# The Core of Bayesian Persuasion*

## Prepared for the CEPR IO Seminar–Please do not circulate

Laura Doval[†] Ran Eilat[‡]

March 15, 2024

**Abstract**

An analyst observes the frequency with which an agent takes actions, but not the frequency with which she takes actions conditional on a payoff relevant state. In this setting, we ask when the analyst can rationalize the agent's choices as the outcome of the agent learning something about the state before taking action. Our characterization marries the obedience approach in information design (Bergemann and Morris, 2016) and the belief approach in Bayesian persuasion (Kamenica and Gentzkow, 2011) relying on a theorem by Strassen (1965) and Hall's marriage theorem. We apply our results to ring-network games and to identify conditions under which a data set is consistent with a public information structure in first-order Bayesian persuasion games.

KEYWORDS: *Bayes correlated equilibrium, Bayesian persuasion, information design, stochastic choice, distributions with given marginals, cooperative games, set functions, core*

# 1 Introduction

Given a primitive payoff structure, information design provides a framework for rationalizing outcomes as the result of non-cooperative play *without* having to specify the players' information structure. For this reason, the seminal work of Bergemann and Morris (2016) has spurred renewed interest among empirical scholars wishing to

*Click here for the latest version. We are grateful to Yaron Azrieli, Denniz Kattwinkel, Emir Kamenica, Shengwu Li, Elliot Lipnowski, Meg Meyer, Stephen Morris, Jacopo Perego, Andrea Prat, John Rehbeck, and Vasiliki Skreta for thought-provoking questions and insightful discussions. We owe special thanks to Quitzé Valenzuela-Stookey, who suggested ideas that facilitated the proof of Theorem 1.

[†]Columbia Business School and CEPR. Email: laura.doval@columbia.edu

[‡]Department of Economics, Ben-Gurion University. E-mail: eilatr@bgu.ac.il

obtain identification and estimation results under a weaker set of informational assumptions (see, for instance, Syrgkanis et al., 2017; Magnolfi and Roncoroni, 2023; Koh, 2023; Gualdani and Sinha, 2019; Rambachan, 2022).

However, the weaker set of assumptions on the information structure comes at the cost of increasing demands on the data set available to the analyst. Indeed, information design presumes the analyst is given a *joint* distribution over payoff relevant states and action profiles. For example, the literatures on rational inattention and stochastic choice usually assume that the analyst observes an agent's choices conditional on the state of the world (e.g., Caplin and Dean, 2015; Aguiar et al., 2018). Given this data set, *Bayes correlated equilibrium* provides an easy to test set of conditions the joint distribution over states and action profiles must satisfy in order to be consistent with the outcome of non-cooperative play under some information structure.

Oftentimes, however, the analyst's data set is more limited. The analyst may observe the distribution over the payoff relevant states of the world and the distribution over action profiles, but not the distribution over action profiles conditional on the state of the world.[1] Given the primitive payoff structure, we can then ask which marginal distributions can be rationalized as the outcome of non-cooperative play under some information structure. We refer to such marginals as *BCE-consistent* because they satisfy that a joint distribution over states and action profiles exists that is consistent with the marginals *and* is a Bayes correlated equilibrium. Characterizing the set of BCE-consistent marginal distributions can only increase the practical applicability of Bayes correlated equilibrium.

The set of BCE-consistent marginal distributions is of interest for two other reasons. First, the analyst is interested not just in the existence of an information structure that rationalizes the (marginal) distribution of play, but one that satisfies certain properties. For instance, the analyst may want to test whether the agents have private information. As we explain below, our characterization result provides us with a test for the existence of a public information structure that rationalizes the observed distribution of play. The second reason is related to reduced-form implementation in mechanism design (Matthews, 1984; Border, 1991). Whenever the information designer only cares about the agents' action profiles, but not the state of the world, the information designer's problem can be expressed as the choice out of the set of BCE-consistent marginals.

In this paper, we take the first step towards characterizing the set of BCE-consistent marginals by considering the single-agent case. We take as given the ingredients of a

[1] Whereas state-dependent stochastic choice data is useful to guide the design and interpretation of experiments, this data is oftentimes hard to come by outside the experimental setting. Dardanoni et al. (2020) provides an eloquent discussion of the *data voracity* of stochastic choice.

single-agent decision problem: a finite set of states of the world, a finite set of actions, and the agent's utility function over action-state tuples. Given these primitives, we seek to understand which pairs of distributions over states of the world and actions are consistent with the agent learning something about the state of the world before taking her action.

Building on a theorem in Strassen (1965), Theorem 1 characterizes the set of BCE-consistent marginals in terms of a finite system of inequalities the primitives–the agent's utility function, the prior distribution over the states, and the marginal over the actions–must satisfy. Furthermore, these inequalities are linear in the prior distribution over the states and the marginal over the actions. Whereas we take as given the agent's utility function, the characterization in Theorem 1 can also be seen as characterizing the pair of agent utility functions and prior distributions over states of the world that are consistent (via information) with the (marginal) distribution over actions. This perspective is useful as it is natural to assume the analyst may have easier access to the marginal distribution over actions, which they then use to estimate the agent's prior and utility function.

Because the proof of Theorem 1 is not constructive, it leaves open the question of which information structures make the marginals BCE-consistent. We address this in Proposition 1: marrying the obedience approach in information design with the belief approach in Kamenica and Gentzkow (2011), Proposition 1 characterizes the Bayes plausible distributions over posteriors that implement a given marginal over actions. We provide two network-based proofs of Proposition 1. Relying on recent extensions of Hall's marriage theorem in Barseghyan et al. (2021) and Azrieli and Rehbeck (2022), the first characterization uncovers a connection between BCE-consistency and the *core* of the game induced by loosely speaking, some (Bayes plausible) posterior distribution (see Remark 2 and Grabisch et al., 2016). The second proof relies on the demand problem of Gale (1957). We show one can interpret the BCE-consistency problem as a supply-demand problem in a persuasion economy, in which the marginal action distribution describes the demand and a Bayes plausible distribution over posteriors describes the supply. We then rely on the results in Gale (1957) to determine when the demand is feasible given the supply.

Section 4 illustrates three applications of Theorem 1 to multi-agent settings. Section 4.1 applies Theorem 1 to the first-order Bayesian persuasion setting of Arieli et al. (2021) to characterize the subset of BCE-consistent marginals that are consistent with a *public* information structure. Section 4.2 applies Theorem 1 to characterize BCE-consistent marginals in ring-network games as in Kneeland (2015). Finally, Section 4.3 provides a test for when the distribution over action profiles is consistent with the players having complete information about the state of the world.

**Related literature** The two closest papers to ours are Rehbeck (2023) and Azrieli and Rehbeck (2022). Rehbeck (2023) studies the same question as us when the analyst has access to a decision maker's unconditional stochastic choices, possibly out of different menus. For the case of a single menu, the characterization in Rehbeck (2023) is stated in terms of the non-existence of a possibly mixed deviation.[2] Instead, the characterization in Theorem 1 is in terms of a finite system of inequalities the marginals must satisfy. Azrieli and Rehbeck (2022) study a similar question to ours in the context of stochastic choice out of menus. In their setting, the analyst has access to a marginal distribution over a decision maker's choices and a marginal distribution over the menus out of which the decision maker made her choices. Azrieli and Rehbeck (2022) show the marginal distributions are consistent if and only if the marginal over choices is in the core of the game induced by the marginal over menus.

A literature in decision theory and experimental economics studies when choices can be rationalized via costly information acquisition and whether the choices can be used to identify the information acquisition costs (see, e.g., Caplin and Dean (2015), Caplin et al. (2017), Chambers et al. (2020), Dewan and Neligh (2020), Denti (2022)). Like we do, many of these papers assume the decision maker's utility is known. More recently, assuming the analyst has access to state-dependent stochastic choice data, Caplin et al. (2023) study when choices can be rationalized as if the agent has access to some information before choosing her actions. Whereas their analyst has access to a richer data set, they require consistency of the information structure across a family of decision problems.

The paper also contributes to the empirical literature that relies on information design to recover parameter estimates in single-agent decision problems (Gualdani and Sinha, 2019) and games (Syrgkanis et al., 2017; Magnolfi and Roncoroni, 2023; Koh, 2023). A key step in these papers is the characterization of the identified set of parameters, where a parameter indexes the prior and the agents' payoffs. Typically these characterizations are of the form "a parameter belongs in the identified set if, given the corresponding prior and payoffs, a Bayes correlated equilibrium exists whose marginal over the action profiles coincides with that in the data." In the single-agent case, Theorem 1 provides an alternative route to characterizing the identified set: A parameter–i.e., prior and utility function the analyst deems possible–belongs in the identified set if and only if it satisfies the finite system of inequalities in Theorem 1. That is, rather than searching over the whole set of Bayes correlated equilibria for each parameter, Theorem 1 reduces the question of identification to the verification of a finite system of inequalities.

[2] De Oliveira and Lamba (2022) study a similar question to us and Rehbeck (2023) in dynamic settings. Like Rehbeck (2023), their characterization is in terms of the non-existence of a possibly mixed deviation.

Arieli et al. (2021) and Morris (2020) characterize joint distributions over posterior beliefs that are consistent with some information structure.[3] Both papers cast the problem as one of distributions with given marginals: they take as given a profile of marginal distributions over posterior beliefs with the same mean and characterize when a joint distribution with the given marginals exists that is consistent with information.

Finally, Vohra et al. (2023) study reduced-form implementation in a Bayesian persuasion in which the sender and the receiver care only about the posterior mean of the states. Leveraging the mean preserving contraction property, the authors show the sender's problem can be written as a linear programming problem that only depends on the marginal distribution over actions. Because the characterization in Theorem 1 considers an arbitrary utility function for the agent, our result could be used to study reduced-form implementation in Bayesian persuasion beyond the posterior mean setting.

# 2 Model

Anticipating our multi-agent results in Section 4, our notation below presumes multiple agents. We then specialize it to the single-agent case in Section 3:

**Base game** An incomplete information *base game*, $G$, is defined as follows. We are given a set of $N$ players, $[N]$=$\{1,\dots,N\}$. Each player $i \in [N]$ chooses an action from the finite set $A_i$. Payoffs $u_i(a,\theta)$ depend on the action profiles $a \in A \equiv \times_{i\in[N]} A_i$ and the state of the world, $\theta$, an element of the finite set $\Theta$.[4] The players share a common prior $\mu_0 \in \Delta(\Theta)$ over the state of the world. That is, $G = \langle \Theta, (A_i, u_i)_{i\in[N]}, \mu_0\rangle$.

**Bayes correlated equilibrium** An *outcome* is a joint distribution over action profiles and states of the world, $\pi \in \Delta(A\times\Theta)$. We are concerned with those outcomes that are consistent with non-cooperative play of the base game, where the solution concept is Bayes Nash equilibrium. The notion of Bayes correlated equilibrium in Bergemann and Morris (2016) captures the set of outcomes that are consistent with (Bayes Nash) equilibrium of the base game under *some* information structure:

**Definition 1** (Bayes correlated equilibrium). *An outcome distribution* $\pi \in \Delta(A\times\Theta)$ *is a* Bayes correlated equilibrium *of base game* $G = \langle \Theta, (A_i, u_i)_{i\in[N]}, \mu_0\rangle$*, if for all agents*

[3] Whereas Arieli et al. (2021) study the binary-state case, the characterization in Morris (2020) requires no such assumption.

[4] As we explain in Section 3 our single-agent characterization extends to the case in which $\Theta$ and $A$ are infinite (see Remark 1). However, the set of finitely many states and actions allows us to provide a *sharper* characterization.

$i \in [N]$, *actions* $a_i, a_i' \in A_i$, *the following holds*

$$\sum_{(a_{-i},\theta)} \pi(a_i, a_{-i}, \theta) \left[u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta)\right] \geq 0, \tag{O}$$

*and for all* $\theta \in \Theta$

$$\sum_{a \in A} \pi(a, \theta) = \mu_0(\theta). \tag{M$_\Theta$}$$

*Let* $\mathrm{BCE}(\mu_0)$ *denote the set of Bayes correlated equilibria.*

In words, a Bayes correlated equilibrium is an outcome distribution that satisfies a series of *obedience* constraints (O) and a *martingale* condition ($\mathsf{M}_\Theta$). The first ensures each player's best response condition under *some* information structure, whereas the second ensures the existence of an information structure that is consistent with the players' prior information. Note that any Bayes correlated equilibrium $\pi \in \Delta(A \times \Theta)$ induces two marginal distributions, $(\pi_\Theta, \pi_A) \in \Delta(\Theta) \times \Delta(A)$. The definition of Bayes correlated equilibrium implies the primitive base game $G$ pins down $\pi_\Theta$, but not necessarily $\pi_A$.

**Information Design with Given Marginals** We take the point of view of an analyst who knows the base game, but not the information structure under which the base game is played. The analyst is also endowed with information about the actions taken by the players. The analyst's goal is to determine whether this information is consistent with non-cooperative play of the base game under some information structure.

We consider two kinds of information the analyst may have about the players' actions, which are equivalent in the single-agent setting. In the first case, the analyst is endowed with a distribution over action profiles, $\nu_0 \in \Delta(A)$. In the second case, the analyst is endowed with a profile of action distributions, one for each player, that is, $\overline{\nu}_0 = (\nu_{0,1}, \ldots, \nu_{0,N}) \in \times_{i\in[N]}\Delta(A_i)$.

In each of these cases, the analyst wants to ascertain whether a Bayes correlated equilibrium $\pi \in \mathrm{BCE}(\mu_0)$ exists such that $\pi_A$ coincides with the analyst's information about the players' actions (i.e., $\pi_A = \nu_0$ or $(\pi_{A_i})_{i\in[N]} = \overline{\nu}_0$). In this case, we say that the marginals $(\mu_0, \nu_0)$ are BCE-consistent or that the profile of marginal distributions $(\mu_0, \overline{\nu}_0)$ are M-BCE-consistent. Definition 2 records this for future reference:

**Definition 2** (BCE- and M-BCE-consistent marginals)**.** *The pair* $(\mu_0, \nu_0)$ *is* BCE-consistent *if a Bayes correlated equilibrium* $\pi \in \mathrm{BCE}(\mu_0)$ *exists such that* $\pi_A = \nu_0$*. Similarly, the profile* $(\mu_0, \overline{\nu}_0)$ *is* M-BCE-consistent *if a Bayes correlated equilibrium* $\pi \in \mathrm{BCE}(\mu_0)$ *exists such that for all players* $i \in [N]$, $\pi_{A_i} = \nu_{0,i}$.

Note that if the pair $(\mu_0, \nu_0)$ is BCE-consistent, then letting $\nu_{0,i}$ denote the marginal of $\nu_0$ over $A_i$, we have that $(\mu_0, \nu_{0,1}, \ldots, \nu_{0,N})$ are M-BCE-consistent.

**Constrained Optimal Transport** We close this section by noting a connection with optimal transport. Given $(\mu_0, \nu_0)$, let $\Pi(\mu_0, \nu_0)$ denote the set of joint distributions $\pi \in \Delta(A \times \Theta)$ with marginals $(\mu_0, \nu_0)$, i.e., $(\pi_\Theta, \pi_A) = (\mu_0, \nu_0)$. Note that $\Pi(\mu_0, \nu_0)$ is always nonempty, e.g., the joint distribution $\pi(a, \theta) = \nu_0(a)\mu_0(\theta)$ satisfies the marginal constraints. Instead, the set of joint distributions in $\Pi(\mu_0, \nu_0)$ that satisfies the obedience constraints (O), $\Pi_{\mathrm{O}}(\mu_0, \nu_0)$ may be empty. Thus, the characterization of the set of BCE-consistent marginals $(\mu_0, \nu_0)$ is equivalent to the characterization of when the feasible set of a *constrained* optimal transport problem–in this case $\Pi_{\mathrm{O}}(\mu_0, \nu_0)$–is nonempty.[5]

# 3 Single-agent case

In this section we characterize the set of BCE-consistent marginals in the case of a single agent, that is, $N = 1$. For this reason, in what follows we remove the index $i = 1$ from the action set and the utility function. In what follows, we denote by $A$ the agent's action set and by $u$ the agent's utility function.

**The action marginal as a distribution over posteriors** An outcome distribution $\pi \in \Delta(A \times \Theta)$ with marginals $(\mu_0, \nu_0)$ induces a conditional probability system, $\{\mu(\cdot|a) \in \Delta(\Theta) : a \in A\}$, which describes the agent's beliefs conditional on action $a$ and satisfies for all actions $a \in A$,

$$\nu_0(a)\mu(\theta|a) = \pi(a, \theta).$$

In this case, one can view $\nu_0$ as a distribution over posteriors and the belief system $(\mu(\cdot|a))_{a\in A}$ as its support. To see this, note that the problem of determining whether the pair $(\mu_0, \nu_0)$ is BCE-consistent can be cast in terms of determining whether a belief system $(\mu(\cdot|a))_{a\in A}$ exists that satisfies the following. First, for all states $\theta \in \Theta$

$$\sum_{a \in A} \nu_0(a)\mu(\theta|a) = \mu_0(\theta), \tag{BP$_{\mu_0}$}$$

and for all $a, a' \in A$,

$$\sum_{\theta \in \Theta} \nu_0(a)\mu(\theta|a) \left[u(a, \theta) - u(a', \theta)\right] \geq 0. \tag{O$_\mu$}$$

[5] In their study of credible Bayesian persuasion, Lin and Liu (2022) characterize the set of credible outcome distributions by noting a connection with optimal transport. In their case, to check whether a given message distribution $\lambda_M$ is implementable, it must be that no other joint distribution over states and messages that respects the given marginals exists and is preferred by the sender to $\lambda_M$.

For an action $a$, let $\Delta_u^*(a)$ denote the set of beliefs under which $a$ is optimal.[6] Then, Equations $\text{BP}_{\mu_0}$ and $\text{O}_\mu$ require that (i) $\nu_0$ induces a Bayes plausible distribution over posteriors and (ii) for all actions $a$, the *posterior belief* $\mu(\cdot|a)$ is an element of $\Delta_u^*(a)$. Under this interpretation, the action distribution $\nu_0$ describes the frequency with which inducing beliefs in $\Delta_u^*(a)$ is necessary. Unsurprisingly, some of the conditions in Theorem 1 below also check that $\nu_0$ satisfies a version of the martingale condition (Aumann et al., 1995; Kamenica and Gentzkow, 2011).

Theorem 1 characterizes the set of BCE-consistent marginals:

**Theorem 1** (BCE-consistency). *The pair $(\mu_0, \nu_0)$ is BCE-consistent if and only if for all states $\theta \in \Theta$,*

$$\sum_{a\in A} \nu_0(a) \min_{\mu\in\Delta_u^*(a)} \mu(\theta) \leq \mu_0(\theta), \tag{1}$$

*and for all pairs of actions $a', a'' \in A$,*

$$\sum_{a\in A} \nu_0(a) \max_{\mu\in\Delta_u^*(a)} \sum_{\theta\in\Theta} \mu(\theta) \left[u(a',\theta) - u(a'',\theta)\right] \geq \sum_{\theta\in\Theta} \mu_0(\theta) \left[u(a',\theta) - u(a'',\theta)\right]. \tag{2}$$

The proof is in Appendix A. In what follows, we provide intuition for the statement in Theorem 1 and review the main steps of its proof.

Equation 1 can be interpreted through the lens of the martingale property of beliefs. As discussed before Theorem 1, the action distribution $\nu_0$ describes the frequency with which beliefs in $\Delta_u^*(a)$ must be induced to satisfy ($\text{BP}_{\mu_0}$). For a given state $\theta \in \Theta$, the term

$$\underline{\mu}_a(\theta) \equiv \min_{\mu\in\Delta_u^*(a)} \mu(\theta),$$

describes the smallest probability that the agent can assign to state $\theta$ and action $a$ be optimal. Thus, Equation 1 states that for $(\mu_0, \nu_0)$ to be BCE-consistent, it must be that the average under $\nu_0$ of these minimum probabilities, $\underline{\mu}_a(\theta)$, are below the prior probability of $\theta$, $\mu_0(\theta)$. It is immediate that if for some state $\theta$, Equation 1 does not hold, then $(\mu_0, \nu_0)$ cannot be BCE-consistent.

As we argue next, Equation 2 can be interpreted through the lens of a martingale property for the utility differences, $u(a',\theta) - u(a'',\theta)$. That is, for all pairs of actions, $a', a''$, the agent's expected ranking over $a'$ and $a''$ under the experiment that rationalizes $(\mu_0, \nu_0)$ has to coincide with the agent's ex ante ranking over these actions, which is

[6] Formally, $\Delta_u^*(a) = \{\mu \in \Delta(\Theta) : (\forall a' \in A) \sum_{\theta\in\Theta} \mu(\theta) \left(u(a,\theta) - u(a',\theta)\right) \geq 0\}$.

the right hand side of Equation 2. Indeed, because Equation 2 must hold when we exchange the roles of $a'$ and $a''$, we obtain that $(\mu_0, \nu_0)$ must also satisfy that

$$\sum_{a \in A} \nu_0(a) \min_{\mu \in \Delta_u^*(a)} \sum_{\theta \in \Theta} \mu(\theta) \left[u(a', \theta) - u(a'', \theta)\right] \leq \sum_{\theta \in \Theta} \mu_0(\theta) \left[u(a', \theta) - u(a'', \theta)\right]. \quad (3)$$

That is, the ranking at the prior between $a'$ and $a''$ must be in between the worst and best rankings under the "distribution over posteriors" $\nu_0$.

This is most easily seen in the simple case that $a''$ is strictly optimal at the prior and $\{a', a''\}$ are the only actions in the support of $\nu_0$. Because $a'$ is in the support of $\nu_0$, under a BCE $\pi$ that satisfies the marginal constraints the agent must sometimes find it optimal to take action $a'$ instead of $a''$. Note, however, that *on average* it must be the case that the agent finds action $a''$ better than $a'$. Consequently, under $\pi$, when the agent takes $a''$, the agent must prefer $a''$ over $a'$ (weakly) more than at the prior. Because the left hand side of Equation 2 selects beliefs in favor of $a'$, it is immediate that if Equation 2 fails one cannot find an experiment in which the agent would take action $a'$ with sufficiently high probability so as to match $\nu_0$.

So far, we have argued that the conditions in Theorem 1 are necessary for $(\mu_0, \nu_0)$ to be BCE-consistent. To explain why they are also sufficient, it is useful to review the main steps in the proof of Theorem 1. Key to our proof is the following result from Strassen (1965), which we record in present notation:[7]

**Observation 1** (Strassen (1965, Theorem 3 and Corollary 1)). *A conditional probability system $\{\mu(\cdot|a) \in \Delta(\Theta) : a \in A\}$ exists such that*

1. *For all actions $a \in A$, $\mu(\cdot|a) \in \Delta_u^*(a)$, and*
2. *For all states $\theta \in \Theta$, $BP_{\mu_0}$ holds,*

*if and only if for all directions $c \in \mathbb{R}^{|\Theta|}$,*

$$\sum_{a \in A} \nu_0(a) \max\{c^T \mu : \mu \in \Delta_u^*(a)\} \geq c^T \mu_0. \quad (4)$$

Whereas Theorem 3 in Strassen (1965) requires that Equation 4 holds for *all* directions in $\mathbb{R}^{|\Theta|}$, Theorem 1 states that verifying Equation 4 holds for *finitely* many directions is enough to conclude that $(\mu_0, \nu_0)$ are BCE-consistent. To see this, note that Equations 1 and 2 correspond to Equation 4 for specific directions $c \in \mathbb{R}^{|\Theta|}$. Indeed, Equation 1

[7] Under our ongoing assumptions of finitely many states and finitely many actions, Strassen's theorem can also be obtained–after some manipulation–as a consequence of Farkas' lemma. As we explain in Remark 1, Strassen's result allows for more general state and action spaces, which allows us to generalize Theorem 1 beyond the finitely many states and actions case.

corresponds to $c = -e_\theta \in \mathbb{R}^{|\Theta|}$, where $e_\theta$ is the vector with a $1$ in the $\theta$-coordinate and $0$ otherwise. Instead, Equation 2 corresponds to the direction $c = d_{a',a''}$, where $d_{a',a''}$ is the vector with $\theta$-coordinate $d_{a',a''}(\theta) = u(a',\theta) - u(a'',\theta)$.

To see why verifying that Equation 4 holds for directions $\{(-e_\theta)_{\theta\in\Theta}, (-d_{a',a''})_{a',a''\in A}\}$ is enough to determine that Equation 4 holds for all directions $c \in \mathbb{R}^{|\Theta|}$, note the following. First, for a fixed action $a'$, the directions $\{(-e_\theta)_{\theta\in\Theta}, (-d_{a',a''})_{a''\in A}\}$ are the normal vectors that define the polyhedron $\Delta^*_u(a')$. Indeed, the directions $(-e_\theta)_{\theta\in\Theta}$ correspond to the condition that the elements of $\Delta^*_u(a')$ are non-negative, whereas the directions $(-d_{a',a''})_{a''\in A}$ correspond to the condition that action $a'$ is optimal for all beliefs in $\Delta^*_u(a')$. Second, it is immediate that in each of the maximization problems on the left hand side of Equation 4, the maximum is attained at an extreme point of $\Delta^*_u(a)$. Standard results in convex analysis then imply that if Equation 4 holds at all normal directions defining the polyhedra $\{\Delta^*_u(a) : a \in A\}$, then it holds for all directions (cf. Hiriart-Urruty and Lemaréchal, 2004).

We close Section 3 with a remark on the generality of the results in Strassen (1965). It can be skipped with no loss of continuity.

**Remark 1** (Strassen, 1965). *Theorem 3 and Corollary 1 in Strassen (1965) hold more generally than our current assumptions. In present notation, Corollary 1 applies whenever (i) $\Theta$ and $A$ are compact metric spaces and the mapping $a \mapsto \Delta^*(a)$ from $A$ to subsets of $\Delta(\Theta)$ is such that $\cup_{a\in A}\{a\} \times \Delta^*(a)$ is closed within $A \times \Delta(\Theta)$ endowed with the weak*-topology.*[8]

*In other words, under the aforementioned assumptions, (an integral version of) Equation 4 characterizes the set of BCE-consistent marginals.*[9] *The finite model allows us to provide a sharper characterization by reducing the number of directions one needs to consider.*

## 3.1 The core of Bayesian Persuasion

In this section we provide a different perspective on Theorem 1. Together with the marginal distributions, $(\mu_0, \nu_0)$, we are given a distribution over posteriors $\tau \in \Delta(\Delta(\Theta))$ with mean equal to the prior $\mu_0$ (henceforth, a Bayes plausible distribution over posteriors). Proposition 1 below characterizes the set of such distributions over posteriors

[8]Instead, Strassen (1965, Theorem 3) requires that $\Theta$ is Polish, $A$ be a convex compact topological vector space, and an appropriate measurability condition on the mapping $a \mapsto \sup\{\int c(\theta)\mu(d\theta) : \mu \in \Delta^*_u(a)\}$ for any continuous function $c$ on $\Theta$.

[9]To be precise, Equation 4 now becomes for all continuous functions $c : \Theta \mapsto \mathbb{R}$,

$$\int_\Theta c(\theta)\mu_0(d\theta) \leq \int_A \sup\left\{\int_\Theta c(\theta)\mu(d\theta) : \mu \in \Delta^*_u(a)\right\}\nu_0(da)$$

.

that *implement* the marginal $\nu_0$ (see Definition 3 below). Whereas this characterization does not substitute that in Theorem 1, it allows us to illustrate how one would go about constructing an information structure that implements $\nu_0$. Along the way we also establish formal connections with the literature on stochastic choice. Without loss of generality, the analysis that follows assumes the distribution over posteriors $\tau$ has finite support (Myerson, 1982; Kamenica and Gentzkow, 2011); we denote the support of $\tau$ by $M_\tau$.

Definition 3 states what it means for a Bayes plausible distribution over posteriors to implement $\nu_0$. To do so, the following piece of notation is useful: let $a^*(\mu)$ denote the agent's best response correspondence when her belief is $\mu$.[10]

**Definition 3** ($\tau$ implements $\nu_0$). *Given a Bayes plausible distribution over posteriors,* $\tau$ implements $\nu_0$ *if a decision rule* $\alpha : M_\tau \mapsto \Delta(A)$ *exists such that for all* $a \in A$

$$\nu_0(a) = \sum_{\mu \in M_\tau} \tau(\mu)\alpha(a|\mu), \tag{5}$$

*and* $\alpha(a|\mu) > 0$ *only if* $a \in a^*(\mu)$.

Whereas in the previous section we interpreted the marginal distribution $\nu_0$ as a Bayes plausible distribution over posteriors, Equation 5 shows this analogy is perhaps incomplete. Indeed, whenever the distribution over posteriors $\tau$ induces beliefs such that $a^*(\mu)$ is not a singleton, specifying the agent's tie-breaking rule $\alpha$ is necessary to determine whether the frequency with which the agent takes actions under $\tau$ matches that under $\nu_0$.

**A demand and supply problem** The problem of determining whether a decision rule $\alpha$ exists satisfying Equation 5 admits the following interpretation (cf. Gale, 1957): Beliefs $\mu \in M_\tau$ are supplied in quantities $\tau(\mu)$ and demanded by actions $a \in A$ in quantities $\nu_0(a)$. The demand $\nu_0(a)$ can only be satisfied by certain beliefs–those that satisfy $\mu \in \Delta^*_u(a)$. The decision rule $\alpha$ describes how much of a given belief $\mu \in M_\tau$ is allocated to action $a$. That $\tau$ implements $\nu_0$ is equivalent to being able to allocate the supply of beliefs to satisfy the action demands in a *market clearing* way.

We can represent this problem graphically (see Figure 1). Define the graph $G_P(\tau) = (A \cup M_\tau, E)$ as follows. To each action $a \in A$ corresponds a node that demands the marginal probability of $a$. To each belief $\mu \in M_\tau$ corresponds a node that supplies the probability with which $\mu$ is realized in $\tau$. For any belief-action pair $(\mu, a)$, an edge $(\mu, a) \in E$ exists between the nodes $\mu$ and $a$ if and only if action $a \in a^*(\mu)$. Finally, for any edge $(\mu, a) \in E$, the edge's flow capacity is given by $c\,(\mu, a) = \infty$ and $c\,(a, \mu) = 0$. That is, no upper bound on the flow from $\mu$ to $a$ exists, but there cannot be a flow from

[10] Formally, $a^*(\mu) = \arg\max_{a \in A} \sum_{\theta \in \Theta} \mu(\theta)u(a, \theta)$.

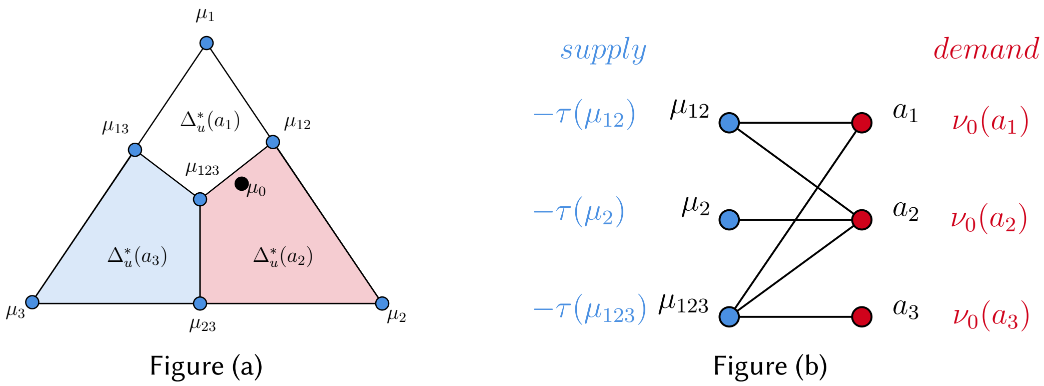


Figure 1: Illustration of the supply-demand proof of Proposition 1 with $|A| = |\Theta| = 3$. The simplex on the left hand side depicts the optimal action(s) for each posterior belief. The graph on the right hand side corresponds to the Bayes plausible distribution over posteriors $\tau$ supported on $M_\tau = \{\mu_{12}, \mu_2, \mu_{123}\}$.

$a$ to $\mu$. To determine whether the (action) demand is feasible given the (belief) supply, we need to find a flow $f(\mu, a)$ that satisfies that the flow along each edge is no larger than its capacity and the net flow into (out of) each node is at least (at most) equal to the demand (supply) at that node. Up to a normalization, this flow is exactly our decision rule.

The above is an instance of supply and demand problem studied in Gale (1957). Building on the main theorem in that paper, Proposition 1 below characterizes when $\tau$ implements $\nu_0$. To state Proposition 1, we need one final piece of notation. Given a Bayes plausible $\tau \in \Delta(\Delta(\Theta))$, one can construct a measure over subsets $B$ of the set of actions $A$ as follows. For each $B \subseteq A$, define $\tau_A(B)$ as

$$\tau_A(B) = \tau\{\mu \in \Delta(\Theta) : a^*(\mu) = B\}. \tag{6}$$

In words, each action subset $B$ has mass equal to the probability that $\tau$ induces a belief under which $B$ is optimal.

Proposition 1 characterizes when the distribution over posteriors $\tau$ implements $\nu_0$:

**Proposition 1.** *Suppose $(\mu_0, \nu_0)$ are BCE-consistent. A Bayes plausible distribution over posteriors, $\tau \in \Delta(\Delta(\Theta))$, implements $\nu_0$ if and only if the following holds*

$$(\forall B \subseteq A) \sum_{a \in B} \nu_0(a) \geq \sum_{C \subseteq B} \tau_A(C). \tag{7}$$

To interpret Equation 7, note the following. The left hand side of Equation 7 is the probability under which the agent takes *some* action $a$ in the set $B$. Instead, the right hand side of Equation 7 is the probability under which the agent finds optimal *some* action in the set $B$ (but no action that is not in $B$). Equation 7 then states the frequency with which the agent takes actions in $B$ has to be at least the frequency with which an action in $B$ is optimal.[11]

**Remark 2** (A core interpretation). *Equation 7 implies that $\nu_0$ is in the* core of the game *induced by the measure $\tau_A$.*[12] *Indeed, given $\tau_A$, define the following cooperative game. The set of players is the set of actions $A$, so that a coalition of players is a subset of actions $B \subset A$. The worth of coalition $B$ is given by $w_{\tau_A}(B) = \sum_{C \subseteq B} \tau_A$. Because $w_{\tau_A} \geq 0$, the core of the game $(A, w_{\tau_A})$ is given by*

$$\operatorname{Core}(w_{\tau_A}) = \left\{ p \in \Delta(A) : (\forall B \subseteq A) \sum_{a \in B} p(a) \geq w_{\tau_A}(B) \right\}.$$

*Equation 7 states that $\nu_0$ is a payment rule for each player in the game that covers the worth of each coalition and hence, belongs to the core of the game.*

The proof of Proposition 1 is in Section A.2 and follows from two steps. First, we show that since $\tau$ and $\nu_0$ are distributions, any feasible flow must clear the market, that is, any feasible flow must satisfy the supply/demand equations with equality. Gale refers to such flows as *maximal*. Second, we show the necessary and sufficient condition for the existence of a feasible and maximal flow in Gale (1957) is equivalent to Equation 7.

**Connection with stochastic choice** We draw now a connection with stochastic choice out of menus which, among other things, allows us to illustrate why Equation 7 implies $\tau$ implements $\nu_0$. We relegate the formal details of the discussion to Section A.3.

As we discuss in Section A.3, the condition in Equation 7 implies a conditional probability system $\{\sigma'(\cdot|B) \in \Delta(A) : B \subset A\}$ exists such that (i) for all $B \subseteq A$, $\sigma'(B|B) = 1$ and (ii)

$$\nu_0(a) = \sum_{B \subseteq A : a \in B} \tau_A(B) \sigma'(a|B). \tag{8}$$

[11] Equation 7 is intimately connected to the Border-Matthews-Maskin-Riley characterization of reduced form implementation in auctions. The latter states that a reduced-form auction (a collection of interim probabilities of trade for each buyer) has an auction implementation if and only if for all subsets of buyer type profiles, the probability the reduced form auction allocates the good to types in that set is no larger than the prior probability of buyer types in that set. Equation 7 is morally related with $\tau_A$ playing the role of the reduced-form auction and $\nu_0$ that of the type distribution.

[12] Azrieli and Rehbeck (2022) also note the connection between consistent marginals in the context of stochastic menu choice and cooperative games.

The conditional probability system $(\sigma'(\cdot|B))_{B\subseteq A}$ can be interpreted as the agent's stochastic choice out of the menus $\{B : B \subseteq A\}$. Equation 8 states the probability the agent chooses action $a$ under marginal is the probability the agent faces a menu $B$ that has $a$ available and the agent chooses $a$ out of $B$.

Consider now the following "experiment": We first draw a menu $B$ using $\tau_A$ and then draw an action $a \in B$ according to $\sigma'(\cdot|B)$. We only inform the agent of the drawn action, but not the menu from which it was drawn. Because we draw menu $B$ only when it is the optimal set of actions, we only recommend $a$ when it is optimal for the agent to follow the recommendation. As we show in Section A.3, Equation 8 then implies that this "experiment" induces the agent to take actions with the desired frequency.

The above still does not describe an experiment–a collection of signal distributions conditional on the state of the world–but this can be done immediately as follows: For each $\theta \in \Theta$ and $a \in A$,

$$\sigma(a|\theta) = \sum_{B:a\in B} \sum_{\{\mu\in M_\tau : a^*(\mu)=B\}} \frac{\mu(\theta)}{\mu_0(\theta)} \tau(\mu)\sigma'(a|B). \tag{9}$$

In this experiment, the agent receives an action recommendation conditional on the state of the world, so that Equation 9 describes the agent's state-dependent stochastic choice.

The previous discussion connects two sets of conditional distribution over choices that arise in the stochastic choice literature: stochastic choices conditional on a state of the world–Equation 9–and stochastic choices out of a menu–$\sigma'$ in Equation 8. Indeed, the measure $\tau_A$ can be interpreted as the frequency with which the agent faces different menus–action subsets in this case–whereas the measure $\nu_0$ represents the frequency with which the agent makes different choices. In other words, the pair $(\tau_A, \nu_0)$ is analogous to the data set in Azrieli and Rehbeck (2022). Our ultimate goal, however, is to obtain the agent's stochastic choice rule, which we obtain relying on the Bayes' plausibility of $\tau$.

# 4 Applications

In this section, we consider three applications of Theorem 1 to simple multi-agent settings. Section 4.1 studies under what conditions a pair of marginal distributions $(\mu_0, \nu_0)$ can be rationalized by a *public* information structure. Section 4.2 shows Theorem 1 characterizes the set of M-BCE-consistent marginals. In what follows, $A$ denotes the set $\times_{i\in[N]} A_i$. Whereas the results in Sections 4.1 and 4.2 are direct applications of Theorem 1, Section 4.3 provides a test of when a pair of marginal distributions is consistent with the players playing the game under complete information building on the result in Strassen (1965).

## 4.1 When is information public?

We consider in this section the following multiplayer game. We assume $N \geq 1$ and each player's utility function depends only on her own action and the state of the world.[13] That is, for all players $i \in \{1, \ldots, N\}$, all action profiles $(a_i, a_{-i}) \in A$, and states of the world $\theta \in \Theta$,

$$u_i\left(a_i, a_{-i}, \theta\right) = u_i\left(a_i, \theta\right).$$

The analyst, who knows the base game $G$ and the marginal distribution of play $\nu_0 \in \Delta(A)$, wants to ascertain whether the distribution of play $\nu_0$ can be rationalized by a *public* information structure (i.e., the players publicly observe the realization of a common signal structure before play).

As we show next, Theorem 1 can be applied to address this question. In what follows, we rely on the following definition:

**Definition 4** (Public BCE-consistency)**.** *The pair* $(\mu_0, \nu_0)$ *is* public BCE-consistent *if: (i)* $(\mu_0, \nu_0)$ *are BCE-consistent, and (ii) a BCE* $\pi \in \mathrm{BCE}\left(\mu_0\right) \cap \Pi(\mu_0, \nu_0)$ *exists, whose information structure uses public signals alone.*

Consider now an auxiliary single-agent base game $\bar{G} = \langle \Theta, (A, \bar{u}), \mu_0 \rangle$. In this game, a player with payoff $\bar{u}(a) = \sum_{i=1}^{N} u_i(a_i, \theta)$ chooses an action $a \in A = \times_{i \in N} A_i$ under incomplete information about $\theta$.

The following result is an immediate corollary of Theorem 1 and the focus on public signals:

**Corollary 1.** $(\mu_0, \nu_0)$ *are public BCE-consistent if and only if* $(\mu_0, \nu_0)$ *are BCE-consistent in base game* $\bar{G}$.

Because of the focus on public signal structures, the analysis of the multi-agent game reduces to the analysis of a single-agent problem. To see this, in a slight abuse of notation, let $A^*(\mu)$ denote the set of actions that the agent with payoff $\bar{u}$ finds optimal when their belief is $\mu$. It is immediate that $A^*(\mu) = \times_{i \in N} a_i^*(\mu)$, where for each player $i$, $a_i^*(\mu)$ denotes the set of actions player $i$ finds optimal when her belief is $\mu$. That is, the profile $a = (a_1, \ldots, a_N) \in A$ is optimal for the agent with payoff $\bar{u}$ if and only if action $a_i$ is optimal for agent $i$, for all $i \in [N]$. And, given a posterior belief $\mu$, any distribution of (optimal) action profiles that the agent with payoff $\bar{u}$ can generate, can also be generated by the players using a public correlation device or by duplicating signal realization, and vice versa.[14] Notice that this equivalence no longer holds if either information is not public, or the players' utilities are interdependent.

[13] Arieli et al. (2021) dub this setting first-order Bayesian persuasion.

[14] For example, suppose that the signal realization $s$ induces the posterior belief $\mu$. Suppose also that under $\mu$, the agent with payoff $\bar{u}$ selects the two optimal action profiles $a, a' \in A^*(\mu)$ with equal

## 4.2 Ring-network games

We consider here ring-network games as in Kneeland (2015), extended to account for incomplete information. A ring-network game is a base game $G$ in which player's payoffs satisfy the following:

$$\begin{aligned} u_1(a,\theta) &= \tilde{u}_1(a_1,\theta) \\ (\forall i \geq 2) u_i(a,\theta) &= \tilde{u}_i(a_{i-1},a_i). \end{aligned} \tag{RN-P}$$

In words, player $1$ cares about their action and the state of the world, whereas for $i \geq 2$ player $i$ cares about their action and that of player $i-1$. Ring-network games are used in the experimental literature that measures players' higher order beliefs to identify departures from Nash equilibrium.

The analyst knows the ring-network base game and for each player $i$, player $i$'s action distribution, $\nu_{0,i} \in \Delta(A_i)$. The analyst wants to ascertain whether $(\mu_0, \overline{\nu}_0)$ is M-BCE-consistent. Relying on Theorem 1 and the ring-network structure, Proposition 2 characterizes the set of M-BCE-consistent marginals:

**Proposition 2** (M-BCE-consistency in ring-network games)**.** *The profile $(\mu_0, \overline{\nu}_0)$ is M-BCE-consistent for the ring-network game $(\tilde{u}_i)_{i=1}^N$ if and only if the following holds:*

1. *$(\mu_0, \nu_{0,1})$ are BCE-consistent in the base game $\langle \Theta, A_1, \tilde{u}_1, \mu_0 \rangle$,*
2. *For all $i \geq 2$, $(\nu_{0,i-1}, \nu_{0,i})$ are BCE-consistent in the base game $\langle A_{i-1}, A_i, \tilde{u}_i, \nu_{0,i-1} \rangle$.*

Similar to Corollary 1, Proposition 2 exploits the structure of the ring-network game to reduce it to a series of single-agent problems in which except for player $1$, the states are given by the actions of the preceding player and the prior distribution over this state space by the marginal over actions of the preceding player. Indeed, for $i \geq 2$, BCE-consistency of $(\nu_{0,i-1}, \nu_{0,i})$ implies that an information structure exists that rationalizes player $i$'s choices as the outcome of some information structure under "prior" $\nu_{0,i-1}$, whereas BCE-consistency of $(\nu_{0,i-2}, \nu_{0,i-1})$[15] guarantees that the "prior" $\nu_{0,i-1}$ is consistent with player $i-1$ observing the outcome of some information structure given their belief $\nu_{0,i-2}$.

probability. The same distribution of actions can be generated by the players: Indeed, one can "split" the signal $s$ into two new signals, $s'$ and $s''$, such that both new signals induce the same posterior belief $\mu$, and each of them is sent with half the probability of the original signal $s$. If whenever $s'$ and $s''$ are realized, each agent acts according to her corresponding optimal action in the profiles $a$ and $a'$, respectively, the distribution over actions will coincide with that of the agent with payoff $\bar{u}$.

[15] With the understanding that $\nu_{0,0} = \mu_0$.

### 4.3 Playing the game under complete information

For our final application, we show that reasoning analogous to that leading to Theorem 1 can be used to derive a test for when the pair $(\mu_0, \nu_0)$ is consistent with the players playing the game under complete information. In what follows, we consider the general multi-agent setting introduced in Section 2 and hence, we lift the payoff restrictions in the previous sections.

We seek to understand when a Bayes correlated equilibrium $\pi \in \mathrm{BCE}(\mu_0)$ exists such that (i) $\pi_A = \nu_0$ and (ii) for all states $\theta \in \Theta$, all players $i \in [N]$, and action pairs $(a_i, a_i')$ the following holds

$$\sum_{a_{-i}} \pi(a_i, a_{-i}, \theta) \left[u_i(a_i, a_{-i}, \theta) - u_i(a_i', a_{-i}, \theta)\right] \geq 0.$$

Condition (ii) means that for all $\theta \in \Theta$ $\pi(\cdot|\theta) \in \Delta(A)$ is a correlated equilibrium of the game with payoffs $(u_i(\cdot, \theta))_{i\in[N]}$. Let $\mathrm{CE}(\theta)$ denote the set of correlated equilibria when the players know the state is $\theta$.

Building on Strassen (1965), Proposition 3 characterizes when the pair $(\mu_0, \nu_0)$ is consistent with the players playing the game under complete information:

**Proposition 3.** *The pair $(\mu_0, \nu_0)$ is consistent with the players playing the game under complete information if and only if for all $c \in \mathbb{R}^{|A|}$ the following holds*

$$\sum_{\theta\in\Theta} \mu_0(\theta) \max\left\{c^T\nu : \nu \in \mathrm{CE}(\theta)\right\} \geq c^T\nu_0. \tag{10}$$

For each $\theta \in \Theta$, the set $\mathrm{CE}(\theta)$ is a polyhedron. Like in Theorem 1, it then follows that checking finitely many directions suffices to check that Equation 10 holds for all directions $c \in \mathbb{R}^{|A|}$.

## References


Victor H Aguiar, Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim. 2018. Does random consideration explain behavior when choice is hard? Evidence from a large-scale experiment. *arXiv preprint arXiv:1812.09619* (2018).

Itai Arieli, Yakov Babichenko, Fedor Sandomirskiy, and Omer Tamuz. 2021. Feasible joint posterior beliefs. *Journal of Political Economy* 129, 9 (2021), 2546–2594.

Robert J Aumann, Michael Maschler, and Richard E Stearns. 1995. *Repeated games with incomplete information.* MIT press.

Yaron Azrieli and John Rehbeck. 2022. Marginal stochastic choice. *arXiv preprint arXiv:2208.08492* (2022).

Levon Barseghyan, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum. 2021. Heterogeneous choice sets and preferences. *Econometrica* 89, 5 (2021), 2015–2048.

Dirk Bergemann and Stephen Morris. 2016. Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics* 11, 2 (2016), 487–522.

Kim C Border. 1991. Implementation of reduced form auctions: A geometric approach. *Econometrica: Journal of the Econometric Society* (1991), 1175–1187.

Andrew Caplin and Mark Dean. 2015. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review* 105, 7 (2015), 2183–2203.

A Caplin, M Dean, and J Leahy. 2017. *Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy*. Technical Report. National Bureau of Economic Research.

Andrew Caplin, Daniel J Martin, and Philip Marx. 2023. *Rationalizable Learning*. Technical Report. National Bureau of Economic Research.

Christopher P Chambers, Ce Liu, and John Rehbeck. 2020. Costly information acquisition. *Journal of Economic Theory* 186 (2020), 104979.

Valentino Dardanoni, Paola Manzini, Marco Mariotti, and Christopher J Tyson. 2020. Inferring cognitive heterogeneity from aggregate choices. *Econometrica* 88, 3 (2020), 1269–1296.

Henrique De Oliveira and Rohit Lamba. 2022. Rationalizing dynamic choices. *Available at SSRN 3332092* (2022).

Tommaso Denti. 2022. Posterior separable cost of information. *American Economic Review* 112, 10 (2022), 3215–3259.

Ambuj Dewan and Nathaniel Neligh. 2020. Estimating information cost functions in models of rational inattention. *Journal of Economic Theory* 187 (2020), 105011.

David Gale. 1957. A theorem on flows in networks. *Pacific J. Math* 7, 2 (1957), 1073–1082.

Michel Grabisch et al. 2016. *Set functions, games and capacities in decision making*. Vol. 46. Springer.

Cristina Gualdani and Shruti Sinha. 2019. Identification in discrete choice models with imperfect information. *arXiv preprint arXiv:1911.04529* (2019).

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

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

Terri Kneeland. 2015. Identifying higher-order rationality. *Econometrica* 83, 5 (2015), 2065–2079.

Paul S Koh. 2023. Stable outcomes and information in games: An empirical framework. *Journal of Econometrics* 237, 1 (2023), 105499.

Xiao Lin and Ce Liu. 2022. Credible persuasion. *arXiv preprint arXiv:2205.03495* (2022).

Lorenzo Magnolfi and Camilla Roncoroni. 2023. Estimation of discrete games with weak assumptions on information. *The Review of Economic Studies* 90, 4 (2023), 2006–2041.

Steven A Matthews. 1984. On the implementability of reduced form auctions. *Econometrica: Journal of the Econometric Society* (1984), 1519–1522.

Stephen E Morris. 2020. *No trade and feasible joint posterior beliefs*. Technical Report. Massachusetts Institute of Technology.

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

Ashesh Rambachan. 2022. *Identifying Prediction Mistakes in Observational Data*. Technical Report. National Bureau of Economic Research.

John Rehbeck. 2023. Revealed Bayesian expected utility with limited data. *Journal of Economic Behavior & Organization* 207 (2023), 81–95.

Volker Strassen. 1965. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics* 36, 2 (1965), 423–439.

Vasilis Syrgkanis, Elie Tamer, and Juba Ziani. 2017. Inference on auctions with weak assumptions on information. *arXiv preprint arXiv:1710.03830* (2017).

Akhil Vohra, Juuso Toikka, and Rakesh Vohra. 2023. Bayesian persuasion: Reduced form approach. *Journal of Mathematical Economics* (2023), 102863.

# A Omitted proofs

## A.1 Proof of Theorem 1

**Preliminaries** Before stating the proof of Theorem 1, we collect some definitions and results from convex analysis that we use in the proof.

Define $C(a') = \{(-e_\theta)_{\theta\in\Theta}, (-d_{a',a''})_{a''\in A}\}$ to be the normal directions to the polyhedron $\Delta^*_u(a')$, which is implicitly defined as the set of vectors $x$ in $\mathbb{R}^{|\Theta|}$ that satisfy:

$$
\begin{aligned}
(\forall\theta \in \Theta)(-e_\theta)^T x \leq 0 \\
(\forall a'' \in A)(-d_{a',a''})^T x \leq 0.
\end{aligned}
\tag{11}
$$

We are omitting the condition that $\sum_{\theta\in\Theta} x(\theta) = 1$, but this is irrelevant in what follows.

Recall that for $x \in \mathbb{R}^{|\Theta|}$, the normal cone of $\Delta^*_u(a)$ at $x$, $N(x|\Delta^*_u(a))$, is defined as

$$
N(x|\Delta^*_u(a)) = \{c \in \mathbb{R}^{|\Theta|} : (\forall x' \in \Delta^*_u(a))c^T x' \leq c^T x\}. \tag{12}
$$

That is, the normal cone of $\Delta^*_u(a)$ at $x$ is the set of directions $c$ for which $x$ solves $\max\{c^T x' : x' \in \Delta^*_u(a)\}$. Importantly, the normal cone of a polyhedron, like $\Delta^*_u(a)$, satisfies the following property. To state it, recall that given a set of points $C$, the cone of $C$ is defined as $\operatorname{cone}(C) = \{\sum_{j=1}^J \alpha_j c_j : J < \infty, c_j \in C, \alpha_j \geq 0\}$.

**Lemma 1** (Hiriart-Urruty and Lemaréchal (2004, Example 5.2.6(b))). *Suppose $\mu \in \Delta^*_u(a)$ and let $B(\mu) = \{c \in C(a) : c^T\mu = 0\}$. Then, $N(\mu|\Delta^*_u(a)) = \operatorname{cone}(B(\mu))$.*

*Proof of Theorem 1.* Necessity of Equations 1 and 2 follows from Strassen (1965, Theorem 3).

We now argue sufficiency. Given Observation 1, it suffices to show Equations 1 and 2 imply Equation 4 holds for all $c \in \mathbb{R}^{|\Theta|}$.

For fixed $c$, we can write Equation 4 as follows:

$$
\sum_{a\in A} \nu_0(a) \max_{\mu\in\Delta^*_u(a)} c^T(\mu - \mu_0) \geq 0. \tag{13}
$$

Thus, Equation 4 holds for all directions $c \in \mathbb{R}^{|\Theta|}$ if and only if

$$
\min_{c\in\mathbb{R}^{|\Theta|}} \sum_{a\in A} \nu_0(a) \max_{\mu\in\Delta^*_u(a)} c^T(\mu - \mu_0) \geq 0. \tag{14}
$$

Note that we can replace $\Delta_u^*(a)$ for the set of extreme points of $\Delta_u^*(a)$, $(\Delta_u^*)^E(a)$ in Equation 14. That is,

$$\min_{c\in\mathbb{R}^{|\Theta|}} \sum_{a\in A} \nu_0(a) \max_{\mu\in(\Delta_u^*)^E(a)} c^T(\mu-\mu_0) \geq 0. \tag{15}$$

Now, let $E = \prod\{(\Delta_u^*)^E(a) : a \in A\}$. For $\bar{\mu}_e \equiv (\mu_{e,a})_{a\in A} \in E$, let

$$C(\bar{\mu}_e) = \{c \in \mathbb{R}^{|\Theta|} : (\forall a \in A)\ c^T\mu_{e,a} = \max_{\mu\in\Delta_u^*(a)} c^T\mu\}.$$

Then, we can write the left hand side of Equation 15 as follows:

$$\min_{\bar{\mu}_e\in E} \min_{c\in C(\bar{\mu}_e)} \sum_{a\in A} \nu_0(a) \max_{\mu\in(\Delta_u^*)^E(a)} c^T(\mu-\mu_0). \tag{16}$$

Note that for each $\bar{\mu}_e \in E$

$$C(\bar{\mu}_e) = \cap_{a\in A} N\left(\mu_{e,a}|\Delta_u^*(a)\right), \tag{17}$$

and by Lemma 1, $N\left(\mu_{e,a}|\Delta_u^*(a)\right) \subseteq C(a)$. Thus, Equations 1 and 2 ensure that the term inside $\min_{\bar{\mu}_e\in E}$ is non-negative, so that Equation 4 holds for all $c \in \mathbb{R}^{|\Theta|}$.

□

## A.2 Proof of Proposition 1

The problem in Gale (1957) can be described as follows. Given a graph $(V, E)$, suppose that to each node $v \in V$ corresponds a real number $d(v)$. If $d(v) > 0$ we interpret $|d(v)|$ as the *demand* of node $v$ for some homogeneous good. If $d(v) < 0$ we interpret $|d(v)|$ as the *supply* of the good by $v$. To each edge $(v, v') \in E$ correspond two non-negative real numbers $c(v, v')$ and $c(v', v)$, the capacity of this edge, which assign an upper bound to the possible flow of the good from $v$ to $v'$ and from $v'$ to $v$, respectively. The demand $d$ is called *feasible* if a flow on the graph exists such that the flow along each edge is no greater than its capacity, and the net flow into (out of) each node is at least (at most) equal to the demand (supply) at that node. Gale (1957) characterizes when a given demand $d$ is feasible in the graph.

Given $\tau$, we can define the graph $G_P(\tau)$ as we did in the main text. Namely, the nodes are $A \cup M_\tau$ and an edge $(\mu, a)$ exists if and only if $a \in a^*(\mu)$. Given $\tau$ and $\nu_0$, define the demand $d_{(\nu_0,\tau)}$ as follows: to each belief $\mu \in M_\tau$ corresponds the node $\mu$ that supplies the probability with which $\mu$ is realized in $\tau$, i.e. $d_{(\nu_0,\tau)}(\mu) = -\tau(\mu)$. To each $a \in A$, corresponds the node $a$ that demand the probability with which $a$ is taken under $\nu_0$, i.e., $d_{(\nu_0,\tau)}(a) = \nu_0(a)$. Finally, for any edge $(\mu, a) \in E$, the edge's flow capacity is given by $c(\mu, a) = \infty$ and $c(a, \mu) = 0$.

Proposition 4 motivates the connection between our problem and that in Gale (1957).

**Proposition 4** (Feasibility and BCE-consistency). *The Bayes plausible distribution over posteriors $\tau$ implements $\nu_0$ if and only if $d_{(\nu_0,\tau)}$ is feasible on $G_P(\tau)$.*

The proof of Proposition 4 relies on the following lemma:

**Lemma 2** (Market clearing). *If $d_{(\nu_0,\tau)}$ is feasible on $G_P(\tau)$, then the flow out of any supply node $\mu \in M_\tau$ is exactly $\tau(\mu)$ (and not less), and the flow into any demand node $a \in A$ is exactly $\nu_0(a)$ (and not more).*

*Proof of Lemma 2.* Suppose that $d_{(\nu_0,\tau)}$ is feasible. We show the flow into any demand node $a \in A$ is exactly $\nu_0(a)$. Towards a contradiction, suppose that $\sum_{\mu \in M_\tau} f(\mu, a) \geq \nu_0(a)$ for all $a \in A$, with strict inequality for some $a$. Summing over all actions on both sides of the inequality yields

$$\sum_{a \in A} \sum_{\mu \in M_\tau} f(\mu, a) > \sum_{a \in A} \nu_0(a) = 1.$$

On the other hand, because $d_{(\nu_0,\tau)}$ is feasible, then the flow out of each $\mu \in M_\tau$ is at most $\tau(\mu)$, and therefore for all $\mu \in M_\tau$

$$\sum_{a \in A} f(\mu, a) \leq \tau(\mu).$$

Summing again over all actions on both sides yields

$$\sum_{\mu \in M_\tau} \sum_{a \in A} f(\mu, a) \leq \sum_{\mu \in M_\tau} \tau(\mu) = 1,$$

a contradiction. The proof that the flow out of any supply node $\mu$ is exactly $\tau(\mu)$ is analogous and hence omitted. □

*Proof of Proposition 4.* Suppose first the Bayes plausible distribution over posteriors $\tau$ is such that $d_{(\nu_0,\tau)}$ is feasible on $G_P(\tau)$ and let $f$ denote the feasible flow. Consider a decision rule $\alpha : \Delta(\Theta) \mapsto \Delta(A)$ such that the agent takes action $a \in A$ when the belief is $\mu \in M_\tau$ with probability $\alpha(a \mid \mu) = f(\mu, a) / \tau(\mu)$. This correctly defines a decision rule as

$$\sum_{a \in A} \alpha(a|\mu) = \frac{\sum_{a \in A} f(\mu, a)}{\tau(\mu)} = 1$$

where the second equality is implied by Lemma 2. Furthermore, $\alpha$ is optimal for the agent because $\mu$ and $a$ are connected with an edge only if $a$ is optimal under $\mu$, i.e. $a \in a^*(\mu)$.

To verify that $(\tau, \alpha)$ induce $\nu_0$, note that for all $a \in A$

$$\sum_{\mu \in M_\tau} \tau(\mu)\, \alpha(a|\mu) = \sum_{\mu \in M_\tau} f(\mu, a) = \nu_0(a).$$

where the second equality follows again from Lemma 2. Thus, $\nu_0$ is consistent with $\tau$.

Conversely, suppose that $(\mu_0, \nu_0)$ are BCE-consistent. Then, a Bayes plausible distribution over posteriors $\tau$ and a decision rule $\alpha$ exists that induce an obedient experiment.[16] Define the graph $G_P(\tau)$ and the demand $d_{(\nu_0,\tau)}$. Note that the demand $d_{(\nu_0,\tau)}$ is feasible on $G_P(\tau)$ by defining the flow $f(\mu, a) = \alpha(a|\mu)\, \tau(\mu)$ for all $(\mu, a) \in M_\tau \times A$. $\square$

Proposition 4 implies that verifying that $\tau$ implements $\nu_0$ is equivalent to verifying the feasibility of the demand $d_{(\nu_0,\tau)}$ for the graph $G_P$. The main theorem in Gale (1957) provides necessary and sufficient conditions under which $d_{(\nu_0,\tau)}$ is feasible. Adapted to our setting, the conditions in Gale (1957) can be stated as follows:

**Proposition 5** (Gale, 1957). *The demand $d_{(\nu_0,\tau)}$ is feasible on graph $G_P(\tau)$ if and only if for every set $B \subseteq A$ a flow $f_B$ exists such that:*

1. $\sum_{a \in B} f_B(\mu, a) \leq \tau(\mu)$ *for all $\mu \in M_\tau$, and*
2. $\sum_{a \in B} \sum_{\mu \in M_\tau} f_B(\mu, a) \geq \sum_{a \in B} \nu_0(a)$.

Because of Lemma 2, given a set $B \subseteq A$ items 1 and 2 in Proposition 5 are satisfied for some flow $f_B$ if and only if they are satisfied when the out flow from every supply node that is connected to nodes in $B$ is maximal. Denote the set of all posterior beliefs in $M_\tau$ for which some action in $B$ is optimal (and perhaps also actions that are not in $B$) by $M^*(B) = \{\mu \in M_\tau \mid \exists a \in B, a \in a^*(\mu)\}$. Thus, in the graph we constructed, all and only beliefs (i.e., supply nodes) in $M^*(B)$ are connected to actions (i.e., demand nodes) in $B$. The next corollary follows immediately:

**Corollary 2.** *The Bayes plausible distribution over posteriors $\tau$ implements $\nu_0$ if and only if for every subset $B \subseteq A$,*

$$\sum_{\mu \in M^*(B)} \tau(\mu) \geq \sum_{a \in B} \nu_0(a). \tag{18}$$

[16] Namely, BCE-consistency implies the existence of an obedient experiment from which we can infer the following distribution over posteriors. First, let

$$\mu_a(\theta) = \frac{\mu_0(\theta)\pi(a|\theta)}{\sum_{\theta' \in \Theta} \mu_0(\theta')\pi(a|\theta')},$$

and let $\tau(\{\mu_a\}) = \sum_{\theta \in \Theta} \mu_0(\theta)\pi(a|\theta)$. The decision rule $\alpha(\cdot|\mu_a) = \mathbb{1}[a' = a]$ completes the construction.

To see that the condition in Corollary 2 is equivalent to that in Proposition 1, note first that because $\tau, \nu_0$ are measures (and hence add up to 1), Equation 18 can be equivalently written as follows:

$$\sum_{a\in\overline{B}} \nu_0(a) \geq \sum_{\mu\in\overline{M^*(B)}} \tau(\mu), \tag{19}$$

where the upper-bar notation denotes the complement of a set–for instance, $\overline{B} = A \setminus B$.

Note that

$$\overline{M^*(B)} = \{\mu \in M_\tau | a^*(\mu) \cap B = \emptyset\} = \bigcup_{C\subseteq\overline{B}} \{\mu \in M_\tau | a^*(\mu) = C\}.$$

Hence, we can write Equation 19 as follows

$$\sum_{a\in\overline{B}} \nu_0(a) \geq \sum_{C\subseteq\overline{B}} \sum_{\mu\in M_\tau : a^*(\mu)=C} \tau(\mu) = \sum_{C\subseteq\overline{B}} \tau_A(C), \tag{20}$$

which is the equation in Proposition 1.

## A.3 Menu-choice proof of Proposition 1

We provide here the omitted detail from the discussion about the connection between Proposition 1 and stochastic choice out of menus.

Suppose that Equation 7 holds for all $B \subseteq A$. Azrieli and Rehbeck (2022, Proposition 9) implies that a conditional probability system $\sigma' : 2^A \mapsto \Delta(A)$ exists such that for all $a \in A$

$$\nu_0(a) = \sum_{B:a\in B} \tau_A(B)\sigma'(a|B). \tag{21}$$

The intuition for such a result follows from an alternative graphical representation of the problem, depicted in Figure 2. Consider the following graph. Nodes are (i) the actions $a \in A$, (ii) the (non-empty) action subsets $B \subseteq A$ (i.e., the elements of $2^A \setminus \{\emptyset\}$), (iii) a source node $s$, and (iv) a sink node $t$. Edges are as follows. There is an edge of weight one between $a \in A$ and $B \subseteq A$ if and only if $a \in B$. There is an edge with weight $\nu_0(a)$ between the source $s$ and $a$. Finally, there is an edge between $B \subseteq A$ and the sink $t$ with weight $\tau_A(B)$. The condition in Equation 7 ensures that a feasible flow exists throughout the network.

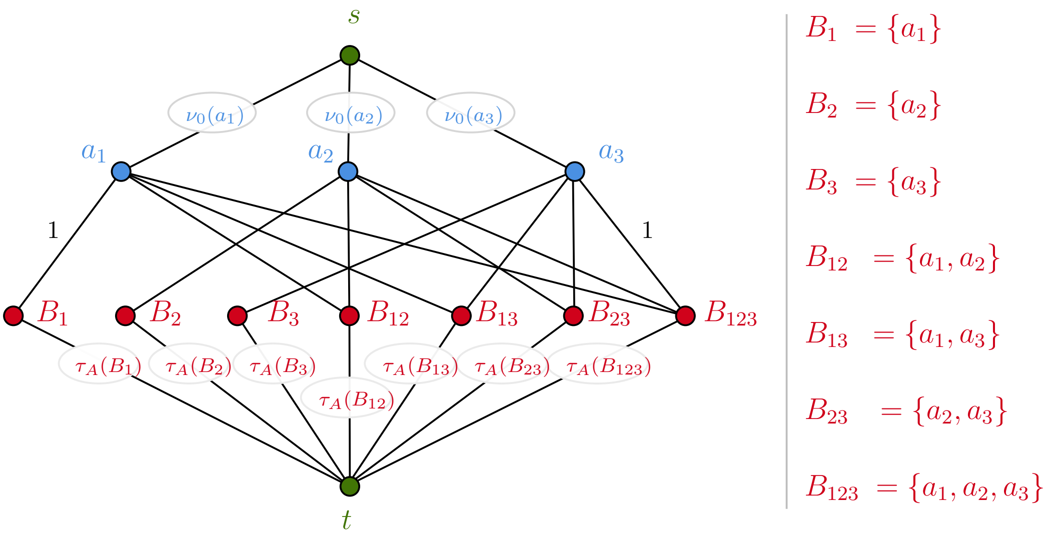


Figure 2: Graphical representation of the BCE-consistency problem with $3$ actions.

We use the conditional probability system to create a s stochastic choice rule $\sigma : \Theta \mapsto \Delta(A)$ as follows:

$$\sigma(a|\theta) = \sum_{B:a\in B} \sum_{\mu:a^*(\mu)=B} \frac{\mu(\theta)}{\mu_0(\theta)} \tau(\mu)\sigma'(a|B).$$

The experiment has an intuitive explanation: We first draw a subset of actions $B$ using the measure $\tau_A$ and then recommend to the agent which particular action she must take using the conditional probability system $\alpha'(\cdot|B)$.

Define the information structure, $\pi \in \Delta(A \times \Theta)$ by letting $\pi(a, \theta) = \mu_0(\theta)\sigma(a|\theta)$. To see that it has the desired properties, note first that

$$\sum_{a\in A} \sigma(a|\theta) = \sum_{a\in A} \sum_{B:a\in B} \sum_{\mu:a^*(\mu)=B} \frac{\mu(\theta)}{\mu_0(\theta)} \tau(\mu)\sigma'(a|B)$$
$$= \sum_{B\subseteq A} \left( \sum_{a\in B} \sigma'(a|B) \right) \sum_{\mu:a^*(\mu)=B} \tau(\mu) \frac{\mu(\theta)}{\mu_0(\theta)} = \sum_{B\subseteq A} \sum_{\mu:a^*(\mu)=B} \tau(\mu) \frac{\mu(\theta)}{\mu_0(\theta)} = 1$$

Second, note that

$$\sum_{\theta\in\Theta} \pi(a, \theta) = \sum_{\theta\in\Theta} \mu_0(\theta)\sigma(a|\theta) = \sum_{\theta\in\Theta} \sum_{B:a\in B} \sum_{\mu:a^*(\mu)=B} \mu(\theta)\tau(\mu)\sigma'(a|B)$$
$$= \sum_{B:a\in B} \sum_{\mu:a^*(\mu)=B} \left( \sum_{\theta\in\Theta} \mu(\theta) \right) \tau(\mu)\sigma'(a|B) = \nu_0(a),$$

by Equation 21.

Finally, note that the experiment is obedient: If $a$ is recommended with positive probability, then a set $B$ exists such that $a \in B$ and $\mu$ such that $a^*(\mu) = B$ is in the support of $\tau$, under which $a$ is optimal. Because $\sigma(a|\theta)$ is obtained by averaging over beliefs in which $a$ is optimal, it remains optimal.

## A.4 Proof of Proposition 2

*Proof of Proposition 2.* In the ring-network base game, for a joint distribution $\pi \in \Delta(A \times \Theta)$, the obedience constraints can be written as follows:

$$(\forall a_1, a_1' \in A_1) \sum_{\theta \in \Theta} \pi_{\Theta \times A_1}(a_1, \theta) \left(\tilde{u}(a_1, \theta) - \tilde{u}(a_1', \theta)\right) \geq 0$$

$$(\forall i \in \{2, \ldots, N\})(\forall a_i, a_i' \in A_i) \sum_{a_{i-1} \in A_{i-1}} \pi_{A_{i-1}, i}(a, \theta) \left(\tilde{u}(a_{i-1}, a_i) - \tilde{u}(a_{i-1}, a_i')\right) \geq 0,$$

where $\pi_{\Theta \times A_1}$ is the marginal of $\pi$ over $\Theta \times A_1$ and similarly for $i \geq 2$, $\pi_{A_{i-1} \times A_i}$ is the marginal of $\pi$ over $A_{i-1} \times A_i$. Thus, it is immediate that the conditions in Proposition 2 are necessary for $(\mu_0, \overline{\nu}_0)$ to be M-BCE-consistent.

For sufficiency, note that Theorem 1 implies that under the conditions of Proposition 2, $\left(\pi_{\Theta \times A_1}, \ldots, \pi_{A_{N-1} \times A_N}\right)$ exist each of which satisfy the respective marginal conditions and obedience constraints.

Given these distributions, define $\hat{\pi} \in \Delta(A \times \Theta)$ as follows: for each $(a, \theta) \in A \times \Theta$

$$\hat{\pi}(a, \theta) = \pi_{A_1 \times \Theta}(a_1, \theta) \pi_{A_1 \times A_2}(a_2 | a_1) \times \ldots \pi_{A_{N-1} \times A_N}(a_N | a_{N-1}), \tag{22}$$

where abusing notation we let for $i \geq 2$, $\pi_{A_{i-1} \times A_i}(\cdot | a_{i-1})$ denote the distribution $\pi_{A_{i-1} \times A_i}$ conditional on $a_{i-1}' = a_{i-1}$.

Note that $\hat{\pi}(a, \theta)$ satisfies the obedience constraints of player $1$ if and only if $\pi_{A_1 \times \Theta}(\cdot)$ does. Indeed, for all $a_1, a_1'$, we have

$$\begin{aligned}
&\sum_{a_{-1}, \theta} \hat{\pi}(a_1, a_{-1}, \theta) \left(\tilde{u}_1(a_1, \theta) - \tilde{u}_1(a_1', \theta)\right) = \\
&\sum_{\theta} \pi_{A_1 \times \Theta}(a_1, \theta) \left(\tilde{u}_1(a_1, \theta) - \tilde{u}_1(a_1', \theta)\right) \sum_{(a_2, \ldots, a_N)} \prod_{i=2}^{N} \pi_{A_{i-1} \times A_i}(a_i | a_{i-1}) = \\
&\sum_{\theta} \pi_{A_1 \times \Theta}(a_1, \theta) \left(\tilde{u}_1(a_1, \theta) - \tilde{u}_1(a_1', \theta)\right).
\end{aligned} \tag{23}$$

Consider now player $i \geq 2$. For simplicity, fix $i = 2$–the rest of the players follow immediately. Then, let $a_2, a_2' \in A_2$. We want to check that $\pi$ satisfies the obedience constraint of player 2 if and only if $\pi_{A_1 \times A_2}$ does.

$$
\begin{aligned}
&\sum_{a_{-2},\theta} \hat{\pi}(a_2, a_{-2}, \theta) \left(\tilde{u}_2(a_1, a_2) - \tilde{u}_2(a_1, a_2')\right) = \\
&\sum_{a_1,\theta} \pi_{A_1\times\Theta}(a_1,\theta)\pi_{A_1\times A_2}(a_2|a_1) \left(\tilde{u}_2(a_1, a_2) - \tilde{u}_2(a_1, a_2')\right) \sum_{(a_3,\ldots,a_N)} \prod_{i=3}^{N} \pi_{A_{i-1}\times A_i}(a_i|a_{i-1}) = \\
&\sum_{a_1\in A_1} \left(\sum_{\theta} \pi_{A_1\times\Theta}(a_1,\theta)\right) \pi_{A_1\times A_2}(a_2|a_1) \left(\tilde{u}_2(a_1, a_2) - \tilde{u}_2(a_1, a_2')\right) = \\
&\sum_{a_1\in A_1} \nu_{01}(a_1)\pi_{A_1\times A_2}(a_2|a_1) \left(\tilde{u}_2(a_1, a_2) - \tilde{u}_2(a_1, a_2')\right) = \\
&\sum_{a_1\in A_1} \pi_{A_1\times A_2}(a_1, a_2) \left(\tilde{u}_2(a_1, a_2) - \tilde{u}_2(a_1, a_2')\right),
\end{aligned}
$$

where the third equality follows from the assumption that $\pi_{A_1\times\Theta}$ satisfies the marginal constraints for player 1. $\square$