Preliminary Draft: please do not cite or circulate.

Coordination Mechanisms with Partially Specified Probabilities

Francesco Giordano*

June 22, 2025

Abstract

People make decisions based on the data they observe. However, the nature of the true data-generating process is often only partially known: we model such partial knowledge as a set of moment conditions. Given the partial information available, we consider an heuristic model of belief formation derived from the maximization of the Shannon entropy. This paper characterizes the outcomes that can be implemented through partial information disclosure both under canonical and generic information structures.

*HEC Paris, Economics and Decision Sciences, PhD Student - francesco.giordano@hec.edu

1 Introduction

This paper investigates some implications of information misspecification in games. Often, people make decisions based on observed data, while they only have partial knowledge on how these data are produced. Partial disclosure of the mechanisms underlying data generation is a pervasive and well-recognized across many economic and organizational contexts: in entrepreneurial finance, for instance, venture capital funds may withhold key information about one portfolio company from others. Similarly, central banks routinely communicate to shape market expectations about future interest rates, yet the models, assumptions, and internal forecasts informing such communication are rarely fully disclosed. In corporate governance, top management teams often rely on board discussions informed by summary statistics of financial accounts. While executive summaries provide a partial reflection of the firm's financial state, they omit the full complexity of how such data is produced, offering only a selective view of underlying processes. Moreover, formal non-disclosure agreements frequently institutionalize these limits by explicitly regulating and constraining the flow of information. These examples emphasize the broad implications of partial information disclosure and raise a central question: what economic outcomes can arise when data is available, but disclosure rules limit the extent to which the data-generating process is revealed?

This paper models an idealized environment that captures the information misspecification as a design problem, where a hypothetical information provider generates data according to a true data-generating process – specified by a set of messages and a probability distribution over them – and an information structure, represented by a disclosure policy that reveals only partial information about how the data is produced. Before taking any action, each decision maker receives some data, which may be correlated across decision makers. However, the joint distribution of these data is only partially revealed by the disclosure policy, leaving ambiguity about the true underlying process. Data are not payoff-relevant per se; therefore, faced with a set of plausible data-generating processes consistent with the disclosed information, decision makers use a heuristic criterion to identify a belief that aligns with their limited knowledge of the true process. The interaction between available data and the induced beliefs gives rise to endogenous patterns of correlated behavior. The main body of the paper develops a formal framework for analyzing such partially specified data-generating processes, focusing on a specific yet sufficiently general class of disclosure rules, and characterizes the set of implementable outcomes under both canonical and arbitrary information structures.

1.1 Outline of the paper

After some clarifying examples and a review of the main literature, Section 2 introduces the primitives of the model, discussing the information environment, partially specified data-generating processes, and players' belief formation. Section 3 presents the relevant notion of implementation and contains the main results. Section 4 concludes. All the proofs and some additional examples are relegated to the Appendix.

1.2 Examples

Example 1. Correlation neglect.

Correlation neglect is a well established bevahioral anomaly (Enke and Zimmermann (2019), Levy et al. (2022)). Consider a simplified setup where two agents play a coordination game and payoffs are described by

	b_1	b_2
a_1	$2,1$	$0,0$
a_2	$0,0$	$1,2$

Figure 1: A coordination game

The implementation of an outcome, described more precisely in Section 3, occurs through the provision of data and partial disclosure on the data generating process. In this example, if players exhibit correlation neglect, a randomization on the anti-diagonal can be implemented as an outcome of the coordination game, despite being far from any best response condition. Consider indeed a data generating process corresponding to table (a) and a disclosure policy corresponding to table (b), where only the marginals are known.

	m_1'	m_2'
m_1	0	2/3
m_2	1/3	0

(a) True DGP

	1/3	2/3
2/3	?	?
1/3	?	?

(b) Information disclosed

	1/3	2/3
2/3	2/9	4/9
1/3	1/9	2/9

(c) Belief on the DGP

Figure 2: Information disclosure and belief formation

To accommodate correlation neglect, suppose that agents use maximum entropy to form a belief on the data generating process, given the information available. The joint distribution that maximizes the Shannon entropy given two fixed marginals is indeed the product of the marginals (see Cover (1999), pag 421). Therefore, the belief on the data generating process is the probability distribution on the right table. Consider a strategy for player 1 that maps message m_1 to action a_1 and message m_2 to action a_2; similarly, the strategy of player 2 maps m_1' to action b_1 and m_2' to b_2. Given the belief on the data generating process, playing such strategies is a best response; indeed, the message distribution, when messages are interpreted as action recommendations, constitutes a Nash equilibrium of the game. We conclude that the outcome that randomizes over the action profiles according to the data generating process can be implemented.

Example 2. Generic information structures and naive belief formation.

Consider a matching pennies game with the following payoffs.

	b_1	b_2
a_1	$2, -2$	$0\ ,\ 0$
a_2	$0\ ,\ 0$	$1, -1$

(a) Game payoffs

	b_1	b_2
a_1	1/9	2/9
a_2	2/9	4/9

(b) Correlated equilibrium

Figure 3: Matching pennies.

The game has a unique correlated equilibrium and, as argued in detail in Section 3, an outcome $\mu \in \Delta(\{a_1, a_2\} \times \{b_1, b_2\})$ is implementable if

$$2\mu_{a_1,b_1} + \mu_{a_1,b_2} + \mu_{a_2,b_1} = \frac{2}{3}.$$

The condition is required to guarantee the distortion of a belief from a true data generating process to the unique correlated equilibrium. However, more outcomes are implementable with larger messages set. For example, consider a degenerate outcome such that $\mu_{a_1,b_1} = 1$. Consider the message set on the table (a), associated to the corresponding strategy profile: player 1 maps message m_1 to a_1 and messages $\{m_2, m_3\}$ to a_2; symmetrically for player 2.

	m_1'	m_2'	m_3'
m_1	a_1, b_1	a_1, b_2	a_1, b_2
m_2	a_2, b_1	a_2, b_2	a_2, b_3
m_3	a_2, b_1	a_2, b_2	a_2, b_2

(a) Map messages to actions

	m_1'	m_2'	m_3'
m_1	1/9	1/9	1/9
m_2	1/9	1/9	1/9
m_3	1/9	1/9	1/9

(b) Target belief

Figure 4: Mapping messages to actions and corresponding target belief.

As the maximum entropy distribution with no constraints in the uniform (see Cover (1999), pag 29), the target belief on the left arises if no information on a data generating process is provided. Given such belief, the strategies of players are best responses. A data generating process that always draws m_1, m_1' implements the target outcome.

Example 3. Pareto improvement and correlation neglect.

Consider the game defined by the following payoffs.

	b_1	b_2	b_3
a_1	$3,3$	$4,0$	$0,4$
a_2	$0,4$	$3,3$	$4,0$
a_3	$4,0$	$0,4$	$3,3$

Figure 5: Pareto improvement through correlation neglect

The game has a unique correlated equilibrium, where each action profile is played with equal probability. However, through a data generating process that couples the mixed Nash equilibrium strategies and discloses only the marginals, a randomization on the main diagonal is implementable. Such randomization achieves a payoff for both players that is higher than the correlated equilibrium payoffs.

Example 4. The Chicken-Game and naive belief formation.

Consider the following Chicken game (table (a)), where the correlated equilibrium that maximizes social welfare is shown on the table (b).

	b_1	b_2
a_1	$5,5$	$2,7$
a_2	$7,2$	$0,0$

(a) Chicken game

	b_1	b_2
a_1	$1/3$	$1/3$
a_2	$1/3$	0

(b) Optimal CE

Figure 6: Optimal correlation device in a Chicken game

Consider the data generating process represented on table (a) and the disclosure policy of table (b). Consistent with the information available, the distribution that maximizes the Shannon entropy is represented on table (c).

	m_1'	m_2'
m_1	1	0
m_2	0	0

(a) True DGP

	m_1'	m_2'
m_1	?	?
m_2	?	0

(b) Information disclosed

	m_1'	m_2'
m_1	$1/3$	$1/3$
m_2	$1/3$	0

(c) Belief on the DGP

Figure 7: Information disclosure and belief formation

Notice that the common belief constitutes a correlated equilibrium of the game. Therefore, the degenerate outcome (a_1, b_1) is implementable, guaranteeing for both player an higher payoff than the welfare maximizing correlated equilibrium.

1.3 Literature

This paper complements a longstanding literature on games with misspecified information environments by connecting the classical studies on coordination mechanisms and correlation devices with more recent research on games and decisions involving partially specified probabilities, incomplete data access, and model misspecification. The starting point lies in the literature on subjectivity and correlation in games, beginning with the foundational works of Aumann (1974) and Aumann (1987), with a recent comprehensive review provided by Forges and Ray (2024). In particular, a mechanism design perspective is introduced in the seminal work of Myerson (1982), who develops a generalized principal-agent model in which a principal can coordinate players' actions through coordination devices;

when players have a single type, the principal's decision problem reduces to choosing a canonical correlated equilibrium. Our work integrates this classical framework with more recent advances concerning partial specification of underlying data-generating processes, as proposed by Lehrer (2012) and Spiegler (2021).

Lehrer (2012) proposes a theory of decision making under partially specified probabilities and introduces an equilibrium concept in which players best respond to the worst-case strategies consistent with a partially specified distribution over others' actions. In that model, the mixed strategies of other players are only partially specified, and players evaluate their actions reflecting ambiguity aversion. In contrast, our model considers a similar framework of partially specified information but assumes a heuristic belief formation process that fills in the missing information. Players then respond optimally by maximizing their expected utility based on this partial information. This distinction is motivated by the fact that, in our setup, messages are not inherently payoff-relevant; rather, an equilibrium condition is defined given a belief over these messages. Furthermore, Lehrer and Solan (2007) proposes a notion of correlated equilibrium with partially specified probabilities, but this again differs from our approach because players are pessimistic. Relatedly, Koessler and Pahlke (2023) investigates the design of feedbacks under information misspecification, using the notion of maxmin self-confirming equilibrium (MSCE). The authors characterize MSCE in the presence of coarse feedback on equilibrium strategies; unlike Koessler and Pahlke (2023), our equilibrium notion is derived from correlated equilibrium conditions, allowing correlation through an external device and resolving ambiguity through an heuristic.

A key conceptual contribution of Spiegler (2021) lies in modeling players who have limited archival access to the information on the steady-state distribution of some endogenous and exogenous variables, and form beliefs by maximizing entropy subject to the available information. The use of maximum entropy (Shannon (1948)) as a rule for prior selection follows the foundational work of Jaynes (1968) and Jaynes (1982). Some basic results are derived from Cover (1999). While maximum entropy methods are have been proposed to select priors in Bayesian models, they have well-known limitations and criticisms – see, for example, Kass and Wasserman (1996); we neither normatively endorse nor reject the method but consider it a relevant case study given its prominence in the literature.

A growing body of work on information misspecification is related to our analysis, particularly in the applications to persuasion. For instance, Spiegler et al. (2021) models strategic communication where a sender provide a receiver with some interpretations of messages, while Eliaz et al. (2021) study bounded rationality in persuasion through the pooling of messages, that in turn affects posterior beliefs and persuasion probabilities. Relatedly, Epstein and Halevy (2024) examine decision-making under signals that have several possible interpretations; also, several studies highlight the effects of correlation neglect in belief formation (Epstein and Halevy (2019), Enke and Zimmermann (2019)) and in persuasion contexts (Levy et al. (2022)). Studies of mechanism design involving ambiguity have been proposed by Bose and Renou (2014) and Dütting et al. (2023). One application that we propose focuses on a Bayesian persuasion example in the spirit of Kamenica and Gentzkow (2011), but with partially specified information policies where messages, while not directly payoff-relevant, convey informative signals about the state. Unlike models that emphasize receiver ambiguity aversion (e.g., Hedlund et al. (2021); Beauchêne et al. (2019); Cheng et al. (2024b) and Cheng et al. (2024a)), we consider generic message structures and heuristic belief formation under partial specification.

2 Information Environment

We consider a simultaneous move finite game $\Gamma = (\mathrm{N}, \mathrm{A}, \mathrm{u})$ extended with a generic finite set of messages M. The set of players, denoted by N, is finite and has generic element $i \in \mathrm{N}$. The set of action profiles is a finite set $\mathrm{A} = \times_i \mathrm{A}_i$ with generic element $(a_1, ..., a_i, ..., a_N) \in \times_i \mathrm{A}_i$. The symbol A_i denotes the set of actions of player $i \in \mathrm{N}$. The set of utility functions is $\mathrm{u} = \{u_i\}_{i \in \mathrm{N}}$, where $u_i : \mathrm{A} \to \mathbb{R}$ is the utility function of player $i \in \mathrm{N}$. A correlated strategy is denoted by $\mu \in \Delta\mathrm{A}$. Throughout the analysis we assume that $\mathrm{M} = \times_i \mathrm{M}_i$, where M_i denotes the set of messages that can be received by player $i \in \mathrm{N}$. As standard, the set obtained by excluding the i component is $\mathrm{M}_{-i} = \{\mathrm{M}_1, ..., \mathrm{M}_{i-1}, \mathrm{M}_{i+1}, ..., \mathrm{M}_N\}$. Before taking an action, each player i observes some data, represented by an element of the message set M_i. The data itself is not directly payoff-relevant. Although the data received by players may be correlated, the joint distribution is only partially determined by a disclosure policy. Confronted with a set of plausible data-generating processes consistent with the available partial information, players employ the principle of maximum entropy as their belief formation rule.

2.1 Partially Specified Data Generating Processes

A partially specified data generating process is a couple consisting of data generating process and an information structure. Given a finite set M, a data generating process is a probability distribution $\eta \in \Delta\mathrm{M}$. An information structure $\mathcal{F} \subseteq \{f : \mathrm{M} \to \mathbb{R}\}$ is a collection of real valued random variables on M. Players do not know directly the data generating process, but are informed about the expectations of such random variables. An information structure represents the disclosure rule on the data generating process.

Definition 1 (Partially Specified Data Generating Process)**.** *A partially specified data generating process is a couple $D = (\eta, \mathcal{F})$, where $\eta \in \Delta M$ is a data generating process and $\mathcal{F} \subseteq \{f : M \to \mathbb{R}\}$ is an information structure.*

Fixed a data generating process $\eta \in \Delta\mathrm{M}$ and an information structure $\mathcal{F}$, all players know

$$\mathbb{E}_{m\sim\eta}\left[f_m\right], \quad \forall\, f \in \mathcal{F}.$$

Two partially specified data generating process $D = (\eta, \mathcal{F})$ and $D' = (\eta', \mathcal{F}')$ are informationally equivalent if they share the same information structure and induce the same moment constraints[1]. Revisiting the examples of above, in example 1 the set of messages is $\mathrm{M} = \{m_1m_1', m_1m_2', m_2m_1', m_2m_2'\}$, and the information structure is

$$\mathcal{F} = \left(\mathbb{1}_{\{m_1m_1',m_2m_1'\}}, \mathbb{1}_{\{m_1m_2',m_2m_2'\}}, \mathbb{1}_{\{m_1m_1',m_1m_2'\}}, \mathbb{1}_{\{m_2m_1',m_2m_2'\}}\right).$$

For example, the random variable $\mathbb{1}_{\{m_1m_1',m_2m_1\}}$ induces the explicit moment constraint

$$\sum_{m\in\mathrm{M}} \eta_m f(m) = \eta_{m_1m_1'} + \eta_{m_2m_1'} = \frac{1}{2}.$$

In the second part of example 2 the set of messages is $\mathrm{M} = \{m_1, m_2, m_3\} \times \{m_1', m_2', m_3'\}$ and the information structure is the empty set. In example 4, the set of signals is once again $\mathrm{M} = \{m_1m_1', m_1m_2', m_2m_1', m_2m_2'\}$, while for both players the information structure is a singleton consisting of

$$f(m) = \begin{cases} 1, & \text{if } m = (m_2m_2') \\ 0, & \text{otherwise.} \end{cases}$$

The explicit moment constraint is

$$\eta_{m_2m_2'} = 0.$$

Notice that a data generating process is fully revealed if the set $\mathcal{F}$ is large enough. Namely, if the moment constraints allow to perfectly identify the probability distribution $\eta \in \Delta\mathrm{M}$. As an example, a fully revealing information structure is $\mathcal{F} = \{\mathbb{1}_{\{m\}},\ m \in \mathrm{M}\}$, the set of all the indicator functions.

Despite being an unconventional way of modeling and reproducing information, the formulation is not new and a general discussion is presented in Lehrer (2012). The notion of partially specified data generating process is indeed derived by the notion of partially specified probabilities used in the decision model of Lehrer (2012). The change of wording is due to the fact that, in the current setup, the information structure does not convey information on other players' strategies, but on generic data that are not directly payoff relevant yet correlate players' behavior. As concluding remark of this section, we point out that in the above examples all the partial information disclosed is public: all players have access to the same partial information according to a common information structure.

2.2 Belief formation

Through a partially specified data generating process, players have partial knowledge about how data is produced. Given such partial knowledge available, each player uses a heuristic rule to form a belief about the data generating process. Let $\mathcal{D}$ be the set of all partially specified data generating processes, a belief formation rule

$$q : \mathcal{D} \to \Delta(\mathrm{M})$$

[1]Namely, we D is informationally equivalent to D' if $\mathcal{F} = \mathcal{F}'$ and

$$\sum_m \eta_m f(m) = \sum_m \eta_m' f(m) \quad \forall\, f \in \mathcal{F}.$$

is a function that to partially specified data generating process associates a belief on the message set. We consider a heuristic process of belief formation that follows the maximization of the Shannon entropy under the moment constraints implied by the partially specified data generating process. Under data generating process $D = (\eta, \mathcal{F})$, the information known by players is defined by the moment constraints, and the set of plausible data generating processes is

$$\Delta_D = \left\{ q \in \Delta\mathrm{M} : \quad \sum_m q_m f(m) = \sum_m \eta_m f(m), \quad \forall\, f \in \mathcal{F} \right\}.$$

Players maximize a function $\mathcal{H} : \Delta(\mathrm{M}) \to \mathbb{R}$ over the constraint set, where $\mathcal{H}(q) = -\sum_m q_m \log q_m$. The belief formation rule $q : \mathcal{D} \to \Delta\mathrm{M}$ arises thus as the solution of

$$\begin{aligned} \max_{q \in \Delta\mathrm{M}} \quad & \mathcal{H}(q) \\ \text{subject to} \quad & \sum_m q_m f(m) = \sum_m \eta_m f(m), \quad \forall\, f \in \mathcal{F}. \end{aligned} \tag{B}$$

Although no restrictions are made on the set $\mathcal{F}$, in this setup the information structure can be assumed to be finite without loss of generality (see Lemma 1 in the appendix).

Despite being far from immune to criticism[2], the maximum entropy belief formation rule has several behavioral implications, above all including correlation neglect, and a naïve and conservative[3] attitude towards uncertainty, which may be justified by the fact that in our setup messages are not directly payoff relevant. An implicit behavioral consequence of this formulation is that, since the loss is steep at the boundary of the simplex, decision makers do not assign zero probability to any event unless this is explicitly required by the moment constraints. This condition can be interpreted as a "grain of truth" assumption in players' beliefs on the true data generating process. A second implicit assumption is that beliefs accurately reflect the information available to decision makers. Finally, beliefs remain invariant across two informationally equivalent partially specified data-generating processes. Also, analytically, the maximum entropy rule exhibits several well-known properties. In particular, if the marginals of a joint distribution are known, the maximum entropy distribution is the product of the marginals - thereby illustrating correlation neglect; furthermore, if only the support is known, the maximum entropy distribution is the uniform (see Theorem 2.6.4 of Cover (1999)).

3 Implementation

Consider a generic message set $\mathrm{M} = \times_i \mathrm{M}_i$ and a profile of strategies $\sigma = \{\sigma_i\}_{i \in \mathrm{N}}$ consisting of a map $\sigma_i : \mathrm{M}_i \to \Delta\mathrm{A}_i$ for any $i \in \mathrm{N}$. Conditioning on the realization of a message profile $m = (m_1, ..., m_n) \in \mathrm{M}$, players joint play is the probability distribution $\sigma(m) \in \times_i \Delta\mathrm{A}_i$ defined as

$$\sigma(a_1, ..., a_n | m) = \prod_{i \in \mathrm{N}} \sigma_i(a_i | m_i), \quad \forall\, (a_1, ..., a_n) \in \mathrm{A}.$$

Outcome $\mu \in \Delta\mathrm{A}$ is derived from a data generating process $\eta \in \Delta\mathrm{M}$ and a profile of strategies σ through the pushforward measure

$$\mu = \eta \circ \sigma$$

defined as

$$\mu(a_1, ..., a_n) = \sum_{m \in \mathrm{M}} \eta_m \sigma(a_1, ..., a_n | m), \quad \forall\, (a_1, ..., a_n) \in \mathrm{A}.$$

The following definition states the notion of implementation with endogenous the belief formation. For clarity, we denote an information structure $\mathcal{F}$ defined on a message set M by $\mathcal{F}^{\mathrm{M}}$. Let the symbol η_{m_i} denotes the marginal probability of η on message m_i.

[2] See, for example, Kass and Wasserman (1996) for a discussion.

[3] As claimed in Jaynes (1968), the maximum entropy distribution *"[...] is the one which is, in a certain sense, spread out as uniformly as possible without contradicting the given information, i.e., it gives free rein to all possible variability allowed by the constraints. Thus it accomplishes, in at least one sense, the intuitive purpose of assigning a prior distribution; it agrees with what is known, but expresses a "maximum uncertainty" with respect to all other matters, and thus leaves a maximum possible freedom for our final decisions to be influenced by the subsequent sample data."*

Definition 2 (Implementation). *An outcome $\mu \in \Delta A$ is implemented by a partially specified data generating process $D = (\eta, \mathcal{F}^M)$ if there exists a profile of strategies $\sigma = \{\sigma_i\}_{i \in N}$ such that*

1. $q^* = \arg\max_{q \in \Delta_{\mathcal{D}}} \mathcal{H}(q)$
2. $\sum_{m_{-i} \in M_{-i}} q^*_{m_i, m_{-i}} \left[u_i(\sigma_i(m_i), \sigma_{-i}(m_{-i})) - u_i(a'_i, \sigma_{-i}(m_{-i}))\right] \geq 0, \quad \forall\, i \in N,\ a'_i \in A_i,$ *and* m_i *with* $q^*_{m_i} > 0$
3. $\mu = \eta \circ \sigma$

The first condition requires that belief formation follows from the maximization of entropy, subject to the information available to each player. The second condition requires that each player adopts a standard expected utility best response[4] with respect to her belief and the strategies of other players[5]. The true data-generating process $\eta \in \Delta \mathrm{M}$ and the belief $q^* \in \Delta \mathrm{M}$ may have different support: the notion of implementation requires that the incentive compatibility condition is satisfied along the belief path. In other words, the players' best response condition must hold whenever a message m_i is assigned positive probability by player i's belief. The third condition requires that the composition of message probabilities and players' strategies yields the intended target outcome. A key notion behind the following characterization is the one of correlated equilibrium, that is reviewed in the appendix.

Proposition 1. *Let Γ admit a rational correlated equilibrium $p \in \Delta A$. Then there exists a finite set of messages M and a partially specified data generating process $D = (\eta, \mathcal{F})$ that implements any outcome μ such that*

$$\mathrm{supp}(\mu) \subseteq \mathrm{supp}(p).$$

The proof is contained in the appendix.

The idea of the proof is the following: we construct an information structure through an auxiliary binary hypercube of arbitrary but finite dimension; this construction allows to identify a suitable message set and an information structure that replicates the incentive compatibility conditions. The information structure is simple in the sense that it relies solely on indicator random variables.

An outcome is directly implementable if $\mathrm{M} = \mathrm{A}$ and $\sigma_i : \mathrm{A}_i \to \Delta \mathrm{A}_i$ is obedient, namely $\sigma_i(a_i) = \mathbb{1}_{a_i}$ for each $i \in \mathrm{N}$. The following proposition characterizes the set of outcome implementable directly.

Proposition 2. *The following are equivalent:*

- *$\mu \in \Delta A$ is directly implementable*
- *There exists a correlated equilibrium distribution $q \in \Delta A$ such that* $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(q)$ *and*

$$\mathcal{KL}(\mu || q) = \mathcal{H}(q) - \mathcal{H}(\mu), \tag{BD}$$

where $\mathcal{H}(\cdot)$ denotes the Shannon entropy and $\mathcal{KL}(\cdot||\cdot)$ the Kullback-Leibler divergence.

The proof is contained in the appendix.

The idea of the proof is the following: the incentive compatibility condition in the canonical setting requires that the recommended action constitutes a best response to the player's belief about the behavior of their opponents; in other words, the requirement corresponds to a correlated equilibrium with respect to these beliefs. The interpretation of condition (BD) pertains to the distortion of beliefs: it provides a necessary and sufficient condition for inducing belief distortion through the selection of an appropriate information structure. The distortion of beliefs is discussed in greater detail in the appendix.

The next simplified examples provide an idea on how the information structure of Proposition 1 is constructed to induce an incentive compatible belief.

Example 1. Information structure and incentive compatibility.

Given a correlated distribution over action profiles –interpreted as a target belief that satisfies incentive compatibility – the objective is to construct a message set, a belief over this message set, and a strategy profile that together

[4] In particular, a profile $\sigma : M \to \Delta \mathrm{A}$ that is constant over messages and constitutes a Nash equilibrium of the game automatically satisfies the incentive compatibility conditions.

[5] Notice that the implementation condition requires an equilibrium where each player knows joint strategy profile, while possessing only partial knowledge of other players' information.

induce the target belief. The information structure consists of a collection of indicator random variables defined over the message set, indexed by messages to which the belief assigns probability zero.

	b_1	b_2
a_1	1/4	1/4
a_2	2/4	0

(a) Target belief

	m'_1	m'_2
m_1	1/4	1/4
m_2	1/4	0
m_3	1/4	0

(b) Message and belief

Figure 8: Incentive compatibility, messages and beliefs

The strategy of player 1 maps message m_1 to a_1, and messages $\{m_2, m_3\}$ to a_2. The strategy of player 2 maps m'_1 to b_1 and m'_2 to b_2.

	b_1	b_2	b_3
a_1	2/5	1/5	0
a_2	1/5	0	1/5

(a) Target belief

	m'_1	m'_2	m'_3	m'_4	m'_5	m'_6
m_1	1/10	1/10	1/10	0	0	0
m_2	1/10	1/10	0	1/10	0	0
m_3	1/10	0	0	0	1/10	0
m_4	0	1/10	0	0	0	1/10

(b) Message and belief

Figure 9: Incentive compatibility, messages and beliefs

The strategy of player 1 maps messages $\{m_1, m_2\}$ to a_1, and messages $\{m_3, m_4\}$ to a_2. The strategy of player 2 maps $\{m'_1, m'_2\}$ to b_1, $\{m'_3, m'_4\}$ to b_2, and $\{m'_5, m'_6\}$ to b_3.

	b_1	b_2
a_1	2/5	0
a_2	0	3/5

(a) Target belief

	m'_1	m'_2	m'_3	m'_4	m'_5
m_1	1/5	0	0	0	0
m_2	0	1/5	0	0	0
m_3	0	0	1/5	0	0
m_4	0	0	0	1/5	0
m_5	0	0	0	0	1/5

(b) Message and belief

Figure 10: Incentive compatibility, messages and beliefs

The strategy[6] of player 1 maps messages $\{m_1, m_2\}$ to a_1, and messages $\{m_3, m_4, m_5\}$ to a_2. The strategy of player 2 maps $\{m'_1, m'_2\}$ to b_1, $\{m'_3, m'_4, m'_5\}$ to b_2.

3.1 An application to persuasion

We now consider an application to a classical persuasion problem involving a Sender and a Receiver. The Sender has access to a set of messages and can choose a partially specified data-generating process. To align with the analytical techniques employed in Section 2, the Sender is permitted to communicate a bundle of messages to the Receiver: the economic interpretation is that while information may be produced in a granular manner, the Receiver interprets this information at the level of concepts or categories, collecting messages that convey similar meanings. Informally, once a message is realized, the Receiver may observe a bundle of messages that contains the realized one. This formulation is aligned with other models appeared in the literature, such as Spiegler et al. (2021) and Eliaz et al. (2021). As in a standard persuasion problem, there is a prior distribution over the state space. However,

[6] Notice that the incentive compatibility requirements is about the conditional distributions.

in contrast to the standard setting, the conditional probabilities over messages may be only partially specified. In a canonical information structure, the set of messages coincides with the set of actions available to the Receiver. Before providing a precise formalization, we present several illustrative examples to clarify the framework.

Example 2. Canonical information structure.

Consider a binary set of states $\Theta = \{\theta_1, \theta_2\}$ and actions $\mathrm{A} = \{c, a\}$. The underlying data generating process is $\eta \in \Delta(\Theta \times A)$, with the constraint that the marginal on the set of states is known and consistent with a prior $p = \text{proba}(\theta = \theta_1)$. A data generating process thus respects

$$\eta(\theta_1, a) + \eta(\theta_1, c) = p.$$

As in Kamenica and Gentzkow (2011), the Receiver is interpreted as a judge who must decide whether to convict (c) or acquit (a) a defendant. There are two possible states of nature: guilty and innocent. The Sender is interpreted as a prosecutor, who can send conditionally informative messages to the judge. The game is represented as follows, where the first entry denotes the payoff of the judge and the second entry denotes the payoff of the prosecutor.

	c	a
θ_1	$1, 1$	$0, 0$
θ_2	$0, 1$	$1, 0$

Figure 11: Payoff matrix in a simple persuasion example

It is well known that in the standard setup the optimal information provision of the prosecutor leads to the concavification of the Sender utility (Kamenica and Gentzkow (2011)). We show in this very simple example that it might not the case in our setup. The prosecutor utility as a function of the prior and the best response of the judge is

$$U(p) = \begin{cases} 1, & \text{if } p \geq \frac{1}{2} \\ 0, & \text{if } p < \frac{1}{2} \end{cases}.$$

Let the prior be $p = 1/3$ and suppose the judge use the belief formation rule described in Section 2. A canonical information structure is defined on $\mathrm{M} = \{\theta_1, \theta_2\} \times \{a, c\}$. The only constraint for the Sender is that the joint distribution $\eta \in \Delta(\Theta \times A)$ satisfies $\sum_{k=a,c} \eta(\theta_1, k) = 1/3$. Consider the data generating process described by table (a), while the disclosure policy is represented in table (b).

	c	a
θ_1	$1/3$	0
θ_2	$2/3$	0

(a) True DGP

	c	a
p	$1/3$	0
$1-p$	?	?

(b) Information disclosed

	c	a
p	$1/3$	0
$1-p$	$1/3$	$1/3$

(c) Belief on the DGP

The belief on the true data generating process is described on the table on the right: conditioning on the received data, the posterior of the Receiver is

$$q^*(\theta_1|c) = \frac{1}{2}.$$

The optimal revelation under partially specified information can be generalized. The strategy of the prosecutor is the following: the judge is always recommended to convict when the defendant is innocent, but the true (conditional) probabilities are not disclosed. On the other hand, the conditional probability $\eta(c|\theta_1)$ is chosen in order to guarantee incentive compatibility. In the setup of Section 2 and $p \geq 1/3$, the optimal data generating process is

	c	a
p	p	0
$1-p$	$1-p$	0

Figure 13: Optimal data generating process

while the information structure is $\mathcal{F} = \{\mathbb{1}_{\{\theta_1 a, \theta_1 c\}}, \mathbb{1}_{\{\theta_1 c\}}\}$. On the other hand, if $p < 1/3$ the prosecutor cannot do any better than the concavification. The above partially specified data generating process achieves an indirect utility that lies above the concavification for some intermediate beliefs.

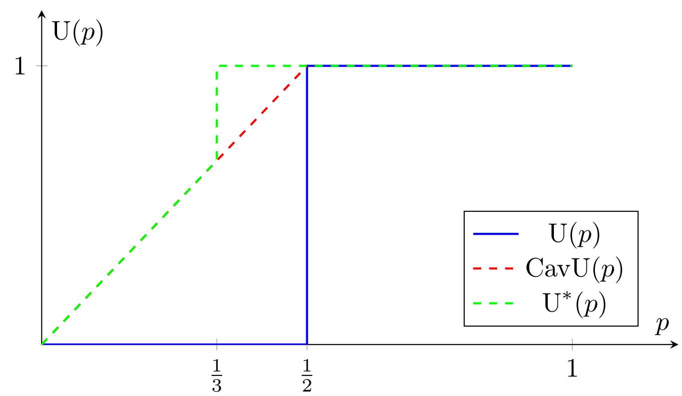

Figure 14: Optimal persuasion with partially specified DGPs

Example 3. Generic information structure.

Consider a revised version of the previous example, in which the cost of a wrong conviction is higher than that of a wrong acquittal. The prior is $p = 1/3$ and the payoff matrix is

	c	a
θ_1	$1,1$	$0,0$
θ_2	$0,1$	$2,0$

Figure 15: Payoff matrix in a revisited persuasion example

The judge will convict if he has belief $p \geq \frac{2}{3}$. Consider the target incentive compatible belief of table (a), and the message set and belief of the table (b)

	c	a
p	$2/9$	$1/9$
$1-p$	$1/9$	$5/9$

(a) Target incentive compatible belief

	m_1	m_2	m_3	m_4	m_5	m_6	m_7
p	$1/9$	$1/9$	$1/9$	0	0	0	0
$1-p$	$1/9$	0	$1/9$	$1/9$	$1/9$	$1/9$	$1/9$

(b) Message and belief

Figure 16: Target incentive compatible beliefs.

The data generating process is $\eta(m_1|\theta_1) = \eta(m_1|\theta_2) = 1$, and the information structure is

$$\mathcal{F} = \{\mathbb{1}_{m_2,\theta_2}, \mathbb{1}_{m_4,\theta_1}, \mathbb{1}_{m_5,\theta_1}, \mathbb{1}_{m_6,\theta_1}, \mathbb{1}_{m_7,\theta_1}\}.$$

While the induced posterior belief from message m_1 is $p(\theta_1|m_1) = 1/2$, revealing a bundle of messages will replicate the incentive compatibility condition. For example, conditioning the bundle $\{m_1, m_2\}$ – that is truthful since it contains m_1, induces posterior

$$p(\theta_1|\{m_1, m_2\}) = \frac{p(\theta_1, m_1) + p(\theta_1, m_2)}{p(m_1) + p(m_2)} = \frac{2}{3}.$$

Formally, a data generating process is $\eta \in \Delta(\Theta \times \mathrm{M})$. The Sender chooses $\eta : \Theta \to \Delta\mathrm{M}$, an information structure $\mathcal{F} \subset \{f : \Theta \times \mathrm{M} \to \mathbb{R}\}$, and discloses either the realized message $m \in \mathrm{M}$ or a bundle $\mathbf{m} \subseteq \mathrm{M}$ such that $m \in \mathbf{m}$. As in Section 2, the Receiver knows

$$\sum_\theta p(\theta) \sum_m \eta(m|\theta) f(\theta, m) \quad \forall\, f \in \mathcal{F}.$$

Consider a persuasion problem defined as $\Gamma = (\Theta, p, u, \mathrm{A})$; as standard, $a \in BR(p_a)$ if $a = \arg\max_A \sum_{a' \in A} p_a(\theta) u(a', \theta)$. To exclude some pathological conditions, the following mild assumption is stated.

Assumption 1. $\exists\, \{p_a\}_{a \in A}$ *such that*

1. $p_a \in int\Delta(\Theta)$ *for each* $a \in A$
2. $p \in co\left(\{p_a\}_{a \in A}\right)$
3. $a \in BR(p_a)$ *for each* $a \in A$

Moreover, all probabilities are rational.

The assumption states that there exists a rational set of posteriors p_a that justifies each action; therefore, it is possible to build a rational joint probability[7] on $\Theta \times \mathrm{A}$ where the conditional probabilities are indeed $\{p_a\}_{a \in A}$.

Proposition 3. *Suppose Assumption 1 holds. Then any outcome* $\mu : \Theta \to \Delta A$ *is implementable by a finite message set* M *and a partially specified data generating process* $D = (\eta, \mathcal{F})$.

The proof is contained in the appendix and it is a specific case of the proof of Proposition 1.

4 Concluding remarks

This paper studies novel implications of information misspecification in games. Under a constrained information framework (e.g., a canonical framework where messages are actions recommendations and strategies are obedient), the implementability of outcomes hinges critically on the extent of belief distortion induced by partial specification. In contrast, when considering more generic message structures, nearly all outcomes that constitute best responses to some correlated belief can be implemented, demonstrating a much broader scope for strategic information design under partial specification. While our analysis focuses on a maximum entropy heuristic and a particular class of disclosure policies, further research is required that extends to alternative heuristics that may better capture relevant behavioral anomalies.

References

Robert J Aumann. Subjectivity and correlation in randomized strategies. *Journal of mathematical Economics*, 1 (1):67–96, 1974.

[7] Indeed, there are indeed rational coefficients $x \in \Delta\mathrm{A}$ such that

$$\sum_a p_a x_a = p.$$

One can thus define

$$p(\theta, a) = p_a(\theta) x_a.$$

Robert J Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica*, pages 1–18, 1987.

Dorian Beauchêne, Jian Li, and Ming Li. Ambiguous persuasion. *Journal of Economic Theory*, 179:312–365, 2019.

Subir Bose and Ludovic Renou. Mechanism design with ambiguous communication devices. *Econometrica*, 82(5): 1853–1872, 2014.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Xiaoyu Cheng, Peter Klibanoff, Sujoy Mukerji, and Ludovic Renou. Persuasion with ambiguous communication. *Working Paper*, 2024a.

Xiaoyu Cheng, Peter Klibanoff, Sujoy Mukerji, and Ludovic Renou. Persuasion with ambiguous communication. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 1201–1202, 2024b.

Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

Paul Dütting, Michal Feldman, Daniel Peretz, and Larry Samuelson. Ambiguous contracts. *arXiv preprint arXiv:2302.07621*, 2023.

Kfir Eliaz, Ran Spiegler, and Heidi C Thysen. Persuasion with endogenous misspecified beliefs. *European Economic Review*, 134:103712, 2021.

Benjamin Enke and Florian Zimmermann. Correlation neglect in belief formation. *The review of economic studies*, 86(1):313–332, 2019.

Larry G Epstein and Yoram Halevy. Ambiguous correlation. *The Review of Economic Studies*, 86(2):668–693, 2019.

Larry G Epstein and Yoram Halevy. Hard-to-interpret signals. *Journal of the European Economic Association*, 22 (1):393–427, 2024.

Françoise Forges and Indrajit Ray. "subjectivity and correlation in randomized strategies": Back to the roots. *Journal of Mathematical Economics*, 114:103044, 2024.

Jonas Hedlund, T Florian Kauffeldt, and Malte Lammert. Persuasion under ambiguity. *Theory and Decision*, 90 (3):455–482, 2021.

Edwin T Jaynes. Prior probabilities. *IEEE Transactions on systems science and cybernetics*, 4(3):227–241, 1968.

Edwin T Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, 1982.

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

Robert E Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American statistical Association*, 91(435):1343–1370, 1996.

Frédéric Koessler and Marieke Pahlke. Feedback design in strategic-form games with ambiguity averse players. 2023.

Ehud Lehrer. Partially specified probabilities: decisions and games. *American Economic Journal: Microeconomics*, 4(1):70–100, 2012.

Ehud Lehrer and Eilon Solan. Learning to play partially-specified equilibrium. *mimeo*, 2007.

Gilat Levy, Inés Moreno de Barreda, and Ronny Razin. Persuasion with correlation neglect: a full manipulation result. *American Economic Review: Insights*, 4(1):123–138, 2022.

Roger B Myerson. Optimal coordination mechanisms in generalized principal–agent problems. *Journal of mathematical economics*, 10(1):67–81, 1982.

Herbert John Ryser. *Combinatorial mathematics*, volume 14. American Mathematical Soc., 1963.

Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3): 379–423, 1948.

R Spiegler, K Eliaz, and H Thysen. Strategic interpretations. *Journal of Economic Theory*, 192, 2021.

Ran Spiegler. Modeling players with random "data access". *Journal of Economic Theory*, 198:105374, 2021.

5 Appendix

5.1 Notation

- $\mathrm{M} = \times_i \mathrm{M}_i$: generic message set
- $\eta \in \Delta\mathrm{M}$: data generating process
- $\mathcal{F} \subseteq \{f : \mathrm{M} \to \mathbb{R}\}$: information structure
- $D = (\eta, \mathcal{F})$: partially specified data generating process
- $\mathcal{D}$: set of partially specified data generating processes
- $q : \mathcal{D} \to \Delta\mathrm{M}$: belief formation rule
- $\mu \in \Delta\mathrm{A}$: outcome
- $\mathcal{H} : \Delta\mathrm{M} \to \mathbb{R}$: Shannon entropy
- $\mathcal{KL} : \Delta\mathrm{M} \times \Delta\mathrm{M} \to \mathbb{R}$: Kullback-Leibler divergence

The definition of Shannon Entropy, Kullback-Leibler divergence and correlated equilibrium are reviewed.

Definition 3 (Shannon Entropy)**.** *The entropy of a random variable with probability mass function $q \in \Delta M$ is*

$$\mathcal{H}(q) = -\sum_m q_m \log q_m.$$

Definition 4 (Kullback-Leibler Divergence)**.** *Let $\eta \in \Delta M$ and $q \in \Delta M$ be discrete probability distributions. The Kullback–Leibler divergence from η to q is*

$$\mathcal{KL}(\eta || q) = \sum_m \eta_m \log \frac{\eta_m}{q_m}.$$

Notice that the Kullback-Leibler divergence is only defined if, for all $m \in \mathrm{M}$, $q_m = 0$ implies $\eta_m = 0$.

Definition 5 (Correlated Equilibrium)**.** *A probability distribution $q \in \Delta(A)$ is a correlated equilibrium of $\Gamma = (N, A, u)$ if*

$$\sum_{a_{-i}} q_{a_i, a_{-i}} \left[u_i(a_i, a_{-i}) - u_i(b, a_{-i})\right] \geq 0 \quad \forall\ i \in N,\ b \in A_i.$$

5.2 Belief distortion

An information provider sends messages to an information receiver from a finite set M using a probability distribution $\eta \in \Delta\mathrm{M}$. The information provider reveals only partial information about the data generating process: (s)he discloses information by revealing the expected value of some random variables. In other words, the information provider chooses a set of functions $\mathcal{F} \subseteq \{f : \mathrm{M} \to \mathbb{R}\}$ and the receiver knows $\mathbb{E}_\eta[f]$ for any $f \in \mathcal{F}$.

The couple $\mathcal{D} = (\eta, \mathcal{F})$ defines a set of plausible probability distributions

$$\Delta_{\mathcal{D}} = \left\{ q \in \Delta\mathrm{M} : \ \sum_m q_m f_m = \sum_m \eta_m f_m, \ \forall\, f \in \mathcal{F} \right\}.$$

The receiver, knowing that the true data generating process belongs to the set $\Delta_{\mathcal{D}}$, forms his belief according to the principle of maximum entropy, solving the following optimization problem

$$\max_{q \in \Delta_{\mathcal{D}}} -\sum_m q_m \log q_m.$$

We ask the following question: under which conditions the information provider can distort the receiver belief, from the true data generating process $\eta \in \Delta\mathrm{M}$ to a given target belief $q^* \in \Delta\mathrm{M}$?

Definition 6. *$\eta \in \Delta M$ can be distorted into $q^* \in \Delta M$ if there exists a set of functions $\mathcal{F} \subseteq \{f : M \to \mathbb{R}\}$ such that*

$$q^* = \arg\max_{\Delta_{\mathcal{D}}} \sum_m -q_m \log q_m.$$

We first show that the information structure is finite without loss of generality.

Lemma 1. *Let M be a finite set of messages. Then for any $\bar{\mathcal{F}} \subset \{f : M \to \mathbb{R}\}$ there exists a finite $\mathcal{F} \subset \{f : M \to \mathbb{R}\}$ such that*

$$q(\mathcal{F}, \eta) = q(\bar{\mathcal{F}}, \eta)$$

for any $\eta \in \Delta M$.

Proof. $\mathcal{D} = (\eta, \mathcal{F})$ and $\bar{\mathcal{D}} = (\eta, \bar{\mathcal{F}})$ define two linear system of equations, the constraints set in the optimization problem (B). It suffices to notice that, fixed M and $\eta \in \Delta$M, there are at most |M| linearly independent vectors. □

The following lemma provides necessary and sufficient conditions for belief distortion.

Lemma 2. *$\eta \in \Delta M$ can be distorted into $q \in \Delta M$, with $\operatorname{supp}(\eta) \subseteq \operatorname{supp}(q)$, if and only if*

$$\mathcal{KL}(\eta || q) = \mathcal{H}(q) - \mathcal{H}(\eta),$$

where $\mathcal{H}(\cdot)$ denotes the Shannon entropy and $\mathcal{KL}(\cdot||\cdot)$ denotes the Kullback-Leibler divergence.

First, notice that

$$\begin{aligned} &\mathcal{KL}(\eta||q) = \mathcal{H}(q) - \mathcal{H}(\eta) \\ \iff &\sum_m \eta_m \log \frac{\eta_m}{q_m} = \sum_m \eta_m \log \eta_m - \sum_m q_m \log q_m \\ \iff &\sum_m \eta_m \log q_m = \sum_m q_m \log q_m. \end{aligned} \tag{1}$$

Consider the optimization program that describes the belief formation process

$$\begin{aligned} \max_{q \in \Delta(\mathrm{M})} \quad & -\sum_m q_m \log(q_m) \\ \text{subject to} \quad & \sum_m f_m (q_m - \eta_m) = 0, \quad \forall\, f \in \mathcal{F}. \end{aligned} \tag{B}$$

The KKT system gives necessary and sufficient condition for the unique solution (Boyd and Vandenberghe (2004), pag. 244), since it is a maximization problem of a strictly concave function with linear constraints. Recall that the set $\mathcal{F}$ is finite, hence we index its generic element as $f^k : \mathrm{M} \to \mathbb{R}$. Recall that the Lagrangean of B is

$$\mathcal{L}(q) = \sum_m q_m \log(q_m) - \sum_k \lambda_k \sum_m f^k_m (q_m - \eta_m).$$

Proof (Belief Distorsion Lemma). One implication is an immediate consequences of the optimality conditions of B. The other implication is an explicit construction of the constraint set.

$\Longleftarrow$:
Suppose there exists a set $\mathcal{F}$ such that q^* is a solution of B. Then,

$$\begin{aligned} &\nabla \mathcal{L}(q^*) = 0 \\ &\iff \log(q^*_m) + 1 - \sum_k \lambda_k f^k_m = 0 \quad \forall\, m \in M \\ &\Longrightarrow \sum_m (q^*_m - \eta_m) \log(q^*_m) = 0, \end{aligned} \tag{2}$$

where the last equality comes from multiplying by $(q^*_m - \eta_m)$, summing over all $m \in$ M, and noticing that $\sum_m f^k_m (q^*_m - \eta_m) = 0$ for each $f^k \in \mathcal{F}$.

$\Longrightarrow$:
Suppose it holds $\sum_m \log(q^*_m)(\eta_m - q^*_m) = 0$. Consider a unique random variable $f : \mathrm{M} \to \mathbb{R}$ defined as

$$f_m = \log q^*_m + 1 \quad \forall\, m \in \mathrm{M}.$$

The KKT system is given by

$$\begin{aligned} &\log q_m + 1 = \lambda f_m \quad \forall\ m \in \mathrm{M} \\ &\sum_m \log(q_m)(q_m - \eta_m) = 0, \end{aligned} \tag{3}$$

and it is solved for $q = q^*$ and $\lambda = 1$. □

5.3 Proofs

Proof of Proposition 2. By definition, $\mu \in \Delta\mathrm{A}$ is implementable directly if it can be distorted into a common belief $q \in \Delta\mathrm{A}$ that respects the correlated equilibrium conditions. By lemma 2, the distortion is possible if and only if $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(q)$ and

$$\mathcal{KL}(\mu||q) = \mathcal{H}(q) - \mathcal{H}(\mu).$$

□

To prove Proposition 1, we need some preliminary lemmata. For two players, the following simple lemma is enough.

Lemma 3 (Matrices)**.** *Let $n, r \in \mathbb{N}$ with $n \geq r \geq 0$. Then there exists a binary matrix $C \in \{0,1\}^{n\times n}$ whose rows and columns sum up to r.*

Proof. Special case of Theorem 1.1, Chapter 6 of Ryser (1963). □

For an n players game, the following generalization is needed.

Lemma 4. *Let $d, n, r \in \mathbb{N}$ with $n \geq r \geq 0$. Then there exists a binary hypercube $C = (c_{i1,\dots,id})_{1\leq i1,\dots,id\leq n}$ such that*

$$\sum_{j=1}^{n} c_{i1,\dots,i(k-1),j,i(k+1),\dots,id} = r \tag{P}$$

for $1 \leq k \leq d$ and any $1 \leq i1, \dots, i(k-1), i(k+1), \dots, id \leq n$.

The proof is by induction and the idea is that, from an hypercube $\mathrm{C} \in \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{d \text{ times}}$ that respects P, we can construct an hypercube $\tilde{\mathrm{C}} \in \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{d+1 \text{ times}}$ that also respects P. The construction is the following: fixing one direction, we collect n hypercubes $\{\mathrm{C}_j \in \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{d \text{ times}}\}_{j=1,\dots,n}$ by shifting all the elements of C in the fixed direction. $\tilde{\mathrm{C}} \in \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{d+1 \text{ times}}$ is obtained by concatenating all such hypercubes.

Proof. We proceed in three steps:

1. Notice that with $d = 1$ one can consider the binary vector whose first r entries are one.
2. Let $\mathrm{C} = (c_{i1}, \dots, c_{i(d-1)})_{1\leq i1,\dots,i(d-1)\leq n}$ respect P and $\mathrm{C}^1 = (c_{i1}, \dots, c_{i(d-1)})_{1\leq i1,\dots,i(d-1)\leq n}$ be defined as[8]

$$c^1_{i1,\dots,i(d-1)} = c_{(i1-1),\dots,i(d-1)}.$$

   Then C^1 respects property P . Indeed, let $1 \leq k \leq d$ and $1 \leq i1, \dots, ik, \dots, i(d-1) \leq n$. Then,

$$\sum_{j=1}^{n} c^1_{i1,\dots,i(k-1),j,i(k+1),\dots,in} = \sum_{j=1}^{n} c_{(i1-1),\dots,i(k-1),j,i(k+1),\dots,in} = r.$$

[8]Where we substitute the first index $i1$ with the preceding index $i1 - 1$, and consider the addition modulo n, namely with $0 = n$.

Recursively, given $\mathrm{C} = (c_{i1}, ..., c_{i(d-1)})_{1 \le i1,...,i(d-1) \le n}$, define C^s as

$$c^s_{i1,...,i(d-1)} = c_{(i1-s),...,i(d-1)}.$$

3. If $\mathrm{C} = (c_{i1,...,id})_{1 \le i1,...,id \le n}$ respects property P, then $\tilde{\mathrm{C}} = (c_{i1,...,i(d+1)})_{1 \le i1,...,i(d+1) \le n}$ with[9]

$$\tilde{c}_{i1,...,i(d+1)} = c^{i(d+1)-1}_{i1,...,id},$$

respects property P with $d+1, n, r \in \mathbb{N}$. Indeed, let $1 \le k \le d+1$ and $1 \le i1, ..., i(k-1), i(k+1), ..., i(d+1) \le n$. We distinguish two cases:

(a) For $k = d+1$, we have

$$\begin{aligned}
\sum_{j=1}^{n} \tilde{c}_{i1,...,id,j} &= \sum_{j=1}^{n} c^{ij-1}_{i1,...,id} \\
&= \sum_{j=1}^{n} c_{i1-j+1,i2,...,id} \\
&= \sum_{j'=1}^{n} c_{j',i2,...,id} = r.
\end{aligned} \tag{4}$$

(b) For $k \in \{1, ..., d\}$,

$$\begin{aligned}
\sum_{j=1}^{n} \tilde{c}_{i1,...,j,...,d+1} &= \sum_{j=1}^{n} c^{i(d+1)-1}_{i1,...,j,...,id} \\
&= \sum_{j=1}^{n} c_{i1-i(d+1)-1,,...,j,...,id} = r.
\end{aligned} \tag{5}$$

□

Proof of Proposition 1. Let $p_a = \frac{k_a}{k}$, where k is the greatest common divisor. We construct each ingredient needed for the implementation of a target $\mu \in \Delta \mathrm{A}$ step by step.

1. **Message set:** Assign to each player i a finite message set M_i of dimension $k|\mathrm{A}_i|$, the total dimension of the message set is thus $|\mathrm{M}| = k^N |\mathrm{A}|$. The message set M is built as a collection of hypercubes: to each action profile $(a_1, ..., a_N) \in \mathrm{A}$, associate a k^N hypercube $\mathrm{M}^a = \times_i \mathrm{M}_i^{a_i}$ of messages, where $|\mathrm{M}_i^{a_i}| = k$.
2. **Strategy profile:** Define the strategy profile of player i as

$$\sigma_i(m_i) = a_i, \quad m_i \in \mathrm{M}_i^{a_i}. \tag{S}$$

3. **Information structure:** To build the information structure, for each M^a consider an auxiliary binary hypercube C^a of dimension k^N such that the sum over each dimension is k_a. By Lemma 4, such hypercube exists and can be built iteratively given $d = |\mathrm{N}|$, $n = k$ and $r = k_a$. The elements of M^a and C^a are related by a bijection. Generally, let $c(m)$ be the element of the binary hypercube corresponding to action profile $a = \sigma(m)$ and message m. The information structure is built as follows:

$$\mathbb{1}_{\{m\}} \in \mathcal{F} \iff c(m) = 0.$$

4. **Data generating process (i):** Consider a data generating process $\eta \in \Delta \mathrm{M}$ that assigns probability zero to messages corresponding to the zero elements of the auxiliary binary hypercube. Namely,

$$c = 0 \implies \eta(c^{-1}(m)) = 0. \tag{DGP}$$

[9]Namely, the sequence of the last dimension is the shifted sequence of the preceding dimension.

5. **Belief formation:** Given these constraints, the belief derived from maximum entropy is the uniform over its support, namely
$$q_m^* = \begin{cases} \frac{1}{\sum_a k^{n-1}k_a}, & \text{if } c(m) = 1, \\ 0 & \text{if } c(m) = 0 \end{cases}.$$
Furthermore, by the construction of each binary hypercube C^a, the belief assigned to each block M^a is the ratio of the number of ones in C^a over the total number of ones,
$$q^*(M^a) = \frac{k^{N-1}k_a}{\sum_a k^{N-1}k_a} = \frac{k_a}{k} = p_a.$$

6. **Incentive compatibility:** Given the strategy profiles as in (S), one has
$$\begin{aligned} q^*(\sigma_{-i}(m_{-i})|m_i) &= \frac{k_a}{\sum_{a:a_i \in a} k_a} \\ &= \frac{k_a}{k} \frac{k}{\sum_{a:a_i \in a} k_a} \\ &= \frac{p_a}{p_{a_i}} = p(a_{-i}|a_i). \end{aligned} \tag{6}$$
Therefore, σ_i as in (S) is incentive compatible since $p \in \Delta\mathrm{A}$ is a correlated equilibrium.

7. **Data generating process (ii) and Implementation:** Notice that all data generating processes $\eta \in \Delta\mathrm{M}$ that respects condition DGP are informationally equivalent and induce the same incentive compatible belief. For a target outcome $\mu \in \Delta\mathrm{A}$ such that $\mathrm{supp}(\mu) \subseteq \mathrm{supp}(p)$, we have
$$\begin{aligned} \mu_a > 0 &\implies p_a > 0 \\ &\implies \frac{k_a}{k} > 0 \\ &\implies \exists\, m \in \mathrm{M} \, : c(m) = 1. \end{aligned} \tag{7}$$
For such $m \in \mathrm{M}$, choose $\eta_m = \mu_a$.

□

Proof of Proposition 3. We replicate the steps of the proof Proposition 1. Consider the target $p \in \Delta(\Theta \times \mathrm{A})$, which can be expressed as
$$p(a, \theta) = \frac{k_{a,\theta}}{k}.$$
Notice that $p(\theta|a) = \frac{k_{a,\theta}}{\sum_\theta k_{a,\theta}}$. Consider a set of messages M with size size $k|\mathrm{A}|$. The data generating process is a stochastic matrix of dimension $|\Theta| \times k|\mathrm{A}|$ constructed as follows: each row is composed by $|\mathrm{A}|$ blocks of messages of size k. Each of such blocks, indexed by a generic element $a \in \mathrm{A}$, is denoted by $\mathbf{m}_a$. To build the information structure, consider the auxiliary matrix C of size $|\Theta| \times k|\mathrm{A}|$ constructed as follows: for every row $\theta \in \Theta$, consider the horizontal concatenation of $|\mathrm{A}|$ vectors of dimension k. Each of these vectors is indexed by a couple $(\theta, a) \in \Theta \times \mathrm{A}$ and has the first $k_{a,\theta}$ entries equal to one and the remaining entries equal to zero[10].
There is a bijection between couples $(\theta, m) \in \Theta \times \mathrm{M}$ and elements of C: denote by $c(\theta, m) \in \mathrm{C}$ the value of the element of C corresponding to indices (θ, m). The information structure is build as follows:
$$\mathbb{1}_{\{\theta,m\}} \in \mathcal{F} \iff c_{\theta,m} = 0.$$
The constraint on the data generating process is that, if $c(\theta, m) = 0$, then $\eta(\theta, m) = 0$. The belief derived from maximum entropy is thus
$$q_{\theta,m}^* = \begin{cases} \frac{1}{k}, & \text{if } c(\theta, m) = 1, \\ 0 & \text{if } c(\theta, m) = 0 \end{cases}.$$

[10]The auxiliary matrix has in total $\sum_{a,\theta} k_{a,\theta} = k$ ones

A message m is drawn from $\eta : \Theta \to \Delta \mathrm{M}$ and the Receiver in informed about corresponding block $\mathbf{m}_a$. Hence, the posterior is

$$p^*(\theta|\mathbf{m}_a) = \frac{k_{a,\theta}}{k} \frac{k}{\sum_\theta k_{a,\theta}} = \frac{k_{a,\theta}}{\sum_\theta k_{a,\theta}}$$

By construction, the best response to belief $p^*(\theta|\mathbf{m}_a)$ is action a.

To construct the data generating process, recall that any $\eta : \Theta \to \Delta \mathrm{M}$ can be chosen as long as it never assigns positive probability to messages such that $c(\theta, m) = 0$. Since $p \in \Delta(\mathrm{A} \times \Theta)$ has full support, $k_{a,\theta} > 0$ for any $(a, \theta) \in \mathrm{A} \times \Theta$. Hence, for any θ there is at least one element $m \in \mathbf{m}_a$ such that $c(\theta, m) = 1$. To implement the target outcome $\mu : \Theta \to \Delta(A)$, for such $m \in \mathbf{m}_a$ it suffices to impose

$$\eta(m|\theta) = \mu(a|\theta)$$

and zero to the other elements in the block $\mathbf{m}_{a,\theta}$ □