Incentive Compatible Surveys via Posterior Probabilities

Jakša Cvitanić [†], Dražen Prelec [‡], Sonja Radas [§] and Hrvoje Šikić [¶]

November 1, 2016

Abstract

We consider the problem of eliciting truthful responses to a survey question, when the respondents share a common prior about which the survey planner is agnostic. The planner would therefore like to have a universal mechanism, which would induce honest answers for all possible priors. If the planner also requires a locality condition that ensures that the mechanism payoffs are determined by the respondents' posterior probabilities of the true state of nature, we prove

[*] We are grateful to George Georgiadis, Matthew Elliott, Katrina Ligett, Patrick Ray, as well as conference audiences at the Southwest Economic Theory Conference 2014, 13th Conference on Research on Economic Theory & Econometrics for useful comments and suggestions. A previous version of this article circulated under the title "Mechanism design for an agnostic planner: universal mechanisms, logarithmic equilibrium payoffs and implementation".

[†] 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, and MIT, Sloan School of Management. E-mail: sradas@eizg.hr; sradas@mit.edu. 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: hsikic@math.hr. Research supported in part by the MZOS grant 037-0372790-2799 of the Republic of Croatia and in part by Croatian Science Foundation under the project 3526.

that, under additional smoothness and sensitivity conditions, the payoff in the truth-telling equilibrium must be the logarithmic function of the posterior probabilities. Moreover, the players are necessarily ranked according to those probabilities. Finally, we discuss implementation issues.

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

JEL codes: C11, D82, D83, M00

1 Introduction

Consider a problem of truthful elicitation of responses in a population of Bayesian agents who share the common prior, in which the survey planner is "agnostic" about the prior (although the planner may have some beliefs about the prior, she may prefer to keep these beliefs private). Prelec (2004) introduces an algorithm for giving scores to the respondents which is easily implementable and requires minimal knowledge on the part of the planner; it is known as *Bayesian Truth Serum* (BTS). Prelec (2004) proves that BTS has two important properties. First, it is strictly incentive compatible (IC), i.e., strict *truth-telling* is an equilibrium, so that the agent types are fully revealed [1]. Second, the BTS equilibrium score of a respondent is, up to a linear transformation, equal to the logarithm of his posterior, so that BTS *ranks respondents by posteriors*, henceforth PstR. Lets call this second property *logarithmic scoring*. The motivating questions for our paper are: Under which conditions do equilibrium payoffs necessarily correspond to logarithmic scoring? Under which conditions a strict IC equilibrium has to satisfy PstR? That is, are all equilibrium payoffs under natural conditions essentially equivalent to BTS payoffs?

Our main results are the following. We identify two conditions on equilibrium payoffs that we name "posterior locality", PstLoc, and a "separation of variables" property, SepVr. The main theorem says

$$\text{IC} \ + \ \text{PstLoc} \ + \ \text{SepVr} \longrightarrow \text{logarithmic scoring.}$$

The second result says

$$\text{IC} \ + \ \text{PstLoc} \ \longrightarrow \ \text{PstR.}$$

To make these statements more precise, we consider our setting in more detail. There is a state of the world drawn from a finite set, and an infinite population of players and each player observes a private signal from a finite set [2]. The signals are conditionally i.i.d. with respect to the state of the world, so that there is a single probability distribution, the "prior", that describes the joint distribution of states of the world and signals. We assume that the prior is common knowledge among players, but

[1] In fact, we show that all strict equilibria in the BTS framework are either truth-telling or a types-permutation of truth-telling, and the scores are unique up to a linear transformation.

[2] We expect that our results would still hold approximately for finite, but large sample sizes. The exact theory for the finite case is very different and left for future research.

unknown to the planner ex-ante. The planner asks each player to submit a response, which typically includes a declaration of the respondent's "type" and "something else" (as we show, reporting only types cannot lead to a truth–telling equilibrium). Each player then earns a score based on his own response and on the responses of everyone else. The score is computed via a scoring function f.

Let us now recall the notion of proper scoring rules in a framework with only one respondent. Consider a random variable Ω, taking values in $\{1, \ldots, N\}$, $N > 1$, representing the state of the world. A respondent is asked to declare his belief about the distribution of Ω. If the outcome $\Omega = i$ is observed, the respondent is paid $F_i(p)$, where $p = (p^1, \ldots, p^N)$ is the probability distribution declared by the respondent. A family of functions $\{F_i\}_{i=1,\ldots,N}$ is called a strictly proper scoring rule if it is incentive compatible for truth-telling, that is, the respondent's expected payoff is maximized at his true belief, the respondent's posterior. More precisely, for all probability vectors $q \neq p$, we have

$$\sum_{i=1}^{N} p^i F_i(p) > \sum_{i=1}^{N} p^i F_i(q) \tag{1.1}$$

There are many proper scoring rules. A general characterization with many examples is provided in Gneiting and Raftery (2007)[3]. An important special case arises if $F_i(p) = F_i(p_i)$ depends only on the local posterior, which is the probability p_i the respondent assigns to the realized outcome $\Omega = i$, and does not depend on how probabilities are divided among the remaining counterfactual outcomes. In that case the scoring rule is necessarily equal to a linear transformation of the logarithm of p_i (Savage (1971), Bernardo (1979)). Such a rule is a natural choice if the local posterior is interpreted as a measure of respondent's expertise.

In our multi-player setting, the values of Ω can be thought of as distributions of types in the infinite population, and we call "posteriors" the probabilities that a type assigns to those values. Because there are many respondents and we will allow the score of a respondent to depend on the responses of others, providing responses becomes a game. Given a scoring function f such that there exists a (strictly) separating equilibrium in which different types choose different answers, we derive properties of the payoffs in that equilibrium. In particular, if $\Omega = i$ is observed[4], we denote by F_i the value of the payoff score f in that equilibrium. Condition PstLoc posits that F_i are functions of local posteriors, that

[3]Offerman, Sonnemans, Van De Kuilen and Wakker (2009) consider the case in which the respondents may have non-expected utilities, with only two possible states of the world.

[4]By de Finetti's theorem and our interpretation of the values of Ω, the value of Ω is known when all the players' responses are known.

is, of posterior probabilities of the event $\Omega = i$. With infinitely many players the form of equilibrium payoff does not change when a player of one type mimics the equilibrium strategy of another type. For this reason, we are able to derive results about equilibrium payoffs even though the scores that players would receive out of equilibrium are never specified.

More precisely, we assume that the realized ex-post payoffs in equilibrium are of the form $F_i(p^i_k, p^i_{-k}; s)$, that is, the payoff of the respondent of type k in state i depends on his posterior p^i_k of state i, possibly also on the posteriors of state i of other types, collected in the vector p^i_{-k}, and on the probability vector s representing the (ex-ante) distribution of types.

We first show, under mild smoothness conditions, that, if the equilibrium payoffs F_i satisfy incentive compatibility, the difference in the state i scores of two respondents with posteriors p^i and q^i, respectively, has to be approximately proportional to $\log(p^i) - \log(q^i)$ for $q \approx p$, up to the first order. Our first theorem says: if we add to incentive compatibility mild requirements on payoff smoothness and sensitivity on other players of the difference in equilibrium payoffs of two respondents, then the difference in incentive compatible scores of the two respondents is exactly proportional to the difference in logarithms of the declared posteriors, rather than only approximately.

Our second theorem says that any incentive compatible ex-post payoff $F_i(p^i_k, p^i_{-k})$ is non-decreasing in the declared probability p^i_k. Consequently, the ranking of experts in equilibrium, if we consider p^i_k to be the measure of expertise, is the same given any incentive compatible mechanism. This result is a generalization of results in the literature on the case of one respondent, on the monotonicity being implied by incentive compatibility of proper scoring rules. See, e.g., McCarthy (1956), Savage (1971), Schervish (1989) and Schlag and van der Weele (2013). The result is very general, proven by purely algebraic methods.

Finally, we also discuss implementation issues. Observe that in general, while a particular ex-post payoff of the form $F_i(p^i_k, p^i_{-k})$ may arise in equilibrium in theory, it is not necessarily simple to implement it in practice. That is, the problem is how to implement 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. Under an assumption somewhat stronger than PstLoc, but without assuming SepVr, we show that the payoffs of all strictly-separating equilibria in our framework can be implemented by particular questionnaires, but the latter may be complex, except for the logarithmic, BTS case. In this context, let us recall that Prelec (2004) shows that promising the respondents the BTS scores ex-ante, results, in equilibrium, in the ex-post scores of the form $\log(p^i_k)$ (plus a term that does not vary with a respondent). We revisit this result and provide a detailed proof thereof. We also show that truth-telling is essentially the only budget-balanced equilibrium under BTS. [5]

Relationship to existing literature. Proper scoring rules in the game-theoretic context have been studied extensively in the case in which the planner knows the prior distribution of the player types. We mention a few works: Miller, Resnick, and Zeckhauser (2005) design a clever use of proper scoring rules in such a framework; Witkowski and Parkes (2012) study a framework with only two types but without common prior, and Waggoner and Chen (2013) consider a general framework without assumptions on information structure. Our contributions to this literature are to analyze truth-inducing payoffs that can be implemented even when the planner is agnostic about the prior, and to show that logarithmic scoring is the only possible equilibrium payoff form under certain assumptions.

The problem we tackle in the paper can also be considered as one of mechanism design, since we seek to describe mechanisms that are both incentive-compatibile and have attractive features for opinion elicitation applications. In one way, our approach is more general than typical mechanism design models [6] 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. Our basic assumption is that the players have a common prior on this joint distribution, but that the prior is not used by the 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 an 'universal' mechanism, one that would work for all priors without any input from her side apart from the initial formulation of the multiple-choice

[5] See also Prelec, Seung and McCoy (2013) who define and test experimentally a broader class of algorithms to produce a ranking of experts according to their posteriors. Within this class, only BTS is known to be incentive-compatible. Additional incentive-compatible mechanisms that are not admissible in the framework of this paper are studied in Cvitanić and Prelec (2015).

[6] See, e.g., Maskin and Sjöström (2001), Bergemann and Morris (2012), Börgers (2013). We refer the reader to Bergemann and Morris (2012) for a detailed literature survey.

question. [7] 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 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.

The rest of the paper is organized as follows: Section 2 introduces the model, Section 3 presents the main theoretical results, Section 4 discusses implementation issues, and we conclude in Section 5. The first appendix provides a careful description of condition PstLoc. The second appendix presents the proofs.

2 Model, Definitions and Assumptions

In our model, a mechanism consists in giving scores to the players (respondents) of different types. [8] 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 wine bottle some years into the future, and asks experts to respond to appropriately designed questions. Broader applications include voting in elections, predicting political events, product market research, online product reviews, and any other application that involves a survey with a multiple choice question. [9]

[7] In theory, using the "majority rule" mechanism that would ask for the common prior to be declared may result in an equilibrium that reveals the common prior; however, such a rule would not be implementable in practice, as we discuss in the paper.

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

[9] Prelec (2004) and Prelec, Seung and McCoy (2013) provide many more examples.

2.1 The model

The players are indexed by $r \in R$, where R is infinite and countable.[10] The state of nature is a random variable Ω, taking values in $\{1, \ldots, N\}$, $N > 1$. [11] 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 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"). A pure strategy for player r is a map σ^r that maps a player's type to his response choice a^r. We allow only for pure strategies.

A response a^r would typically include a declaration of a respondent's type (choosing an answer to a multiple choice question), and it would also include responses to some other questions in order to be truth-inducing.[12] It could also include a declaration of the respondent's prior distribution of types and states of nature, as introduced below; that is, the respondent could be asked to state what his prior is. We posit the following

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

(ii) If respondent r chooses response a^r, and the remaining responses are represented by a^{-r}, then his score is given by function $f(a^r, a^{-r})$, where the order of different respondents' responses in a^{-r} does not matter, that is, f is symmetric in those.

Condition (i) implies that the order in which we consider our players is irrelevant (from the point

[10] 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.

[11] 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 wine bottle's value, the state of nature could be the percentage of experts who believe the bottle is worth more than one thousand dollars.

[12] 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.

of view of the probability distribution of the entire sequence). Moreover, by de Finetti's theorem, the exchangeability assumption actually implies the second part of Assumption 2.1, that there exists a random variable Ω such that T^r's are conditionally i.i.d. with respect to Ω; see, e.g., Aldous (1985), or Chow and Teicher (1997).

The symmetry property in condition (ii) is a natural restriction considering that the planner does not make a distinction between different types, assumed exchangeable by condition (i).

From now on, we assume the players are risk-neutral, that is, each player maximizes his expected payoff. [13]

2.2 The prior and the posteriors

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 suppose 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 that the planner uses is M. For example, M is needed for implementation using a multiple choice question – the planner has to offer exactly as many possible choices as there are types. [14]

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

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

[13] 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.

[14] To get around the problem of not knowing the common prior the planner could ask each player to state the whole prior distribution and harshly penalize the player who gives a response different from others. However, 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.

and

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

We assume that the marginal probabilities of types and of states of nature are all strictly positive. The posteriors form a matrix $Z = [z_k^i]_{k=1,i=1}^{M,\ \ N}$. Note that z_k^i does not depend on r, that for every $k \in \{1, ..., M\}$, we have $\sum_{i=1}^N z_k^i = 1$, 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 $(s_1, ..., s_M)$ by S.

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 2.1 *Suppose there exists a Bayesian Nash equilibrium for our game of respondents.* [15] *Then, under the symmetry assumptions on f and with infinite number of players, if a respondent of type k deviates from the equilibrium by using the strategy of type $j \neq k$, his deviation payoff is equal to the equilibrium payoff of type j.*

This holds because every type is represented by infinitely many players, and the score function f is symmetric in their responses. The proof is in Appendix.

2.2.1 Equilibrium payoff and incentive compatibility

In the standard literature on scoring rules, there is only one respondent, asked to declare his posterior belief about the distribution of Ω, that is, to declare z^i's. If the outcome $\Omega = i$ is observed, the respondent is paid $F_i(z)$. A family of functions $\{F_i\}_{i=1,...,N}$ is called a strictly proper scoring rule if it is incentive compatible for truth-telling, that is, the respondent's expected payoff is maximized at his true belief, meaning, for all probability vectors $q \neq p$, we have $\sum_{i=1}^N p^i F_i(p) > \sum_{i=1}^N p^i F_i(q)$.

In our framework with infinitely many respondents, we consider only the payoff mechanisms that allow for a Bayesian Nash equilibrium in which the equilibrium payoffs are functions $F_i : (0,1)^{2M} \to \mathbb{R}$, of the form $F_i(z_k^i, z_{-k}^i; s_k, s_{-k})$ where, for example, $z_{-k}^i = (z_1^i, ..., z_{k-1}^i, z_{k+1}^i, ..., z_M^i)$.

[15] A Bayesian Nash equilibrium is a family of strategies σ^r, one for each player r, such that, if player r deviated to a strategy different from σ^r, his expected payoff would not be greater. If it is strictly less, we say that the equilibrium is strict.If the players of different types submit different responses, we say that that the equilibrium is separating. For a detailed definition, see Appendix A.

The details of the relationship between the ex-ante payoff f and the ex-post payoff $\{F_i\}$, and the technical conditions needed, are presented in Appendix A; see also Section 2.2.2 below. Here, for the purposes of stating our results, we only need to assume that, if the realized state is i, and the players play the equilibrium strategy, a player's realized score is given by function F_i that depends on posteriors z_j^i of the realized state and type probabilities s_j. Moreover, we require that the realized equilibrium payoffs satisfy the conditions in the following definition.

Definition 2.1 *The family $\{F_i\}$ of functions of the form $F_i(z_k^i, z_{-k}^i; s_k, s_{-k})$ is called a Posterior-Local Equilibrium Payoff System (PLEPS) if the following is satisfied:*

- *(i)* **Symmetry:** $(\forall x, y \in (0,1))$ $(\forall z_2, ...z_M, s_2, \ldots, s_M \in (0,1))$ $(\forall$ *permutation* Π *of* $\{2, ..., M\})$, *we have*

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

- *(ii)* **Incentive compatibility, strict separation inequality:**

$(\forall Z-\textit{matrix})$ $(\forall S-\textit{vector})$ $(\forall k, j \in \{1, ..., M\}$ *such that* $(z_k^1, ..., z_k^N) \neq (z_j^1, ..., z_j^N))$

$$\sum_{i=1}^N z_k^i F_i(z_k^i, z_{-k}^i; s_k, s_{-k}) > \sum_{i=1}^N z_k^i F_i(z_j^i, z_{-j}^i; s_j, s_{-j}) \tag{2.1}$$

Assumption (i) on symmetry means that the equilibrium score of type k does not depend on the order of other types, and is consistent with Assumption 2.1 (ii) on the symmetry of scoring function f. Assumption (ii) implicitly assumes that the players are risk-neutral and maximize the expected score. By Proposition 2.1, it is automatically satisfied if F_i are the equilibrium payoffs (as formalized by Assumption 6.1 in Appendix A) in a truth-telling equilibrium.

We now elaborate more on the assumed form of equilibrium payoffs F_i.

Remark 2.1 The crucial assumption for the results of this paper is that the score of a player in equilibrium depends on the player's posterior z^i of the realized state of nature i, called local posterior. This is justified if the posterior is a good measure of the player's expertise. There are cases in which the planner clearly wants to know about the distribution of types, such as elections or product market research, trying to estimate what percentage of population will vote for each candidate, or is likely to buy a product. In such cases, it is intuitive that a respondent with higher posterior is a better expert – he has the highest probability of being right about the actual distribution of responses, reminiscent of the concept of maximum likelihood estimators that maximize the probability of the event that does actually occur. Moreover, if the survey study has more than one stage, for example, in market research, a mechanism that results in PLEPS payoffs could be used to identify experts in the first stage, and then only the experts could be used for further surveys, thus reducing the cost of the study. It is primarily these applications we have in mind. In other applications, such as, for example, surveying economists on whether this year's inflation will be higher than a certain level, it is less clear that a higher posterior on the distribution of types means a higher expertise, and scoring rules other than those with ex-post PLEPS payoffs might be appropriate. In particular, if the planner is not concerned with identifying experts, but only with truth-telling, the assumption may exclude perfectly reasonable scoring rules. For example, it can be shown that the following is an incentive compatible payoff, paid to the player of type k in state i (see Cvitanić and Prelec (2015)):

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

However, this payoff is not monotone in the local posterior.

We also note that we look for the simplest possible equilibrium payoffs that describe players' expertise, which is why the payoff F is not allowed to depend on other local probabilities that can be derived from the prior. On the other hand, the reason why we allow dependence on ex-ante type probabilities s_k, s_{-k} is because these, in implementation, translate to type frequencies, which may be used to make a mechanism budget-balanced, as defined below. Actually, for budget balance, it is sufficient to have dependence on local conditional probabilities $s_k^i = Pr(T^r = k \mid \Omega = i)$, but we allow dependence on s_k, s_{-k} for generality (except in the implementation section), as discussed next.

A natural question to ask is whether for any PLEPS F there exists a scoring rule f that implements it in equilibrium. In the implementation section below we argue that this is, indeed, the case, under the assumption that, instead of on possibly all s_k, s^{-k}, F_i depends only on $s_k^i = Pr(T^r = k \mid \Omega = i)$. It is also natural to ask if, for a given f, the equilibrium that implements F is unique. We show later below that this is essentially true for the benchmark example of the Bayesian Truth Serum scoring rule.

We will often restrict the payoffs to those for which the planner pays zero in aggregate, in which case we say that the mechanism is budget balanced. More precisely, we have

Definition 2.2 *A payoff mechanism is said to be budget-balanced if the sum of the scores of all the players is equal to zero, with probability one.*[16]

We have the following negative result, proved in Appendix, when the number of players is finite.[17]

Proposition 2.2 *Assume (only in this proposition) a finite number of players. Then, there exists no budget-balanced PLEPS.*

2.2.2 Ex-ante vs. ex-post payoff: implementation

Even when identifying states of nature with possible empirical frequencies of responses, asking about posterior probabilities of state of nature is likely to be prohibitively complex in practice, because it would require respondents to provide a distribution over all possible empirical frequencies. Thus, in practice, the planner who wants the mechanism to result in ex-post payoffs F_i when the players play the truth-telling equilibrium, would like to find a way to induce those ex-post payoffs by promising to pay the players based on ex-ante scores that require much simpler inputs than the players' beliefs about the distribution of the empirical frequencies. We will discuss this issue in the implementation section, and here we just mention the following. Our benchmark example of a PLEPS is the classical logarithmic scoring rule payoff

$$F_i(z_k^i, z_{-k}^i; s_k, s_{-k}) = \log(z_k^i) \ .$$

Prelec (2004) showed that the budget-balanced version of this payoff can be implemented by, in addition to asking (infinitely many) respondents to declare their own type – the multiple choice question – also asking them what they think is the percentage of other types in the population, that is, the empirical frequencies of each choice in the multiple choice question. It is much easier for the respondents to provide their estimates of empirical frequencies than their estimates of the probability distribution of the empirical frequencies. In the logarithmic case, this means that a respondent of type k is not asked for z_k^i's and is not promised $\log(z_k^i)$, but he is asked for simpler inputs that determine his promised

[16] It should be mentioned that in a budget-balanced game the players know they may receive negative "payments", and some players may not be willing to participate. In practice, the "payments" will often not be monetary, but used as score points, and every respondent might be paid a non-negative amount, that may consist of a fixed fee and a variable fee that depends on the respondent's score, or his ranking according to the scores. That is, what is used may not be a budget-balanced scoring rule, but a modification thereof.

[17] We leave for a future study a more thorough analysis of the case with the finite number of players.

score via a specific function f (called Bayesian Truth Serum), and the value of the score will turn out to be equal to $\log(z_k^i)$ when the players play the truth-telling equilibrium.

3 Possible equilibrium payoffs

In this subsection we present examples of PLEPS's and address the question whether logarithmic equilibrium payoffs (EP's) or simple modifications thereof are the only possible PLEPS's.

3.1 Logarithmic equilibrium payoffs

3.1.1 The benchmark example – the logarithmic function

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

$$F_i(z_k^i, z_{-k}^i; s_k, s_{-k}) = \log(z_k^i)$$

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 (2.1). This is because of the well known *Gibbs inequality* which says that for a probability vector $(p^1, \ldots, p^N)$, we have

$$0 = \min_{q^i \geq 0, \sum_i q^i = 1} \sum_{i=1}^{N} p^i [\log(p^i) - \log(q^i)] \tag{3.1}$$

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

$$0 = \min_{q^i} \left\{ \sum_{i=1}^{N} p^i [\log(p^i) - \log(q^i)] + \lambda \sum_i q^i \right\} \tag{3.2}$$

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 show that under mild additional conditions logarithmic equilibrium payoffs are, in fact, the only possible PLEPS's.

3.1.2 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 (3.2). 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

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

Define the following function:

$$F(p, q, r) = Klog(p) + p^4 - 2p^3(q + r) - 6p(qr^2 + q^2 r)$$

It is straightforward to verify that, for large enough K, this function satisfies the strict separation inequality (2.1). This is because the first order conditions (FOC's) for the Lagrangian optimization problem

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

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

$$p^i [F_p(q^i, p^i, r^i) - F_q(p^i, q^i, r^i)] = \lambda \tag{3.3}$$

for some Lagrange multiplier λ, 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 (2.1) is satisfied.

Remark 3.1 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, denoting by ∂_i the partial derivative with respect to the $i-$th argument, we have

$$F(p,q,r) - F(q,p,r) \approx [\partial_2 F(p,p,r) - \partial_1 F(p,p,r)](q-p)$$

To evaluate the right-hand side, note that since F is incentive compatible, the solution to the problem $\min_{q^i}\left\{\sum_i p^i[F(p^i,q^i,r^i) - F(q^i,p^i,r^i)] + \lambda \sum_i q^i\right\}$ is $q^i = p^i$, where λ is a Lagrange multiplier. The first order condition for this problem gives (see Lemma 3.1 below)

$$\partial_2 F_q(p,p,r) - \partial_1 F(p,p,r) = -\frac{\lambda}{p}$$

Combining the above equations and using that the first order Taylor expansion of the log function around $q = p$ is $\log(p) - \log(q) = \frac{1}{p}(p-q)$, we get

$$F(p,q,r) - F(q,p,r) \approx \lambda(1 - \frac{q}{p}) \approx \lambda(\log(p) - \log(q))$$

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.

3.2 When are equilibrium payoffs logarithmic?

We assume in this section that $N \geq 3$. [18] As we have just shown, the difference in the ex-post 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 PLEPS is essentially a logarithmic EP.

[18]It is well known that there are quadratic scoring rules that are strictly separating when $N = 2$, for all priors.

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. This is without loss of generality, the same proof works for more than three types.

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. In particular, it is weaker than the assumption that the difference in equilibrium payoffs of two types does not depend on other types.

Assumption 3.1 *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 not affected by other types.

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

Proposition 3.1 *Consider a PLEPS system $\{F_i\}$ such that Assumption 3.1 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 $(p^i, q^i, r^i) = (p^0, \dots, p^0) \in (0,1)^M$, then, necessarily, the following Additive Representation (AR) holds:*

$$F_i(p^i, q^i, r^i; s_p, s_q, s_r) = G_i(p^i; s_p, s_q, s_r) + H_i(p^i, q^i, r^i; s_p, s_q, s_r) \tag{3.4}$$

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 3.1 *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 3.1. Assume also that F_i is such that G_i is symmetric in all s_k variables, for every fixed p^i, $i = 1, \ldots, N$. Then, we have, for some functions λ 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 with posterior p_i is given by

$$F_i(p^i, q^i, r^i; s_p, s_q, s_r) = \lambda(S) \log p^i - \lambda(S) \sum_{t=p,q,r} s_t^i \log t^i \tag{3.5}$$

where s_t^i is the conditional probability of the type with posterior t in state i.

Remark 3.2 We emphasize again that this result is obtained by restricting only equilibrium properties of a scoring rule, without restrictions on the off-equilibrium properties.

Proof: Since F_i is a PLEPS, it satisfies separation property (2.1). By the stated symmetry of H_i, function G_i also satisfies the same type of inequality, which can be written as

$$0 = \min_{q^i} \left\{ \sum_i p^i G_i(p^i; s_p, s_q, s_r) - \sum_i p^i G_i(q^i; s_p, s_q, s_r) \right\}, \tag{3.6}$$

As shown in Savage (1971), this property implies that G_i is continuously differentiable in the p^i variable, $i = 1, \ldots, N$. Then, 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 (3.5) is then straightforward to verify.

■

The following "Lagrange optimization" lemma is proved in Appendix. It gives the first order condition for the IC minimization problem in (3.7)below.

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_t variables. Recall the strict separation inequality* (2.1)*, written in the form*

$$0 = \min_{q^i} \left\{ \sum_i p^i F_i(p^i, q^i, r^i; s_p, s_q, s_r) - \sum_i p^i F_i(q^i, p^i, r^i; s_p, s_q, s_r) \right\}, \tag{3.7}$$

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 $i, p^i, q^i, r^i, s_p, s_q, s_r$,

$$\lambda(S) = p^i[\partial_p F_i(p^i, p^i, r^i, s_p, s_q, s_r) - \partial_q F_i(p^i, p^i, r^i, s_p, s_q, s_r)] \quad . \tag{3.8}$$

3.3 Ranking by posteriors

We now show the following result: PLEPS payoffs necessarily rank the players according to the relative ranking of the corresponding posteriors. That is, when using a scoring system resulting in an equilibrium with PLEPS payoffs, the planner will know which players are better experts than others, if she considers the level of the posterior equivalent to the level of expertise. [19] We emphasize that for this result it is crucial to assume that the equilibrium scores depend only on the posteriors of the realized state of nature.

The main result of this section is

Theorem 3.2 *PLEPS payoffs $\{F_i\}$ are strictly increasing in the posterior probabilities of the true state of nature. That is, functions F_i satisfy (for any prior distribution matrix Q),*

$$\textit{If } j, k \in \{1, ..., M\} \textit{ and } z^i_k > z^i_j, \textit{ then } F_i(z^i_k, z^i_{-k}; s_k, s_{-k}) > F_i(z^i_j, z^i_{-j}; s_j, s_{-j}) \tag{3.9}$$

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 (2.1) not only guarantees strict separation of types, but also has the posterior-based ranking as a direct consequence.

This theorem is a generalization of the results in the literature on the monotonicity being implied by incentive compatibility of proper scoring rules. See, e.g., McCarthy (1956), Savage (1971), Schervish

[19] If they were not ranked by their posteriors, then, in the pre-game phase, they might want to avoid collecting information about the true state of nature, which is undesirable.

(1989) and Schlag and van der Weele (2013). Those papers consider only the non-game version of the problem with one respondent only. Moreover, they mostly use analytic methods to prove it, while our proof is completely algebraic.

Theorem 3.1 is also in the spirit of the theorems that relate incentive compatibility to monotonicity in types, if we equate types with posterior probabilities; see Myerson (1981) for an early theorem of that type, and Vohra (2007) for a comprehensive treatment. However, our framework is different from the standard mechanism design framework, in that we have random states of nature, so that incentive compatibility is a property of a weighted sum (expected value conditional on type), not on the value itself. As a consequence, the methodology of those papers does not work,

The intuition behind the 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 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 PLEPS 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 equilibrium, 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.

4 Implementation

In this section we first show how to implement any PLEPS, up to an additional mild restriction, 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 (3.5) (under standard Bayesian and rationality assumptions); see also Prelec, Seung and McCoy (2013). We also comment on the uniqueness of equilibrium under the BTS scoring rule.

4.1 Implementing PLEPS

We continue to 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 take values in the set of probability distributions of the responses to the multiple choice questions.

Proposition 4.1 *Consider a scoring function f such that there exists an SNE with ex-post payoffs given by a PLEPS with type k payoff in state i given by $F_i(z^i_k, z^i_{-k}; s^i_k, s^i_{-k})$; that is, in addition to z^k_j's, the payoff depends only on local conditional type probabilities $s^i_k = Pr(T^r = k \mid \Omega = i)$ instead of on possibly all ex-ante type probabilities contained in vector s_k. Then, this PLEPS can be implemented by the agnostic planner. More precisely, there exist questions that the planner can ask from which she can form estimates $\hat{i}$, $\hat{z}^{\hat{i}}_k$ and $\hat{s}^{\hat{i}}_k$ of the true state of nature i and the true probabilities z^k_i and s^k_i, and such that, if the planner announces that a player who declares type k will receive $F_{\hat{i}}(\hat{z}^{\hat{i}}_k, \hat{z}^{\hat{i}}_{-k}; \hat{s}^{\hat{i}}_k, \hat{s}^{\hat{i}}_{-k})$, then, truth-telling is an equilibrium.*

Proof: Suppose the planner asks the following from the respondents:

- (a) to choose the correct answer to the multiple choice question;

- (b) 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.

To guarantee truth-telling is an equilibrium, the planner announces she will compute the values F_i as follows. Her estimate $\hat{i}$ of the true state of nature i will be given by the frequencies by which each particular answer to the multiple choice question has been chosen by the respondents. She will also make the estimates $\hat{s}^{\hat{i}}_k$ of the type probabilities equal to those frequencies. Having estimated state i, she will then choose the corresponding $\hat{z}^{\hat{i}}_k$ from all the z^j_k's, $j = 1, \ldots, N$, that a player provides as the answer to (b). Having all of these, the planner will compute the corresponding values of $F_{\hat{i}}$'s.

Suppose now that all players other than player r of type k play the truth-telling strategy. If player r also plays the truth-telling strategy, his payoff in state i is $F_i(z^i_k, z^i_{-k}; s^i_k, s^i_{-k})$, because i, z's and s's are correctly estimated by the planner. If player r of type k declares type $j \neq k$, his payoff in state i is $F_i(z^i_j, z^i_{-j}; s^i_j, s^i_{-j})$, because, with infinite number of players and all except player r being honest, i, z's and s's are again correctly estimated by the planner. By IC inequality (2.1), player's r expected value of the payoff when he is dishonest is less than the expected value of the payoff when he is honest, and he would not deviate.

■

Remark 4.1 The above implementation procedure is not robust – in practice, there will be more different outcomes of responses to question (b) than the number of types, and different respondents will consider different distributions of the responses to (a) as the possible outcomes for the states of nature. Thus, some approximate grouping of the responses would have to be done. Moreover, responding to (b) puts a large burden on the subjects, because they have to provide possible frequencies of the responses to (a) and distributions over those frequencies. For budget-balanced logarithmic equilibrium payoffs 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 using inputs that are simpler than those obtained from the responses to (b), and a procedure which is robust (that is, no grouping of similar responses is necessary).

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 $(X^r; Y^r) = ((X^r_1, \ldots, X^r_M); (Y^r_1, \ldots, Y^r_M))$ of random variables. Here, X^r_i'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^r_i's take values in $[0, 1]$ and $\sum_{i=1}^{M} Y^r_i = 1$. 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

$$\bar{X} = \lim_n \frac{1}{n} \sum_{r=1}^{n} X^r$$

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 Ω.

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 can state a new uniqueness result, namely, that with the BTS mechanism any budget-balanced strict (Bayesian) Nash equilibrium, henceforth SNE, is separating, as defined in Appendix A.

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*[20]*. Under these conventions, the only possible budget-balanced 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

[20]This is because increasing y_j does not change the score, while decreasing y_k for $k \neq j$ would lower the score, if $\bar{x}_k > 0$.

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 deviated to type k, by above natural conventions his BTS score would be zero, which is weakly better than not deviating, so the equilibrium could not 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 recall the Prelec (2004) result that, in such a truth-telling equilibrium, the BTS score is equal to the budget-balanced logarithmic payoff, and we provide a detailed proof in Appendix.

Theorem 4.1 *(Prelec 2004; Theorem 2) Under the above assumptions, when the players play the truth-telling equilibrium, BTS scoring results in budget-balanced logarithmic EP's. More precisely, in the equilibrium we have*

$$BTS^r = \log Pr(\bar{X} = \bar{x}|X^r = x^r) - \lim_{n\to\infty} \frac{1}{n} \sum_{s=1}^{n} \log Pr(\bar{X} = \bar{x}|X^s = x^s) \tag{4.1}$$

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

$$BTS^r = \log(Pr(\Omega = i \mid T^r = k)) - \sum_{j=1}^{M} Pr(T^r = j \mid \Omega = i) \log(Pr(\Omega = i \mid T^r = j)) \tag{4.2}$$

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.

To conclude, the main message of this section is the one confirming the superiority of BTS because of the following three properties: BTS leads to a (essentially) unique equilibrium, it results in the benchmark, logarithmic ex-post scores, and it is easily implementable. We know of no other PLEPS that has these properties.

5 Conclusions

We consider the problems of extracting true opinions from a large group of respondents and of ranking them according to their posteriors on state of nature, 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 for equilibrium payoffs that are determined only by the posteriors and type probabilities: (i) under assumptions on the sensitivity of score differences, the incentive compatible budget-balanced equilibria necessarily result in logarithmic payoffs; (ii) for arbitrary mechanisms, any incentive compatible equilibrium necessarily ranks the respondents according to the relative size of their posterior probabilities of the realized state of nature. 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. Furthermore, it may be of interest to study scoring rules that do not satisfy posterior locality, such as the rules studied in Cvitanić and Prelec (2015). Finally, ours is a static game, while many applications are dynamic by nature.

6 Appendix A

6.1 Definition of equilibrium and assumptions

In this section we identify sufficient conditions to have well-defined equilibrium payoffs F_i given a scoring function f.

Recall that 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 $f(\sigma^r(t^r), \sigma^{-r}(t^{-r}))$, where f is a scoring function that takes the responses to the set of real numbers. Function $f(\cdot,\cdot)$ is of the same functional form for all N and Q.

We assume that the players are risk-neutral and maximize the expected score value.

Definition 6.1

- *(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 response choice $a^r \neq \sigma^r(t^r)$, we have, with expectation taken with respect to the (conditional) distribution of Ω,

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

The strategy profile σ 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 σ a Strictly Separating (Bayesian) Nash Equilibrium (SSNE).

- *(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 σ_Q, if every budget-balanced SNE is an SSNE, and if any two budget-balanced SSNE's σ_Q and σ'_Q result in the same scores: $f(\sigma_Q) = f(\sigma'_Q)$.*

Remark 6.1 *Condition (ii) essentially assumes uniqueness of the budget-balanced SSNE σ_Q. We have shown that the budget-balanced logarithmic scoring can be implemented by an USSR (that is, by BTS) 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.*

Note that by the above definition, an USSR score of a certain type is the same in any budget-balanced SNE. We use notation F_i for such equilibrium score in state i, and the same notation for non budget-balanced versions of the equilibrium payoffs. In this paper, we always impose the following assumption on the equilibrium payoffs:

Assumption 6.1

- **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* $F_i : (0,1)^{2M} \to \mathbb{R}$,

$$f(\sigma_Q^r(k), \sigma_Q^{-r}(T^{-r})) = F_i(z_k^i, z_{-k}^i; s_k, s_{-k}) \ ,$$

where, for example, $z_{-k}^i = (z_1^i, ..., z_{k-1}^i, z_{k+1}^i, ..., z_M^i)$.

7 Appendix B

Proof of Proposition 2.1:

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 $f(\sigma^r(t^r), \sigma^{-r}(t^{-r}))$. Let us denote by σ the equilibrium strategy profile of all the respondents, and define ρ to be the strategy profile that is identical to σ, except that a specific player r of type $k \neq j$ plays the strategy $\sigma^r(j)$ corresponding to type j. Let s denote a player of type j. Then, we have that the payoff to the mimicry strategy, when r plays j is

$$\begin{aligned} f(\rho^r(k), \rho^{-r}(T^{-r})) &= f(\rho^s(j), \rho^{-s}(T^{-s})) \\ &= f(\sigma^s(j), \rho^{-s}(T^{-s})) \\ &= f(\sigma^s(j), \sigma^{-s}(T^{-s})) \end{aligned}$$

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, equilibrium payoffs corresponding to σ determine the payoff of a mimicry deviation by player r.

■

Proof of Proposition 2.2:

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, any PLEPS functions F_i would 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 would be

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

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

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

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 7.1 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.

■

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 $\lambda(\vec{p}, \vec{r}, s_p, s_q, s_r)$, where, for example, $\vec{p} = (p^1, \ldots, p^N)$, such that

$$p^i[\partial_p F_i(p^i, p^i, r^i, s_p, s_q, s_r) - \partial_q F_i(p^i, p^i, r^i, s_p, s_q, s_r)] = \lambda(\vec{p}, \vec{r}, s_p, s_q, s_r) \tag{7.1}$$

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 $\lambda(\vec{p}, \vec{r}, S)$ 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 3.1:

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

$$c^r_{m,n} = d_{m,n} \ , \quad e_{l,m,n} = e_{m,l,n}$$

because then we can write

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

where H is symmetric in all the pairs $(p^i, s_p), (q^i, s_q), (r^i_j, s_{r_j})$.

Let us consider the consequences of strict separation inequality (3.7), using Lemma 3.1. We have

$$\partial_q F(p,p,r) - \partial_p F(p,p,r)$$
$$= \sum_{n=1}^{\infty} n b^q_n \bar{p}^{n-1} + \sum_{m,n=1}^{\infty} c^q_{m,n} n \bar{p}^{m+n-1} + \sum_{m,n=1}^{\infty} d_{m,n} m \bar{p}^{m-1}\bar{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m \bar{p}^{l+m-1}\bar{r}^n$$
$$-\sum_{n=0}^{\infty} n a_n \bar{p}^{n-1} - \sum_{m,n=1}^{\infty} m\bar{p}^{m-1}(c^q_{m,n}\bar{p}^n + c^r_{m,n}\bar{r}^n) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} l \bar{p}^{l+m-1}\bar{r}^n$$

We can then write

$$p\partial_q F(p,p,r) - p\partial_p F(p,p,r)$$
$$= \sum_{n=1}^{\infty} n b^q_n \bar{p}^{n} + \sum_{m,n=1}^{\infty} c^q_{m,n} n \bar{p}^{m+n} + \sum_{m,n=1}^{\infty} d_{m,n} m \bar{p}^{m}\bar{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m \bar{p}^{l+m}\bar{r}^n$$
$$-\sum_{n=0}^{\infty} n a_n \bar{p}^{n} - \sum_{m,n=1}^{\infty} m\bar{p}^{m}(c^q_{m,n}\bar{p}^n + c^r_{m,n}\bar{r}^n) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} l \bar{p}^{l+m}\bar{r}^n$$
$$+\sum_{n=1}^{\infty} n b^q_n p^0 \bar{p}^{n-1} + \sum_{m,n=1}^{\infty} n c^q_{m,n} p^0 \bar{p}^{m+n-1} + \sum_{m,n=1}^{\infty} d_{m,n} m p^0 \bar{p}^{m-1}\bar{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m p^0 \bar{p}^{l+m-1}\bar{r}^n$$
$$-\sum_{n=0}^{\infty} n a_n p^0 \bar{p}^{n-1} - \sum_{m,n=1}^{\infty} m p^0 \bar{p}^{m-1}(c^q_{m,n}\bar{p}^n + c^r_{m,n}\bar{r}^n) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} l p^0 \bar{p}^{l+m-1}\bar{r}^n$$

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

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

$$c^r_{1,n} = d_{1,n} \tag{7.2}$$

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

$$0 = c^r_{1,n} - d_{1,n} + c^r_{2,n} - d_{2,n} \tag{7.3}$$

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

$$0 = 2(d_{2,n} - c^r_{2,n}) + 3p^0(d_{3,n} - c^r_{3,n}) + p^0(e_{1,2,n} - e_{2,1,n}) \tag{7.4}$$

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

$$0 = 3(d_{3,n} - c^r_{3,n}) + (e_{1,2,n} - e_{2,1,n}) + 4p^0(d_{4,n} - c^r_{4,n}) + 2p^0(e_{1,3,n} - e_{3,1,n}) \tag{7.5}$$

And so on.

So, it is sufficient to show $e_{l,m,n} = e_{m,l,n}$ This follows directly from Assumption 3.1, 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,

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

This completes the proof.

■

Then following lemma is the key ingredient in proving Theorem 3.2. It is a slight extension of Lemma A.1 in Schervish (1989).

Lemma 7.1 (Schervish 1989). *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 $0 < q(-A) + (a-q)(-B) < q(-A) + \frac{a-q}{a-p}pA = Aa\frac{p-q}{a-p} < 0$, which is impossible. This contradiction proves $A > 0$.

■

Proof of Theorem 3.2:

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 $q_{ki} = z_k^i s_k$.

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 $\begin{bmatrix} z_1^1 & z_1^2 \\ z_2^1 & z_2^2 \end{bmatrix}$. 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 $1 - q > 1 - p$). The IC property (2.1) 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

$qF_1(q, p) + (1 - q)F_2(1 - q, 1 - p) > qF_1(p, q) + (1 - q)F_2(1 - p, 1 - q)$.

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

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

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 7.1 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.

Case 2: Assume $M \geq 3, N = 2$. The matrix Z can be written as $\begin{bmatrix} z_1^1 & z_1^2 \\ z_2^1 & z_2^2 \\ \vdots & \vdots \\ z_M^1 & z_M^2 \end{bmatrix}$.

The matrix entries satisfy $z_k^2 = 1 - z_k^1, k = 1, ..., M$. Take any $k, j \in \{1, ..., M\}$ 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 $z_{-j,k}^i$ for the $(N - 2)$-tuple which consists of $\{z_1^i, ..., z_M^i\} \setminus \{z_j^i, z_k^i\}$, from (2.1) we obtain that the following two equations hold:

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

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

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

$$B = F_2(1-p, 1-q, z^2_{-(j,k)}) - F_2(1-q, 1-p, z^2_{-(j,k)}) = F_2(z^2_k, z^2_{-k}) - F_2(z^2_j, z^2_{-j})$$

we are again within the framework of Lemma 7.1, and we conclude that $A > 0$ and $B < 0$, which proves (3.9) for both $i = 1$ and $i = 2$.

Case 3: Assume $M \geq 2, N \geq 3$ and denote $Z = \begin{bmatrix} z^1_1 & z^2_1 & \dots & z^N_1 \\ z^1_2 & z^2_2 & \dots & z^N_2 \\ \vdots & \vdots & \vdots & \vdots \\ z^1_M & z^2_M & \dots & z^N_M \end{bmatrix}$. 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, $\begin{bmatrix} z^1_1 \\ z^1_2 \\ \vdots \\ z^1_M \end{bmatrix}$ and $\begin{bmatrix} z^i_1 \\ z^i_2 \\ \vdots \\ z^i_M \end{bmatrix}$. We take any rows $j, k \in \{1, ..., M\}$ 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 $p := z^i_k$ and $q := z^i_j$, with $1 > p > q > 0$, for any choice of $z^i_{-(k,j)} \in (0,1)^{M-2}$ (if $M = 2$ this last requirement is unnecessary), and for any choice of corresponding S_i, we have

$$F_i(p, q, z^i_{-(k,j)}) > F_i(q, p, z^i_{-(k,j)}) \tag{7.6}$$

Observe that other entries of the matrix Z (in other than the i-th column) do not enter into (7.6), and therefore can be adjusted accordingly, as long as we have a Z-matrix.

Because we can always choose Q by setting $q_{ki} = z^i_k s_k$, 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

$$\tilde{Z}_l^t := \begin{cases} z_l^t, & \text{if } l \in \{1, ..., M\} \setminus \{j, k\} \\ p, & \text{if } l = k, t = i \\ q, & \text{if } l = j, t = i \\ a - p, & \text{if } l = k, t = 1 \\ a - q, & \text{if } l = j, t = 1 \\ \frac{\epsilon}{N-2}, & \text{otherwise} \end{cases}$$

where p, q are arbitrary values in $(0, a)$ with $p > q$. Then for every choice of ϵ 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 IC property (2.1) applied now to j and k yields

$$\sum_{t=1}^{N} \tilde{z}_k^t F_t(\tilde{z}_k^t, \tilde{z}_{-k}^t) > \sum_{t=1}^{N} \tilde{z}_k^t F_t(\tilde{z}_j^t, \tilde{z}_{-j}^t)$$

$$\sum_{t=1}^{N} \tilde{z}_j^t F_t(\tilde{z}_j^t, \tilde{z}_{-j}^t) > \sum_{t=1}^{N} \tilde{z}_j^t F_t(\tilde{z}_k^t, \tilde{z}_{-k}^t)$$

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

$$(a-p)F_1(a-p, a-q, z_{-(j,k)}^1) + pF_i(p, q, z_{-(j,k)}^i) > (a-p)F_1(a-q, a-p, z_{-(j,k)}^1) + pF_i(q, p, z_{-(j,k)}^i)$$

$$(a-q)F_1(a-q, a-p, z_{-(j,k)}^1) + qF_i(q, p, z_{-(j,k)}^i) > (a-q)F_1(a-p, a-q, z_{-(j,k)}^1) + qF_i(p, q, z_{-(j,k)}^i)$$

If we set A and B as $A = F_i(p, q, z_{-(j,k)}^i) - F_i(q, p, z_{-(j,k)}^i)$ and $B = F_1(a-p, a-q, z_{-(j,k)}^1) - F_1(a-q, a-p, z_{-(j,k)}^1)$, we are again within the framework of Lemma 7.1. Therefore $A > 0$, which proves inequality (7.6) for $a > p > q > 0$. By letting $\epsilon \to 0$, we obtain (7.6) for $1 > p > q > 0$.

■

Proof of Theorem 4.1:

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

$$Pr(X^r = x^r | X^s = x^s) = \frac{p_{ij}}{\sum_{k=1}^M p_{kj}} \tag{7.7}$$

We will need the following three properties.

- Property I: $y_j^r = \sum_{i=1}^M x_i^r \frac{p_{ij}}{\sum_{k=1}^M p_{ki}}$

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

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

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 (7.7). For Property III, let ℓ be such that $x_\ell^r = 1$. De Finetti's theorem implies

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

The sum on the right always has only one term different from zero. Therefore, taking the log implies Property III.

Next, let x^s be any values such that

$$\bar{x}_k = \lim_n \frac{1}{n} \sum_s x_k^s$$

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 a lower bound A and an upper bound B such that $0 < A \leq Pr(X^s = x^s) \leq B < 1$. Then it follows that $A = \sqrt[n]{A^n} \leq \sqrt[n]{\prod_{s=1}^n Pr(X^s = x^s)} \leq \sqrt[n]{B^n} = B$.

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

$$\sum_{k=1}^{M} \bar{x}_k \log y_k^r = \lim_n \frac{1}{n} \sum_s \log Pr(X^s = x^s | X^r = x^r)$$

and

$$\sum_{k=1}^{M} x_k^r \log \bar{y}_k = \lim_n \frac{1}{n} \sum_s \log Pr(X^r = x^r | X^s = x^s)$$

and so, using Bayes rule,

$$\begin{aligned}
BTS^r &= \sum_{k=1}^{M} x_k^r \log \frac{\bar{x}_k}{\bar{y}_k} + \sum_{k=1}^{M} \bar{x}_k \log y_k^r \\
&= \log Pr(X^r = x^r | \bar{X} = \bar{x}) + \lim_n \frac{1}{n} \sum_s \log Pr(X^s = x^s | X^r = x^r) - \lim_n \frac{1}{n} \sum_s \log Pr(X^r = x^r | X^s = x^s) \\
&= \log \left(Pr(X^r = x^r | \bar{X} = \bar{x}) \lim_n \Pi_{s=1}^{n} \frac{Pr^{1/n}(X^s = x^s | X^r = x^r)}{Pr^{1/n}(X^r = x^r | X^s = x^s)} \right) \\
&= \log \left(Pr(X^r = x^r | \bar{X} = \bar{x}) \frac{\lim_n \Pi_{s=1}^{n} Pr^{1/n}(X^s = x^s)}{Pr(X^r = x^r)} \right) \\
&= \log Pr(\bar{X} = \bar{x} | X^r = x^r) - \log Pr(\bar{X} = \bar{x}) + \lim_n \frac{1}{n} \sum_s \log Pr(X^s = x^s)
\end{aligned}$$

Since the last two terms do not depend on r, and $\sum_r BTS^r = 0$, we get equation (4.1). 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 (4.1) as

$$BTS^r = \log Pr(\bar{X} = \bar{x} | X^r = x^r)$$

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

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

Since

$$\lim_{n\to\infty} \frac{n_j}{n} = Pr(T^r = j \mid \bar{X} = \bar{x})$$

we prove equation (4.2).

■

References

[1] Aldous, D.J. (1985) ”Exchangeability and related topics”. École d'été de probabilités de Saint-Flour, XIII - 1983, 1 - 198, Lecture Notes in Mathematics, 1117, Springer, Berlin.

[2] Bergemann, D. and Morris S. (2012) An Introduction to Robust Mechanism Design. In “Foundations and Trends in Microeconomics” 8, 169–230.

[3] Bernardo, J.M. (1979) ”Expected information as expected utility”. *The Annals of Statistics* 7, 686–690.

[4] Börgers, T. (with a chapter by Krähmer,D. and Strausz, R.) (2013) “An Introduction to the Theory of Mechanism Design”. Lecture Notes, U. of Michigan.

[5] Chow, Y.S. and Teicher, H. (1997) ”Probability Theory. Independence, Interchangeability, Martingales”. 3rd ed. Springer Texts in Statistics. Springer-Verlag, New York.

[6] Cvitanić, J. and Prelec, D. (2015) “Truth-Inducing Scoring Rules”. Working paper.

[7] Feller, W. (1971). ”An Introduction to Probabiliy Theory and its Applications”. Vol II, 2nd ed. John Wiley and Sons, Inc., New York.

[8] Gneiting, T. and Raftery, A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* 102, 359–378.

[9] Maskin, E. and Sjöström, T. (2002) Implementation theory. In “Handbook of Social Choice and Welfare”; K. J. Arrow, A. K.Sen and K. Suzumura (ed.) 1, 237–288, Elsevier.

[10] McCarthy, J. (1956) Measures of the value of information. *Proc Natl Acad Sci USA* 42, 654–655.

[11] Miller, N., Resnick, P. and Zeckhauser, R. (2005) Eliciting Informative Feedback: The Peer-Prediction Method. *Management Science* 51, 1359–1373.

[12] Myerson, R. B. (1981) Optimal Auction Design *Mathematics of Operations Research* 6, 58–73.

[13] Offerman, T., Sonnemans, J., Van De Kuilen, G. and Wakker, P.P. (2009) A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *Review of Economic Studies* 76, 1461–1489.

[14] Prelec, D. (2004) A Bayesian Truth Serum for Subjective Data. *Science* 306, 462–466.

[15] Prelec, D., Seung, H.S. and McCoy, J. (2013) Finding truth even if the crowd is wrong. Working paper, MIT.

[16] Savage, L.J. (1971) Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association* 66, 783–801

[17] Schervish, M. (1989) A General Method for Comparing Probability Assessors. *The Annals of Statistics* 17, 1856–1879.

[18] Schlag, K.H. and van der Weele, J.J. (2013) Eliciting Probabilities, Means, Medians, Variances and Covariances without assuming Risk Neutrality. *Theoretical Economics Letters* 2, 38–42.

[19] Vohra, R.V. (2007) Paths, Cycles and Mechanism Design. Preprint.

[20] Waggoner, B., and Chen, Y. (2013) Information Elicitation Sans Verification. In Proceedings of the 3rd Workshop on Social Computing and User Generated Content (SC13).

[21] Witkowski, J., and Parkes, D. (2012) Peer Prediction Without a Common Prior. In Proceedings of the 13th ACM Con-ference on Electronic Commerce (EC12).