Mechanism Design for an Agnostic Planner: universal mechanisms, logarithmic equilibrium payoffs and implementation

Jakša Cvitanić *, Dražen Prelec †, Sonja Radas ‡ and Hrvoje Šikić §

April 25, 2015

Abstract

We consider the problem of Bayesian mechanism design when the respondents share a common prior about which the survey planner is agnostic. This assumption is natural when designing incentives for surveys of public or expert opinion, because the survey planner may be a relatively ignorant outsider, or, even if informed, might prefer not to insert personal beliefs about the prior into the incentive structure. The planner would therefore like to have a universal mechanism, which would induce honest answers for all possible priors. Evidently, such mechanisms must

---

*Division of the Humanities and Social Sciences, Caltech. E-mail: cvitanic@hss.caltech.edu. Research supported in part by NSF grant DMS 10-08219.

†MIT, Sloan School of Management, Department of Economics, Department of Brain and Cognitive Sciences. E-mail: dprelec@mit.edu. Supported by Intelligence Advanced Research Projects Activity (IARPA) via the Department of Interior National Business Center contract number D11PC20058. Disclaimer: The views and conclusions expressed herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

‡The Institute of Economics, Zagreb. E-mail: sradas@eizg.hr. This research was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme, PIOF-GA-2013-622868 - BayInno.

§University of Zagreb, Faculty of Science, Department of Mathematics. E-mail: lsikic@math.hr. Research supported in part by the MZOS grant 037-0372790-2799 of the Republic of Croatia an in part by Croatian Science Foundation under the project 3526.
request from players more information than merely reports of their types. We prove that, if a mechanism results in a type-separating equilibrium and satisfies a locality condition, then the players are necessarily ranked according to their posterior on the states of nature. We further identify conditions under which payoffs in equilibrium must be logarithmic functions of the posterior probabilities, even when those probabilities are not explicitly elicited by the mechanism. Logarithmic equilibrium payoffs can be implemented by the agnostic planner via the Bayesian Truth Serum algorithm of Prelec (2004).

*Key words:* robust/universal mechanisms, proper scoring rules, Bayesian Truth Serum, mechanism implementation, ranking experts
1 Introduction

Consider a multiple-choice question that might be posed in a study of public or expert opinion, such as a survey of voting intentions, product ratings, or expert scientific forecasts. In interpreting the survey result, two generic issues arise: First, are respondents providing careful and honest answers? Second, even if everyone answers honestly, are some answers more informative than others? We may refer to the first issue as one of truthfulness, and the second as one of aggregating answers to arrive at (impersonal) truth. In this paper, we are largely concerned with the first issue, which connects our results to the literature on robust mechanism design. However, our results are also relevant to the second issue, in that we will show that incentive-compatibility under some reasonable conditions leads to equilibrium payoff systems that necessarily reward the more informed players.

Our basic assumption is that the prior over types is common knowledge among the players, but is not used by the survey planner in designing the survey. We present this as a methodological rather than a substantive requirement: Although the planner may have some beliefs about the prior, she may prefer to keep these beliefs private and adopt the position of an agnostic/neutral outsider, not imposing her conjectures on the survey respondents. Thus, she is interested in a 'universal' mechanism, one that would work for all priors, in the absence of any input from the planner, apart from the initial formulation of the multiple-choice question.

We also assume that the common prior over players’ types is exchangeable with respect to player identities, which in the large sample limit implies that types are independent conditional on a state-of-nature variable (this follows from de Finetti’s theorem). In an infinite sample, players’ posteriors over possible distributions of types may therefore be identified with their posteriors over hypothesized states-of-nature.

The first result we obtain is that strictly separating equilibria that satisfy a locality condition necessarily rank players according to their posteriors over states-of-nature, as defined above. That is, incentive-compatibility requires that the rules of the game produce a result (in equilibrium) that favors players with superior information, where quality of information is defined by their posteriors.\footnote{The locality condition requires that, in equilibrium, a player’s ex-post payoff depends only on the posteriors and the probabilities of players’ types.}

The second question we analyze is: under what conditions only logarithmic payoffs arise in a strictly
separating equilibrium? We show that this is the case under an assumption on the sensitivity of the payoff difference of two players on other players. We point out that we get this uniqueness result by restricting effects of the mechanism on the respondents only in equilibrium, while we impose no assumptions on the off-equilibrium effects.

While a particular payoff form based on the players' posteriors, such as logarithm, may arise in equilibrium, it is not necessarily simple to implement it in practice. That is, the problem is how to compute the theoretically optimal payoff score using only the players’ responses to a questionnaire designed by the agnostic planner, while having the questionnaire as simple as possible. We show that the payoffs of all strictly-separating equilibria can be implemented by particular questionnaires, but the latter may be complex. However, Prelec (2004) introduced an algorithm, called Bayesian Truth Serum (BTS), for incentivizing honest answers on the basis of simple inputs: along with their personal answer, the respondents are also asked to predict the percentage distribution of answers in the sample. The article also showed that the BTS scores (arising in equilibrium) are equivalent to budget-balanced logarithmic equilibrium payoffs, but the connection between this result and individual expertise was left open. Subsequently, Prelec, Seung and McCoy (2013) defined and tested experimentally a broader class of "least-surprised-by-truth" (LST) algorithms that all use predicted distributions of answers to produce a ranking of experts according to their posteriors. Within this class, only BTS was known to be incentive-compatible. Additional incentive-compatible mechanisms are studied in Cvitanić and Prelec (2014).

The problem we tackle in the paper can be considered as one of mechanism design, since we seek to describe mechanisms that are both incentive-compatible and have attractive features for opinion elicitation applications. In one way, our approach is more general than typical mechanism design models because we allow for both uncertainty regarding the players information (type), and uncertainty regarding the true state of nature, and those two may be correlated in a nontrivial way. It is exactly the joint distribution of the two that drives all the results. We do not assume that the planner knows the joint distribution, so that the mechanism has to work for any such distribution. In this sense, ours is a study of robust Bayesian mechanisms. On the other hand, our setup is less general in another way – the players do not choose actions other than reporting their responses, which is assumed to be costless. Thus, there is no modeling of utility/disutility drawn from actions, the

\footnote{See, e.g., Maskin and Sjöström (2001), Bergemann and Morris (2011), Börgers (2013).}
only utility the players draw is from the expected payoff they attain. Moreover, our framework is less general than some models of robust mechanism design that, unlike ours, do not assume common knowledge of the prior distribution by all the players. (In our case only the planner may be ignorant.) We discuss in the conclusions section in what directions one could try to extend our results.

Our paper is also related to the literature on proper scoring rules, that is, the rules for rewarding players to predict probabilities of particular events, and to do that truthfully. Logarithmic scoring has been studied extensively as a proper scoring rule in the case in which the planner knows the prior distribution of the player types; see Miller, Resnick, and Zeckhauser (2005) for a clever use of logarithmic scoring in such a framework, and Offerman, Sonnemans, Van De Kuilen and Wakker (2009) for further references. Our contributions to this literature are to show that truth-inducing payoffs can be implemented even when the planner is agnostic about the prior, and that logarithmic scoring is the only possible equilibrium payoff form under weak assumptions.

The rest of the paper is organized as follows: Section 2 introduces the model, Section 3 introduce the concept of posterior locality and presents the main theoretical results, Section 4 discusses implementation issues, and we conclude in Section 5. Most of the proofs are delegated to the appendix.

## 2 Model, Definitions and Assumptions

In our model mechanisms consist in giving scores to the players (respondents) of different types. Applications we have in mind are of the polling type: the respondents are asked to provide responses to queries assigned by a survey planner. The planner is interested in eliciting truthful opinions to a multiple choice question, and ranking players according to the quality of their information, which in our framework, will mean according to their posterior probabilities of the true state of nature. For instance, the planner might be interested in the value of a certain painting some years into the future, and asks experts to respond to appropriately designed questions.

---

3Recently, Witkowski and Parkes (2012) study a framework with only two types and without common prior, and Waggoner and Chen (2013) a general framework without assumptions on information structure.

4A negative of a score is usually called a transfer in the mechanism design literature; see, e.g., Börgers (2013).

5Prelec (2004) and Prelec, Seung and McCoy (2013) provide many more examples.
2.1 Goals and difficulties

We have two goals in mind: (i) to study implications of the truth-telling property (more generally, the separation property) on equilibrium payoffs, for which the assumption of infinite number of players will help us to obtain a surprisingly precise description; (ii) to study implementation issues.

The respondents are assumed to have a common prior, but the survey questions are not allowed to depend on the prior. Let us start by showing that if the planner only asks the respondents to declare their type, and nothing else, then we cannot have truth-telling for all priors. Suppose there are only two choices, “yes”, Y, and “no”, N. Suppose x% of the respondents choose Y, and (1-x)% choose N, and suppose the corresponding payoffs are $f_Y(x)$ and $f_N(x)$. Let us assume that these are truth-telling scoring rules for all priors. Denote by $p(x)$ the prior on state $x$. The posterior distribution is

$$Pr(\text{state}|\text{type}) = Pr(\text{type}|\text{state})Pr(\text{state})/Pr(\text{type})$$

If state $x$ is the realized state, then, if we assume that the payoffs are truth-telling, for type Y this is equal to $xp(x)/Pr(Y)$, and for type N it is equal to $(1 - x)p(x)/Pr(N)$. Then, in order for a respondent who sees Y not to report N, we need to have (where the integral can be a discrete sum)

$$\int_0^1 x[f_Y(x) - f_N(x)]p(x)dx \geq 0$$

and in order for a respondent who sees N not to report Y, we need to have

$$-\int_0^1 (1 - x)[f_Y(x) - f_N(x)]p(x)dx \geq 0$$

Take a prior that is not zero only for $x$ for which $f_Y(x) - f_N(x) > 0$. Then, the former inequality is satisfied, but the latter is not. Consequently, we cannot have scoring rules that are truth-telling for all priors if they are based only on type declaration. The players need to provide some additional information to the scoring system.

In the next couple of sections, we will not restrict in any way the questions the planner can ask to obtain this additional information. For example, the planner could ask each player to state the whole prior distribution, and harshly penalize the player who gives a response different from others. What we will restrict is the form of the equilibrium payoffs, and we will provide an analysis of their properties, regardless of the mechanism that leads to equilibrium. Then, in the implementation section, we will address difficulties of implementing equilibrium payoffs. For example,
asking for the common prior is unlikely to work in practice - more likely than not, most responses would be different from each other, and the planner would have to penalize harshly most respondents. Similarly, even when identifying states of nature with possible empirical frequencies of responses, asking only about posterior probabilities of state of nature is prohibitively complex, because it would require respondents to provide a density over all possible empirical frequencies. Fortunately, as we shall see in the implementation section, there are equilibrium payoffs that can be implemented by asking respondents only what they think is the percentage of other types in the population, in addition to asking them to declare their own type.

2.2 The model

The players are indexed by \( r \in R \), where \( R \) is infinite and countable.\(^6\) The state of nature is a random variable \( \Omega \), taking values in \( \{1, \ldots, N\} \), \( N > 1 \).\(^7\)

The players can be of \( M > 1 \) different types, that can be interpreted as random signals the players receive about the state of nature. Player \( r \)'s type is a random variable denoted \( T^r \), and it takes values \( t^r \in \{1, \ldots, M\} \).

We posit the following

**Assumption 2.1** The family of signals \( T^r, r \in R \), is a family of exchangeable random variables. Moreover, random variables \( T^r, r \in R \), are i.i.d. conditional on the state of nature \( \Omega \).

Observe that one obvious consequence of this assumption is that the order in which we consider our players is irrelevant (from the point of view of the probability distribution of the entire sequence).

\(^6\)We need the assumption that there are infinitely many players for several reasons: first, we don’t want to impose assumptions on the form of the payoffs outside of equilibrium; for this, we will use the fact that, with infinite number of players, the form of equilibrium payoff does not change when a player of one type mimics the equilibrium strategy of another type; second, achieving truth-telling of types is much harder with finitely many players, and so is the implementation of equilibrium payoffs using practical inputs. We postpone to future research the analysis of the setup with finitely many players; finally, we need the infinite number of players because we invoke de Finetti’s theorem in our model setup.

\(^7\)Strictly speaking, this is only an approximation for most applications, in which the state of nature could naturally have a continuous range of values. For instance, in the example about a painting’s value, the state of nature could be the percentage of experts who believe the painting is worth more than one million dollars.
Moreover, by de Finetti’s theorem, the exchangeability assumption actually implies the second part of Assumption 2.1, that there exists a random variable $\Omega$ such that $T^r_k$’s are conditionally i.i.d. with respect to $\Omega$; see, e.g., Aldous (1985), or Chow and Teicher (1997).

The joint distribution of types and states of nature is given by an $M \times N$ matrix $Q = [q_{ki}]$, where

$$q_{ki} = Pr(T^r = k, \Omega = i).$$

Note that $Q$ does not actually depend on $r$, a consequence of the exchangeability assumption.

We assume that the matrix $Q$ is common knowledge among the players, but not used by the planner when designing the survey. In fact, the planner does not even need to know the number of the states of nature $N$. The only thing we assume is that the planner uses is $M$, needed for implementation using a multiple choice question – the planner has to offer exactly as many possible choices as there are types.

Matrix $Q$ determines the marginal probabilities of types, referred to as type probabilities, and the probabilities of states of nature given the type, referred to as posteriors. They are denoted

$$s_k = Pr(T^r = k)$$

and

$$z^i_k = Pr(\Omega = i \mid T^r = k).$$

The posteriors form a matrix $Z = [z^i_k]_{k=1}^M_{i=1}^N$. Note that $z^i_k$ does not depend on $r$, that for every $k \in \{1, \ldots, M\}$, we have $\sum_{i=1}^N z^i_k = 1$, and that any matrix with this property can be represented as a $Z$-matrix of posteriors for some joint distribution $Q$. We denote the vector $(s_1, \ldots, s_M)$ by $S$.

### 2.3 Definitions and assumptions

Henceforth, we fix the number of types $M$. We consider scoring mechanisms in which, for a given fixed positive integer $K$, player $r$ submits as a response a $K$-dimensional value $a^r \in \mathbb{R}^K$ ($a$ for “action”), from a set of at least $M$ possible responses.\(^8\) A response $a^r$ would typically include a declaration of a respondent’s type (choosing an answer to a multiple choice question), and, as the discussion of Section 2.1 shows, it would have to include responses to some other questions in order to be truth-inducing.\(^9\)

---

\(^8\) We allow only for pure strategies.

\(^9\) In the section on implementation, we will see that another question might be about the percentage of other respondents choosing a specific choice from the multiple choice list.
A pure strategy for player \( r \) is a map \( \sigma^r(t^r) \), that maps a player’s type to his response choice \( a^r \). The profile of all respondents’ pure strategies is denoted \( \sigma(t) \), with entries \( \sigma^r(t^r) \), and the profile excluding player \( r \) is denoted \( \sigma^{-r}(t^{-r}) \). The score for player \( r \) is given by a scoring function \( f(\sigma^r(t^r), \sigma^{-r}(t^{-r})) \) that takes the responses to the set of real numbers, and \( f(\cdot, \cdot) \) is of the same functional form for all \( N \) and \( Q \). Moreover, we assume that if respondent \( r \) chooses response \( a^r \), and the remaining responses are represented by \( a^{-r} \), then his score is given by \( f(a^r, a^{-r}) \), where the order of different respondents’ responses in \( a^{-r} \) does not matter, that is, \( f \) is symmetric in those. This condition is a natural restriction considering that the planner does not make a distinction between different types, assumed exchangeable.

We will often assume that the planner pays zero in total, that is, that \( f \) satisfies

**Definition 2.1** The scoring function \( f \) is said to be budget-balanced if the sum of the scores of all the players is equal to zero, with probability one.

We assume that the players are risk-neutral and maximize the expected score value. We will consider only those scoring functions \( f \) that allow for a unique equilibrium, as defined next.

**Definition 2.2**

- (i) Given a prior matrix \( Q \), we say that a scoring function \( f \) allows a Strict (Bayesian) Nash Equilibrium (SNE) if there exists a strategy \( \sigma = \sigma_Q \) such that for all \( r, t^r, t^{-r}, t^s \), we have: for an arbitrary potential response choice \( a^r \neq \sigma^r(t^r) \), we have, with expectation taken with respect to the (conditional) distribution of \( \Omega \),

\[
E[f(a^r; \sigma^{-r}(t^{-r})) \mid T^r = t^r] < E[f(\sigma^r(t^r), \sigma^{-r}(t^{-r})) \mid T^r = t^r]
\]

The strategy profile \( \sigma \) is called an SNE. If the equilibrium is also separating, that is, if, in addition to the above, we also have \( \sigma^r(t^r) = \sigma^s(t^s) \Rightarrow t^r = t^s \), we call \( \sigma \) a Strictly Separating (Bayesian) Nash Equilibrium (SSNE).

\(^{10}\)Typically, mechanism design models consider only the types as being random, according to a prior which is known also to the planner. Our model is more general by considering random states of natures in addition to random types, with a non-degenerate correlation between the two. On the other hand, it is less general in that the players do not choose actions other than choosing a response, and thus, they draw no utility/disutility other than from the expected score value.
- (ii) We say that a scoring function $f$ is an Universal Separating Scoring Rule (USSR), if for all
$Q$ it allows at least one budget-balanced SNE $\sigma_Q$, if every budget-balanced SNE is an SSNE, and if any
two budget-balanced SSNE's $\sigma_Q$ and $\sigma'_Q$ result in the same scores: $f(\sigma_Q) = f(\sigma'_Q)$.

**Remark 2.1** Condition (ii) essentially assumes uniqueness of the budget-balanced SSNE $\sigma_Q$. We will
show later below that the budget-balanced logarithmic scoring can be implemented by an USSR for which
the above uniqueness holds. We also note that when there is more than one SSNE, our results still
hold for those SSNE's that satisfy the assumptions we impose.

From now on, we only consider USSR functions $f$.

### 3 Properties of Equilibrium Payoffs Based on Local Posteriors

#### 3.1 Posterior locality

We now further restrict the family of scoring functions, by imposing conditions on what kind of payoffs
they allow in equilibrium, but we impose no constraints on the off-equilibrium payoffs to the players
(including potentially high penalties). With that in mind, note that by the above definition, an USSR
score of a certain type is the same in any budget-balanced SNE. In what follows we will use notation
$F_i$ for such equilibrium score in state $i$, and the same notation for non budget-balanced versions of the
equilibrium payoffs. We introduce the following assumption on the equilibrium payoffs:

**Assumption 3.2**

- (i) **Posterior Locality.** $(\forall k \in \{1, ..., M\})$ and $(\forall i \in \{1, ..., N\})$, and $\forall j \neq k$, if $T^r = k$, and
$\Omega = i$, the equilibrium score of player $r$ has the representation, with $F_i : (0,1)^{2M} \rightarrow \mathbb{R},$

$$f(\sigma^r_Q(k), \sigma^{-r}_Q(T^{-r})) = F_i(z^i_k, z^i_{-k}; s_k, s_{-k})$$

where, for example, $z^i_{-k} = (z^i_1, ..., z^i_{k-1}, z^i_{k+1}, ..., z^i_M)$.

- (ii) **Symmetry.** The following symmetry holds: $(\forall x, y \in (0,1)) (\forall z_2, ..., z_M, s_2, ..., s_M \in (0,1))$
$(\forall$ permutation $\pi$ of $\{2, ..., M\}$), we have

$$F_i(x, z_2, z_3, ..., z_M; y, s_2, ..., s_M) = F_i(x, z_{\Pi(2)}, z_{\Pi(3)}, ..., z_{\Pi(M)}; y, s_{\Pi(2)}, s_{\Pi(3)}, ..., s_{\Pi(M)})$$

8
We call the family \( \{F_i\} \) a Posterior-Local Equilibrium Payoff System (PLEPS), and we say that an SSNE \( \sigma_Q \) is "realized" via the PLEPS \( \{F_i\} \).

The terminology "locality" refers to the fact that the above assumption implies that the scores of two players differ in state \( i \) that does occur if they have different beliefs about state \( i \), but not if they have different beliefs about other, counterfactual states. This is partly motivated by the desire to have payoffs that are implementable by simple inputs; see the implementation section below. The conditions require that the score of a player in equilibrium depends crucially on the player’s local posterior. This is justified if the planner considers the local posterior as being a measure of the player’s expertise, and if the planner wants to assign higher scores to better experts: as we shall see in the next subsection, under our assumptions the players’ scores will be increasing in their local posteriors. Also, we look for the simplest possible equilibrium payoffs that describe players’ expertise, which is why the payoff \( F \) does not depend on other local probabilities that can be derived from the prior. This is also reminiscent of the concept of maximum likelihood estimators, that maximize the probability of the event that does actually occur.

The symmetry assumption is consistent with the exchangeability.

Our benchmark example of an SSNE realized by a PLEPS will be

\[
F_i(z^i_k, z^i_{-k}; s_k, s_{-k}) = \log(z^i_k)
\]

Note that there might exist SSN equilibrium payoffs that do not satisfy posterior locality; for example, it can be shown that the following is an SSNE payoff, paid to the player of type \( k \) in state \( i \):

\[
\sum_{j=1}^M \Pr(T^s = j \mid \Omega = i) \log \Pr(T^s = j \mid T^r = k)
\]

However, this payoff is not monotone in the local posterior; see Cvitanić and Prelec (2014).

### 3.2 Mimicking Payoffs and Type Separation in Equilibrium

The following result is simple, but crucial for our results. It tells us what the score looks like for the type who mimics another type’s equilibrium strategy. We emphasize that we need infinite number of players for this result.
Proposition 3.1 Consider an SSNE $\sigma$ realized via a PLEPS $\{F_i\}$. Then, if a respondent of type $k$ deviates from the equilibrium by using the strategy of type $j \neq k$, his payoff is equal to $F_i(z^i_j; z^i_{\{\bar{j}\}}; s_j, s_{\{\bar{j}\}})$. That is, if a player of type $k$ mimics the equilibrium strategy of type $j$, then his payoff is given by the equilibrium evaluation corresponding to type $j$.

Proof: Let $\sigma$ be an SSNE strategy profile, and define $\rho$ to be the strategy profile that is identical to $\sigma$, except that a specific player $r$ plays the strategy $\sigma^r(j)$ corresponding to type $j$, independent of his type. Suppose $k \neq j$, and consider a player $s$ of type $j$. Then, we have that the payoff to the mimicry strategy, when $r$ plays $j$ even though his type is $k$, is

$$f(\rho^r(k), \rho^{-r}(T^{-r})) = f(\rho^s(j), \rho^{-s}(T^{-s}))$$

$$= f(\sigma^s(j), \rho^{-s}(T^{-s}))$$

$$= f(\sigma^s(j), \sigma^{-s}(T^{-s}))$$

because $\sigma^{-s}(T^{-s})$ and $\rho^{-s}(T^{-s})$ differ only in $r$’s response, and this does not matter with infinitely many players. This is because every type will be represented by infinitely many players, and $f$ is symmetric in their responses. Hence, SSNE payoffs set by $\sigma$ determine the payoff of a mimicry deviation by player $r$.

\[\blacksquare\]

This form of the mimicking payoff in equilibrium, implies that, since we consider strict equilibria, the PLEPS $\{F_i\}$ has to satisfy the following strict separation inequality between different types:

$(\forall Z - \text{matrix}) (\forall S - \text{vector}) (\forall k, j \in \{1, ..., M\} \text{ such that } (z^1_k, ..., z^N_k) \neq (z^1_j, ..., z^N_j))$

$$\sum_{i=1}^{N} z^i_k F_i(z^i_k, z^i_{\{\bar{k}\}}; s_k, s_{\{\bar{k}\}}) > \sum_{i=1}^{N} z^i_j F_i(z^i_j, z^i_{\{\bar{j}\}}; s_j, s_{\{\bar{j}\}}) \quad (3.1)$$

All of our results in present Section 3 require this inequality. However, we reiterate once again that the scoring rule $f(\cdot)$ is not required to assign scores of the form $F_i(z^i_k, z^i_{\{\bar{k}\}}; s_k, s_{\{\bar{k}\}})$ outside of equilibrium.

We emphasize that the above inequality relies crucially on assuming infinite number of players, because that allowed us to conclude that the payoff of a type mimicking another type is still given in the form $F_i$. Without infinite number of players, the payoff of a mimicking type may have a much less
simple representation. Moreover, we have the following negative result when the number of players is finite.  

**Proposition 3.2** Assume a finite number of players. Then, there exists no budget balanced PLEPS that strictly separates the types.

### 3.3 SSNE rankings

We now state our first main result: in any SSN equilibrium realized via a PLEPS, the equilibrium payoffs necessarily rank the players according to the relative ranking of the corresponding posteriors. That is, when using a scoring system resulting in such an equilibrium, the planner’s objective is automatically satisfied: she will know which players are better experts than others, if she considers the level of the posterior equivalent to the level of expertise. If they were not ranked by their posteriors, before the game players might want to avoid collecting information before the game starts to improve their knowledge about the true state of nature, which is undesirable.

In the SSNE associated with a PLEPS \( \{F_i\} \), the types will be ranked according to the size of the values \( F_i(z^i_k, z^i_{-k}; s_k, s_{-k}), k = 1, \ldots, M \). Thus, the following statement is equivalent to ranking according to the posteriors.

\[
\text{If } j, k \in \{1, \ldots, M\} \text{ and } z^j_k > z^j_j, \text{ then } F_i(z^j_k, z^j_{-k}; s_k, s_{-k}) > F_i(z^j_j, z^j_{-j}; s_j, s_{-j}) \quad (3.2)
\]

The main result of this section is

**Theorem 3.1** In an SSNE realized by a PLEPS \( \{F_i\} \), the players’ scores are strictly increasing in their posterior probabilities of the true state of nature. That is, functions \( F_i \) satisfy inequality (3.2) (for any prior distribution matrix \( Q \)).

Put differently, if the planner wants to determine relative expertise of players receiving exchangeable signals, it is sufficient to design a scoring system which allows only for equilibria that are realized via a PLEPS.

Thus, inequality (3.1) not only guarantees strict separation of types, but also has the posterior-based ranking as a direct consequence. The intuition behind this result is that if the posterior probability of type \( A \) of a state was higher than the one of type \( B \), but type \( A \)'s score in that state was

\[^{11}\text{We leave a more thorough analysis of that case for a future study.}\]
lower, then, he would be better off pretending to be type $B$. To be more precise, consider the case with only two types, $A$ and $B$, and two states of the world, 1 and 2. Denote by $p_A$ and $p_B$ the posterior probabilities of state 1, and suppose, without loss of generality, $p_A > p_B$. There are only two possible SSNE scores in each state $i$, denoted $F_i(p_A, p_B)$ and $F_i(p_B, p_A)$ (suppressing dependence on $S$ vector). Denote by $D_i$ the difference in scores, $D_i^A = F_i(p_A, p_B) - F_i(p_B, p_A)$. The claim is that, in an SSNE, type $A$’s higher posterior probability of state $i$ implies higher score in that state, that is, positive $D_i^A$. To argue this, note first that by the strict separation inequality, player $A$’s expected value of the differences in scores, that is, the weighted average of $D_1^A$ and $D_2^A$ with weights $p_A$ and $1 - p_A$, is positive. By the same token, the weighted average of $D_1^A$ and $D_2^A$ with weights $p_B$ and $1 - p_B$, is negative. The only way this can be possible when $p_A > p_B$ (thus also $1 - p_A < 1 - p_B$) is that $D_1^A > 0$ and $D_2^A < 0$. Thus, indeed, the type with higher posterior probability of a state receives higher score in that state. Or, put differently, if the type with higher posterior probability of a state does not receive higher score in that state, he would adopt the other type’s strategy. In Appendix, we state and prove the above simple argument in a lemma, and extend it to any number of types and states.

## 3.4 Possible equilibrium payoffs

In this subsection we present examples of PLEPS’s and identify assumptions under which logarithmic equilibrium payoffs (EP’s) are the only ones arising in SSNE’s.

### 3.4.1 Logarithmic equilibrium payoffs

The canonical example of a PLEPS (ignoring budget-balancing) is the logarithmic function:

$$F_i(z^i_k; z^i_{-k}; s_k, s_{-k}) = \log(z^i_k)$$

More precisely, a player’s equilibrium payoff is the logarithm of the posterior probability of the state of nature given his type. It is well known and straightforward to verify that this, indeed, satisfies the strict separation inequality (3.1). This is because of the well known Gibbs inequality which says that for a probability vector $(p^1, \ldots, p^N)$, we have

$$0 = \min_{q' \geq 0, \sum_q' = 1} \sum_{i=1}^N p^i [\log(p^i) - \log(q^i)]$$

(3.3)
This can be verified by noting that, with $\lambda$ being a Lagrange multiplier for the constraint $\sum_i q^i = 1$, the first order conditions for the problem

$$0 = \min_{\mathbf{q}} \left\{ \sum_{i=1}^{N} p^i [\log(p^i) - \log(q^i)] + \lambda \sum_i q^i \right\}$$

(3.4)

are $p^i/q^i = \lambda$, thus satisfied with $q^i = p^i$.

The question arises whether the log function is the only PLEPS (modulo budget balancing). The answer is negative in general, and we present a counterexample in what follows. Later below, we provide conditions under which logarithmic equilibrium payoffs are, in fact, the only possible payoffs of an SSNE.

### 3.4.2 Other examples of PLEPS’s

Let us first note that there are variations of the log EP’s that produce equivalent scores when we require budget balance. For instance, if we set, for some function $G$ symmetric in all the arguments, and some constant $K$, suppressing the dependence on the state of nature $i$,

$$F(z_k, z_{-k}) = \log(z_k) - K \sum_{j \neq k} \log(z_j) + G(z_1, \ldots, z_M)$$

then, function $F$ corresponds to a PLEPS, as can be verified in the same way as for the problem (3.4). However, it is not really different from logarithmic EP’s if we insist on budget balance, because, as is straightforward to check, if we add the constant term that makes it budget-balanced, we get the same EP’s as for the budget-balanced logarithmic EP’s.

We now present a PLEPS that has higher order terms that make it distinct from the logarithmic PLEPS, even if we make it budget-balanced.

#### Example 3.1

Consider the case with three types, $M = 3$, and denote

$$p^i = z^i_k, \quad (q^i, r^i) = z_{-k}^i$$

Define the following function:

$$F(p, q, r) = K \log(p) + p^4 - 2p^3(q + r) - 6p(qr^2 + q^2 r)$$
It is straightforward to verify that, for large enough $K$, this function satisfies the strict separation inequality (3.1). This is because the first order conditions (FOC’s) for the Lagrangian optimization problem

$$\min_q \left\{ \sum_i p^i [F(p^i, q^i, r^i) - F(q^i, p^i, r^i)] + \lambda \sum_i q^i \right\}$$

are, denoting with $F_x$ the derivative with respect to $x$ argument,

$$p^i [F_p(q^i, p^i, r^i) - F_q(p^i, q^i, r^i)] = \lambda$$

for some Lagrange multiplier $\lambda$, and these FOC’s are satisfied for the above function with $q^i = p^i$. For large enough $K$, the FOC’s are also sufficient conditions for optimality because the second order optimality conditions will also be satisfied, which implies that (3.1) is satisfied.

**Remark 3.2** We make an important observation here that, even if a PLEPS does not lead to logarithmic EP’s, the difference in equilibrium scores of two players with posteriors $p$ and $q$, respectively, has to be proportional to $\log(p) - \log(q)$ for $q \approx p$, up to the first order. To explain what we mean by that, consider, for simplicity of notation, the case with three types, and use the same notation $p^i, q^i, r^i$ as above. For fixed $p$ and $r$, suppressing dependence on $i$, expanding the score difference up to the first order as a function of $q$ around the point $p$, we have, for a Lagrange multiplier $\lambda$,

$$F(p, q, r) - F(q, p, r) \approx [F_q(p, p, r) - F_p(p, p, r)](q - p) = \lambda(1 - \frac{q}{p}) \approx \lambda(\log(p) - \log(q))$$

where the two approximate equalities come from the first order Taylor expansion terms, and the equality is due to the appropriate version of the first order conditions as stated in Lemma 3.1 below, for the problem $\min_q \{ \sum_i p^i [F(p^i, q^i, r^i) - F(q^i, p^i, r^i)] + \lambda \sum_i q^i \}$. Thus, even though there are “strange” PLEPS functions $F$ as in the example above, for all of them the difference in two EP’s is proportional to the difference of logarithmic payoffs, up to the first order. This is also true if $F$ depends on type probabilities $s_k$, under the conditions of Lemma 3.1.

We next identify conditions under which there can be no second-order terms, and the budget-balanced logarithmic EP is the only budget-balanced PLEPS.
3.4.3 When are equilibrium payoffs logarithmic?

We assume in this section that \( N \geq 3 \).\(^{12}\) As we have just shown, the difference in SSNE scores of two types is equal to the difference of the log scores up to the first order. We will now find conditions under which the higher order terms cannot appear, and under which any SSNE realized by a PLEPS results in log EP’s.

We do the following:

- (i) we first state an assumption on the second order mixed derivative of the difference in equilibrium scores of two types;
- (ii) we then show that the assumption implies an additive representation of the EP of a given type – the EP is a sum of a term that does not depend on the posteriors of other types and a term that is symmetric in types.
- (iii) finally, we show that such additive representation is sufficient to imply log EP’s, under a smoothness assumption.

For ease of notation we continue assuming \( M = 3 \), and use the above notation \( p^i, q^i, r^i \) for the posteriors of the three types. Also denote by \( s_p, s_q, s_r \) the corresponding type probabilities.

The following is the assumption we need; not surprisingly, in light of the first-order approximation above, it is an assumption on the second-order properties of the equilibrium payoffs.

**Assumption 3.3** For all \( i \), and all type probabilities \( s_p, s_q, s_r \), the second mixed derivative (assumed to exist)

\[
\partial_{pq} \left[ F_i(p^i, q^i, r^i; s_p, s_q, s_r) - F_i(q^i, p^i, r^i; s_q, s_p, s_r) \right]
\]

of the difference in scores of two types with posteriors \( p^i \) and \( q^i \) respectively, does not depend on other type’s posteriors \( r^i \).

The assumption says that the (mixed) sensitivity of the difference in EP’s to the corresponding types is independent of other types. The assumption is, of course, satisfied if the difference in EP’s of two types does not depend on other types.

We now state the following additive representation result, proved in Appendix.

---

\(^{12}\)It is well known that there are quadratic scoring rules that are strictly separating when \( N = 2 \), for all priors.
Proposition 3.3 Consider a PLEPS system \( \{F_i\} \) such that Assumption 3.3 holds. Then, if, for some \( p^0 \in (0,1) \) and for any fixed type probabilities \( s_p, s_q, s_r \) the function \( F_i(p^i, q^i, r^i; s_p, s_q, s_r) \) can be expanded as an infinite Taylor series around the point \( (p^i, q^i, r^i) = (p^0, \ldots, p^0) \in (0,1)^M \), then, necessarily, the following Additive Representation (AR) holds:

\[
F_i(p^i, q^i, r^i; s_p, s_q, s_r) = G_i(p^i; s_p, s_q, s_r) + H_i(p^i, q^i, r^i; s_p, s_q, s_r)
\]

where \( H_i \) is a function that is symmetric in all the pairs \( (p^i, s_p), (q^i, s_q), (r^i, s_r), i = 1, \ldots, n \).

The main result of the section is the following:

Theorem 3.2 Consider a PLEPS consisting of functions \( F_i(p^i, q^i, r^i; s_p, s_q, s_r), i = 1, 2, \ldots, N \), that satisfy the assumptions of Proposition 3.3. Assume also that \( G_i \) is continuously differentiable in the \( p^i \) variable and symmetric in all \( s_k \) variables, for every fixed \( p^i, i = 1, \ldots, n \). Then, we have, for some functions \( \lambda \) and \( B \) of type probabilities \( S = (s_p, s_q, s_r) \),

\[
G_i(p^i, s_p, s_q, s_r) = \lambda(S) \log p^i + B_i(S)
\]

In particular, if the corresponding PLEPS is budget-balanced, the EP of type \( k \) is given by

\[
F_i(p^i, q^i, r^i; s_p, s_q, s_r) = \lambda(S) \log p^i - \lambda(S) \sum_{t=p,q,r} s^i_t \log t^i
\]

where \( s^i_t \) is the conditional probability of the type with posterior \( t \) in state \( i \).

Remark 3.3 We emphasize again that this result is obtained by restricting only equilibrium properties of a scoring rule, without any restrictions on the off-equilibrium properties.

Proof: Since \( F_i \) is an SSNE function, it satisfies separation property (3.1). By the stated symmetry of \( H_i \), function \( G_i \) also satisfies the same type of inequality, which can be written as

\[
0 = \min_{q^i} \left\{ \sum_i p^i G_i(p^i; s_p, s_q, s_r) - \sum_i p^i G_i(q^i; s_p, s_q, s_r) \right\},
\]

By Lemma 3.1 below that identifies the first order condition for this minimization problem, there exists a Lagrange multiplier \( \lambda(S) \) independent of \( p \), such that, suppressing dependence on \( i \),

\[
\lambda(S) \frac{1}{p^i} = \partial_p G(p^i; s_p, s_q, s_r)
\]
The above implies the statement about the logarithmic form of \( G_i \). Equation (3.7) is then straightforward to verify.

The following “Lagrange optimization” lemma is proved in Appendix.

**Lemma 3.1** Consider, in the above notation, functions \( F_i(p^i, q^i, r^i; s_p, s_q, s_r) \), \( i = 1, 2, \ldots, N \), that are continuously differentiable in the \( p^i \) and \( q^i \) variables, and, for every fixed \( p^i, q^i, r^i \), symmetric in all values of \( s_i \) variables. Recall the strict separation inequality (3.1), written in the form

\[
0 = \min_{q^i} \left\{ \sum_i p^i F_i(p^i, q^i, r^i; s_p, s_q, s_r) - \sum_i p^i F_i(q^i, p^i, r^i; s_p, s_q, s_r) \right\},
\]

that is, the minimum over probabilities \( q^i \) is obtained at \( q^i = p^i \). Then, there exists a function \( \lambda(S) = \lambda(s_p, s_q, s_r) \) such that, for all \( i, p^i, q^i, r^i, s_p, s_q, s_r, \)

\[
\lambda(S) = p^i [\partial_p F_i(p^i, p^i, r^i; s_p, s_q, s_r) - \partial_q F_i(p^i, p^i, r^i; s_p, s_q, s_r)].
\]

4 Implementation

In this section we first show how to implement any PLEPS, and then we elaborate on the Prelec (2004) result that the Bayesian Truth Serum algorithm provides a feasible implementation of budget-balanced logarithmic EP’s of (3.7) (under standard Bayesian and rationality assumptions); see also Prelec, Seung and McCoy (2013). We also comment on on uniqueness of equilibrium under the BTS scoring rule.

4.1 Implementing any PLEPS

We still assume there are infinitely many respondents. Consider the case in which the respondents are asked to choose the correct answer to a multiple choice question, and assume that the possible states of nature are elements of the set of finitely valued distributions of the responses to the multiple choice questions. Consider a PLEPS that depends on posteriors \( z_k^i \), and, instead of on ex-ante type probabilities \( s_k \), it depends only on local conditional type probabilities \( s_k^i = Pr(T^r = k \mid \Omega = i) \). Then, in principle, an agnostic planner could implement any given PLEPS as follows. In order to infer
The planner could ask the following from the (infinitely many) respondents:

- (i) to choose the correct answer to the multiple choice question;
- (ii) to state the possible states of nature, that is, to declare what the set of the possible distributions of the responses to (i) is, AND to state their perceived probability for each of those distributions.

The planner decides what the true state of nature \( \Omega = i \) is, she can then use the corresponding \( z_k^i \) from all the \( z_k^j \)'s, \( j = 1, \ldots, N \), that a player provides as the answer to (ii), and she can compute the corresponding value of \( F_i \). The planner could then do the following: she could assign a payoff to a player which is the weighted average of that value and of the average of such values of all the players who provide the same response to (i) as the player in question\(^\text{13}\). Since the player’s expected value (over the states of nature \( i \)) of the payoffs \( F_i \) is the same as for all other players of his type, and higher for him than his expected value of the payoffs of other types, he has incentives to choose the same response to (i) as the other players of his type, thus effectively declaring the latter. He also has incentives to provide the same response to (ii) as the other players of his type, because \( F \) is a PLEPS.

**Remark 4.1** The above implementation procedure is not robust – in practice, there will be more different outcomes of responses to question (ii) than the number of types, and different respondents will consider different distributions of the responses to (i) as the possible outcomes for the states of nature. Thus, some approximate grouping of the responses would have to be done. Moreover, responding to (ii) puts a large burden on the subjects, because they have to provide possible frequencies of the responses to (i) and distributions over those frequencies. For budget-balanced logarithmic EP’s the story is different, as discussed in the next section: the Bayesian Truth Serum (BTS) scoring rule of Prelec (2004) implements budget-balanced logarithmic EP’s; moreover, BTS uses inputs that are simpler than those obtained from the responses to (ii), and a procedure which is robust (that is, no grouping of similar responses is necessary).

\(^\text{13}\)Other variations are possible for the second term of the weighted average; for example, it could be the \( F_i \) payoff of a randomly chosen player with the same response to (i) as the player in question.
4.2 Implementing logarithmic equilibrium payoffs by the Bayesian Truth Serum

We first recall the definition of the Bayesian Truth Serum (BTS). We specify the model in the notation of Section 2. We assume that there are infinitely (countably) many respondents, labeled \( r \in R \). The truthful opinion of respondent \( r \) is represented by a pair of \( M \)-tuples \((X^r; Y^r) = ((X_1^r, \ldots, X_M^r); (Y_1^r, \ldots, Y_M^r))\) of random variables. Here, \( X_i^r \)'s take values zero or one, and only one is equal to one. This is interpreted as choosing an answer from a set of \( M \) possible answers. Random variables \( Y_i^r \)'s take values in \([0, 1]\) and \( \sum_{i=1}^{M} Y_i^r = 1 \). The latter represent the declared opinion that respondent \( r \) has on what percentage of respondents will choose \( i \) as the correct answer.

As in Section 2, we assume that the infinite sequence \((X^r, r \in R)\) is exchangeable. Then, by de Finetti’s theorem, there is an \( M \)-dimensional (potentially random) vector

\[
\bar{X} = \lim_{n \to \infty} \frac{1}{n} \sum_{r=1}^{n} X^r
\]

taking values in \([0, 1]^M\), such that \( X^r \)'s are conditionally independent given \( \bar{X} \). We interpret \( \bar{X} \) to be the true state of nature, denoted previously by \( \Omega \).

Denote by \( \bar{x}_j \) the sample mean of the declared values \( x_j^r \) of \( X_j^r \) over all respondents \( r \), and by \( \log \bar{y}_j \) the sample mean of all the declared values \( \log y_j^r \) of \( \log Y_j^r \) (so that \( \bar{y}_j \) is their geometric mean):

\[
\log \bar{y}_j := \lim_{n \to \infty} \frac{1}{n} \sum_{r=1}^{n} \log y_j^r
\]

**Definition 4.1** The Bayesian Truth Serum (BTS) score function for respondent \( r \) is given by

\[
BTS^r = \sum_{j=1}^{M} x_j^r \log \frac{\bar{x}_j}{\bar{y}_j} + \sum_{j=1}^{M} \bar{x}_j \log \frac{y_j^r}{\bar{x}_j}
\]

Prelec (2004) proved that BTS is an incentive compatible mechanism, in the sense that a respondent’s payoff is maximized by declaring the true opinion, if everyone else does. Moreover, we note that we have uniqueness:

**Remark 4.2 (Uniqueness of equilibrium.)** It is a natural convention to define \( \log(\bar{x}_j/\bar{y}_j) = 0 \) if \( \bar{x}_j = \bar{y}_j = 0 \), as well as to define \( \bar{x}_j \log(y_j^r/\bar{x}_j) = 0 \) if \( \bar{x}_j = 0 \). Note that if \( x_j^r = 0 \) for all but a finite number of \( r \)'s, so that \( \bar{x}_j = 0 \), then it is optimal for every player \( r \) to correctly predict \( y_j^r = 0 \), so that
\( \bar{y}_j \) is naturally defined to be zero\(^{14}\). Under these conventions, the only possible budget-balanced pure strategy SNE's are those which are separating. Indeed:

-(i) First, it is impossible to have an SNE in pure strategies in which two individuals of the same type choose different strategies and hence have different expected scores: suppose they have different strategies in this SNE. If player 1 switched to strategy 2, he would have a strictly lower value, by definition of "strict", and this value would be the same as player 2's value, because with infinite number of players, the value of one player is not affected by what another player does. For the same reason, if player 2 switched to strategy 1, his value would be equal to the original player 1's value, which we argued above is strictly larger. This means that player 2 was not playing an equilibrium strategy to start with. A contradiction.

- (ii) Second, two individuals of different types cannot have the same strategies in an SNE: if they did, by (i) all other players of their types also would choose the same strategy, which means that there would be a type \( k \) that nobody would "claim", that is a \( k \) such that \( x_k^r = 0 \) for all \( r \). Because we assume budget balance, there is a player with a non-positive score. If that player deviates to type \( k \), by above natural conventions his BTS score would be zero, which is weakly better than not deviating, so the equilibrium cannot be strict.

Because of this, and since the truth-telling equilibrium is focal among strictly separating equilibria, from now on we consider \( x_i \)'s and \( y_i \)'s to be the truthful responses.

For the reader's convenience and to provide additional details, we present a proof of the Prelec (2004) result that, in such a truth-telling equilibrium, the BTS score is equal to the budget-balanced logarithmic payoff. Let us denote

\[
p_{ij} = Pr(X_i^r = 1, X_j^s = 1)
\]

where we use the fact that, by exchangeability, the right-hand side does not depend on the choice of \( r \neq s \). Thus, we also have

\[
Pr(X^r = x^r | X^s = x^s) = \frac{p_{ij}}{\sum_{k=1}^M p_{kj}} \quad (4.1)
\]

We will need the following three properties.

\(^{14}\text{This is because increasing } y_j \text{ does not change the score, while decreasing } y_k \text{ for } k \neq j \text{ would lower the score, if } \bar{x}_k > 0.\)
- Property I: \( y_i^r = \sum_{i=1}^{M} x_i^r \frac{p_{ij}}{\sum_{k=1}^{M} p_{ki}} \)

- Property II: \( \log P(X^s = x^s \mid X^r = x^r) = \sum_{j=1}^{M} x_j^s \log y_j^r \), where conditioning indicates conditioning on the truthful response, hence on the signal.

**Proposition 4.1** The following holds:

- Property III: \( \log P_r(X^r = x^r | \bar{X} = \bar{x}) = \sum_{k=1}^{M} x_k^r \log \bar{x}_k. \)

Property I is assumed because we assume a Bayesian game: the respondents compute conditional probabilities in a Bayesian fashion. Property II is a consequence of Property I and equation (4.1). Property III is proved in the appendix, and using Properties I-III we also provide a detailed proof of the following theorem of Prelec (2004) in the appendix.

**Theorem 4.1** (Prelec 2004) Under the above assumptions, BTS scoring is equivalent to the budget-balanced logarithmic EP’s. More precisely, we have

\[
BTS^r = \log P_r(\bar{X} = \bar{x} | X^r = x^r) - \lim_{n \to \infty} \frac{1}{n} \sum_{s=1}^{n} \log P_r(\bar{X} = \bar{x} | X^s = x^s) \tag{4.2}
\]

or, denoting \( x^r = k, x^s = j, \bar{x} = i, \)

\[
BTS^r = \log(Pr(\Omega = i | T^r = k)) - \sum_{j=1}^{M} Pr(T^r = j | \Omega = i) \log(Pr(\Omega = i | T^r = j)) \tag{4.3}
\]

Thus, the BTS score corresponds to the PLEPS function \( F_i \) that is logarithmic. Put differently, BTS implements budget-balanced logarithmic EP’s by asking the players only two things: to choose an answer from the multiple choice list, and to predict what percentage of players will choose a particular answer.

## 5 Conclusions

We consider the problems of extracting true opinions from a large group of respondents and of ranking them according to their expertise, in the case in which the planner is agnostic about the distribution of the states of nature and the respondents’ types. Thus, the planner has to design a universal mechanism, that would work for all such distributions. One such mechanism is the one that is based on ex-post logarithmic payoffs. We prove the following results: (i) for arbitrary mechanisms, any strictly
separating equilibrium satisfying a locality condition necessarily ranks the respondents according to the relative size of their posterior distributions; (ii) under additional assumptions on the sensitivity of score differences, the strictly separating budget-balanced equilibria necessarily result in logarithmic payoffs. We elaborate on the result from Prelec (2004) that the logarithmic equilibrium payoffs can be implemented using the BTS algorithm, and we note that other equilibrium payoff rules can also be implemented, but may require responses to more complex questions.

Our setup does not allow for players’ actions other than costless expressing of their opinions. Thus, developing a more general analysis of robust mechanisms in our framework, in which the players also would draw utility from costly actions, is an unfinished task. In our model the experts have no reason to lie, but need positive incentive to tell the truth. One could envision a framework in which players have some reason to lie, for example they do not care about their own payoff, but want to manipulate the results so as to have some other type have the highest score. Or, a framework with known utilities and unknown correlation of types, in which the planner wants to elicit information about the correlations without disturbing the stated utilities; for example, the case in which the planner wants to ask players to predict what others will do, but she doesn’t want the payoff they get for making these predictions to change any of the other incentives in the game. Finally, ours is a static game, while many applications are dynamic by nature.
6 Appendix

Proof of Proposition 3.2:

For notational simplicity, we consider the case $M = 2$ with two types only, type 1 and type 2, and with $N = 3$, the states of nature 1 being $(2, 0)$ (two of type 1, zero of type 2), state 2 being $(1, 1)$, and state 3 being $(0, 2)$. We consider a $Z$ matrix of the form

$$\begin{pmatrix}
    p & 1-p & 0 \\
    0 & q & 1-q \\
\end{pmatrix}$$

where $0 < p, q < 1$ (notice that the rows correspond to types and columns to states of nature).

With finitely many players, the PLEPS functions $F_i$ depend on the posteriors based on the state of nature $i$ corresponding to the declared types. For example, if the true state is $(2, 0)$, but one respondent declares herself as type 2, then the payoffs correspond to state $(1, 1)$.

The expected score of the truthful response for type 1 is then

$$pF_1(p, p) + (1 - p)F_2(1 - p, q)$$

If one respondent lies and declares his type 1 as type 2, then the expected value is

$$pF_2(q, 1 - p) + (1 - p)F_3(1 - q, 1 - q)$$

Therefore, the separating inequality is

$$pF_1(p, p) + (1 - p)F_2(1 - p, q) > pF_2(q, 1 - p) + (1 - p)F_3(1 - q, 1 - q)$$

This becomes

$$p[F_1(p, p) - F_2(q, 1 - p)] + (1 - p)[F_2(1 - p, q) - F_3(1 - q, 1 - q)] > 0$$

Similarly, when one type 2 respondent lies we get the following:

$$qF_2(q, 1 - p) + (1 - q)F_3(1 - q, 1 - q) > qF_1(p, p) + (1 - q)F_3(1 - p, q)$$

This becomes

$$q[F_1(q, 1 - p) - F_1(p, p)] + (1 - q)[F_3(1 - q, 1 - q) - F_2(1 - p, q)] > 0$$
Suppose now that \( p \neq q \). Without loss of generality we consider the case \( p > q \), and apply Lemma 6.1 in Appendix on the inequalities above. We obtain

\[
F_1(p, p) - F_2(q, 1 - p) > 0
\]
\[
F_2(1 - p, q) - F_3(1 - q, 1 - q) < 0
\]

Assuming budget balance holds, we must have \( F_1(p, p) = 0 = F_3(1 - q, 1 - q) \) and so

\[
(1 - p)F_2(1 - p, q) + qF_2(q, 1 - p) = 0
\]

Note that \( F_1(p, p) = 0 \) leads to \( F_2(q, 1 - p) < 0 \), while \( F_3(1 - q, 1 - q) = 0 \) leads to \( F_2(1 - p, q) < 0 \). This is in clear contradiction with the last equality.

\[\blacksquare\]

Then following lemma is also the key ingredient in proving Theorem 3.1:

**Lemma 6.1** Let \( 0 < a \leq 1 \), \( p, q \in (0, a) \), and \( p > q \). If \( A, B \) are real numbers such that

\[
pA + (a - p)B > 0
\]
\[
q(-A) + (a - q)(-B) > 0
\]

then \( A > 0 \) and \( B < 0 \).

**Proof:**

Notice that \( A \neq 0 \). If not, then the two above inequalities become \((a-p)B > 0\) and \((a-q)(-B) > 0\), a contradiction. In order to prove the lemma we only need to prove that \( A > 0 \). Suppose to the contrary that \( A < 0 \). Then \( B > 0 \). From \((a-p)B > -pA\) it follows that \( B > -\frac{p}{a-p}A > 0 \). We then get \( 0 < q(-A) + (a-q)(-B) < q(-A) + \frac{a-q}{a-p}pA = Aa\frac{a-q}{a-p} < 0 \), which is impossible. This contradiction proves \( A > 0 \).

\[\blacksquare\]

**Proof of Theorem 3.1:**
We suppress the dependence on $s_k$’s in our notation. This is justified because fixing $s_k$’s does not restrict the choice of any two rows of the $Z$-matrix, because we can always define $Q$ by $q_{ki} = z_k^1 s_k$.

We consider three cases separately according to the values of $M$ and $N$.

**Case 1:** Assume $M = 2, N = 2$. The matrix $Z$ can be written then as \[
\begin{bmatrix}
1 & 2 \\
2 & 1
\end{bmatrix}.
\]
If we denote $p := z_1^1, q := z_2^1$, then the matrix $Z$ becomes $Z = \begin{bmatrix} p & 1-p \\ q & 1-q \end{bmatrix}$. Suppose $p > q$ (which is equivalent to $1 - q > 1 - p$). SSNE property (3.1) of $F_i$ implies

\[pF_1(p, q) + (1 - p)F_2(1 - p, 1 - q) > pF_1(q, p) + (1 - p)F_2(1 - q, 1 - p)\]
and

\[qF_1(p, q) + (1 - q)F_2(1 - q, 1 - p) > qF_1(q, p) + (1 - q)F_2(1 - p, 1 - q).\]

This leads to

\[p[F_1(p, q) - F_1(q, p)] + (1 - p)[F_2(1 - p, 1 - q) - F_2(1 - q, 1 - p)] > 0\]

and

\[q[F_1(p, q) - F_1(q, p)] + (1 - q)[F_2(1 - q, 1 - p) - F_2(1 - p, 1 - q)] > 0.\]

We set $a = 1, A = F_1(p, q) - F_1(q, p)$ and $B = F_2(1 - p, 1 - q) - F_2(1 - q, 1 - p)$, and apply Lemma 6.1 in Appendix to the above equations. We obtain that $F_1(p, q) > F_1(q, p)$ and $F_2(1 - p, 1 - q) > F_2(1 - q, 1 - p)$, which proves the theorem in this case.

\[\begin{bmatrix}
z_1^1 & z_1^2 \\
z_2^1 & z_2^2 \\
\vdots & \vdots \\
z_M^1 & z_M^2
\end{bmatrix}.
\]

**Case 2:** Assume $M \geq 3, N = 2$. The matrix $Z$ can be written as

\[
\begin{bmatrix}
z_k^1 \\
z_k^2 \\
\vdots \\
z_M^1 \\
z_M^2
\end{bmatrix}.
\]

The matrix entries satisfy $z_k^2 = 1 - z_k^1, k = 1, \ldots, M$. Take any $k, j \in \{1, \ldots, M\}$ such that $z_k^1 > z_j^1$ (which is equivalent to $z_j^2 > z_k^2$). Using the notation $p := z_k^1, q := z_j^1$, and the notation $z_{-j,k}^1$ for the $(N - 2)$-tuple which consists of $\{z_i, \ldots, z_M\} \setminus \{z_j, z_k\}$, from (3.1) we obtain that the following two equations hold:

\[pF_1(p, q, z_{-j,k}^1) + (1 - p)F_2(1 - p, 1 - q, z_{-j,k}^2) > pF_1(q, p, z_{-j,k}^1) + (1 - p)F_2(1 - q, 1 - p, z_{-j,k}^2)\]

and

\[qF_1(p, q, z_{-j,k}^1) + (1 - q)F_2(1 - q, 1 - p, z_{-j,k}^2) > qF_1(q, p, z_{-j,k}^1) + (1 - q)F_2(1 - p, 1 - q, z_{-j,k}^2).\]

Hence, if we define $A$ and $B$ in the following way,

\[A = F_1(p, q, z_{-j,k}^1) - F_1(q, p, z_{-j,k}^1) = F_1(z_k^1, z_{-k}^1) - F_1(z_j^1, z_{-j}^1)\]

and

\[B = F_2(1 - p, 1 - q, z_{-j,k}^2) - F_2(1 - q, 1 - p, z_{-j,k}^2) = F_2(z_k^2, z_{-k}^2) - F_2(z_j^2, z_{-j}^2)\]

and
we are again within the framework of Lemma 6.1, and we conclude that \( A > 0 \) and \( B < 0 \), which proves (3.2) for both \( i = 1 \) and \( i = 2 \).

**Case 3:** Assume \( M \geq 2, N \geq 3 \) and denote
\[
Z = \begin{bmatrix}
  z_1^1 & z_1^2 & \ldots & z_1^N \\
  z_2^1 & z_2^2 & \ldots & z_2^N \\
  \vdots & \vdots & \ddots & \vdots \\
  z_M^1 & z_M^2 & \ldots & z_M^N
\end{bmatrix}.
\]

We consider the \( i \)-th column in the matrix \( Z \). We can pair it up with any other column, so, without loss of generality we consider \( i \neq 1 \) and we focus on the first and the \( i \)-th column,
\[
\begin{bmatrix}
  z_1^1 \\
  z_2^1 \\
  \vdots \\
  z_M^1
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
  z_i^1 \\
  z_i^2 \\
  \vdots \\
  z_i^M
\end{bmatrix}.
\]

We take any rows \( j, k \in \{1, \ldots, M\} \) where \( j \neq k \). Since the only requirement for the matrix \( Z \) is that its rows are non-degenerate probability distributions, and since the values of \( F_i \) will depend only on the quantities in the \( i \)-th column, then in order to complete our proof we need to prove only that for every \( p := z_k^i \) and \( q := z_j^i \), with \( 1 > p > q > 0 \), for any choice of \( z_{-i(k,j)}^i \in (0, 1)^{M-2} \) (if \( M = 2 \) this last requirement is unnecessary), and for any choice of corresponding \( S_i \), we have
\[
F_i(p, q, z_{-i(k,j)}^i) > F_i(q, p, z_{-i(k,j)}^i) \quad (6.1)
\]

Observe that other entries of the matrix \( Z \) (in other than the \( i \)-th column) do not enter into (6.1), and therefore can be adjusted accordingly, as long as we have a \( Z \)-matrix.

Because we can always choose \( Q \) by setting \( q_{ki} = z_k^is_k \), the following matrix can be taken as a \( Z \)-matrix with unchanged original type probabilities \( s_k \)’s: Take \( 0 < \epsilon < 1 \) and \( a := 1 - \epsilon \); we adjust the matrix \( \tilde{Z} \) so that
\[ \tilde{Z}^t_l := \begin{cases} 
  z^t_l, & \text{if } l \in \{1, \ldots, M\} \setminus \{j, k\} 
  p, & \text{if } l = k, t = i 
  q, & \text{if } l = j, t = i 
  a - p, & \text{if } l = k, t = 1 
  a - q, & \text{if } l = j, t = 1 
  \frac{\epsilon}{N - 2}, & \text{otherwise} 
\end{cases} \]

where \( p, q \) are arbitrary values in \((0, a)\) with \( p > q \). Then for every choice of \( \epsilon \) and \( p \) and \( q \), we have that \( \tilde{Z} \) is a \( Z \)-matrix which differs from \( Z \) only in the \( j \)-th and \( k \)-th row and these rows are

\[
\begin{bmatrix}
  a - p & \frac{\epsilon}{N - 2} & \cdots & \frac{\epsilon}{N - 2} & p & \frac{\epsilon}{N - 2} & \cdots & \frac{\epsilon}{N - 2} \\
  a - q & \frac{\epsilon}{N - 2} & \cdots & \frac{\epsilon}{N - 2} & q & \frac{\epsilon}{N - 2} & \cdots & \frac{\epsilon}{N - 2}
\end{bmatrix}.
\]

The SSNE property (3.1) applied now to \( j \) and \( k \) yields

\[
\sum_{t=1}^{N} \tilde{z}^t_k F_t(\tilde{z}^t_k, \tilde{z}^t_{-k}) > \sum_{t=1}^{N} \tilde{z}^t_j F_t(\tilde{z}^t_j, \tilde{z}^t_{-j})
\]

and

\[
\sum_{t=1}^{N} \tilde{z}^t_j F_t(\tilde{z}^t_j, \tilde{z}^t_{-j}) > \sum_{t=1}^{N} \tilde{z}^t_k F_t(\tilde{z}^t_k, \tilde{z}^t_{-k})
\]

Observe that for \( t \in \{1, \ldots, N\} \setminus \{1, i\} \) we have \( \tilde{z}^t_j = \tilde{z}^t_k = \frac{\epsilon}{N - 2} \). Hence on both sides of the above inequalities we have terms \( \frac{\epsilon}{N - 2} F_t(\frac{\epsilon}{N - 2}, \frac{\epsilon}{N - 2}, z^t_{(j,k)}) \) and they cancel each other. Therefore, the inequalities take the following form:

\[ (a - p) F_1(a - p, a - q, z^1_{(j,k)}) + p F_1(p, q, z^i_{(j,k)}) > (a - p) F_1(a - q, a - p, z^1_{(j,k)}) + p F_1(q, p, z^i_{(j,k)}) \]

\[ (a - q) F_1(a - q, a - p, z^1_{(j,k)}) + q F_1(q, p, z^i_{(j,k)}) > (a - q) F_1(a - p, a - q, z^1_{(j,k)}) + q F_1(p, q, z^i_{(j,k)}) \]

If we set \( A \) and \( B \) as \( A = F_i(p, q, z^i_{(j,k)}) - F_i(q, p, z^i_{(j,k)}) \) and \( B = F_i(a - p, a - q, z^1_{(j,k)}) - F_i(a - q, a - p, z^1_{(j,k)}) \), we are again within the framework of Lemma 6.1. Therefore \( A > 0 \), which proves inequality (6.1) for \( a > p > q > 0 \). By letting \( \epsilon \to 0 \), we obtain (6.1) for \( 1 > p > q > 0 \).
Proof of Lemma 3.1: By the standard result on optimization under constraints (in our case the constraint being \( \sum_i q^i = 1 \)), there exists a Lagrange multiplier function \( \lambda(\bar{p}, \bar{r}, s_p, s_q, s_r) \), where, for example, \( \bar{p} = (p^1, \ldots, p^N) \), such that

\[
p^i [\partial_p F_i(p^i, \bar{p}, \bar{r}, s_p, s_q, s_r) - \partial_q F_i(p^i, \bar{p}, \bar{r}, s_p, s_q, s_r)] = \lambda(\bar{p}, \bar{r}, s_p, s_q, s_r) \quad (6.2)
\]

Fix an arbitrary value of \( i \) and \( p^i, r^i \). Since \( N > 2 \), we can set \( p^j = x, r^j = y \), for a fixed, but arbitrary \( j \neq i \), for any \( 0 < x < 1 - p^i, 0 < y < 1 - r^i \). By the above equality we have that \( \lambda(\bar{p}, \bar{r}, S) \) is a function \( \lambda(p^i, r^i, S) \) of \( p^i, r^i, S \), only, and we have

\[
x[\partial_p F_j(x, x, y, S) - \partial_q F_j(x, x, y, S)] = \lambda(p^i, r^i, S),
\]

for all \( 0 < x < 1 - p^i, 0 < y < 1 - r^i \). Since we can choose \( p^i, r^i \) arbitrarily small, we have then, for fixed \( S \), that the left-hand side is constant across all values of \( x, y \) in \( (0, 1) \), and because \( i \) is arbitrary we get that \( \lambda(S) \) does not depend on any of the values \( p^i, r^i, i = 1, \ldots, N \).

Proof of Proposition 3.3:

We suppress dependence on \( i \) in this proof, and on \( s_p, s_q, s_r \). We want to show that

\[
F(p, q, r) = G(p) + H(p, q, r)
\]

where \( H \) is symmetric in all the pairs \( (p, s_p), (q, s_q), (r^j, s_r^j) \).

For \( p^0 \in (0, 1) \) denote

\[
\bar{p} = p - p^0, \bar{q} = q - p^0, \bar{r} = r - p^0
\]

From the smoothness and the symmetry property of \( F \), we can write, for some functions \( a, b, c, d, e \) of the type probabilities,

\[
F(p, q, r) = \sum_{n=0}^{\infty} a_n \bar{p}^n + \sum_{n=1}^{\infty} (b_n^q \bar{q}^n + b_n^r \bar{r}^n) + \sum_{m,n=1}^{\infty} \bar{p}^m (c_{m,n}^q \bar{q}^n + c_{m,n}^r \bar{r}^n) + \sum_{m,n=1}^{\infty} d_{m,n} \bar{q}^m \bar{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} \bar{p}^l \bar{q}^m \bar{r}^n
\]
where, by the symmetry property,

\[ b_n^q(s_p, s_q, s_r) = b_n^r(s_p, s_r, s_q), \quad c_{m,n}^q(s_p, s_q, s_r) = c_{m,n}^r(s_p, s_r, s_q) \]

\[ d_{m,n}(s_p, s_q, s_r) = d_{n,m}(s_p, s_r, s_q), \quad e_{l,m,n}(s_p, s_q, s_r) = e_{l,n,m}(s_p, s_r, s_q) \]

Note that it is sufficient to show that

\[ c_{m,n}^r = d_{m,n}, \quad e_{l,m,n} = e_{m,l,n} \]

because then we can write

\[ F(p, q, r) = \sum_{n=0}^{\infty} (a_n - b_n^q)\bar{p}^n + H(p, q, r) \]

where \( H \) is symmetric in all the pairs \((p^i, s_p), (q^j, s_q), (r^k, s_r)\).

Let us consider the consequences of strict separation inequality (3.9), using Lemma 3.1. We have

\[
\frac{\partial_q F(p, p, r)}{\partial_p F(p, p, r)} = \sum_{n=1}^{\infty} n b_n^q \bar{p}^{n-1} + \sum_{m,n=1}^{\infty} c_{m,n}^q n \bar{p}^{m+n-1} + \sum_{m,n=1}^{\infty} d_{m,n} m \bar{p}^{m-1} r^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m l \bar{p}^{l+m-1} r^n
\]

\[ - \sum_{n=0}^{\infty} n a_n \bar{p}^{-n} - \sum_{m,n=1}^{\infty} m \bar{p}^{m-1} (c_{m,n}^q + c_{m,n}^r) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} m l \bar{p}^{l+m-1} r^n \]

We can then write

\[
p \frac{\partial_q F(p, p, r)}{\partial_p F(p, p, r)} = \sum_{n=1}^{\infty} n b_n^q \bar{p}^{n} + \sum_{m,n=1}^{\infty} c_{m,n}^q n \bar{p}^{m+n} + \sum_{m,n=1}^{\infty} d_{m,n} m \bar{p}^{m} r^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m l \bar{p}^{l+m} r^n
\]

\[ - \sum_{n=0}^{\infty} n a_n \bar{p}^{-n} - \sum_{m,n=1}^{\infty} m \bar{p}^{m} (c_{m,n}^q + c_{m,n}^r) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} m l \bar{p}^{l+m} r^n \]

\[ + \sum_{n=1}^{\infty} n b_n^q \bar{p}^{n-1} + \sum_{m,n=1}^{\infty} c_{m,n}^q n \bar{p}^{m+n-1} + \sum_{m,n=1}^{\infty} d_{m,n} m \bar{p}^{m-1} r^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m l \bar{p}^{l+m-1} r^n
\]

\[ - \sum_{n=0}^{\infty} n a_n \bar{p}^{-n-1} - \sum_{m,n=1}^{\infty} m \bar{p}^{m-1} (c_{m,n}^q + c_{m,n}^r) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} m l \bar{p}^{l+m-1} r^n \]

By Lemma 3.1, to have an SSNE this has to be equal to \((-\lambda)\) for all \(p, r\), which is possible only if

- from \(\bar{r}^n\) terms:

\[ c_{1,n}^r = d_{1,n} \]

(6.3)

29
- from $\bar{p}r^n$ terms:

$$0 = c_{1,n}^r - d_{1,n} + c_{2,n}^r - d_{2,n} \quad (6.4)$$

- From $\bar{p}^2r^n$ terms:

$$0 = 2(d_{2,n} - c_{2,n}^r) + 3p^0(d_{3,n} - c_{3,n}^r) + p^0(e_{1,2,n} - e_{2,1,n}) \quad (6.5)$$

- From $\bar{p}^3r^n$ terms:

$$0 = 3(d_{3,n} - c_{3,n}^r) + (e_{1,2,n} - e_{2,1,n}) + 4p^0(d_{4,n} - c_{4,n}^r) + 2p^0(e_{1,3,n} - e_{3,1,n}) \quad (6.6)$$

And so on.

So, it is sufficient to show $e_{l,m,n} = e_{m,l,n}$. This follows directly from Assumption 3.3, because then the third mixed derivative of the difference $F(p, q, r) - F(q, p, r)$ in scores is zero for all $p, q, r$, that is,

$$0 = \sum_{l,m,n=1}^{\infty} lmn \ e_{l,m,n} p^{l-1} q^{m-1} r^{n-1} - \sum_{l,m,n=1}^{\infty} lmn \ e_{l,m,n} q^{l-1} p^{m-1} r^{n-1}$$

This completes the proof.

Proof of Proposition 4.1: Let $t$ be such that $x_t^r = 1$. De Finetti’s theorem implies

$$Pr(X^r = x^r | \bar{X} = \bar{x}) = \bar{x}_t = \sum_{k=1}^{M} x_k^r \bar{x}_k .$$

The sum on the right always has only one term different from zero. Therefore, taking the log completes the proof.

Proof of Theorem 4.1: Let $x^s$ be any values such that

$$\bar{x}_k = \lim_{n \to \infty} \frac{1}{n} \sum_{s} x_k^s$$

Note that we can use exchangeability to reorder the respondents so that $r = 1$ and $s = 2, ..., n + 1$. For such choice of $r$ and $s$ we have $\log Pr(X^r = x^r | X^s = x^s) = \sum_{j=1}^{M} x_j^r \log(y_j^s)$. We may always omit
those \( s \) such that \( Pr(X^s = x^s) = 0 \). Thus, we actually have only finitely many choices for an \( M \)-tuple \( x^s \) such that \( 0 < Pr(X^s = x^s) < 1 \), and there is an upper bound \( A \) and a lower bound \( B \) such that \( 0 < A \leq Pr(X^s = x^s) \leq B < 1 \). Then it follows that \( A = \sqrt[A] n \leq \sqrt[n] {\prod_{s=1}^n Pr(X^s = x^s)} \leq B = \sqrt[B] n \). The log function is continuous, so \( \log(\lim(f)) = \lim(\log(f)) \) as long as \( f \) and \( \lim(f) \) are both finite and strictly positive. We conclude that the limit \( \lim_{n \to \infty} \prod_{s=1}^n Pr(X^s = x^s) \) exists, that it is not zero, and that we can take the log outside or inside the limit.

Next, using the above conclusion, from Properties I-III we get

\[
\sum_{k=1}^M \bar{x}_k \log y_k^r = \lim_{n \to \infty} \frac{1}{n} \sum_s \log Pr(X^s = x^s | X^r = x^r)
\]

and

\[
\sum_{k=1}^M x_k^r \log \bar{y}_k = \lim_{n \to \infty} \frac{1}{n} \sum_s \log Pr(X^r = x^r | X^s = x^s)
\]

and so, using Bayes rule,

\[
B T S^r = \sum_{k=1}^M x_k^r \log \frac{\bar{x}_k}{\bar{y}_k} + \sum_{k=1}^M \bar{x}_k \log y_k^r
\]

\[
= \log Pr(X^r = x^r | \bar{X} = \bar{x}) + \lim_{n \to \infty} \frac{1}{n} \sum_s \log Pr(X^s = x^s | X^r = x^r) - \lim_{n \to \infty} \frac{1}{n} \sum_s \log Pr(X^r = x^r | X^s = x^s)
\]

\[
= \log \left( Pr(X^r = x^r | \bar{X} = \bar{x}) \lim_{n \to \infty} \frac{\Pi_{s=1}^n Pr^{1/n} (X^s = x^s | X^r = x^r)}{\Pi_{s=1}^n Pr^{1/n} (X^r = x^s)} \right)
\]

\[
= \log \left( Pr(X^r = x^r | \bar{X} = \bar{x}) \frac{\lim_{n \to \infty} \Pi_{s=1}^n Pr^{1/n} (X^s = x^s)}{Pr(X^r = x^r)} \right)
\]

\[
= \log Pr(\bar{X} = \bar{x}|X^r = x^r) - \log Pr(\bar{X} = \bar{x}) + \lim_{n \to \infty} \frac{1}{n} \sum_s \log Pr(X^s = x^s)
\]

Since the last two terms do not depend on \( r \), and \( \sum_r B T S^r = 0 \), we get equation (4.2). Next, for fixed \( n \) and \( \bar{x} \), denote by \( n_j \) the number of respondents who have type \( j \), so that

\[
\sum_j n_j = n
\]

Then we can write equation (4.2) as

\[
B T S^r = \log Pr(\bar{X} = \bar{x}|X^r = x^r)
\]

\[
- \lim_{n \to \infty} \frac{1}{n} \left[ \sum_{s=1}^{n_1} \log Pr(\bar{X} = \bar{x}|X^r = x^1) + \ldots + \sum_{s=n_{M-1}+1}^{n_M} \log Pr(\bar{X} = \bar{x}|X^r = x^M) \right]
\]

31
\[ \log Pr(\bar{X} = \bar{x}|X^r = x^r) = \lim_{n \to \infty} \left[ \frac{n_1}{n} \log Pr(\bar{X} = \bar{x}|X^r = x^1) + \ldots + \frac{n_m}{n} \log Pr(\bar{X} = \bar{x}|X^r = x^M) \right] \]

Since
\[ \lim_{n \to \infty} \frac{n_j}{n} = Pr(T^r = j | \bar{X} = \bar{x}) \]

we prove equation (4.3).

References


