the sender (S, he/him) learns the state of the world. Second, the sender chooses a message, which is a verifiable statement about the state of the world, and sends it to the receiver (R, she/her). Verifiability means that any feasible message must contain the truth (the true state of the world), but not necessarily the whole truth, as it may include other states. Upon observing the message, R takes an action from a finite set. S's preferences are state-independent and strictly increasing in R's action, while R's preferences depend on both her action and the state.

Seminal papers in this literature (e.g., Grossman, 1981, Milgrom, 1981) establish an "unraveling" result, which states that S fully reveals the state in every equilibrium. In these papers, S's preferences are strictly monotone in R's action (e.g., he is maximizing quantity sold) and R's action space is rich (e.g., she is choosing a perfectly divisible quantity to buy). The argument goes as follows: the sender who is privately informed about the quality of his product always wants to separate himself from all lower-quality senders, as that convinces R to purchase a strictly higher quantity of the product. We note that if R's action space is finite, S may not fully reveal the state in every equilibrium. This is easiest to see when R's action space is binary, such as when she is choosing between buying and not buying. Then the high-quality senders may not mind pooling with some lower-quality senders, as long as R chooses to buy.

Our first result characterizes (perfect Bayesian) equilibrium outcomes, which we define as mappings from the state space to a distribution over R's actions. In Theorem 1, we show that every equilibrium outcome must be incentive-compatible (for S, IC for short) and obedient (for R). We say that an outcome is IC if S receives at least his complete information payoff in each state; otherwise, he would have a profitable deviation toward fully revealing the state. Obedience requires that if R takes an action with positive probability in some states, it must maximize her expected utility. The second part of Theorem 1 adds that if an outcome is deterministic, IC, and obedient, then it is an equilibrium outcome. A deterministic outcome is one in which R takes some action with probability one in every state. Although not all equilibrium outcomes are deterministic, we show in Lemma 1 that all equilibria in which R does not mix (e.g., if R uses a predetermined tie-breaking rule) induce deterministic outcomes.

In our model, the sender does not have commitment power: he learns the state and then chooses a verifiable message that maximizes his expected payoff in that state. Our second goal is to understand when S can achieve the same payoff in equilibrium as he does in information design (e.g., Kamenica and Gentzkow, 2011). In information design, S commits to a disclosure strategy before learning the state; a commitment outcome is an obedient outcome that maximizes the sender's ex-ante utility. Our second main result (Theorem 2) states that the commitment payoff is achievable in equilibrium if and only if there exists a deterministic and IC commitment outcome. The intuitive reason that IC but non-deterministic commitment outcomes are generally not equilibrium outcomes is that R typically breaks ties in favor of the S-preferred action in a commitment outcome. However, as we mentioned earlier, all equilibria in which R does not mix induce deterministic outcomes.

The question then becomes: when does a deterministic commitment outcome exist? Our answer is "always" if the state space is rich (Proposition 2). When the state space is finite, however, it can be that *all* commitment outcomes are non-deterministic (e.g., in the seminal example of Kamenica and Gentzkow, 2011). We show that modifying our game to one in which the set of available verifiable messages is determined stochastically allows us to implement any IC commitment outcome (see Section 5). The second difference between equilibrium and commitment outcomes is that S faces additional incentive-compatibility concerns when he does not have commitment power. Our second main result states that a (deterministic) commitment outcome is an equilibrium outcome if and only if S obtains at least his complete-information payoff in every state.

Throughout the paper, we consider a special case of the model in which R chooses between two actions, a setting commonly used in applications.[1] In that special case, an incentive-compatible commitment outcome always exists (Proposition 1). We thus show that verifiability and commitment assumptions are interchangeable when the state space is sufficiently rich (Propositions 3 and 4).

[1]See, for example, Kolotilin (2015), where pharmaceutical companies persuade the Food and Drug Administration to approve drugs; Ostrovsky and Schwarz (2010) and Boleslavsky and Cotton (2015) where schools persuade employers to hire their graduates; Alonso and Câmara (2016) and Bardhi and Guo (2018) where politicians persuade voters; Gehlbach and Sonin (2014) where governments persuade citizens.

Related Literature

The literature on verifiable disclosure (games in which the sender learns the state and then chooses a message out of a state-dependent message space) was pioneered by Grossman and Hart (1980), Grossman (1981), and Milgrom (1981); this paper uses the same mapping from states to available messages as in Milgrom and Roberts (1986), except in Section 5.[2]

A few recent papers also characterize the equilibrium set (or the set of equilibrium payoffs of the sender) and assess the value of commitment in various verifiable disclosure models. Zhang (2022) focuses on a special case of our model, further assuming that the state space is a unit interval, the receiver has monotone preferences and her optimal action only depends on the expected state; under these assumptions, the information design problem is known to have a bi-pooling solution, which always induces a deterministic commitment outcome. Zhang (2022) provides conditions under which that solution is implementable in equilibrium. Ali, Kleiner, and Zhang (2024) focus on settings where the sender favors uncertainty: his preferences are state-dependent and deviations to full revelation are never profitable. They provide conditions under which the sets of equilibrium payoffs of the sender are virtually the same in the disclosure game as in information design. Gieczewski and Titova (2024) consider a generalized disclosure game with an arbitrary message mapping and focus on coalition-proof equilibria.

Outside of verifiable disclosure models, our paper also relates to the informed information design (IID) literature, pioneered by Perez-Richet (2014), especially Koessler and Skreta (2023; KS henceforth) and Zapechelnyuk (2023; Z henceforth). In IID, the sender chooses a Blackwell experiment like in information design, except he observes the state of the world before making the choice. Therefore, an IID sender faces additional incentive-compatibility constraints relative to (uninformed) information design, much like in verifiable disclosure. The key difference between IID and disclosure games is that the sender can use stochastic evidence in IID, while his evidence in verifiable disclosure is deterministic. Thus, the differences in equilibrium sets between IID and our model highlight the value of stochastic evidence.[3] In unconstrained IID (KS), an

[2] For detailed surveys of this literature, see, for example, Milgrom (2008) and Dranove and Jin (2010).

[3] Equilibrium concepts differ across all aforementioned papers; for a direct comparison of our results to KS and Z, we will use perfect Bayesian equilibrium (PBE) with the refinement of the principle of preeminence of tests (PPT), which requires that "every out-of-equilibrium posterior belief must assign probability one to each event that is revealed as certain by the test" (Z, page 1061). PPT rules out non-IC PBE because R learns the state when S sends a fully informative experiment, and that deviation must be unprofitable. Note that PBE without refinements has no predictive power in IID, meaning that every obedient outcome is a PBE outcome (KS, page 3197).

obedient outcome is an equilibrium outcome if and only if it is incentive-compatible.[4] In IID constrained to non-degenerate experiments (Z), every obedient outcome is an equilibrium outcome. We show that in Milgrom-Roberts verifiable disclosure, an obedient outcome is an equilibrium outcome if and only if it is incentive-compatible and *deterministic* (further assuming that the receiver uses a pure strategy as in KS and Z). Thus, the sender values stochastic evidence when the state space is finite but not when it is rich. An IID problem can also be interpreted as a verifiable disclosure game with random certification (where the randomization between messages is done by a machine, not the sender).[5] We formalize this observation in Section 5 by introducing a verifiable disclosure game with a stochastic message mapping and showing that its equilibrium set is the same as in unconstrained IID. We describe the relationship between our results and KS in more detail throughout the paper.

While we study when the sender does not benefit from commitment power, a growing literature examines how much the receiver gains from having commitment power by comparing equilibrium outcomes with those of optimal mechanisms in sender-receiver games with verifiable information. When the sender's preferences are state-independent, Glazer and Rubinstein (2004, 2006) and Sher (2011) find that the receiver does not need commitment to reach the optimal mechanism outcome. Hart, Kremer, and Perry (2017) and Ben-Porath, Dekel, and Lipman (2019) provide conditions for the equivalence of the equilibrium and optimal mechanism outcomes.

Chakraborty and Harbaugh (2010), Lipnowski and Ravid (2020), and Lipnowski (2020) study cheap-talk games with state-independent preferences of the sender; the latter two compare equilibrium outcomes in one-shot cheap-talk games with commitment outcomes. In cheap-talk games, the sender's messages are not verifiable: in every state, the sender has access to the same (sufficiently rich) set of messages. The verifiability requirement faced by our sender significantly impacts the set of equilibrium outcomes.[6] Kamenica and Lin (2024) show that in generic cheap-talk games, the commitment payoff is achieved in an equilibrium if and only if there exists a deterministic commitment outcome; our Theorem 2 provides a similar result for verifiable disclosure games.

---

[4]KS focuses on interim optimal (IO) outcomes, which are PBE outcomes with a restriction on R's off-path beliefs: any such belief must assign positive probability only to states in which the sender strictly benefits from the deviation.

[5]We thank Frédéric Koessler for pointing this out.

[6]Verifiability of his messages may help or hurt the sender, depending on the preferences of the players. In fact, every equilibrium of the verifiable information game may be ex-ante better for the sender than every cheap-talk equilibrium, and vice versa.

2. Model

We study a game of persuasion with verifiable information between a sender (S, he/him) and a receiver (R, she/her). Below we describe the timing of the game along with the assumptions:[7]

1. S observes the state of the world $\theta \in \Theta$.

   The state space Θ is either finite ($\Theta = \{1, \ldots, N\}$, $N \geq 2$) or rich (Θ is a convex and compact subset of $\mathbb{R}^n$). The state of the world is drawn from a common prior $\mu_0 \in \Delta(\Theta)$ with supp $\mu_0 = \Theta$. If the state space is rich, we further assume that the prior is atomless.
2. S sends message $m \in M$ to the receiver, where M is the collection of nonempty Borel subsets of Θ. Since each message is a subset of the state space, we interpret it as a statement about the state of the world. The sender's messages are *verifiable* in the sense that every message must contain the truth: the set of messages available to S in state $\theta \in \Theta$ is $\{m \in M \mid \theta \in m\}$.[8]
3. R observes the message (but not the state) and takes an action from a finite set $J := \{1, \ldots, K\}$ with $K \geq 2$.
4. Game ends, payoffs are realized.

   S's payoff $v : J \to \mathbb{R}$ depends only on R's action. Without loss, we assume that actions are ordered such that v is increasing in $j \in J$. For ease of exposition, we further assume that v is strictly increasing.

   R's preferences are described by a bounded measurable utility function $u : J \times \Theta \to \mathbb{R}$. We define R's *complete information action-j set* as $A_j := \{\theta \in \Theta \mid u(j, \theta) \geq u(j', \theta)$ for all $j' \in J \setminus \{j\}\}$ to include all the states of the world in which she prefers to take action j under complete information.

We consider perfect Bayesian equilibria (henceforth equilibria) of this game. Firstly, S's strategy is a function $\sigma : \Theta \to \Delta_0 M$, where $\Delta_0 M$ is the set of proba-

---

[7]For a topological space Y, let $\Delta(Y)$ denote the set of Borel probability measures on Y. For $\gamma \in \Delta Y$, let supp γ denote the support of γ. We say that $\gamma \in \Delta Y$ is degenerate if supp γ is a singleton, and non-degenerate otherwise.

[8]We borrow from Milgrom and Roberts (1986) the definition of a verifiable message as a subset of the state space that includes the realized state. This method satisfies normality of evidence (Bull and Watson, 2007), which makes it consistent with both major ways of modeling hard evidence in the literature.

bility measures on M with a finite support.[9] Secondly, R's strategy is a function $\tau : M \to \Delta J$. Finally, R's belief system $q : M \to \Delta\Theta$ describes R's beliefs about the state after any observed message.

Definition 1. *A triple* (σ, τ, q) *is an* equilibrium *if*

(i) For all $\theta \in \Theta$, $\sigma(\cdot \mid \theta)$ *is supported on* $\arg\max\limits_{\{m \in M \mid \theta \in m\}} \sum\limits_{j \in J} v(j)\, \tau(j \mid m)$*;*

(ii) For all $m \in M$, $\tau(\cdot \mid m)$ *is supported on* $\arg\max\limits_{j \in J} \int\limits_{\Theta} u(j, \theta)\, \mathrm{d}q(\theta \mid m)$*;*

(iii) q *is obtained from* μ_0*, given* σ*, using Bayes rule whenever possible;*[10]

(iv) For all $m \in M$, $q(\cdot \mid m) \in \Delta m$.

In words, in equilibrium, (i) S chooses verifiable messages that maximize his expected utility in every state $\theta \in \Theta$; (ii) R maximizes her expected utility given her posterior belief; (iii) R uses Bayes' rule to update her beliefs whenever possible; and (iv) R's posteriors are consistent with disclosure on and off the path.

To analyze the model, we use the following approach. Let Ψ be the set of all Borel measurable functions from Θ to ΔJ. We refer to any $\alpha \in \Psi$ as an *outcome*; it specifies, for each state $\theta \in \Theta$, the probability $\alpha(j \mid \theta)$ that R takes action $j \in J$. Given a pair of strategies (σ, τ) of S and R, we let $M_j(\sigma, \tau) := \{m \in M \mid m \in \operatorname{supp} \sigma(\cdot \mid \theta) \text{ for some } \theta \in \Theta \text{ and } \tau(j \mid m) > 0\}$ be the set of messages that convince R to take action $j \in J$, sent with a positive probability in some state $\theta \in \Theta$. We say that $\alpha \in \Psi$ is an *equilibrium outcome* if there exists an equilibrium (σ, τ, q) that *induces* it, meaning that $\alpha(j \mid \theta) = \sum\limits_{m \in M_j(\sigma,\tau)} \tau(j \mid m)\sigma(m \mid \theta)$ for all $j \in J$ and $\theta \in \Theta$.

We say that an outcome $\alpha \in \Psi$ is *deterministic* if $\alpha(\cdot \mid \theta)$ is degenerate for each $\theta \in \Theta$. For a deterministic outcome α, we refer to the collection of sets $\{W_j\}_{j \in J}$, where $W_j := \{\theta \in \Theta \mid \alpha(j \mid \theta) = 1\}$, as the *outcome partition* (into subsets W_j of the state space in which R takes action $j \in J$ with probability one) of α.

Given an outcome α, we let $v_\alpha(\theta) := \sum\limits_{j \in J} v(j)\, \alpha(j \mid \theta)$ be S's interim (expected) payoff in state $\theta \in \Theta$ and $V_\alpha := \int\limits_{\Theta} v_\alpha(\theta)\, \mathrm{d}\mu_0(\theta)$ be S's ex-ante utility.

[9] That is, we assume that S mixes between finitely many messages. This assumption imposes no restriction when Θ is finite. When Θ is rich, it guarantees that $\sigma(\cdot \mid \theta)$ is well-defined, and the restriction does not affect the set of achievable equilibrium payoffs.

[10] That is, q is a regular conditional probability system.

3. Equilibrium Analysis

We begin by establishing the lower bound on S's payoff in an equilibrium outcome α. One thing that S can do in state θ is fully reveal it by sending message $\{\theta\}$ with probability one. Upon receiving message $\{\theta\}$, R learns that the state is θ and takes an action that is a best response under complete information. Thus, S's equilibrium payoff in state θ is bounded below by $\underline{v}(\theta) := \min_{j \in J \text{ s.t. } \theta \in A_j} v(j)$. We refer to this condition as S's incentive-compatibility constraint:[11]

$$v_\alpha(\theta) \geq \underline{v}(\theta). \tag{IC$_\theta$}$$

Definition 2. *An outcome α is* incentive-compatible *(IC) if it satisfies* (IC$_\theta$) *for each state $\theta \in \Theta$.*

In information design, $\alpha(j \mid \theta)$ is interpreted as the probability that S recommends action j in state θ; R's best response is to follow the recommendation if

$$\int_\Theta (u(j,\theta) - u(j',\theta))\alpha(j \mid \theta)\, \mathrm{d}\mu_0(\theta) \geq 0, \quad \text{for all } j' \in J \setminus \{j\}. \tag{obedience$_j$}$$

Definition 3. *An outcome α is* obedient *if it satisfies* (obedience$_j$) *for each action $j \in J$.*

Naturally, our setting does not allow for action recommendations since S's message space is not J. Nevertheless, we will soon show that all equilibrium outcomes must be obedient.

If α is a deterministic outcome with partition $\{W_j\}_{j \in J}$, then (IC$_\theta$) becomes $\theta \in W_j \implies v(j) \geq \underline{v}(\theta) \iff j \geq \min_{i \in J \text{ s.t. } \theta \in A_i} i$, indicating that the action taken in state θ must be no lower than R's worst best response under complete information. The obedience constraint for action j simplifies to $\int_{W_j} (u(j,\theta) - u(j',\theta))\, \mathrm{d}\mu_0(\theta) \geq 0$ for all $j' \in J \setminus \{j\}$.

Our first result establishes that every equilibrium outcome is incentive-compatible and obedient. For deterministic outcomes, these two properties are necessary and sufficient for equilibrium implementation.

[11] In fact, $\underline{v}(\theta)$ is the lower bound on S's equilibrium payoff in state θ, meaning an equilibrium exists where S's interim payoff is exactly $\underline{v}(\theta)$ for each $\theta \in \Theta$. In that equilibrium, S fully reveals every state, R takes the lowest action that is a best response under complete information, and R's beliefs are skeptical off-path (we define R's skeptical beliefs in the proof of Theorem 1).

Theorem 1.

(a) Every equilibrium outcome is IC and obedient.

(b) If a deterministic outcome is IC and obedient, then it is an equilibrium outcome.

Proof. [Part (a)] Consider an equilibrium (σ, τ, q) with outcome $\alpha \in \Psi$. Observe that α must be incentive-compatible, or else there exists a state θ in which S has a profitable deviation to fully revealing the state. Next, we show that α is also obedient. Consider any action $j \in J$. By the equilibrium condition (ii), we have

$$\text{for all } m \in M_j(\sigma,\tau) \text{ and } j' \in J \setminus \{j\}, \quad \int_\Theta (u(j,\theta) - u(j',\theta)) \, \mathrm{d}q(\theta \mid m) \geq 0$$
$$\implies \int_\Theta (u(j,\theta) - u(j',\theta))\tau(j \mid m) \, \mathrm{d}q(\theta \mid m) \geq 0,$$

where the second inequality follows because $\tau(j \mid m) > 0$ for all $m \in M_j(\sigma, \tau)$. Using the Bayes rule, the above inequality implies that

$$\text{for all } j' \in J \setminus \{j\}, \quad \int_\Theta (u(j,\theta) - u(j',\theta)) \sum_{m \in M_j(\sigma,\tau)} \tau(j \mid m)\sigma(m \mid \theta) \, \mathrm{d}\mu_0(\theta) \geq 0,$$
$$\implies \int_\Theta (u(j,\theta) - u(j',\theta))\alpha(j \mid \theta) \, \mathrm{d}\mu_0(\theta) \geq 0,$$

where the last inequality is (obedience$_j$). Since j was chosen arbitrarily, α is obedient.

[Part (b)] Consider a deterministic outcome α that is IC and obedient and denote its outcome partition by $\{W_j\}_{j \in J}$. We construct an equilibrium (σ, τ, q) that induces α. Let S's strategy be $\sigma(m \mid \theta) = \mathbb{1}(m = W_j \text{ and } \theta \in W_j)$, which reveals which element of the outcome partition the realized state belongs to. When R receives an on-path message W_j, she learns that $\theta \in W_j$ and nothing else; by (obedience$_j$), playing action j is a best response; thus, we let $\tau(j \mid W_j) = 1$ for all $j \in J$. For off-path messages, assume R is "skeptical" and believes that any unexpected message comes from the state in which R prefers to take the lowest action under complete information. Formally, for all $m \notin \{W_j\}_{j \in J}$, let $q(\cdot \mid m) \in \Delta(m \cap A_{\underline{j}})$, where $\underline{j} \in J$ is the lowest action $i \in J$ such that the set $m \cap A_i$ is non-empty. Then, playing action $\underline{j}$ with probability one is a best response to message m, so we let $\tau(\underline{j} \mid m) = 1$.

We now show that S has no profitable deviations using the fact that $\{W_j\}_{j \in J}$ is a partition of the state space. Consider a state $\theta \in \Theta$, which is in W_j for some action

$j \in J$. S cannot send any other on path message because $\theta \in W_j$ implies $\theta \notin W_i$ for any $i \neq j$, so W_i is not a verifiable message in state θ. If S deviates to an off-path (verifiable) message $m \notin \{W_j\}_{j \in J}$, then S's payoff is $v(\underline{j}) \leq \underline{v}(\theta)$, and this deviation is unprofitable by (IC$_\theta$). Therefore, (σ, τ, q) is an equilibrium that induces α. ■

Part (a) of Theorem 1 confirms that every equilibrium outcome must be obedient, and the idea behind the proof is similar to Theorem 1(a) in Zapechelnyuk (2023). Consider all on-path messages after which R plays j with a positive probability. If the same action is a best response after all these messages, then she should choose the same action without knowing which of these messages was sent. Since R chooses the same optimal action after all these messages, we can "bundle" them into a single "recommendation" to take action j. Part (b) of Theorem 1 characterizes the set of deterministic equilibrium outcomes, and its proof suggests a simple way of implementing them in a pure-strategy equilibrium with at most K on-path messages that essentially serve as action recommendations. Specifically, if $\{W_j\}_{j \in J}$ is an outcome partition, then W_j serves as both the set of states in which R plays action j and as the on-path message recommending action j in the constructed equilibrium inducing this outcome.

While Theorem 1 fully characterizes the set of deterministic equilibrium outcomes, it does not provide a full characterization of the entire set of equilibrium outcomes. In general, an IC and obedient non-deterministic outcome may or may not be an equilibrium outcome. For instance, consider the seminal example from Kamenica and Gentzkow (2011).

Example 1. Suppose S is a prosecutor and R is a judge. The state of the world is binary: $\Theta = \{1, 2\} = \{\text{innocent}, \text{guilty}\}$, R's action space is binary: $J = \{1, 2\} = \{\text{acquit}, \text{convict}\}$, and the prior is $\mu_0(1) = 0.7$. S's preferences are $v(1) = 0$ and $v(2) = 1$, while R's objective is to "match the state": $u(1, 1) = u(2, 2) = 1$, and $u(1, 2) = u(2, 1) = 0$. Consider an outcome α^* where $\alpha^*(2 \mid 2) = 1$ and $\alpha^*(2 \mid 1) = 3/7$. It is easy to verify that α^* is both IC and obedient. However, α^* is not an equilibrium outcome: when $\theta = 1$, R convicts with probability $3/7$ and acquits with probability $4/7$. Since S strictly prefers conviction, he has a profitable deviation to sending the message after which R convicts when $\theta = 1$.

Example 1 illustrates that (IC and obedient) outcomes in which S receives different payoffs from different messages in the same state cannot be equilibrium outcomes. Once the state is realized, S's message space becomes fixed and known. Thus, if S mixes between multiple messages in the same state, he must receive the same payoff from all these messages. In Section 5, we show that once S's set of available messages is rich and determined stochastically, IC and obedience become necessary and sufficient for equilibrium implementation of a commitment outcome, because the requirement for S's payoff to remain constant in state θ is lifted.

Of course, if S does receive the same payoff in every state, then an IC, obedient and non-deterministic outcome could be an equilibrium outcome. However, in any such equilibrium, R must play a mixed strategy:

LEMMA 1. *Suppose that α is a non-deterministic outcome induced by an equilibrium (σ, τ, q). Then in each state $\theta \in \Theta$ such that $\alpha(\cdot \mid \theta)$ is non-degenerate, R is playing a mixed strategy (meaning $\tau(\cdot \mid m)$ is non-degenerate) for some $m \in \operatorname{supp} \sigma(\cdot \mid \theta)$.*

Proof. Let $\theta \in \Theta$ be a state such that $\alpha(\cdot \mid \theta)$ is non-degenerate. By contradiction, suppose that $\tau(\cdot \mid m)$ is degenerate for all $m \in \operatorname{supp} \sigma(\cdot \mid \theta)$. By equilibrium condition (i), for any pair of messages $m, m' \in \operatorname{supp} \sigma(\cdot \mid \theta)$, we have $\sum_{j \in J} v(j)\tau(j \mid m) = \sum_{j \in J} v(j)\tau(j \mid m')$, implying that there exists an action $j^* \in J$ such that $\tau(j^* \mid m) = \tau(j^* \mid m') = 1$. In other words, if R is not mixing, every message sent by S in state θ leads R to take the same action. Therefore, $\alpha(j^* \mid \theta) = \sum_{m \in M_j(\sigma,\tau)} \tau(j^* \mid m)\sigma(m \mid \theta) = 1$, which is a contradiction. ∎

The contrapositive of Lemma 1 also tells us that if R is not mixing in an equilibrium (e.g., if she uses an exogenously-given tie-breaking rule like in informed information design), then an obedient outcome is an equilibrium outcome *if and only if* it is IC and deterministic. Theorem 1 and Lemma 1 together highlight the difference in equilibrium sets between our verifiable disclosure game and informed information design (KS and Z). KS's characterization (Proposition 2) states that an outcome is interim optimal (IO) if and only if it is obedient and IOC, where IOC essentially requires that for every set of states Q, and for every state in Q, S does not strictly prefer R having a belief supported on Q.[12] Naturally, the first difference—IOC in their setting versus IC in ours—arises from the difference in equilibrium selection, as they impose a stronger restriction on off-path beliefs than we do. The second difference is that IOC and obedience are necessary and sufficient for an outcome to be IO, while for us IC and obedience are not sufficient. Since in our model the sender chooses messages, an additional restriction applies: S can mix between different messages only if they yield the same expected payoffs—a constraint absent in informed information design. For this reason, some

[12]In contrast, IC only requires that in any given state, S does not strictly prefer inducing the degenerate belief at that state. Interestingly, KS also show that IOC and IC are equivalent if S's value function is quasiconvex in R's belief (Proposition 3) or if R chooses between two actions (Lemma B.2).

non-deterministic IO, and thus IC, outcomes are not equilibrium outcomes in our game (e.g., one from Example 1).

Theorem 1 characterizes all pure-strategy equilibria of the game, as these equilibria are deterministic. Koessler and Renault (2012) find that IC and obedience are necessary and sufficient for a pure-strategy outcome to be an equilibrium outcome in a setting where S has state-independent preferences, sends verifiable messages, and sets a price, and R chooses between two actions. Theorem 1 highlights that their result: (1) extends to cases where R has more than two actions, and (2) is not driven by S's additional choice variable (price).[13]

4. Value of Commitment

In this section, we ask: when is a commitment outcome, a solution to the information design problem, also an equilibrium outcome? In the information design problem, stage 1 of the game (in which the sender learns the state) is removed, and stage 2 of the game (in which the sender chooses a verifiable message) is replaced by S committing to an experiment that sends signals depending on state realizations. [14] Importantly, when S has commitment power, he no longer faces an incentive-compatibility constraint: he does not need to maximize his utility state-by-state, nor do his signals need to be verifiable.

Following Kamenica and Gentzkow (2011), we focus on straightforward signals that R interprets as action recommendations. Therefore, an (optimal) *commitment outcome* $\overline{\psi} \in \Psi$ solves

$$\max_{\psi \in \Psi} V_\psi \quad \text{subject to, for each action } j \in J,$$
$$\int_\Theta (u(j,\theta) - u(j',\theta))\psi(j \mid \theta)\, \mathrm{d}\mu_0(\theta) \geq 0 \quad \text{for all } j' \in J \setminus \{j\}. \tag{CO}$$

Simply put, a commitment outcome is an obedient outcome that maximizes S's ex-ante utility. We refer to the value of problem (CO) as the *commitment payoff*. Our second result shows that a commitment outcome must be deterministic and incentive-compatible to be an equilibrium outcome.

[13] However, as the authors point out, the price choice in their setting ensures that R plays a pure strategy in equilibrium.

[14] An experiment $(\mathcal{S}, \chi)$ consists of a compact metrizable space $\mathcal{S}$ of signals and a Borel measurable function $\chi : \Theta \to \Delta\mathcal{S}$. R observes the choice of the experiment and a signal realization $s \in \mathcal{S}$ drawn from $\chi(\cdot \mid \theta)$, where θ is the realized state.

THEOREM 2. *Consider a commitment outcome $\overline{\psi} \in \Psi$.*

(a) If $\overline{\psi}$ is IC and deterministic, then it is an equilibrium outcome.

(b) If $\overline{\psi}$ is an equilibrium outcome, then it is IC and μ_0-almost everywhere deterministic.

Part (a). Recall that every commitment outcome is obedient. Therefore, if $\overline{\psi}$ is IC and deterministic, Part (b) of Theorem 1 implies that it is an equilibrium outcome.

[Part (b)] Suppose a commitment outcome $\overline{\psi}$ is an equilibrium outcome, meaning that there exists an equilibrium (σ, τ, q) that induces it. By Theorem 1, $\overline{\psi}$ is incentive-compatible. We will now show that $\overline{\psi}$ is deterministic μ_0-almost everywhere. Define $T := \{\theta \in \Theta \mid \overline{\psi}(\cdot \mid \theta) \text{ is non-degenerate}\}$ as the set of states where R plays multiple actions and suppose, by contradiction, that $\mu_0(T) > 0$. By Lemma 1, for each $\theta \in T$, there exists a message $m \in \operatorname{supp} \sigma(\cdot \mid \theta)$ such that $\tau(\cdot \mid m)$ is non-degenerate. Let $\widetilde{M} := \{m \in M \mid \tau(\cdot \mid m) \text{ is non-degenerate}\}$ be the set of messages after which R plays a mixed strategy. Define $\widetilde{\tau}(j^* \mid m) := \mathbb{1}(j^* = \max_{j \in \operatorname{supp} \tau(\cdot \mid m)} j)$ for all $m \in M$ as R's strategy that breaks all ties in τ in favor of S. Denote the outcome from the strategy profile $(\sigma, \widetilde{\tau})$ by $\widetilde{\psi}$.

We derive a contradiction by showing that $\widetilde{\psi}$ is an obedient outcome with $V_{\widetilde{\psi}} > V_{\overline{\psi}}$, which implies that $\overline{\psi}$ is not a commitment outcome. Indeed, we have $v_{\widetilde{\psi}}(\theta) > v_{\overline{\psi}}(\theta)$ for all $\theta \in T$ (since there exists an $m \in \widetilde{M}$ with $\sigma(m \mid \theta) > 0$), while $v_{\widetilde{\psi}}(\theta) = v_{\overline{\psi}}(\theta)$ for all $\theta \notin T$. Therefore, $V_{\widetilde{\psi}} - V_{\overline{\psi}} = \int_T (v_{\widetilde{\psi}}(\theta) - v_{\overline{\psi}}(\theta)) \, \mathrm{d}\mu_0(\theta) > 0$ since $\mu_0(T) > 0$. To prove that $\widetilde{\psi}$ is obedient, we apply equilibrium condition (ii) for the equilibrium (σ, τ, q) and follow the steps in the proof of Theorem 1(a), replacing τ with $\widetilde{\tau}$ and noting that $M_j(\sigma, \widetilde{\tau}) \subseteq M_j(\sigma, \tau)$. ■

The non-trivial part of Theorem 2 involves proving that if $\overline{\psi}$ is both an equilibrium and a commitment outcome, then it is deterministic almost everywhere. This is equivalent to showing that a non-deterministic equilibrium outcome cannot be a commitment outcome. Indeed, by Lemma 1, in the equilibrium inducing $\overline{\psi}$, R must play a mixed strategy following some on-path messages from a positive measure of states. However, breaking those ties in favor of the S-preferred action strictly increases S's ex-ante utility, which implies that $\overline{\psi}$ is not a commitment outcome.

In many relevant settings, R chooses between two actions. In this case, the analysis vastly simplifies. From R's perspective, there are "bad" states $\theta \in A_1$, in which R prefers the low action 1, and "good" states $\theta \notin A_1$, in which she prefers the high action 2. The highest payoff that S can achieve is $v(2)$ (when R takes action 2 with probability one), and the lowest is $v(1)$. To state that an outcome $\psi \in \Psi$ is incentive-compatible, it suffices to show that $\theta \notin A_1$ implies that $v_\psi(\theta) = v(2)\psi(2 \mid \theta) + v(1)\psi(1 \mid \theta) \geq v(2)$, which is equivalent to $\psi(2 \mid \theta) = 1$. The IC condition for $\theta \in A_1$ is not relevant because $v(1)$ is already the lowest payoff in the game. In words, an outcome is IC if and only if R plays action 2 with probability one in all states where action 2 is the unique best response under complete information. The following result establishes the existence of an incentive-compatible commitment outcome when R chooses between two actions.

Proposition 1. *If $|J| = 2$, then there exists an IC commitment outcome.*

Proof. Since Θ is a compact subset of $\mathbb{R}^n$, a commitment outcome exists by Proposition 3 in the Online Appendix of Kamenica and Gentzkow (2011) and Theorem 1 in Terstiege and Wasser (2023). Let $\overline{\psi} \in \Psi$ be a commitment outcome and let $\widetilde{\psi} \in \Psi$ be an outcome such that $\widetilde{\psi}(\cdot \mid \theta) = \overline{\psi}(\cdot \mid \theta)$ for all $\theta \notin A_2$ and $\widetilde{\psi}(2 \mid \theta) = 1$ for all $\theta \in A_2$. By construction, $\widetilde{\psi}$ is incentive-compatible and weakly increases S's ex-ante utility over $\overline{\psi}$. Define $\delta(\theta) := u(2, \theta) - u(1, \theta)$ and observe that

$$\int_\Theta \delta(\theta)\widetilde{\psi}(2 \mid \theta)\,\mathrm{d}\mu_0(\theta) = \int_\Theta \delta(\theta)\overline{\psi}(2 \mid \theta)\,\mathrm{d}\mu_0(\theta) + \int_{A_2} \delta(\theta)(1 - \overline{\psi}(2 \mid \theta))\,\mathrm{d}\mu_0(\theta),$$

where the last term is non-negative because $\delta(\theta) \geq 0$ for all $\theta \in A_2$. Consequently, obedience of $\overline{\psi}$ (for both actions) implies obedience of $\widetilde{\psi}$. Hence, $\widetilde{\psi}$ is also a commitment outcome. ∎

The existing literature provides additional insights into commitment outcomes when $|J| = 2$ and Θ is finite. Alonso and Câmara (2016) show that each commitment outcome is characterized by a cutoff state θ^*, with all states satisfying $\delta(\theta) > \delta(\theta^*)$ pooled together to recommend action 2. In particular, all good states $\theta \notin A_1$ recommend action 2, which implies that every commitment outcome is incentive-compatible (see also Lemma B.2 in Koessler and Skreta, 2023). Our Proposition 1 also deals with the case when Θ is rich, in which case some commitment outcomes are not IC (although they are IC μ_0-almost everywhere), and its proof outlines how to make an existing commitment outcome incentive-compatible.

Returning to the more general case with $J \geq 2$, Theorem 2 is useful for verifying whether an existing commitment outcome $\overline{\psi}$ is an equilibrium outcome. The answer is affirmative if and only if $\overline{\psi}$ is deterministic μ_0-a.e. and incentive-compatible. Although verifying incentive-compatibility may be straightforward, a deterministic commitment outcome is not guaranteed to exist. In the remainder of this section, we consider the cases where Θ is rich and finite separately. We show that when Θ is rich, a deterministic commitment outcome always exists. Furthermore, if $|J| = 2$, the commitment payoff is always attained in equilibrium. When Θ is finite, we derive an approximation result.

4.1. Rich State Space

When the state space Θ is rich (a convex and compact subset of $\mathbb{R}^n$) and the prior μ_0 is atomless, the existence of a deterministic commitment outcome is guaranteed.

Proposition 2. *If Θ is rich, then a deterministic commitment outcome exists. Furthermore, a deterministic commitment outcome is an equilibrium outcome if and only if it is IC.*

Proof. The existence of a commitment outcome $\overline{\psi}$ follows using the same argument as in the proof of Proposition 1. Furthermore, $\overline{\psi}(j \mid \cdot) : \Theta \to [0, 1]$ is Borel measurable for every $j \in J$, and $\sum_{j \in J} \overline{\psi}(j \mid \theta) = 1$ for all $\theta \in \Theta$. Let μ_j be such that $\mathrm{d}\mu_j := u(j, \cdot)\,\mathrm{d}\mu_0$ for each $j \in J$.

Since μ_0 is a finite and atomless positive measure and u is bounded, μ_j is a finite and atomless signed measure for each $j \in J$. By Theorem 2.1 in Dvoretzky, Wald, and Wolfowitz (1951), since J is finite, there exist Borel measurable functions $\widetilde{\psi}(j \mid \cdot) : \Theta \to \{0, 1\}$ for all $j \in J$, with $\sum_{j \in J} \widetilde{\psi}(j \mid \cdot) = 1$, such that (I) $\int_\Theta \widetilde{\psi}(j \mid \theta)\,\mathrm{d}\mu_0 = \int_\Theta \overline{\psi}(j \mid \theta)\,\mathrm{d}\mu_0$ and (II) $\int_\Theta \widetilde{\psi}(j \mid \theta)\,\mathrm{d}\mu_j = \int_\Theta \overline{\psi}(j \mid \theta)\,\mathrm{d}\mu_j$ for all $j \in J$. Condition (I) implies that

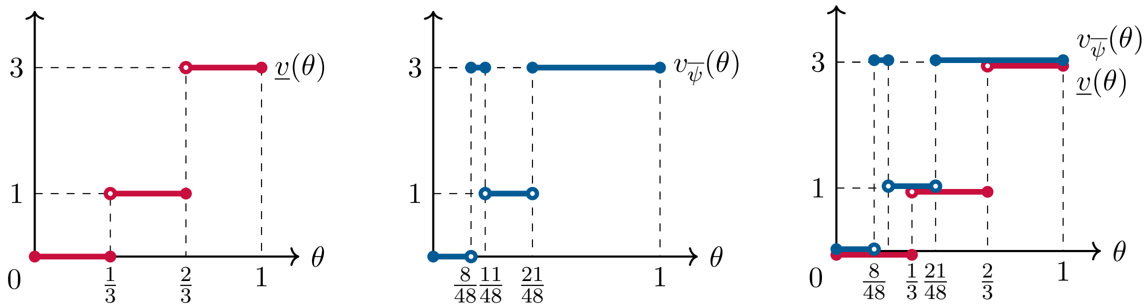
$$V_{\widetilde{\psi}} = \int_\Theta \sum_{j \in J} v(j)\widetilde{\psi}(j \mid \theta)\,\mathrm{d}\mu_0 = \int_\Theta \sum_{j \in J} v(j)\overline{\psi}(j \mid \theta)\,\mathrm{d}\mu_0 = V_{\overline{\psi}}.$$

Condition (II) implies that $\widetilde{\psi}$ is obedient, as $\overline{\psi}$ is. Hence, $\widetilde{\psi}$ is a deterministic commitment outcome. The second part follows from Theorem 2. ■

Verifying whether a deterministic commitment outcome with partition $\{W_j\}_{j \in J}$ is incentive-compatible (and therefore an equilibrium outcome) is straightforward. It requires checking that $\theta \in W_j$ implies $v(j) \geq \underline{v}(\theta)$ for all $\theta \in \Theta$. Consider the following example from Gentzkow and Kamenica (2016).

Example 2. Suppose R has three actions, $J = \{1, 2, 3\}$, and the prior is uniform on $\Theta = [0, 1]$. S's payoffs are given by $v(1) = 0$, $v(2) = 1$, and $v(3) = 3$. R's preferences depend only on the posterior mean. Given belief $\mu \in \Delta\Theta$, action 1 is optimal if and only if $\mathbb{E}_\mu[\theta] \leq 1/3$; action 2 is optimal if and only if $\mathbb{E}_\mu[\theta] \in [1/3, 2/3]$; action 3 is optimal if and only if $\mathbb{E}_\mu[\theta] \geq 2/3$. Therefore, R's complete-information action sets are

$A_1 = [0, 1/3]$, $A_2 = [1/3, 2/3]$, and $A_3 = [2/3, 1]$. Gentzkow and Kamenica (2016) identify a deterministic commitment outcome $\overline{\psi}$ with an outcome partition $\overline{W}_1 = [0, 8/48)$, $\overline{W}_2 = (11/48, 21/48)$, and $\overline{W}_3 = [8/48, 11/48] \cup [21/48, 1]$. This outcome is incentive-compatible, which we illustrate in Figure 1. Since $\overline{\psi}$ is a deterministic and IC commitment outcome, it is an equilibrium outcome by Proposition 2.



(a) *S's lowest payoff under complete information.* ***(b)*** *S's expected payoff in the commitment outcome.* ***(c)*** *$\overline{\psi}$ is IC since $v_{\overline{\psi}}(\theta) \geq \underline{v}(\theta)$ for all $\theta \in [0, 1]$.*

Figure 1. *Commitment outcome $\overline{\psi}$ is incentive-compatible since S receives at least his complete information payoff in every state of the world.*

When R chooses between two actions, S always attains his commitment payoff in equilibrium.

PROPOSITION 3. *If Θ is rich and $|J| = 2$, then there exists a commitment outcome that is an equilibrium outcome.*

Proof. By Proposition 2, there exists a deterministic commitment outcome $\overline{\psi}$. Using the same argument as in the proof of Proposition 1, we construct a deterministic commitment outcome $\widetilde{\psi}$ that is IC. By Proposition 2, $\widetilde{\psi}$ is an equilibrium outcome. ■

4.2. FINITE STATE SPACE

When the state space is finite, i.e., $\Theta = \{1, \ldots, N\}$, a deterministic commitment outcome may not exist. For instance, in Example 1, the *unique* commitment outcome is not deterministic. As a result, S may not be able to achieve the commitment payoff in equilibrium.

However, here we show that when the state space is sufficiently rich (in the sense that $\mu_0(\theta)$ is sufficiently small for each $\theta \in \Theta$), then S's equilibrium payoff approaches his commitment payoff. For a concise argument, we adopt the assumptions of Alonso and Câmara (2016): R has a binary action and

$$\theta' \neq \theta'' \implies \delta(\theta') \neq \delta(\theta''), \tag{RU}$$

where $\delta(\theta) = u(2, \theta) - u(1, \theta)$ for all $\theta \in \Theta$.

PROPOSITION 4. *Suppose that Θ is finite, $|J| = 2$, (RU) holds, and S's payoffs are normalized to $v(2) = 1$ and $v(1) = 0$.*[15] *Let V^* be S's commitment payoff. For every $\varepsilon > 0$, there is $\gamma > 0$ such that if $\mu_0(\theta) < \gamma$ for all $\theta \in \Theta$, then there exists an equilibrium outcome α with $|V^* - V_\alpha| < \varepsilon$.*

Proof. If $A_2 = \Theta$ then let $\alpha(2 \mid \theta) = 1$ for all $\theta \in \Theta$ so that $V_\alpha = V^*$. Thus, we assume for the remainder of the proof that A_2 is a proper subset of Θ. Since (RU) holds, we can use Proposition 2 in Alonso and Câmara (2016) to find a cutoff state $\theta^* \in \Theta$ such that $\delta(\theta^*) < 0$ and, for every commitment outcome ψ, we have $\psi(2 \mid \theta) = 1$ ($\psi(1 \mid \theta) = 1$) for all $\theta \in \Theta$ such that $\delta(\theta) > \delta(\theta^*)$ ($\delta(\theta) < \delta(\theta^*)$). Now, consider a deterministic outcome α with partition $\{W_1, W_2\}$ such that $W_2 = \{\theta \in \Theta \mid \psi(2 \mid \theta) = 1\}$ and $W_1 = \Theta \setminus W_2$. It is easy to see that α is IC and obedient, and therefore it is an equilibrium outcome by Theorem 1. The difference in S's ex-ante payoffs is zero if $\psi(2 \mid \theta^*) = 1$; otherwise, we have $V^* - V_\alpha = \psi(2 \mid \theta^*)\mu_0(\theta^*) < \mu_0(\theta^*) < \gamma := \varepsilon$. ∎

Thus, when R chooses between two actions, S can attain a payoff arbitrarily close to his commitment payoff in equilibrium, as long as the prior probability of each state is sufficiently small.

5. A Model with a Stochastic Message Mapping

In the main model, IC and obedience were not sufficient for an outcome to be induced by an equilibrium; non-deterministic but IC and obedient outcomes exist in which S effectively recommends multiple actions, leading to different expected payoffs in the same state. This violates equilibrium condition (i). The reason why (i) is violated is that the mapping $E : \Theta \rightrightarrows M$, a correspondence that determines the set of messages available in state θ, is deterministic. This assumption is standard in the literature on verifiable disclosure and cheap talk.[16] In some cases, however, it is reasonable to assume that the mapping $E(\theta)$ is stochastic: for example, there may be different labels for the same state, and S can make statements about the label rather than the state. In this section, we introduce a verifiable disclosure game with a stochastic message mapping

[15] Normalizing S's payoffs is without loss of generality; condition (RU) simplifies the proof, but the result remains true without it.

[16] For example, in Grossman (1981), Milgrom (1981), and Milgrom and Roberts (1986), $E(\theta)$ includes subsets of Θ that contain θ. In Dye (1985), $E(\theta)$ is binary, S can reveal θ or say nothing. In Hart, Kremer, and Perry (2017) and Ben-Porath, Dekel, and Lipman (2019), $E(\theta)$ is a partial order on Θ. In cheap talk, $E(\theta)$ is the same for all $\theta \in \Theta$.

(henceforth, the SMM game) and show that IC and obedience are sufficient for an outcome to be an equilibrium outcome of that game.

The SMM game has the same timeline and players' objectives as our main model, with the only modification occurring in Stage 2, where S communicates with R. Specifically, we assume that along with the state of the world $\theta \in \Theta$, where the state space $\Theta = \{1, \ldots, N\}$ is finite, S also observes a *label* $x \in [0, 1]$, which is payoff-irrelevant to both S and R. The label x is drawn from the uniform distribution on X^θ, where $\{X^\theta\}_{\theta\in\Theta}$ forms a partition of the unit interval such that $\lambda(X^\theta) = \mu_0(\theta)$, where λ is the Lebesgue measure.[17] Having observed θ and x, S sends message $m \in \widehat{M}$ such that $x \in m$, where $\widehat{M}$ is the collection of nonempty Borel subsets of $[0, 1]$. Thus, the set of messages available to S in state θ is now determined stochastically (through x). The *equilibrium of the SMM game* $(\widehat{\sigma}, \widehat{\tau}, \widehat{q})$ is defined analogously to that of the main model, except S's strategy also depends on x.

Definition 4. *A triple $(\widehat{\sigma}, \widehat{\tau}, \widehat{q})$, where $\widehat{\sigma} : \Theta \times [0, 1] \to \Delta_0\widehat{M}$ is S's strategy, $\widehat{\tau} : \widehat{M} \to \Delta J$ is R's strategy and $\widehat{q} : \widehat{M} \to \Delta\Theta$ is R's belief system, is an* equilibrium of the SMM game *if*

(i) For all $\theta \in \Theta$ and $x \in [0, 1]$, $\widehat{\sigma}(\cdot \mid \theta, x)$ is supported on $\underset{\{m\in\widehat{M} \mid x\in m\}}{\arg\max} \sum_{j\in J} v(j)\, \widehat{\tau}(j \mid m)$;

(ii) For all $m \in \widehat{M}$, $\widehat{\tau}(\cdot \mid m)$ is supported on $\underset{j\in J}{\arg\max} \int_\Theta u(j, \theta)\, \mathrm{d}q(\theta \mid m)$;

(iii) $\widehat{q}$ is obtained from μ_0, given $\widehat{\sigma}$, using Bayes rule.

(iv) For all $m \in \widehat{M}$, $\widehat{q}(\cdot \mid m) \in \Delta\{\theta \in \Theta \mid X^\theta \cap m \neq \varnothing\}$.

Since x is payoff-irrelevant, an outcome α of the SMM game is an element of Ψ. An outcome α is an *equilibrium outcome* of the SMM game if an equilibrium $(\widehat{\sigma}, \widehat{\tau}, \widehat{q})$ exists that induces it, i.e., $\alpha(j \mid \theta) = \frac{1}{\mu_0(\theta)} \int_{X^\theta} \sum_{m\in\operatorname{supp} \widehat{\sigma}(\cdot \mid \theta,x)} \widehat{\sigma}(m \mid \theta, x)\widehat{\tau}(j \mid m)\, \mathrm{d}x$.

We derive a sharp characterization of equilibrium outcomes in the SMM game.

Theorem 3. *Let Θ be finite. Then, $\alpha \in \Psi$ is an equilibrium outcome of the SMM game $\iff$ α is IC and obedient.*

Proof. ($\Longrightarrow$) is proved exactly the same way as Theorem 1 (a). An equilibrium outcome must be IC or else S has a profitable deviation to fully revealing x (which also reveals $\theta \in \Theta$ since $x \in X^\theta$). An equilibrium outcome must be obedient by Bayes rule.

---

[17]For example, let $t_0 := 0$, $t_\theta := \sum_{\theta'=1}^{\theta} \mu_0(\theta')$ for all $\theta \in \Theta$; also, let $X^\theta = [t_{\theta-1}, t_\theta)$ for all $\theta \in \{1, \ldots, N-1\}$ and $t_N = [t_{N-1}, 1]$. Then, $\{X^\theta\}_{\theta\in\Theta}$ is a partition of $[0, 1]$ and $\lambda(X^\theta) = \sum_{\theta'=1}^{\theta} \mu_0(\theta') - \sum_{\theta'=1}^{\theta-1} \mu_0(\theta') = \mu_0(\theta)$.

($\Longleftarrow$) Consider an IC and obedient outcome α. For every $\theta \in \Theta$, let $J^\theta := \operatorname{supp} \alpha(\cdot \mid \theta)$ be the set of actions that R takes with a positive probability when the realized state is θ. Next, partition X^θ into a set of intervals $\{X_j^\theta\}_{j \in J^\theta}$ such that $\frac{\lambda(X_j^\theta)}{\lambda(X^\theta)} = \alpha(j \mid \theta)$. Also, for each action $j \in J$, let $W_j := \bigcup_{\theta \in \Theta} X_j^\theta$; by construction, $\{W_j\}_{j \in J}$ is a partition of $[0,1]$.

Now, let S's strategy be $\widehat{\sigma}(m \mid \theta, x) = \mathbb{1}(m = W_j \text{ and } x \in W_j)$. Then, R's posterior after an on-path message W_j is $\widehat{q}(\theta \mid W_j) = \frac{\lambda(X_j^\theta)}{\lambda(W_j)}$. Furthermore, since α is obedient, for every action $j \in J$ such that $\lambda(W_j) > 0$, we have

$$\sum_{\theta \in \Theta} \big(u(j,\theta) - u(j',\theta)\big)\alpha(j \mid \theta)\mu_0(\theta) \geq 0 \iff$$
$$\sum_{\theta \in \Theta} \big(u(j,\theta) - u(j',\theta)\big)\frac{\lambda(X_j^\theta)}{\lambda(W_j)} \geq 0 \quad \text{for all } j' \in J \setminus \{j\},$$

meaning that R prefers to take action j after message W_j, so we let $\widehat{\tau}(j \mid W_j) = 1$. Off the path, let R be "skeptical" and assume that any unexpected message comes from the state in which S benefits from such deviation the most. Formally, for all $m \notin \{W_j\}_{j \in J}$, let $\widehat{q}(\cdot \mid m) \in \Delta A_{\underline{j}}$, where $\underline{j} \in J$ is the lowest action such that $m \cap X^\theta \neq \varnothing$ and $\theta \in A_i$. Then, playing action $\underline{j}$ is a best response to message m, so we let $\widehat{\tau}(\underline{j} \mid m) = 1$. Since α is IC, S does not have profitable deviations by the same argument as in the proof of Theorem 1. Deviations to on-path messages are not available because $\{W_j\}_{j \in J}$ is a partition, while deviations to off-path messages are not profitable since the payoff from any deviation in state θ is at most $\underline{v}(\theta)$, which is below $v_\alpha(\theta)$ by the (IC_θ) constraint. Hence, $(\widehat{\sigma}, \widehat{\tau}, \widehat{q})$ is an equilibrium of the SMM game. ■

In contrast to Theorem 1, incentive-compatibility and obedience are necessary and sufficient for an outcome to be an equilibrium outcome of the SMM game. Two properties of the SMM game ensure that every IC and obedient outcome is an equilibrium outcome. First, S's message space depends on x, which means S may receive different equilibrium payoffs in some state θ (but for different realizations of x). Secondly, the message space is "rich," meaning that for every vector $p = (p_1, \ldots, p_N) \in [0,1]^N$ there exists a message m that is available in state $\theta \in \Theta$ with probability p_θ. This richness allows us to "purify" any non-deterministic outcome: the equilibria that we construct to implement an IC and obedient outcome is in pure strategies of both S and R.

Using the sharp equilibrium characterization of the SMM game, we derive the following results.

Corollary 1. *Let Θ be finite. Then, a commitment outcome is an equilibrium outcome* *of the SMM game if and only if it is IC.*

Corollary 2. *If Θ is finite and $|J| = 2$, then every commitment outcome is an equilibrium outcome of the SMM game.*

Corollary 1 is a direct consequence of Theorem 3. Corollary 2 follows from Alonso and Câmara (2016), who show that every commitment outcome is incentive-compatible (see also our discussion after Proposition 1) and Theorem 3, which ensures that every such outcome is an equilibrium outcome.

The set of equilibrium outcomes in the SMM game coincides with the set of IO outcomes found in KS if S's value function is quasiconvex in R's belief (KS Proposition 3), or when R chooses between two actions, then these two sets coincide (KS Proposition 4). Generally, the set of IO outcomes is a subset of the set of equilibrium outcomes in KS, because interim-optimality is a stronger restriction on off-path beliefs.

6. Conclusion

This paper examines a persuasion game with verifiable information, in which a sender with transparent motives chooses which verifiable messages to send to a receiver in order to convince her to take a particular action from a finite set. We show that every equilibrium outcome must be incentive-compatible for the sender and obedient for the receiver. If an outcome is deterministic, then these conditions are both necessary and sufficient for it to be an equilibrium outcome. We also identify sufficient conditions under which the ex-ante commitment assumption in Bayesian persuasion can be replaced by communication with verifiable information. We show that if the state space is rich, then a deterministic commitment outcome always exists; that commitment outcome is an equilibrium outcome if and only if the sender receives at least his complete information payoff in every state. If the receiver chooses between two actions, this condition is automatically satisfied. We hope these results prove useful in applied settings.

References

Ali, S. Nageeb, Andreas Kleiner, and Kun Zhang (2024), "From Design to Disclosure", working paper. (p. 3.)

Alonso, Ricardo and Odilon Câmara (2016), "Persuading Voters", *American Economic Review*, 106, 11 (Nov. 2016), pp. 3590-3605. (pp. 2, 13, 15, 16, 19.)

Bardhi, Arjada and Yingni Guo (2018), "Modes of Persuasion toward Unanimous Consent", *Theoretical Economics*, 13, 3, pp. 1111-1149. (p. 2.)

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019), "Mechanisms With Evidence: Commitment and Robustness", *Econometrica*, 87, 2, pp. 529-566. (pp. 4, 16.)

Boleslavsky, Raphael and Christopher Cotton (2015), "Grading Standards and Education Quality", *American Economic Journal: Microeconomics*, 7, 2 (May 2015), pp. 248-279. (p. 2.)

Bull, Jesse and Joel Watson (2007), "Hard Evidence and Mechanism Design", *Games and Economic Behavior*, 58, 1, pp. 75-93. (p. 5.)

Chakraborty, Archishman and Rick Harbaugh (2010), "Persuasion by Cheap Talk", *American Economic Review*, 100, 5, pp. 2361-2382. (p. 4.)

Dranove, David and Ginger Zhe Jin (2010), "Quality Disclosure and Certification: Theory and Practice", *Journal of Economic Literature*, 48, 4, pp. 935-963. (p. 3.)

Dvoretzky, A., A. Wald, and J. Wolfowitz (1951), "Elimination of Randomization in Certain Statistical Decision Procedures and Zero-Sum Two-Person Games", *The Annals of Mathematical Statistics*, 22 (1 Mar. 1951), pp. 1-21. (p. 14.)

Dye, Ronald A (1985), "Disclosure of nonproprietary information", *Journal of accounting research*, pp. 123-145. (p. 16.)

Gehlbach, Scott and Konstantin Sonin (2014), "Government Control of the Media", *Journal of Public Economics*, 118 (Oct. 2014), pp. 163-171. (p. 2.)

Gentzkow, Matthew and Emir Kamenica (2016), "A Rothschild-Stiglitz Approach to Bayesian Persuasion", *American Economic Review*, 106, 5 (May 2016), pp. 597-601. (pp. 14, 15.)

Gieczewski, Germán and Maria Titova (2024), "Coalition-Proof Disclosure", working paper. (p. 3.)

Glazer, Jacob and Ariel Rubinstein (2004), "On Optimal Rules of Persuasion", *Econometrica*, 72 (6 Nov. 2004), pp. 1715-1736. (p. 4.)

— (2006), "A Study in the Pragmatics of Persuasion: a Game Theoretical Approach", *Theoretical Economics*, 1, pp. 395-410. (p. 4.)

Grossman, Sanford J. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality", *The Journal of Law and Economics*, 24, 3 (Dec. 1981), pp. 461-483. (pp. 1, 3, 16.)

Grossman, Sanford J. and Oliver D. Hart (1980), "Disclosure Laws and Takeover Bids", *The Journal of Finance*, 35, 2, pp. 323-334. (p. 3.)

Hart, Sergiu, Ilan Kremer, and Motty Perry (2017), "Evidence Games: Truth and Commitment", *American Economic Review*, 107 (3 Mar. 2017), pp. 690-713. (pp. 4, 16.)

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian Persuasion", *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615. (pp. 2, 9, 11, 13.)

Kamenica, Emir and Xiao Lin (2024), "Commitment and Randomization in Communication", *arXiv preprint arXiv:2410.17503*. (p. 4.)

Koessler, Frédéric and Régis Renault (2012), “When Does a Firm Disclose Product Information?”, *The RAND Journal of Economics*, 43, 4, pp. 630-649. (p. 11.)

Koessler, Frédéric and Vasiliki Skreta (2023), “Informed Information Design”, *Journal of Political Economy* (Mar. 28, 2023). (pp. 3, 13.)

Kolotilin, Anton (2015), “Experimental Design to Persuade”, *Games and Economic Behavior*, 90 (Mar. 2015), pp. 215-226. (p. 2.)

Lipnowski, Elliot (2020), “Equivalence of Cheap Talk and Bayesian Persuasion in a Finite Continuous Model”, working paper. (p. 4.)

Lipnowski, Elliot and Doron Ravid (2020), “Cheap Talk With Transparent Motives”, *Econometrica*, 88, 4, pp. 1631-1660. (p. 4.)

Milgrom, Paul R. (1981), “Good News and Bad News: Representation Theorems and Applications”, *The Bell Journal of Economics*, 12, 2, p. 380. (pp. 1, 3, 16.)

— (2008), “What the Seller Won’t Tell You: Persuasion and Disclosure in Markets”, *Journal of Economic Perspectives*, 22, 2 (Mar. 2008), pp. 115-131. (p. 3.)

Milgrom, Paul R. and John Roberts (1986), “Relying on the Information of Interested Parties”, *The RAND Journal of Economics*, 17, 1, p. 18. (pp. 3, 5, 16.)

Ostrovsky, Michael and Michael Schwarz (2010), “Information Disclosure and Unraveling in Matching Markets”, *American Economic Journal: Microeconomics*, 2, 2 (May 2010), pp. 34-63. (p. 2.)

Perez-Richet, Eduardo (2014), “Interim Bayesian Persuasion: First Steps”, *American Economic Review*, 104, 5, pp. 469-474. (p. 3.)

Sher, Itai (2011), “Credibility and Determinism in a Game of Persuasion”, *Games and Economic Behavior*, 71 (2 Mar. 2011), pp. 409-419. (p. 4.)

Terstiege, Stefan and Cédric Wasser (2023), “Experiments Versus Distributions of Posteriors”, *Mathematical Social Sciences*, 125, pp. 58-60. (p. 13.)

Zapechelnyuk, Andriy (2023), “On the Equivalence of Information Design by Uninformed and Informed Principals”, *Economic Theory*, 76, 4 (Nov. 2023), pp. 1051-1067. (pp. 3, 9.)

Zhang, Kun (2022), “Withholding Verifiable Information”, working paper. (p. 3.)