THEORY PROBAB. APPL.
Vol. 65, No. 2, pp. 292–321

INCENTIVE-COMPATIBLE SURVEYS VIA POSTERIOR PROBABILITIES*

J. CVITANIĆ†, D. PRELEC‡, S. RADAS§, AND H. ŠIKIĆ¶

Abstract. We consider the problem of eliciting truthful responses to a survey question when the respondents share a common prior that the survey planner is agnostic about. 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 that, under additional smoothness and sensitivity conditions, the payoff in the truth-telling equilibrium must be a logarithmic function of those posterior probabilities. Moreover, the respondents 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

DOI. 10.1137/S0040585X97T989957

1. Introduction. Consider the problem of truthful elicitation of responses in a population of Bayesian agents who share a common prior.[1] We allow the survey planner to be "agnostic" about the prior (although the planner may have beliefs about the prior, she prefers to keep these beliefs private). The true outcome that responses are elicited about is not verifiable. If the (lack of) respondent honesty or care is an issue, the survey planner may want to implement an incentive-compatible mechanism or a "scoring rule." We say that a multiperson scoring rule is (strictly) incentive-compatible if, for each respondent π, the following condition is satisfied: if every other respondent responds truthfully, then the honest answer of the respondent π (strictly) maximizes its expected score. Incentive-compatible scoring rules play an important role in survey studies in various fields, most notably economics and business. Thus, it would be very valuable to both researchers and practitioners in numerous applications to characterize usable incentive-compatible algorithms. Should we

*Received by the editors October 7, 2018. This paper was presented at the conference "Innovative Research in Mathematical Finance" (September 3–7, 2018, Marseille, France). A previous version of this article circulated under the title "Mechanism design for an agnostic planner: Universal mechanisms, logarithmic equilibrium payoffs and implementation." The first author was supported in part by NSF grant DMS 1810807. The second author was supported by Intelligence Advanced Research Projects Activity (IARPA) via the Department of Interior National Business Center (contract number D11PC20058). The third author was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (grant PIOF-GA-2013-622868-BayInno). The fourth author was 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. Originally published in the Russian journal *Teoriya Veroyatnostei i ee Primeneniya*, 65 (2020), pp. 368–408.
https://doi.org/10.1137/S0040585X97T989957

†Division of the Humanities and Social Sciences, Caltech, Pasadena, CA 91125 (cvitanic@hss.caltech.edu).

‡Sloan School of Management, Department of Economics, Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139 (dprelec@mit.edu).

§The Institute of Economics, Zagreb, Croatia, and Sloan School of Management, MIT, Cambridge, MA 02139 (sradas@eizg.hr, sradas@mit.edu).

¶Faculty of Science, Department of Mathematics, University of Zagreb, Zagreb, Croatia (hsikic@math.hr).

[1]For example, a market survey in which the respondents are asked whether they would buy a particular new product.

expect a variety of such algorithms, or is the incentive-compatibility a rather restrictive property? We show in this paper that, under a fairly natural "equilibrium-locality" condition, the property is more on the restrictive side and guides us to an algorithm that is well known in the literature. Let us now discuss in detail the precise nature of our results.

Prelec [24] introduced the Bayesian truth serum (BTS) algorithm based on two inputs determined from each respondent: a declaration of the respondent's type and his belief about the distribution of the type declarations by other players. BTS has two important properties [24]. First, it is strictly incentive-compatible (IC) (that is, strict truth-telling is an equilibrium), so that the agents' "types" corresponding to their true responses are fully revealed.[2] Second, the BTS equilibrium score of a respondent is, up to a linear transformation, equal to the logarithm of his posterior probability of the true state of nature, so that BTS *ranks respondents by posteriors* (henceforth called posterior ranking). We say that the BTS mechanism results in *logarithmic scoring.* BTS has been successfully applied in various fields, including new product adoption [12], economics and psychology [15], knowledge design [20], and criminology [17].

In the present paper we consider the following motivating questions: what are conditions for equilibrium payoffs generated by a strictly incentive-compatible scoring rule to necessarily correspond to logarithmic scoring? Which conditions are such that a strict incentive-compatible equilibrium must satisfy the posterior locality condition? In other words, which conditions are such that equilibrium payoffs are essentially equivalent to BTS payoffs?

Our main results are as follows. We identify two conditions on equilibrium payoffs that we call the "posterior locality" and the "separation of variables" properties. The main theorem says that

$$\left.\begin{array}{r}\text{incentive-compatible}\\ +\text{ posterior locality}\\ +\text{ separation of variables}\end{array}\right\}\to\text{logarithmic scoring.}$$

The second main result asserts that

$$\left.\begin{array}{r}\text{incentive-compatible}\\ +\text{ posterior locality}\end{array}\right\}\to\text{posterior ranking.}$$

To make these results more precise, we now describe our setting in more detail. There is a state of nature drawn from a finite set and an infinite population (for application purposes, one can think of having a large population). There is a planner who asks each respondent to submit responses to a questionnaire. The respondents are players in a Bayesian game where each player is rewarded by the planner according to a score that depends on his responses and the responses of everyone else. Each player has a type, and all types take values in a finite set. The variation in types can be interpreted as a consequence of players observing varying signals. (As shown in [26] and in the present paper, when using a different proof, a report of just types cannot lead to a truth-telling equilibrium.[3]) The types are conditionally independent and identically distributed (i.i.d.) w.r.t. the state of nature, so that there is a single

[2] 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.

probability distribution (the "prior") that describes the joint distribution of the state of nature and types. We assume that the prior is common knowledge among players but unknown to the planner ex-ante.

Let us now recall the notion of proper scoring rules in a framework with only one respondent. Consider a random variable (r.v.) Ω taking values in $\{1,\ldots,N\}$, $N>1$, representing the state of nature. Based on the respondent's declarations, his belief about the distribution $\widetilde{p} = (\widetilde{p}^1,\ldots,\widetilde{p}^N)$ of Ω is implied. As a consequence, if the outcome $\Omega = i$ occurs, the respondent is paid $F_i(\widetilde{p})$. In the multirespondent setting of this paper, each i itself is a distribution but over the beliefs of the respondents. In this case, if $\widetilde{p}$ is implied by the respondent's answer, and if the outcome $\Omega = i$ occurs, we refer to $\widetilde{p}_i$ as his "local posterior."

With only one respondent, a family of functions $\{F_i\}_{i=1,\ldots,N}$ is called a strictly proper scoring rule if it is incentive-compatible with truth-telling. This means that the respondent's expected payoff is maximized at his true belief, and the maximum is attained at the respondent's posterior (denoted p). More precisely, for all probability vectors $\widetilde{p} \neq p$, we have

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

There are many proper scoring rules. A general characterization with many examples is provided in [11].[4] 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 that the respondent assigns to the outcome $\Omega = i$ that is actually realized, and does not depend on how the 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 (see [28], [4]). Such a rule is a natural choice if the local posterior is interpreted as a measure of the respondent's expertise, that is, if the quality of the respondent's signal/type is measured by the probability assigned to the true state of nature.

As mentioned above, in the multiplayer setting the values of Ω are interpreted as distributions of the types in the infinite population, and posteriors as the probabilities that a type assigns to these values of Ω. We do not allow all distributions but only finitely many. In practice, the respondents would be given a discretized choice of distributions, for example, "Do you think the percentage of votes for the candidate A will fall within the range of 0%–10%, or 10%–20%, ... ?" Assuming that our game has a strictly type-separating equilibrium, our results depend only on the form of the equilibrium payoffs rather than on the actual scoring rules that lead to those payoffs. The posterior locality condition posits that the equilibrium payoffs F_i in the state $\Omega = i$ depend on the local posteriors, that is, on the posterior probabilities of the event $\Omega = i$.

Anticipating the results of the present paper, we show now that, under mild smoothness conditions, if the equilibrium payoffs F_i satisfy a property analogous to (1.1), the difference in the state i scores of two respondents with local 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. To explain what we mean by this, we now assume,

[3] 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. For some elements of the finite case theory within the information-theoretic setup, see [7].

[4] Offerman et al. [23] consider the case in which the respondents may have nonexpected utilities, with only two possible states of nature.

for simplicity of notation, the case with two types with local posteriors p^i and q^i. If a player implies, by her responses, local posterior p^i, she receives $F_i(p^i, q^i)$, and if she implies q^i, she gets $F_i(q^i, p^i)$. The incentive compatibility of the family $\{F_i\}$, as generalized from the one-player incentive compatibility (1.1) of proper scoring rules, means that the solution to the problem

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

is $q^i = p^i$, where λ is a Lagrange multiplier for the constraint $\sum_i q^i = 1$. The first-order condition for the above problem reads as

$$\partial_q F_i(p^i, p^i) - \partial_p F_i(p^i, p^i) = -\frac{\lambda}{p^i},$$

where ∂_x denotes the partial derivative w.r.t. x. In the one player case, this reads $F'(p_i) = \lambda/p_i$ and results in the log function as the only proper scoring rule that satisfies the locality condition. In our multiplayer case, we fix p^i and expand the score difference up to the first order as a function of q^i around the point p^i to get

$$F_i(p^i, q^i) - F_i(q^i, p^i) \approx [\partial_q F_i(p^i, p^i) - \partial_p F_i(p^i, p^i)](q^i - p^i).$$

Finally, combining the above equations and using the fact that the first-order Taylor expansion of the log function around $q_i = p_i$ is $\log p^i - \log q^i = (p^i - q^i)/p^i$, we get

$$F_i(p^i, q^i) - F_i(q^i, p^i) \approx \lambda\left(1 - \frac{q^i}{p^i}\right) \approx \lambda(\log p^i - \log q^i).$$

Based on this approximation, our first theorem asserts the following: if we add to the incentive compatibility mild requirements on payoff smoothness and sensitivity to other players of the difference in equilibrium payoffs of two respondents, then the difference in incentive-compatible scores of the two respondents is exactly, rather than just approximately, proportional to the difference in the logarithms of the implied local posteriors.

Our second theorem asserts that any incentive-compatible equilibrium payoff $F_i(p^i_k, p^i_{-k})$ of the player who implies, via his responses, probability p^i_k corresponding to type k, given that other types imply probabilities p^i_{-k}, is nondecreasing in the local posterior probability p^i_k. Consequently, the ranking of experts in equilibrium, if we consider p^i_k as the measure of expertise, is the same given any incentive-compatible mechanism, and it corresponds to the ranking by posteriors. The result is very general and is proved by purely algebraic methods. It is a generalization of the results in the literature (in the case of one respondent) on monotonicity, which is implied by incentive compatibility of proper scoring rules (see, e.g., [19], [28], [29], and [30]).[5]

We also discuss implementation problems. Observe that in general, while a particular ex-post payoff of the form $F_i(p^i_k, p^i_{-k})$ may arise in theory in equilibrium, it is not necessarily simple to implement 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

[5]See also Prelec, Seung, and McCoy [25] 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.

be as simple as possible. Under an assumption somewhat stronger than posterior locality but without assuming separation of variables, we show that the payoffs of all strictly-separating equilibria in our framework can be implemented by particular questionnaires; however, the latter may be complex, except for the logarithmic, BTS case. In this context, let us recall that Prelec [24] showed 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. We also show that the budget-balanced strict equilibrium under BTS is necessarily separating.

Relationship to existing results. In recent years, after the appearance of [24] proper scoring rules in the game-theoretic context, also called "information elicitation without verification," were studied extensively. In the case where the planner knows the prior distribution of the player types, an early work is [21], where a clever use of proper scoring rules was designed to elicit truthful information. However, because the assumption that the planner knows the prior is unrealistic in practice, alternative approaches have been proposed. Jurca and Faltings [13] use robust optimization to deal with small variations in belief models. Expanding on the framework of Prelec [24], where the planner does not know the prior but the number of players is infinite, Witkowski and Parkes [34] and Witkowski [33] devise mechanisms, under the name Robust Bayesian Truth Serum (RBTS), that worked for a finite number of respondents but with only two types. Waggoner and Chen [32] consider a general framework without any assumptions on information structure. Radanovic and Faltings [26], [27] and Zhang and Chen [36] developed mechanisms that were incentive-compatible for any number of agents and nonbinary player types. Dasgupta and Ghosh [9] and Witkowski and Parkes [35] relaxed the knowledge requirement even more by asking for more extensive reports by respondents in the form of responses to multiple similar questions. Baillon [2] implemented truth-telling equilibria via a "Bayesian market" in the case of binary types. Cvitanić et al. [8] provided additional incentive-compatible mechanisms that were simple to explain to respondents. In [10], the authors characterized "minimal" peer-prediction mechanisms; that is, those where the score depended only on the respondent's reported type and the reported type of another, suitably chosen "peer" respondent. We, on one hand, allow the scores to depend on reported types of all the respondents and on their predictions of those reports, and, on the other hand, we impose the assumption of locality on the scores. The preference for one or the other approach depends on whether the minimality or locality is the preferred feature of the mechanism. The paper [14] developed an informational theoretic paradigm for designing incentive mechanisms, which included, as special cases, many established mechanisms, including BTS, and showed that the properties of BTS could be proved in a simpler way by using a connection to Shannon mutual information. For a different connection between BTS and information theory, see [7]. Liu and Chen [16] designed a "uniform dominant" truth serum when there was a noisy signal of the ground truth and there were sufficiently many agents and tasks. Their scoring depended on whether a report is informative or not, and so, in a sense, they were related to the expertise of the respondent.

However, with the exception of Prelec [24], all of the papers mentioned above were concerned with incentive compatibility rather than with ranking of respondents, so that the proposed mechanisms either do not satisfy our posterior locality condition or require the planner to know the common prior of the respondents. We focus on mechanisms that have all of these three properties: they allow IC equilibria, they can be implemented even when the planner is agnostic about the prior, and they rank the respondents by posteriors. In this case, our main result says that, under relatively weak conditions, logarithmic scoring is the only possible equilibrium payoff form. When ranking by posteriors is not required from the mechanism, the above papers provide many other ways to design IC mechanisms.

The problem we tackle in the paper can also be considered as a mechanism design, since we seek to describe mechanisms that are incentive-compatible and have attractive features for opinion elicitation applications. On the one hand, our approach is more general than typical mechanism design models,[6] because we allow for uncertainty regarding both the players information (type) and the true state of nature, and these uncertainties may be correlated in a nontrivial way. It is exactly the joint distribution of the uncertainties 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 rather than imposing her conjectures on the survey respondents. Thus, she is interested in a "universal" mechanism, which would work for all priors without any input from her side apart from the initial formulation of the multiple-choice questionnaire.[7] From this perspective, the present paper is concerned with robust Bayesian mechanisms. On the other hand, our setup is less general in the following sense: 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 the expected payoff they attain. Moreover, our framework is less general than some models of robust mechanism design, which, unlike ours, do not assume common knowledge of the prior distribution by all players. (In our case only the planner may be ignorant.) In the concluding section 5, we discuss directions wherein 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, and section 5 presents conclusions. The proofs are presented in the appendix (section 6).

2. Model, definitions, and assumptions. In our model, a mechanism gives 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 multiple-choice questions and in ranking the players according to the quality of their information, which, in our framework, means the ranking 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 some years into the future and asks experts to respond to appropriately designed questions. Broader applications include voting in elections, predicting political events, conducting product market research or online product reviews, and any other application that involves a survey with

[6]See, e.g., [18], [3], [5]. We refer the reader to [3] for a detailed literature survey.

[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 is not implementable in practice, as we discuss in this paper.

[8]A negative score is usually called a transfer in the mechanism design literature; see, e.g., [5].

a multiple-choice question.[9] We note that it is not necessary to assume that the true response to a multiple-choice question is verifiable.

2.1. The model. The players are indexed by $\pi \in R$, where R is infinite and countable.[10] The state of nature is an r.v. Ω taking values in $\{1, \dots, N\}$, $N > 1$.[11] The players can be of $M > 1$ different types, which can be interpreted as random signals that the players receive about the state of nature. A player π's type is an r.v. T^π with values $t^\pi \in \{1, \dots, M\}$. We consider scoring mechanisms where, for a given fixed positive integer K, the player π submits as a response a K-dimensional value $a^\pi \in \mathbf{R}^K$ (a stands for "action"), where K is a fixed natural number.

A response a^π 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 his prior. We posit the following.

Assumption 2.1. (i) The family of signals T^π, $\pi \in R$, is a family of exchangeable r.v.'s i.i.d. conditional on the state of nature Ω.

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

Condition (i) implies that the order we use to consider our players is irrelevant (from the point of view of the probability distribution of the entire sequence). Moreover, by de Finetti's theorem, the exchangeability assumption actually implies Assumption 2.1(ii) to the effect that there exists an r.v. Ω such that T^π's are conditionally i.i.d. w.r.t. Ω; see, e.g., [1] or [6].

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

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

[9] Many more examples can be found in [24] and [25].

[10] We need the assumption that there are infinitely many players for several reasons: first, we do not want to impose assumptions on the form of the payoffs outside of equilibrium; for this, we use the fact that, with an 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, as is the implementation of equilibrium payoffs using practical inputs. We postpone for future research the analysis of the setup with finitely many players; finally, we need an 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, where the state of nature can 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 that the bottle is worth more than one thousand U.S. dollars.

[12] In section 4, it is shown that another question might be about the percentage of other respondents choosing a specific choice from the multiple-choice list.

[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 nature in addition to random types, with a nondegenerate correlation between the two. On the other hand, it is less general in that selecting a response to a question is the only action available to a player.

The prior and the posteriors. The joint distribution of types and states of nature is given by an $(M \times N)$-matrix $Q = [q_k^i]$, where

$$q_k^i = \Pr(T^\pi = k,\, \Omega = i).$$

Note that Q does not actually depend on π; this is a consequence of the exchangeability assumption.

We suppose that the matrix Q is common knowledge among the players; we refer to it as their "common prior." However, Q is 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 should know 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]

The 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 by

$$s_k = \Pr(T^\pi = k) \quad \text{and} \quad z_k^i = \Pr(\Omega = i \mid T^\pi = k),$$

respectively.

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 π. Moreover, $\sum_{i=1}^N z_k^i = 1$ for every $k \in \{1, \dots, M\}$, and any matrix with this property can be represented as a Z-matrix of posteriors for some joint distribution Q. We denote the vector $(s_1, \dots, s_M)$ by S.

2.2. Equilibrium payoff and incentive compatibility. In the standard literature on scoring rules, there is only one respondent who is asked to declare his posterior belief about the distribution of Ω; that is, to declare z^i's. If the outcome $\Omega = i$ occurs, the respondent is paid $F_i(z)$. A family of functions $\{F_i\}_{i=1,\dots,N}$ 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 that $\sum_{i=1}^N p^i F_i(p) > \sum_{i=1}^N p^i F_i(\widetilde{p})$ for all probability vectors $\widetilde{p} \neq p$.

In our framework with infinitely many respondents, *we consider only the payoff mechanisms that allow for a strictly separating Bayesian Nash equilibrium* (*SSNE*), *as defined in Definition* 2.1 *below, where the equilibrium payoffs are functions* F_i: $(0,1)^{2M} \to \mathbf{R}$ *of the form* $F_i(z_k^i, z_{-k}^i; s_k, s_{-k})$ where, for example, $z_{-k}^i = (z_1^i, \dots, z_{k-1}^i, z_{k+1}^i, \dots, z_M^i)$. We call this property the *posterior locality.*

We now make this more precise. A pure strategy for a player π is a map σ^π, which sends the player's type to his response choice a^π. We allow only pure strategies. The profile of all respondents' pure strategies is denoted by $\sigma(t)$, with entries $\sigma^\pi(t^\pi)$, and the profile excluding the player π is denoted by $\sigma^{-\pi}(t^{-\pi})$. The score for the player π is given by $f(\sigma^\pi(t^\pi), \sigma^{-\pi}(t^{-\pi}))$, where f is a scoring function, which takes the responses to the set of real numbers. The function $f(\,\cdot\,,\cdot\,)$ is of the same functional form for all N and Q.

[14]To get around the issue 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, the majority of responses would be different from one another, and the planner would have to harshly penalize most respondents.

We assume that the players maximize the expected score value. We mostly restrict the payoff mechanisms to those which are "budget balanced"; that is, such that the sum of the scores of all players is zero with probability one.[15]

Here is the definition of equilibrium.

Definition 2.1. (i) *Given a prior matrix* Q, *we say that a scoring function* f *allows a Strict* (*Bayesian*) *Nash Equilibrium* (*SNE*) *if there exists a pure strategy profile* $\sigma = \sigma_Q$ *such that, for all* π, t^π, $t^{-\pi}$, t^γ, *we have the following*: *for an arbitrary response choice* $a^\pi \neq \sigma^\pi(t^\pi)$,

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

with the expectation taken w.r.t. the (*conditional*) *distribution of* Ω.

In this case, 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* $\sigma^\pi(t^\pi) = \sigma^\gamma(t^\gamma) \Rightarrow t^\pi = t^\gamma$, *we call* σ *a Strictly Separating* (*Bayesian*) *Nash Equilibrium* (*SSNE*).

(ii) *We say that a scoring function* f *is a Universal Separating Scoring Rule* (*USSR*) *if, for all* Q, *it allows at least one budget-balanced SNE* σ_Q, *and if every budget-balanced SNE is an SSNE.*

We show below that the budget-balanced logarithmic scoring can be implemented by a USSR (that is, by BTS), which also satisfies the assumptions below on the equilibrium payoffs.

From now on, we only consider USSR functions f, or nonbudget-balanced versions thereof, so that there exists at least one SSNE. We denote by F_i the state i ex-post payoff in an SSNE corresponding to f, budget-balanced or not. The following is the additional crucial assumption we impose, and it is an assumption on the ex-post, equilibrium payoffs F_i of an SSNE.

Assumption 2.2 (posterior locality). For any $k \in \{1, \dots, M\}$, any $i \in \{1, \dots, N\}$, and any $j \neq k$, if $T^\pi = k$ and $\Omega = i$, then the equilibrium score (but not necessarily the out-of-equilibrium scores) of the player π has the representation

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

where $F_i \colon (0,1)^{2M} \to \mathbf{R}$.

We discuss this assumption at the end of this section. We do not address uniqueness of equilibrium, and we study only such equilibria that the realized equilibrium payoffs (but not necessarily the out-of-equilibrium payoffs) are of the above form. Moreover, we require that the realized equilibrium payoffs satisfy the conditions in the following definition.

Definition 2.2. *A 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 conditions are satisfied*:

[15]It should be mentioned that in a budget-balanced game the players know that they may receive negative "payments," and some players may not be willing to participate. In practice, the "payments" will not often be monetary but used as score points, and every respondent might be paid a nonnegative amount, which 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.

(i) Symmetry: *for all* $x, y \in (0,1)$, *for all* $z_2, \ldots, z_M$, $s_2, \ldots, s_M \in (0,1)$, *and for any permutation* Π *of* $\{2, \ldots, M\}$,

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

(ii) Incentive compatibility, strict separation inequality: *for any* Z*-matrix, for any* S*-vector, and for any* $k, j \in \{1, \ldots, M\}$ *such that* $(z_k^1, \ldots, z_k^N) \neq (z_j^1, \ldots, z_j^N)$,

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

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 the scoring function f. Assumption (ii) implicitly assumes that the players are risk-neutral and maximize the expected score. By Proposition 2.1 below, it is automatically satisfied if F_i are the equilibrium payoffs in a truth-telling equilibrium.

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

Remark 2.1. The crucial assumption in 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 a player's expertise. Note, however, that here we think of expertise regarding the actual state of nature in this one particular survey rather than about average accuracy over many surveys. It is the ex-post expertise resulting from the signal a player receives rather than the ex-ante expertise of obtaining good signals.

There are cases where the planner clearly wants to know about the distribution of types, such as elections or product market research, where the planner is trying to estimate what percentage of the population will vote for each candidate or is likely to buy a product. In such cases, it is intuitive that a respondent with higher local posterior is a better expert — he has the highest probability of being right about the actual distribution of responses, which is 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 local posterior on the distribution of types means a higher expertise. This is because in this example it is not necessarily the case that those who are better at estimating the percentage of their colleagues who will predict high inflation are also better at predicting the inflation. In such cases 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, as in the papers mentioned in the literature survey in the introduction.

We also note that in the present paper we look for the simplest possible equilibrium payoffs that describe players' expertise — this 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 in implementations these probabilities agree with type frequencies, which may be used to make a mechanism budget-balanced. Actually, for budget balance, it is sufficient to have a dependence on the local conditional probabilities $s_k^i = \Pr(T^\pi = k \mid \Omega = i)$, but we allow a dependence on s_k, s_{-k} for generality (except in section 4) 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 and s^{-k}, F_i depends only on $s_k^i = \Pr(T^\pi = k \mid \Omega = i)$. It is also natural to ask if, for a given f, the equilibrium implementing F is unique. We show below that this is essentially true for the benchmark example of the Bayesian Truth Serum scoring rule.

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 an infinite number of players for this result.

PROPOSITION 2.1. *Suppose there exists a strictly separating Bayesian Nash equilibrium for our game of respondents such that the ex-post payoffs are given by PLEPS* $\{F_i\}$. *If a respondent of type* k *deviates from the equilibrium by using the strategy of type* $j \neq k$, *then his deviation payoff is equal to* $F_i(z_j^i, z_{-j}^i; s_j, s_{-j})$. *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.

This is indeed so, because every type is represented by infinitely many players, the equilibrium payoffs are strictly separating, and the scoring function f is symmetric in their responses. See the proof of Proposition 2.1 in section 6.

We have the following negative result, proved in section 6, when the number of players is finite.[16]

PROPOSITION 2.2. *Assume* (*only in this proposition*) *a finite number of players but at least two players. Then there exists no budget-balanced PLEPS.*

We also note that even a nonbudget-balanced version of BTS is not incentive-compatible when there are finitely many players.

Ex-ante versus 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 discuss this issue in the implementation section, section 4, 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 [24] 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

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

(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 $\log z_k^i$; rather, he is asked for simpler inputs that determine his promised score via a specific function f (called Bayesian Truth Serum), and the value of the score turns out to be equal to $\log z_k^i$ when the players play the truth-telling equilibrium.

3. Possible equilibrium payoffs.

3.1. Logarithmic equilibrium payoffs.

In this subsection, we present examples of PLEPSs and consider the question of whether logarithmic equilibrium payoffs or simple modifications thereof are the only possible PLEPSs.

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 so because of the well-known *Gibbs inequality*, which says that, for a probability vector $(p^1, \dots, p^N)$,

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

The last equality can be verified by considering the problem

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

where λ is a Lagrange multiplier for the constraint $\sum_i q^i = 1$. The first-order conditions $p^i/q^i = \lambda$ for the problem are satisfied with $q^i = p^i$.

The question arises of 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. However, we next show that under mild additional conditions logarithmic equilibrium payoffs are, in fact, the only possible PLEPSs.

3.1.2. *Other examples of PLEPSs.* Let us first note that there are variations of the logarithmic equilibrium payoffs that produce equivalent scores when we require budget balance. For instance, for some function G symmetric in all the arguments and for some constant K, consider the function

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

where the dependence on the state of nature i is suppressed. This is a PLEPS function, which can be verified in the same way as for problem (3.2). However, it is not really different from logarithmic equilibrium payoffs 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 equilibrium payoffs as for the budget-balanced logarithmic equilibrium payoffs.

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, $M = 3$, denote

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

and define the 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 (2.1). This is because the first-order conditions (FOCs) for the Lagrangian optimization problem

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

read as

$$p^i[\partial_p F(q^i, p^i, r^i) - \partial_q F(p^i, q^i, r^i)] = \lambda \tag{3.3}$$

for some Lagrange multiplier λ, where $\partial_x F$ is the derivative w.r.t. x. These FOCs are satisfied for the above function with $q^i = p^i$. For large enough K, the above FOCs are also sufficient conditions for optimality, because the second-order optimality conditions are also satisfied, which implies (2.1).

Remark 3.1. Even though there are "strange" PLEPS functions F_i as in the example above, as we explained in the introduction, the difference in two equilibrium payoffs is proportional for all of them to the difference of logarithmic payoffs, up to the first order. This is also true if F_i depends on type probabilities s_k, under the conditions of Lemma 3.1 below.

We next identify conditions under which there can be no second-order terms, and the budget-balanced logarithmic equilibrium payoff is the only budget-balanced PLEPS.

3.2. When are equilibrium payoffs logarithmic? We assume[17] in this section that $N \geqslant 3$. 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 now find conditions such that the higher-order terms cannot appear, and any PLEPS is essentially a logarithmic equilibrium payoff.

We proceed as follows:

(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 this assumption implies an additive representation of the equilibrium payoff of a given type — the equilibrium payoff is a sum of a term that does not depend on the posteriors of other types and a term that is symmetric in type;

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

(iii) finally, we show, under a smoothness assumption, that such additive representation is sufficient to imply logarithmic equilibrium payoffs.

For ease of notation, we continue to assume that $M = 3$ and use the above notation p^i, q^i, r^i for the local posteriors of the three types. We also denote by s_p, s_q, s_r the corresponding type probabilities. This does not impair the generality, and the same proof works for more than three types.

The following is the assumption on "separation of variables" that 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 (which is assumed to exist)

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

of the difference in scores of two types with local posteriors p^i and q^i, respectively, does not depend on other types' local posteriors r^i.

The assumption says that the (mixed) "sensitivity" of the difference in equilibrium payoffs to the corresponding types is not affected by other types.

We now state the following additive representation result (for a proof, see section 6).

PROPOSITION 3.1. *Consider a PLEPS* $\{F_i\}$ *satisfying Assumption* 3.1. *Suppose that, 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 a point* $(p^i,q^i,r^i) = (p^0,\dots,p^0) \in (0,1)^M$. *Then 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 symmetric in all the pairs* (p^i,s_p), (q^i,s_q), (r^i,s_r), $i=1,\dots,N$.

The main result of this section is the following.

THEOREM 3.1. *Consider a PLEPS consisting of functions* $F_i(p^i,q^i,r^i;s_p,s_q,s_r)$, $i=1,2,\dots,N$, *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, $i=1,\dots,N$. *Then, for some functions* λ *and* B *of probabilities of types* $S=(s_p,s_q,s_r)$,

$$G_i(p^i;s_p,s_q,s_r) = \lambda(S)\log p^i + B_i(S).$$

In particular, if the corresponding PLEPS is budget-balanced, the equilibrium payoff to the respondent with local 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_o^i *is the conditional probability of the type* o *in the state* i, $o=p,q,r$.

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 of Theorem 3.1. Since F_i is a PLEPS, it satisfies the separation inequality (2.1). By the stated symmetry of H_i, the function G_i also satisfies the same type of inequality, which can be written as

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

According to [28], this property implies that G_i is continuously differentiable in p^i, $i = 1, \dots, N$. Hence, by Lemma 3.1 below, which identifies the first-order condition for this minimization problem, there exists a Lagrange multiplier $\lambda(S)$ independent of p such that (suppressing the dependence on i)

$$\lambda(S)\frac{1}{p^i} = \partial_p G(p^i; s_p, s_q, s_r).$$

This implies the logarithmic form of G_i. Equation (3.5) is then straightforward to verify. Theorem 3.1 is proved.

The following "Lagrange optimization" lemma is proved in section 6. It gives a first-order condition for the IC minimization problem in (3.7) below.

LEMMA 3.1. *In the above notation, assume that functions* $F_i(p^i, q^i, r^i; s_p, s_q, s_r)$, $i = 1, 2, \dots, N$, *are continuously differentiable in* p^i *and* q^i, *and, for each fixed* p^i, q^i, r^i, *are symmetric w.r.t. any* s. *The above strict separation inequality* (2.1) *can be written as*

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

that is, the minimum over the probabilities q^i *is attained 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, *and* s_r,

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

3.3. Ranking by (local) posteriors. Our next aim is to show that PLEPS payoffs necessarily rank the players according to the relative ranking of the corresponding local posteriors. That is, when using a scoring system resulting in an equilibrium with PLEPS payoffs, the planner knows which players are better experts than others if she considers the level of the local posterior equivalent to the level of expertise.[18] We emphasize that for this result it is crucial to assume that the equilibrium scores depend only on the local posteriors of the realized state of nature.

The main result of this section is as follows.

THEOREM 3.2. *PLEPS payoffs* $\{F_i\}$ *are strictly increasing in the posterior probabilities of the true state of nature*; *that is*, *for any prior distribution matrix* Q,

$$\begin{aligned} &\text{if } j, k \in \{1, \dots, M\} \text{ and } z^i_k > z^i_j,\\ &\quad\text{then } F_i(z^i_k, z^i_{-k}; s_k, s_{-k}) > F_i(z^i_j, z^i_{-j}; s_j, s_{-j}). \end{aligned} \tag{3.9}$$

In other words, if the planner wants to determine the relative expertise of players that receive exchangeable signals, it is sufficient to design a scoring system which

[18] If they are not ranked by their local posteriors, then, in the pregame phase, they might want to avoid collecting information about the true state of nature, which is undesirable.

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.

Theorem 3.2 is a generalization of the results in the literature on the monotonicity implied by incentive compatibility of proper scoring rules (see, e.g., [19], [28], [29], and [30]). Authors of those papers considered only the nongame version of the problem with one respondent only. Moreover, the argument in these papers was based mostly on analytic methods, while our proof is completely algebraic.[19]

The intuition behind the result is that if type A's posterior probability of a state is higher than that of type B, but type A's score in that state is 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 nature, 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 the dependence on S). 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 a higher score in that state, that is, positive D_i^A. To argue this, we first note 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 when $1-p_A < 1-p_B$) is to have $D_1^A > 0$ and $D_2^A < 0$. Thus, indeed, the type with higher posterior probability of a state receives a higher score in that state. In other words, if the type with higher posterior probability of a state does not receive a higher score in that state, he would adopt the other type's strategy. In section 6, the above simple argument is formulated and proved in Lemma 6.1, which is then extended to any number of types and states.

4. Implementation.

We first show how to implement any PLEPS, up to an additional mild restriction, and then we elaborate on result from [24] to the effect that the Bayesian Truth Serum algorithm provides a feasible implementation of budget-balanced logarithmic equilibrium payoffs of (3.5) (under standard Bayesian and rationality assumptions); see also [25]. We also comment on the uniqueness of equilibrium under the BTS scoring rule.

4.1. PLEPS actual implementation.

By implementing an equilibrium payoff system F, we mean designing a questionnaire and a scoring rule such that the associated game allows an equilibrium with payoffs given by F.

As before, we assume that there are infinitely many respondents. We consider the case where the respondents are asked to choose the correct answer to a multiple-choice question (to declare their type) and assume that the possible states of nature take values in the set of probability distributions of the responses to the multiple-choice questions.

[19]Theorem 3.2 is formulated in the spirit of theorems that relate incentive compatibility to monotonicity in types if we equate types with posterior probabilities; see [22] for an early theorem of that type and [31] for a comprehensive treatment. However, our framework is different from the standard mechanism design framework, where we have random states of nature, so that incentive compatibility is a property of a weighted sum (expected value conditional on type) and not of the value itself. As a consequence, the methodology of the aforementioned papers cannot be applied.

The following result shows that any PLEPS $\{F_i\}$ such that F_i depends only on local conditional type probabilities $s_k^i = \Pr(T^\pi = k \mid \Omega = i)$ can be implemented by using F_i to compute the value of the scoring function f in state i.

PROPOSITION 4.1. *Consider a PLEPS with type k payoff in equilibrium state i given by $F_i(z_k^i, z_{-k}^i; s_k^i, s_{-k}^i)$; that is, in addition to z_k^j's, the payoff depends only on local conditional type probabilities $s_k^i = \Pr(T^\pi = k \mid \Omega = i)$ instead of on possibly all ex-ante type probabilities contained in vector s_k. Then this PLEPS can be implemented by an agnostic planner. More precisely, there exist questions that the planner can ask such that she can form estimates $\widehat{i}$, $\widehat{z}_k^{\widehat{i}}$, and $\widehat{s}_k^{\widehat{i}}$ of the true state of nature i and the true probabilities z_k^i and s_k^i from these questions, and, if the planner announces that a player who declares type k receives $F_{\widehat{i}}(\widehat{z}_k^{\widehat{i}}, \widehat{z}_{-k}^{\widehat{i}}; \widehat{s}_k^{\widehat{i}}, \widehat{s}_{-k}^{\widehat{i}})$, then truth-telling represents 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 (a) is, and to state their perceived probability (z's) for each of those distributions.

To guarantee that truth-telling represents an equilibrium, the planner announces that she computes the scores F_i as follows. Her estimate $\widehat{i}$ of the true state of nature i is given by the frequencies for each particular answer to the multiple-choice question to be chosen by the respondents. She also makes the estimates $\widehat{s}_k^{\widehat{i}}$ of the type probabilities equal to those frequencies. The estimates $\widehat{z}_k^{\widehat{i}}$ are chosen among all probabilities z_k^j, $j = 1, \dots, N$, which the player provides as the answer to (b). Having all the required estimates, the planner computes the corresponding values of $F_{\widehat{i}}$'s by using $\widehat{z}_k^{\widehat{i}}$'s and $\widehat{s}_k^{\widehat{i}}$ as the arguments of the function $F_{\widehat{i}}$.

Suppose now that all players, except the player π of type k, play the truth-telling strategy. If the player π also plays the truth-telling strategy, his payoff in the state i is $F_i(z_k^i, z_{-k}^i; s_k^i, s_{-k}^i)$, because i, z's, and s's are correctly estimated by the planner. If the player π of type k declares a type $j \neq k$, his payoff in the state i is $F_i(z_j^i, z_{-j}^i; s_j^i, s_{-j}^i)$, because, with an infinite number of players and all except player π being honest, i, z's, and s's are again correctly estimated by the planner. By the IC inequality (2.1), the player's π 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. Proposition 4.1 is proved.

Remark 4.1. The above implementation procedure is not robust — in practice, the number of different outcomes of responses to question (b) will be higher 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 [24] implements budget-balanced logarithmic equilibrium payoffs using inputs which 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 $\pi \in R$. The truthful opinion of respondent π is represented by a pair of M-tuples $(X^\pi; Y^\pi) = ((X_1^\pi, \dots, X_M^\pi); (Y_1^\pi, \dots, Y_M^\pi))$ of r.v.'s. Here, X_i^π's take value 0 or 1, and only one is equal to 1. This is interpreted as choosing an answer from a set of M possible answers. The r.v.'s Y_i^π's take values in $[0,1]$ and $\sum_{i=1}^M Y_i^\pi = 1$. The latter represent the declared opinion of the respondent π about what percentage of respondents choose i as the correct answer.

As in section 2, we assume that the infinite sequence $(X^\pi, \pi \in R)$ is exchangeable. Hence, by de Finetti's theorem, there is an M-dimensional (potentially random) vector

$$\overline{X} = \lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} X^\pi$$

with values in $[0,1]^M$ and such that the X^π's are conditionally independent given $\overline{X}$. We interpret $\overline{X}$ as the true state of nature, which was denoted previously by Ω.

We denote by $\overline{x}_j$ the sample mean of the declared values x_j^π of X_j^π over all respondents π, and by $\log \overline{y}_j$ the sample mean of all declared values $\log y_j^\pi$ of $\log Y_j^\pi$ (so that $\overline{y}_j$ is their geometric mean)

$$\log \overline{y}_j := \lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} \log y_j^\pi.$$

Definition 4.1. *The Bayesian Truth Serum* (*BTS*) *score function for the respondent π is given by*

$$\mathrm{BTS}^\pi = \sum_{j=1}^{M} x_j^\pi \log \frac{\overline{x}_j}{\overline{y}_j} + \sum_{j=1}^{M} \overline{x}_j \log \frac{y_j^\pi}{\overline{x}_j}.$$

Prelec [24] 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 declares their true opinion. Moreover, we can state a new "uniqueness" result, namely, that with the BTS mechanism any budget-balanced strict (Bayesian) Nash equilibrium is separating.

Remark 4.2. It is a natural convention to define $\log(\overline{x}_j/\overline{y}_j) = 0$ if $\overline{x}_j = \overline{y}_j = 0$, and $\overline{x}_j \log(y_j^\pi/\overline{x}_j) = 0$ if $\overline{x}_j = 0$. Note that if $x_j^\pi = 0$ for all but a finite number of π's, so that $\overline{x}_j = 0$, then it is optimal for every player π to correctly predict $y_j^\pi = 0$, so that $\overline{y}_j$ is naturally defined to be zero.[20] Under these conventions, the only possible budget-balanced SNEs are those that are separating, and where the players of the same type have the same strategies. Let us justify this.

(i) First, it is impossible to have an SNE in pure strategies where 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 switches 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 an infinite number of players, the value of

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

one player is not affected by that of another player. For the same reason, if player 2 switches 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 is not playing an equilibrium strategy to start with, which is a contradiction.

(ii) Second, two individuals of different types cannot have the same strategies in an SNE: if they do, by (i) all other players of their types would also choose the same strategy, which means that there would be a type k that nobody would "claim," that is, a k such that $x_k^\pi = 0$ for all π. Since we assume budget balance, there is a player with a nonpositive score. If this player deviates to type k, by the 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 remark 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 result of Prelec [24] to the effect that, in such a truth-telling equilibrium, the BTS score is equal to the budget-balanced logarithmic payoff (a detailed proof is provided in section 6).

THEOREM 4.1 (see [24, Theorem 2]). *Under the above assumptions, when the players play the truth-telling equilibrium, BTS scoring results in budget-balanced logarithmic equilibrium payoffs. More precisely, in the equilibrium we have*

$$\mathrm{BTS}^\pi = \log \Pr(\overline{X} = \overline{x} \mid X^\pi = x^\pi) - \lim_{n\to\infty} \frac{1}{n} \sum_{s=1}^{n} \log \Pr(\overline{X} = \overline{x} \mid X^\gamma = x^\gamma) \tag{4.1}$$

or (*denoting* $x^\pi = k$, $x^\gamma = j$, $\overline{x} = i$)

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

Thus, the BTS score corresponds to the PLEPS function F_i, which is logarithmic. In other words, BTS implements budget-balanced logarithmic equilibrium payoffs by asking the players just 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, this section confirms the appeal of BTS because of the following three properties: BTS always leads to a strictly separating equilibrium wherein the players of the same type get the same score; it results in the benchmark, logarithmic ex-post scores; and it is easily implemented. As far as we know, no other mechanism has all three 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 the realized state of nature (local posteriors) in the case where 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 based on ex-post logarithmic payoffs. We prove the following results for equilibrium payoffs that are determined only by the local 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 [24] 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 expression of their opinions. Thus, developing a more general analysis of robust mechanisms in our framework, where 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 wherein 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 to obtain the highest score, or a framework with known utilities and unknown correlation of types, where the planner wants to elicit information about the correlations without disturbing the stated utilities; for example, the case where the planner wants to ask players to predict what others will do, but she does not want their payoff 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 2.1. Our aim is to show that, 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.

First, recall that a pure strategy for the player π is a map $\sigma^\pi(t^\pi)$, which sends the player's type to his response choice a^π. The profile of all respondents' pure strategies is denoted by $\sigma(t)$ with entries $\sigma^\pi(t^\pi)$, and the profile excluding player π is denoted by $\sigma^{-\pi}(t^{-\pi})$. The score for the player π is given by $f(\sigma^\pi(t^\pi), \sigma^{-\pi}(t^{-\pi}))$. Let us denote by σ the equilibrium strategy profile of all respondents and define ρ to be the strategy profile that is identical to σ, except that a specific player π of type $k \neq j$ plays the strategy $\sigma^\pi(j)$ corresponding to type j. Let γ denote a player of type j. Then the payoff to the mimicry strategy, when π plays j, is

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

because $\sigma^{-s}(T^{-s})$ and $\rho^{-s}(T^{-s})$ differ only in π'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. More precisely, to justify the last equality above, we argue as follows. We need to prove that if exactly one of $-s$ respondents deviates from the equilibrium response, then the score f of respondent γ does not change. Since σ corresponds to a strictly separating equilibrium, the sequence $\{\sigma^{-s}(T^{-s})\}$ consists of M different K-tuples, each repeated infinitely many times. If a respondent of type k deviates to type j, that means one repetition (among infinitely many) of K-tuples corresponding to type k becomes an additional repetition (among infinitely many) of K-tuples corresponding to type j. We can then define a permutation of the sequence $\{\sigma^{-s}(T^{-s})\}$, which is equal to the deviation sequence $\{\rho^{-s}(T^{-s})\}$, and, by symmetry of f, we prove the last equality in the equation above. Proposition 2.1 is proved.

Proof of Proposition 2.2. Recall that it is required to prove there is no budget-balanced PLEPS for finitely many players.

For notational simplicity, we consider the case $M = 2$ with two types only, type 1 and type 2, with at least two players, 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 $(0, 2)$. The proof adjusts easily to the case $M > 2$, since we can consider only those matrices Q wherein the two types are isolated in a particular block.

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 infinitely 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 is

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

or, equivalently,

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

or

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

Similarly, when one type 2 respondent lies, the separating inequality reads as

$$qF_2(q, 1-p) + (1-q)F_3(1-q, 1-q) > qF_1(p, p) + (1-q)F_2(1-p, q).$$

This becomes

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

Suppose now that $p \neq q$. Without loss of generality, we consider the case $p > q$ and apply Lemma 6.1 to the above two inequalities. We have

$$F_1(p, p) - F_2(q, 1-p) > 0, \qquad F_2(1-p, q) - F_3(1-q, 1-q) < 0.$$

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 clearly contradicts the last equality. Proposition 2.2 is proved.

Proof Lemma 3.1. Let us prove (3.8). By the standard result on optimization under constraints (in our case the constraint is $\sum_i q^i = 1$), 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

$$p^i[\partial_p F_i(p^i, p^i, r^i; s_p, s_q, s_r) - \partial_q F_i(p^i, p^i, r^i; s_p, s_q, s_r)] = \lambda(\vec{p}, \vec{r}, s_p, s_q, s_r). \tag{6.1}$$

We fix arbitrary values 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, $\lambda(\vec{p}, \vec{r}, S)$ is a function $\lambda(p^i, r^i, S)$ of p^i, r^i, S only, and so

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

for all $0 < x < 1 - p^i$, $0 < y < 1 - r^i$. Since we can choose p^i, r^i to be arbitrarily small, it follows that, for fixed S, the left-hand side is constant across all values of x, y in $(0,1)$, and because i is arbitrary, $\lambda(S)$ does not depend on any of the values p^i, r^i, $i = 1, \ldots, N$. Lemma 3.1 is proved.

Proof Proposition 3.1. We suppress the dependence on i, s_p, s_q, and s_r. We want to show that

$$F(p,q,r) = G(p) + H(p,q,r),$$

where H is symmetric in all pairs (p, s_p), (q, s_q), (r^j, s_{r^j}).

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

$$\overline{p} = p - p^0, \quad \overline{q} = q - p^0, \quad \overline{r} = r - p^0.$$

From the smoothness and the symmetry property of F, we can write, for some functions a, b, c, d, e of the type probabilities, by Taylor's expansion,

$$\begin{aligned} F(p,q,r) = \sum_{n=0}^{\infty} a_n \overline{p}^n + \sum_{n=1}^{\infty}(b_n^q \overline{q}^n + b_n^\pi \overline{r}^n) + \sum_{m,n=1}^{\infty} \overline{p}^m (c_{m,n}^q \overline{q}^n + c_{m,n}^\pi \overline{r}^n) \\ + \sum_{m,n=1}^{\infty} d_{m,n} \overline{q}^m \overline{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} \overline{p}^l \overline{q}^m \overline{r}^n, \end{aligned}$$

where, by the symmetry property,

$$\begin{aligned} b_n^q(s_p, s_q, s_r) &= b_n^\pi(s_p, s_r, s_q), & c_{m,n}^q(s_p, s_q, s_r) &= c_{m,n}^\pi(s_p, s_r, s_q), \\ d_{m,n}(s_p, s_q, s_r) &= d_{n,m}(s_p, s_r, s_q), & e_{l,m,n}(s_p, s_q, s_r) &= e_{l,n,m}(s_p, s_r, s_q). \end{aligned}$$

Note that it is sufficient to show that

$$c_{m,n}^\pi = d_{m,n}, \qquad e_{l,m,n} = e_{m,l,n},$$

because then we can write

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

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

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

$$\begin{aligned} \partial_q F(p,p,r) - \partial_p F(p,p,r) = \sum_{n=1}^{\infty} n b_n^q \overline{p}^{n-1} + \sum_{m,n=1}^{\infty} c_{m,n}^q n \overline{p}^{m+n-1} \\ + \sum_{m,n=1}^{\infty} d_{m,n} m \overline{p}^{m-1} \overline{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m \overline{p}^{l+m-1} \overline{r}^n \\ - \sum_{n=0}^{\infty} n a_n \overline{p}^{n-1} - \sum_{m,n=1}^{\infty} m \overline{p}^{m-1} (c_{m,n}^q \overline{p}^n + c_{m,n}^\pi \overline{r}^n) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} l \overline{p}^{l+m-1} \overline{r}^n. \end{aligned}$$

We can then write

$$
\begin{aligned}
p\,\partial_q F(p,p,r) - p\,\partial_p F(p,p,r) &= \sum_{n=1}^{\infty} n b_n^q \overline{p}^n + \sum_{m,n=1}^{\infty} c_{m,n}^q n \overline{p}^{m+n} \\
&+ \sum_{m,n=1}^{\infty} d_{m,n} m \overline{p}^m \overline{r}^n + \sum_{l,m,n=1}^{\infty} e_{l,m,n} m \overline{p}^{l+m} \overline{r}^n - \sum_{n=0}^{\infty} n a_n \overline{p}^n \\
&- \sum_{m,n=1}^{\infty} m \overline{p}^m (c_{m,n}^q \overline{p}^n + c_{m,n}^\pi \overline{r}^n) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} l \overline{p}^{l+m} \overline{r}^n + \sum_{n=1}^{\infty} n b_n^q p^0 \overline{p}^{n-1} \\
&+ \sum_{m,n=1}^{\infty} n c_{m,n}^q p^0 \overline{p}^{m+n-1} + \sum_{m,n=1}^{\infty} d_{m,n} m p^0 \overline{p}^{m-1} \overline{r}^n \\
&+ \sum_{l,m,n=1}^{\infty} e_{l,m,n} m p^0 \overline{p}^{l+m-1} \overline{r}^n - \sum_{n=0}^{\infty} n a_n p^0 \overline{p}^{n-1} \\
&- \sum_{m,n=1}^{\infty} m p^0 \overline{p}^{m-1} (c_{m,n}^q \overline{p}^n + c_{m,n}^\pi \overline{r}^n) - \sum_{l,m,n=1}^{\infty} e_{l,m,n} l p^0 \overline{p}^{l+m-1} \overline{r}^n.
\end{aligned}
$$

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

– for $\overline{r}^n$ terms,

$$c_{1,n}^\pi = d_{1,n}; \tag{6.2}$$

– for $\overline{pr}^n$ terms,

$$0 = c_{1,n}^\pi - d_{1,n} + c_{2,n}^\pi - d_{2,n}; \tag{6.3}$$

– from $\overline{p}^2\overline{r}^n$ terms,

$$0 = 2(d_{2,n} - c_{2,n}^\pi) + 3p^0(d_{3,n} - c_{3,n}^\pi) + p^0(e_{1,2,n} - e_{2,1,n}); \tag{6.4}$$

– for $\overline{p}^3\overline{r}^n$ terms,

$$0 = 3(d_{3,n} - c_{3,n}^\pi) + (e_{1,2,n} - e_{2,1,n}) + 4p^0(d_{4,n} - c_{4,n}^\pi) + 2p^0(e_{1,3,n} - e_{3,1,n}), \tag{6.5}$$

and so on.

So, it is sufficient to show that $e_{l,m,n} = e_{m,l,n}$. 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,

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

This contradiction completes the proof of Proposition 3.1.

The following lemma is the key ingredient in proving Theorem 3.2. This result is a slight extension of Lemma A.1 in [29].

Lemma 6.1 (Schervish [29]). *Let $0 < a \leqslant 1$, $p, q \in (0,a)$, and $p > q$. If A, B are real numbers such that*

$$pA + (a-p)B > 0, \qquad q(-A) + (a-q)(-B) > 0,$$

then $A > 0$ and $B < 0$.

Proof. Notice that $A \neq 0$. If this is false, then the above two inequalities should become $(a-p)B > 0$ and $(a-q)(-B) > 0$, which is a contradiction. In order to prove the lemma, we just need to show that $A > 0$. Suppose on the contrary that $A < 0$. Then $B > 0$. Since $(a-p)B > -pA$, we have $B > -pA/(a-p) > 0$. We then get $0 < q(-A) + (a-q)(-B) < q(-A) + (a-q)pA/(a-p) = Aa(p-q)/(a-p) < 0$, which is impossible. This contradiction shows that $A > 0$. Lemma 6.1 is proved.

Proof of Theorem 3.2. Recall that we want to prove that PLEPS payoffs are strictly increasing in the posterior probabilities of the true state of nature.

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, inasmuch as we can always define Q by $q_k^i = z_k^i s_k$.

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

Case 1. Let $M = 2, N = 2$. The matrix Z can be written 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) implies

$$\begin{aligned} pF_1(p,q) + (1-p)F_2(1-p,1-q) &> pF_1(q,p) + (1-p)F_2(1-q,1-p), \\ qF_1(q,p) + (1-q)F_2(1-q,1-p) &> qF_1(p,q) + (1-q)F_2(1-p,1-q). \end{aligned}$$

Hence

$$\begin{aligned} p[F_1(p,q) - F_1(q,p)] + (1-p)[F_2(1-p,1-q) - F_2(1-q,1-p)] &> 0, \\ q[F_1(q,p) - F_1(p,q)] + (1-q)[F_2(1-q,1-p) - F_2(1-p,1-q)] &> 0. \end{aligned}$$

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 from section 6 to the above equations. We obtain $F_1(p,q) > F_1(q,p)$ and $F_2(1-p,1-q) > F_2(1-q,1-p)$, which proves the theorem in this case.

Case 2. Assume $M \geqslant 3$, $N = 2$. 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 entries of the matrix 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 $z_{-j,k}^i$ for the $(N-2)$-tuple, which consists of $\{z_1^i, \ldots, z_M^i\} \setminus \{z_j^i, z_k^i\}$, from (2.1) we get the two equations

$$\begin{aligned} &pF_1(p,q,z^1_{-(j,k)}) + (1-p)F_2(1-p,1-q,z^2_{-(j,k)}) \\ &\qquad > pF_1(q,p,z^1_{-(j,k)}) + (1-p)F_2(1-q,1-p,z^2_{-(j,k)}), \\ &qF_1(q,p,z^1_{-(j,k)}) + (1-q)F_2(1-q,1-p,z^2_{-(j,k)}) \\ &\qquad > qF_1(p,q,z^1_{-(j,k)}) + (1-q)F_2(1-p,1-q,z^2_{-(j,k)}). \end{aligned}$$

Hence, if we define A and B as

$$A = F_1(p, q, z^1_{-(j,k)}) - F_1(q, p, z^1_{-(j,k)}) = F_1(z^1_k, z^1_{-k}) - F_1(z^1_j, z^1_{-j}),$$
$$B = F_2(1-p, 1-q, z^2_{-(j,k)}) - F_2(1-q, 1-p, z^2_{-(j,k)})$$
$$= F_2(z^2_k, z^2_{-k}) - F_2(z^2_j, z^2_{-j}),$$

then from Lemma 6.1 we conclude that $A > 0$ and $B < 0$, which proves (3.9) for both $i = 1$ and $i = 2$.

Case 3. Let $M \geqslant 2$, $N \geqslant 3$. We put

$$Z = \begin{bmatrix} z^1_1 & z^2_1 & \dots & z^N_1 \\ z^1_2 & z^2_2 & \dots & z^N_2 \\ \vdots & \vdots & \vdots & \vdots \\ z^1_M & z^2_M & \dots & z^N_M \end{bmatrix}.$$

Let us prove (3.9) for the ith column of the matrix Z. We choose any rows (types) $j, k \in \{1, \dots, M\}$, where $j \neq k$. Since the only requirement for the matrix Z is that its rows are nondegenerate probability distributions, and since the values of F_i depend only on the quantities in the ith column, in order to complete the proof we need only show that

$$F_i(p, q, z^i_{-(k,j)}) > F_i(q, p, z^i_{-(k,j)}) \tag{6.6}$$

for every $p := z^i_k$ and $q := z^i_j$, with $1 > p > q > 0$, and for any choice of $z^i_{-(k,j)} \in (0,1)^{M-2}$ (if $M = 2$, this last requirement would be unnecessary).

In order to use (2.1) and apply Lemma 6.1, we modify the matrix Z without changing its ith column by finding a matrix $\widetilde{Q}$ with the same type of probabilities s_ℓ as the original matrix Q. As mentioned above, we can always do that by choosing $q_{\ell i} = z^i_\ell s_\ell$. Moreover, looking at the arguments in Cases 1 and 2, we see that we apply Lemma 6.1 to the payoff differences in two different states which correspond to two different columns of matrix Z. Without loss of generality, we assume $i \neq 1$. More precisely, instead of working with the original matrix Z, we work with the following Z-matrix where the ith column is not changed: given $0 < \varepsilon < 1$ and $a := 1 - \varepsilon$, we define the matrix $\widetilde{Z}$ by

$$\widetilde{Z}^t_l := \begin{cases} z^t_l & \text{if } l \in \{1, \dots, M\} \setminus \{j, k\}, \\ p & \text{if } l = k,\ t = i, \\ q & \text{if } l = j,\ t = i, \\ a - p & \text{if } l = k,\ t = 1, \\ a - q & \text{if } l = j,\ t = 1, \\ \dfrac{\varepsilon}{N-2} & \text{otherwise,} \end{cases}$$

where p, q are arbitrary values in $(0, a)$ with $p > q$. Hence, for every choice of ε and p and q, we see that $\widetilde{Z}$ is a Z-matrix, which differs from Z only in the jth and kth rows, and these rows read as

$$\begin{bmatrix} a - p & \dfrac{\varepsilon}{N-2} & \cdots & \dfrac{\varepsilon}{N-2} & p & \dfrac{\varepsilon}{N-2} & \cdots & \dfrac{\varepsilon}{N-2} \\ a - q & \dfrac{\varepsilon}{N-2} & \cdots & \dfrac{\varepsilon}{N-2} & q & \dfrac{\varepsilon}{N-2} & \cdots & \dfrac{\varepsilon}{N-2} \end{bmatrix}.$$

Applying the IC property (2.1) to the jth and kth rows, we get

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

Observe that $\widetilde{z}_j^t = \widetilde{z}_k^t = \varepsilon/(N-2)$ for $t \in \{1, \dots, N\} \setminus \{1, i\}$. Hence both sides of the above inequalities involve the terms $[\varepsilon/(N-2)]F_t\big(\varepsilon/(N-2), \varepsilon/(N-2), z_{-(j,k)}^t\big)$, which cancel each other. Now these inequalities can be written as

$$\begin{aligned}
&(a-p)F_1(a-p, a-q, z_{-(j,k)}^1) + pF_i(p, q, z_{-(j,k)}^i) \\
&\qquad > (a-p)F_1(a-q, a-p, z_{-(j,k)}^1) + pF_i(q, p, z_{-(j,k)}^i), \\
&(a-q)F_1(a-q, a-p, z_{-(j,k)}^1) + qF_i(q, p, z_{-(j,k)}^i) \\
&\qquad > (a-q)F_1(a-p, a-q, z_{-(j,k)}^1) + qF_i(p, q, z_{-(j,k)}^i).
\end{aligned}$$

If we consider

$$\begin{aligned}
A &= F_i(p, q, z_{-(j,k)}^i) - F_i(q, p, z_{-(j,k)}^i), \\
B &= F_1(a-p, a-q, z_{-(j,k)}^1) - F_1(a-q, a-p, z_{-(j,k)}^1),
\end{aligned}$$

we see from Lemma 6.1 that $A > 0$, which proves inequality (6.6) for $a > p > q > 0$. By letting $\varepsilon \to 0$, we obtain (6.6) for $1 > p > q > 0$.

Theorem 3.2 is proved.

Proof of Theorem 4.1. We need to derive a representation of the BTS score in terms of the logarithms of the local posteriors.

Let us denote

$$p_{ij} = \Pr(X_i^\pi = 1,\, X_j^\gamma = 1),$$

where we use the fact that, by exchangeability, the right-hand side does not depend on the choice of $\pi \neq \gamma$. So,

$$\Pr(X^\pi = x^\pi \mid X^\gamma = x^\gamma) = \frac{p_{ij}}{\sum_{k=1}^{M} p_{kj}}. \tag{6.7}$$

We need the following three properties.

Property I: $y_j^\pi = \sum_{i=1}^{M} \big(x_i^\pi p_{ij} / \sum_{k=1}^{M} p_{ki}\big)$.

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

Property III: $\log \Pr(X^\pi = x^\pi \mid \overline{X} = \overline{x}) = \sum_{k=1}^{M} x_k^\pi \log \overline{x}_k$.

Property I holds as a Bayesian game is assumed: the respondents compute conditional probabilities in a Bayesian fashion. Property II is a consequence of Property I and equation (6.7). For Property III, let ℓ be such that $x_\ell^\pi = 1$. De Finetti's theorem implies

$$\Pr(X^\pi = x^\pi \mid \overline{X} = \overline{x}) = \overline{x}_\ell = \sum_{k=1}^{M} x_k^\pi \overline{x}_k.$$

The sum on the right always has only one nonzero term. Therefore, taking the logarithms of both sides, we arrive at Property III.

Next, let x^γ be such that

$$\overline{x}_k = \lim_n \frac{1}{n} \sum_s x_k^\gamma.$$

Note that we can use exchangeability to reorder the respondents so that $\pi = 1$ and $\gamma = 2, \ldots, n+1$. For such choices of π and γ we have $\log \Pr(X^\pi = x^\pi | X^\gamma = x^\gamma) = \sum_{j=1}^M x_j^\pi \log(y_j^\gamma)$. We may always omit those γ such that $\Pr(X^\gamma = x^\gamma) = 0$. Thus, we actually have only finitely many choices for an M-tuple x^γ such that $0 < \Pr(X^\gamma = x^\gamma) < 1$, and there is a lower bound A and an upper bound B such that $0 < A \leqslant \Pr(X^\gamma = x^\gamma) \leqslant B < 1$. Now it follows that $A = \sqrt[n]{A^n} \leqslant \sqrt[n]{\prod_{s=1}^n \Pr(X^\gamma = x^\gamma)} \leqslant \sqrt[n]{B^n} = B$. The log function is continuous, and 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 $\lim_{n\to\infty} \prod_{s=1}^n \Pr(X^\gamma = x^\gamma)$ exists and is not zero, and we can take the logarithm outside or inside the limit.

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

$$\begin{aligned}
\sum_{k=1}^M \overline{x}_k \log y_k^\pi &= \lim_n \frac{1}{n} \sum_s \log \Pr(X^\gamma = x^\gamma \mid X^\pi = x^\pi),\\
\sum_{k=1}^M x_k^\pi \log \overline{y}_k &= \lim_n \frac{1}{n} \sum_s \log \Pr(X^\pi = x^\pi \mid X^\gamma = x^\gamma).
\end{aligned}$$

Hence, using the Bayes rule,

$$\begin{aligned}
\mathrm{BTS}^\pi &= \sum_{k=1}^M x_k^\pi \log \frac{\overline{x}_k}{\overline{y}_k} + \sum_{k=1}^M \overline{x}_k \log y_k^\pi\\
&= \log \Pr(X^\pi = x^\pi \mid \overline{X} = \overline{x}) + \lim_n \frac{1}{n} \sum_s \log \Pr(X^\gamma = x^\gamma \mid X^\pi = x^\pi)\\
&\qquad - \lim_n \frac{1}{n} \sum_s \log \Pr(X^\pi = x^\pi \mid X^\gamma = x^\gamma)\\
&= \log\left(\Pr(X^\pi = x^\pi \mid \overline{X} = \overline{x}) \lim_n \prod_{s=1}^n \frac{\Pr^{1/n}(X^\gamma = x^\gamma \mid X^\pi = x^\pi)}{\Pr^{1/n}(X^\pi = x^\pi \mid X^\gamma = x^\gamma)} \right)\\
&= \log\left(\Pr(X^\pi = x^\pi \mid \overline{X} = \overline{x}) \frac{\lim_n \prod_{s=1}^n \Pr^{1/n}(X^\gamma = x^\gamma)}{\Pr(X^\pi = x^\pi)} \right)\\
&= \log \Pr(\overline{X} = \overline{x} \mid X^\pi = x^\pi) - \log \Pr(\overline{X} = \overline{x}) + \lim_n \frac{1}{n} \sum_s \log \Pr(X^\gamma = x^\gamma).
\end{aligned}$$

Since the last two terms do not depend on π, and since $\sum_r \mathrm{BTS}^\pi = 0$, we arrive at (4.1). Next, for fixed n and $\overline{x}$, denote by n_j the number of respondents of type j, so that

$$\sum_j n_j = n.$$

Hence we can write (4.1) as

$$\begin{aligned}\mathrm{BTS}^{\pi} &= \log\Pr(\overline{X}=\overline{x} \mid X^{\pi}=x^{\pi}) - \lim_{n\to\infty}\frac{1}{n}\Bigg[\sum_{s=1}^{n_1}\log\Pr(\overline{X}=\overline{x} \mid X^{\pi}=x^{1}) \\ &\quad + \cdots + \sum_{s=n_{M-1}+1}^{n_M}\log\Pr(\overline{X}=\overline{x} \mid X^{\pi}=x^{M})\Bigg] \\ &= \log\Pr(\overline{X}=\overline{x} \mid X^{\pi}=x^{\pi}) - \lim_{n\to\infty}\Bigg[\frac{n_1}{n}\log\Pr(\overline{X}=\overline{x} \mid X^{\pi}=x^{1}) \\ &\quad + \cdots + \frac{n_M}{n}\log\Pr(\overline{X}=\overline{x} \mid X^{\pi}=x^{M})\Bigg].\end{aligned}$$

Now (4.2) follows, since $\lim_{n\to\infty}(n_j/n) = \Pr(T^{\pi}=j \mid \overline{X}=\overline{x})$. Theorem 4.1 is proved.

Acknowledgments. We are grateful to George Georgiadis, Matthew Elliott, Katrina Ligett, and Patrick Ray, the conference audiences at the Southwest Economic Theory Conference 2014, 13th Conference on Research on Economic Theory and Econometrics, and the 2017 Bayesian Crowd Conference, and to the anonymous referees for useful comments and suggestions.

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.

REFERENCES

[1] D. J. Aldous, *Exchangeability and related topics*, in École d'été de probabilités de Saint-Flour XIII — 1983, Lecture Notes in Math. 1117, Springer, Berlin, 1985, pp. 1–198, https://doi.org/10.1007/BFb0099421.

[2] A. Baillon, *Bayesian markets to elicit private information*, Proc. Natl. Acad. Sci. USA, 114 (2017), pp. 7958–7962, https://doi.org/10.1073/pnas.1703486114.

[3] D. Bergemann and S. Morris, *An introduction to robust mechanism design*, Foundations and Trends in Microeconomics, 8 (2012), pp. 169–230, https://doi.org/10.1561/0700000057.

[4] J. M. Bernardo, *Expected information as expected utility*, Ann. Statist., 7 (1979), pp. 686–690, https://doi.org/10.1214/aos/1176344689.

[5] T. Börgers, *An Introduction to the Theory of Mechanism Design*, Oxford Univ. Press, New York, 2015, https://doi.org/10.1093/acprof:oso/9780199734023.001.0001.

[6] Y. S. Chow and H. Teicher, *Probability Theory. Independence, Interchangeability, Martingales*, 3rd ed., Springer Texts Statist., Springer-Verlag, New York, 1997, https://doi.org/10.1007/978-1-4612-1950-7.

[7] J. Cvitanić, D. Prelec, S. Radas, and H. Šikić, *Game of duels: Information-theoretic axiomatization of scoring rules*, IEEE Trans. Inform. Theory, 65 (2019), pp. 530–537, https://doi.org/10.1109/TIT.2018.2867469.

[8] J. Cvitanić, D. Prelec, B. Riley, and B. Tereick, *Honesty via choice-matching*, AER Insights, 1 (2019), pp. 179–192, https://doi.org/10.1257/aeri.20180227.

[9] A. Dasgupta and A. Ghosh, *Crowdsourced judgement elicitation with endogenous proficiency*, in WWW'13: Proceedings of the 22nd ACM International World Wide Web Conference (Rio de Janeiro, 2013), ACM, New York, 2013, pp. 319–330, https://doi.org/10.1145/2488388.2488417.

[10] R. Frongillo and J. Witkowski, *A geometric perspective on minimal peer prediction*, ACM Trans. Econ. Comput., 5 (2017), 17, https://doi.org/10.1145/3070903.

[11] T. Gneiting and A. E. Raftery, *Strictly proper scoring rules, prediction, and estimation*, J. Amer. Statist. Assoc., 102 (2007), pp. 359–378, https://doi.org/10.1198/016214506000001437.

[12] P. J. Howie, Y. Wang, and J. Tsai, *Predicting new product adoption using Bayesian truth serum*, J. Med. Market., 11 (2011), pp. 6–16, https://doi.org/10.1057/jmm.2010.19.

[13] R. Jurca and B. Faltings, *Robust incentive-compatible feedback payments*, in TADA 2006, AMEC 2006: Agent-mediated Electronic Commerce. Automated Negotiation and Strategy Design for Electronic Markets (Hakodate, Japan, 2006), Lecture Notes in Comput. Sci. 4452, Springer, Berlin, 2007, pp. 204–218, https://doi.org/10.1007/978-3-540-72502-2_15.
[14] Y. Kong and G. Schoenebeck, *An information theoretic framework for designing information elicitation mechanisms that reward truth-telling*, ACM Trans. Econ. Comput., 7 (2019), 2, https://doi.org/10.1145/3296670.
[15] A. Kukla-Gryz, J. Tyrowicz, M. Krawczyk, and K. Siwiński, *We all do it, but are we willing to admit? Incentivizing digital pirates' confessions*, Appl. Econ. Lett., 22 (2015), pp. 184–188, https://doi.org/10.1080/13504851.2014.931918.
[16] Y. Liu, J. Wang, and Y. Chen, *Surrogate Scoring Rules*, preprint, https://arxiv.org/abs/1802.09158, 2020.
[17] T. A. Loughran, R. Paternoster, and K. J. Thomas, *Incentivizing responses to self-report questions in perceptual deterrence studies: An investigation of the validity of deterrence theory using Bayesian truth serum*, J. Quant. Criminol., 30 (2014), pp. 677–707, https://doi.org/10.1007/s10940-014-9219-4.
[18] E. Maskin and T. Sjöström, *Implementation theory*, in Handbook of Social Choice and Welfare, Vol. 1, Handbooks in Econom. 19, Elsevier/North-Holland, Amsterdam, 2002, pp. 237–288, https://doi.org/10.1016/S1574-0110(02)80009-1.
[19] J. McCarthy, *Measures of the value of information*, Proc. Natl. Acad. Sci. USA, 42 (1956), pp. 654–655, https://doi.org/10.1073/pnas.42.9.654.
[20] S. R. Miller, B. P. Bailey, and A. Kirlik, *Exploring the utility of Bayesian truth serum for assessing design knowledge*, Human–Computer Interaction, 29 (2014), pp. 487–515, https://doi.org/10.1080/07370024.2013.870393.
[21] N. Miller, P. Resnick, and R. Zeckhauser, *Eliciting informative feedback: The peer-prediction method*, Manag. Sci., 51 (2005), pp. 1359–1373, https://doi.org/10.1287/mnsc.1050.0379.
[22] R. B. Myerson, *Optimal auction design*, Math. Oper. Res., 6 (1981), pp. 58–73, https://doi.org/10.1287/moor.6.1.58.
[23] T. Offerman, J. Sonnemans, G. Van De Kuilen, and P. P. Wakker, *A truth serum for non-Bayesians: Correcting proper scoring rules for risk attitudes*, Rev. Econom. Stud., 76 (2009), pp. 1461–1489, https://doi.org/10.1111/j.1467-937X.2009.00557.x.
[24] D. Prelec, *A Bayesian truth serum for subjective data*, Science, 306 (2004), pp. 462–466, https://doi.org/10.1126/science.1102081.
[25] D. Prelec, H. S. Seung, and J. McCoy, *A solution to the single-question crowd wisdom problem*, Nature, 541 (2017), pp. 532–535, https://doi.org/10.1038/nature21054.
[26] G. Radanovic and B. Faltings, *A robust Bayesian truth serum for non-binary signals*, in AAAI'13: Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI Press, Palo Alto, CA, 2013, pp. 833–839.
[27] G. Radanovic and B. Faltings, *Incentives for truthful information elicitation of continuous signals*, in AAAI'14: Proceedings of the 28th AAAI Conference on Artificial Intelligence, AAAI Press, Palo Alto, CA, 2014, pp. 770–776.
[28] L. J. Savage, *Elicitation of personal probabilities and expectations*, J. Amer. Statist. Assoc., 66 (1971), pp. 783–801, https://doi.org/10.1080/01621459.1971.10482346.
[29] M. J. Schervish, *A general method for comparing probability assessors*, Ann. Statist., 17 (1989), pp. 1856–1879, https://doi.org/10.1214/aos/1176347398.
[30] K. H. Schlag and J. J. van der Weele, *Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality*, Theor. Econ. Lett., 3 (2013), pp. 38–42, https://doi.org/10.4236/tel.2013.31006.
[31] R. V. Vohra, *Paths, Cycles and Mechanism Design*, Tech. rep., Northwestern Univ., Evanston, IL, 2007.
[32] B. Waggoner and Y. Chen, *Information elicitation sans verification*, in Proceedings of the 3rd Workshop on Social Computing and User Generated Content (SC-13) (Philadelphia, PA, 2013), 2013, pp. 39:1–39:18; also available online from http://yiling.seas.harvard.edu/publications/.
[33] J. Witkowski, *Robust Peer Prediction Mechanisms*, Ph.D. Diss., Albert-Ludwigs-Universität, Freiburg, 2014; also available online from https://freidok.uni-freiburg.de/data/10054.
[34] J. Witkowski and D. C. Parkes, *A robust Bayesian truth serum for small populations*, in AAAI'12: Proceedings of the 26th AAAI Conference on Artificial Intelligence, AAAI Press, Palo Alto, CA, 2012, pp. 1492–1498.
[35] J. Witkowski and D. C. Parkes, *Learning the prior in minimal peer prediction*, in Proceedings of the 3rd Workshop on Social Computing and User Generated Content (SC-13) (Philadelphia,

PA, 2013), 2013, pp. 39:2–39:12; also available online from http://econcs.seas.harvard.edu/files/econcs/files/witkowski_ec13.pdf.

[36] P. Zhang and Y. Chen, *Elicitability and knowledge-free elicitation with peer prediction*, in AAMAS'14: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems (Paris, 2014), IFAAMAS, Richland, SC, 2014, pp. 245–252; also available online from http://yiling.seas.harvard.edu/publications/.