Elicitability and Knowledge-Free Elicitation with Peer Prediction

Peter Zhang
Harvard University
peterzhang@post.harvard.edu

Yiling Chen
Harvard University
yiling@seas.harvard.edu

ABSTRACT

The elicitation of private information from individuals is crucially important to many real-world tasks. But elicitation is most challenging when it is most useful: when objective (verifiable) truth is inaccessible or unavailable, and there is no "answer key" available to verify reports. Prior work has designed mechanisms that truthfully elicit private information without verification for some restricted set of possible information structures of the participants (i.e. the common prior joint distributions of participants' signals). In fact, no mechanism can elicit private information truthfully for all information structures without verification. In this paper, we identify the maximal set of information structures that are truthfully elicitable without verification, and provide a mechanism for such elicitation. This mechanism requires that the designer know the information structure of the participants, which is unavailable in many settings. We then propose a knowledge-free peer prediction mechanism that does not require knowledge of the information structure and can truthfully elicit private information for a set of information structures slightly smaller than the maximal set. This mechanism works for both small and large populations in settings with both binary and non-binary private signals, and is effective on a strict superset of information structures as compared to prior mechanisms that satisfy these properties.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Economics, Theory

Keywords

Information Elicitation, Peer Prediction, Proper Scoring Rules

1. INTRODUCTION

The elicitation of private information from individuals is essential to human knowledge gathering and decision making. For example, Yelp.com collects user feedbacks on restaurants and other local businesses; citizen scientists contribute their knowledge to help classify galaxies or find planets around stars at websites like Zooiverse; and the popularity of online labor markets such as Amazon Mechanical Turk allows many people to solicit answers to questions of interest from the crowd.

However, information elicitation is most challenging when it is most useful: when there is no ground truth available to evaluate answers. For example, the information being elicited may be inherently subjective, such as in user satisfaction surveys, or too costly to be practical to verify, such as the answers collected on a citizen science website. These settings, recently termed *information elicitation without verification* by Waggoner and Chen [16], have motivated both the rapid theoretical development of mechanisms – Peer Prediction [9] and Bayesian Truth Serum [12] being the most famous – that elicit private information strictly truthfully in some restricted settings despite the lack of verification, and the deployment of simple and intuitive mechanisms, such as the Output Agreement mechanism used by the ESP game [15], in practice.

While the theoretical development of mechanism design has achieved many successes in strictly truthfully eliciting private information without verification, mechanisms proposed in the literature all have various restrictions. The original Peer Prediction mechanism [9], its improved variants [4, 5, 6], the Bayesian Truth Serum [12], and the more recently developed Robust Bayesian Truth Serum [20] all require that the private information of individuals is identically and independently distributed conditioned on some true outcome of an event. The original Peer Prediction method [9] also requires that the mechanism know the prior distribution of the private signals of all individuals and the event outcome, which we call the *information structure* in this paper, to operate. The Bayesian Truth Serum [12] relaxes this requirement by asking individuals to predict the distribution of all participants' signals in addition to reporting their private signal and effectively obtains the prior distribution empirically. However, the Bayesian Truth Serum incentivizes truthful reporting only when the number of individuals is large, and this number depends on the prior distribution. Robust Bayesian Truth Serum [20] was developed specifically for small populations. While it elicits private signals truthfully from any population with three or more individuals, each individual's signal must be binary. Radanovic and Faltings [13] improved the Robust Bayesian Truth Serum with a mechanism that can handle non-binary signals. But,

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

in doing so, the mechanism places additional restrictions on the information structure, assuming that signals are correlated – that one agent receiving a signal maximizes the probability that another agent receives the same signal.

Clearly, the development of truthful mechanisms for information elicitation without verification has been accompanied by carefully carving out territories for these mechanisms, by placing restrictions on information structures, size of the population, signal spaces, and the knowledge of the mechanism. In an ideal world, we wish to have a mechanism that can strictly truthfully elicit private information without verification for all information structures, all signal spaces, and both small and large populations, and does not require the knowledge of information structure to operate. Not surprisingly, we prove that it is impossible for a mechanism to strictly truthfully elicit private information for all information structures without the knowledge of the information structure [16]. In light of this impossibility result, we ask the following two questions in this paper:

- What is the maximal set of information structures that a mechanism can strictly truthfully elicit private information from, even with knowledge of the information structure?
- Without knowledge of the information structure, can we design a mechanism that strictly truthfully elicits private information for a large set of information structures, both binary and non-binary signals, and both small and large populations?

We study these questions in a general model where the mechanism designer can query any knowledge of individuals that solely depends on their private signal and the information structure. For a class of queries, including the signal elicitation query considered in the above mentioned prior work, we show that no mechanism can truthfully elicit such information in any information structure that does not satisfy a *stochastic relevance* condition, regardless of whether or not the mechanism knows the information structure. For the signal elicitation query – which is the focus of most prior work – we demonstrate that this bound is strict, by providing a Generalized Peer Prediction mechanism that can elicit signals in all stochastically relevant information structures.

To answer our second question, we propose a Knowledge-Free Peer Prediction mechanism. This mechanism does not require knowledge of the information structure and strictly truthfully elicits private information for both small and large populations in settings with both binary and non-binary private signals, without placing significant restrictions on the information structure. It can elicit signals in all information structures that satisfy a *second-order stochastic relevance* condition; this requirement is only slightly more restrictive than stochastic relevance, the minimal requirement for truthful elicitation of signals.

The rest of the paper is organized as follows. We discuss other related work in Section 1.1. In Section 2, we introduce the setting and discuss our model of the information elicitation problem. In section 3, we provide an upper bound for the set of information structures that can be elicited from, and show that this upper bound is strict for signal elicitation. Finally, in section 4, we introduce the Knowledge-Free Peer Prediction Mechanism, and conclude in section 5.

1.1 Other Related Work

In addition to original Peer Prediction [9], Bayesian Truth Serum[12] and Robust Bayesian Truth Serum [20], Jurca and Faltings have a nice line of work proposing mechanisms that improve the original Peer Prediction in many aspects, including reducing the mechanism's total payment using techniques from computational mechanism design [4], and ensuring that truthful reporting is the unique equilibrium and the mechanism is collusion resistant [5, 6].

All mechanisms discussed so far require all agents share a common prior distribution. Witkowski and Parkes [19] presented two related mechanisms that truthfully elicit signals even when agents do not share a common prior for a binary signal space. This is achieved by using the same shadowing technique developed for the Robust Bayesian Truth Serum [20]. We assume the existence of a common prior in this paper.

In terms of elicitability, in addition to Waggoner and Chen [16]'s general impossibility result mentioned above, Radanovic and Faltings [13] also give several impossibility results concerning the design of mechanisms for various information structures. Their results are limited to the setting where agents have conditionally independent and identically distributed signals.

More recently Dasgupta and Ghosh [2] and Witkowski et. al. [18] consider settings where agents may incur a cost to improve their information and study how to incentivize effort in information elicitation without verification. We do not consider effort elicitation in this paper and assume that the private information of agents is given.

2. THE SETTING

We model the general problem of information elicitation without verification as follows. For some event of interest, there are a finite set of possible *outcomes* Ω, from which nature selects the true outcome $\omega \in \Omega$. There are $n \geq 2$ rational and risk-neutral agents that are only interested in maximizing their payout. Each agent i will observe a private *signal* θ_i drawn from a finite set of all possible signals Θ_i. $\Theta = \Theta_1 \times \cdots \times \Theta_n$ is the joint signal space of all agents and $\theta \in \Theta$ is a vector of the signals of all agents. Agents have common knowledge about the event and how it affects their signals in the form of a prior distribution $\Pr[\omega, \theta]$ over the outcome space and signal spaces. For example, if the event is whether a hotel is of high or low quality, i.e. $\Omega = \{\text{high}, \text{low}\}$, and each player may experience either good or bad service at the hotel, i.e. $\Theta_i = \{\text{good}, \text{bad}\}$, then the prior distribution can be viewed as modeling agents' common belief on the distribution of the hotel's quality, $\Pr[\omega]$, and the conditional distribution of all agents' experience given a realized hotel quality, $\Pr[\theta \mid \omega]$, prior to receiving their private signals. Upon receiving a private signal θ_i, agent i updates her belief in a Bayesian fashion.

We refer to this common prior $\Pr[\omega, \theta]$ as an *information structure* I because it describes how agents' private signals relate to each other and to the event outcome. We use $\mathcal{I}$ to denote the set of all possible information structures on Ω and Θ, i.e. $\mathcal{I} \equiv \Delta(\Omega \times \Theta)$.

This model allows individual agents to have different signal spaces and places no restriction on the information structure as long as it is a valid probability distribution over Ω and Θ. This is less restrictive the models used in most prior work. For example, the original Peer Prediction method [9], the Bayesian Truth Serum [12], and the Robust Bayesian Truth Serum [20] all assume that agents have the same signal space and their signals are identically and independently drawn conditional on the realized event outcome.

Similar to Waggoner and Chen [16], we formalize the information sought by the mechanism designer as a *query* (C, T), where

- $C = (C_1, ..., C_n)$, where each C_i is the set of possible answers for agent i.
- $T = (T_1, ..., T_n)$, where $T_i : \mathcal{I} \times \Theta_i \to 2^{\Delta(C_i)}$ maps an information structure and a signal to a set of distributions on C_i. In words, T_i specifies the response of agent i, given her signal and her common prior distribution, that the mechanism designer desires or considers truthful. We allow responses that randomize among answers and more than one truthful response for every unique signal and information structure pair.

Conceptually, the mechanism designer begins by deciding upon a question (where the possible answers are the members of C) and a desired response to that question given a particular signal and information structure (represented by T). These two together constitute a specification of the information that the mechanism designer wants; they specify the question being asked, the possible answers to that question, and the meaning of the answers (in relation to the private information of the agents) to the mechanism designer. Several examples of queries are provided below for additional clarity:

(a) Eliciting the posterior distribution on event outcomes.

One common information elicitation task is to elicit every agent's posterior belief about the event outcome, given their signal. This is represented by the query $C_i = \Delta(\Omega)$ and $T_i(\Pr[\omega, \theta], \theta_i) = \{\Pr[\omega \mid \theta_i]\}$.

(b) Eliciting every agent's signal.

Another common information elicitation task is to elicit every agent's private signal. This is represented by the query $C_i = \Theta_i$ and $T_i(\Pr[\omega, \theta], \theta_i) = \{\theta_i\}$.

(c) Eliciting the category of every agent's signal.

Suppose there was a mapping $G_i : \Theta_i \to S$ that classified every agent's signal into a class $s \in S$, and we wished to elicit the class label of every agent's signal. This can be represented by the query $C_i = S$ and $T_i(\Pr[\omega, \theta], \theta_i) = \{G_i(\theta_i)\}$.

(d) Eliciting one of the attributes of every agent's signal.

Suppose there was a set of attributes S. An agent's signal can be associated with one or more attributes via the mapping $H_i : \Theta_i \to 2^S$. We wish to elicit any attribute of every agent's signal. This can be represented by the query $C_i = S$ and $T_i(\Pr[\omega, \theta], \theta_i) = \Delta(H_i(\theta_i))$ where $\Delta(H_i(\theta_i))$ is the set of all distributions on $H_i(\theta_i)$.

In this paper, we limit our scope to the subset of queries that are *signal specific*, defined below.

Definition 1. *We call a query* (C, T) signal specific *for a set* $\mathcal{I}'$ *of information structures, if, for any information structure* $\Pr[\omega, \theta] \in \mathcal{I}'$ *and* $i \in \{1, ..., n\}$, T *satisifies the following property:*

$$T_i(\Pr[\omega, \theta], \theta'_i) \cap T_i(\Pr[\omega, \theta], \theta''_i) \neq \emptyset \textit{ if and only if } \theta'_i = \theta''_i$$

In some coarse sense, signal specific queries are just queries in which the mechanism designer views different behaviors as being truthful for different signals, thus making each signal distinct in terms of the corresponding response distributions desired by the mechanism designer. Formally, a signal specific query specifies mutually exclusive sets of distributions on answers (rather than exclusively specifying an answer itself) as being truthful for different signals. Accordingly, a signal specific query is not necessarily one in which the mechanism designer can actually distinguish between two different signals received by a particular agent, since the truthful distributions on answers for these signals may have intersecting supports, and thus an agent may report the same answer after receiving either of the two distinct signals by chance.

The signal elicitation query in example (b), which is also widely studied in the literature [9, 12, 20], is signal specific for all information structures $\mathcal{I}$. The posterior elicitation query in example (a) is only signal specific for a subset of $\mathcal{I}$, while the query in examples (c) and (d) are generally not signal specific, unless each class contains only one signal in (c) and each attribute applies to only one signal in (d).

To answer a query (C, T), the mechanism designer constructs a mechanism $M = (A, h^M)$ where:

- $A = (A_1, ..., A_n)$, where each A_i is the set of actions available to agent i in the mechanism. There must exist a mapping $\gamma : A_i \to C_i$, which allows the mechanism designer to read the answers to his query from the actions in the mechanism.
- $h^M = (h^M_1, ..., h^M_n)$ where each $h^M_i : \mathcal{I} \times A \to \mathbb{R}$ is the payout function for agent i under the mechanism. Notably, the payoff functions do not depend on agents' signals, which are not verifiable.

When the mechanism designer directly asks agents to provide an answer to his query, the action sets and the sets of possible answers coincide, $A = C$. But the mechanism designer could instead ask the agents other questions in addition to the question that he cares about in order to obtain good answers to the query. For example, in the scoring rule literature, it is known that the variance of a random variable alone cannot be truthfully elicited, but the mean and the variance can be truthfully elicited together [7]. In order to elicit the variance truthfully, it is necessary for the mechanism designer to ask for both the mean and the variance. We thus allow A to be different from C as long as the mechanism designer can map an action to an answer.

Together, a mechanism and information structure pair (M, I) specifies a Bayesian game. We call the set $\{(M, I) \mid I \in \mathcal{I}\}$ the games *induced* by M over the set of information structures $\mathcal{I}$.

We use the Bayesian Nash equilibrium solution concept to analyze such games. Let $s_i : \Theta_i \to \Delta(A_i)$ denote agent i's strategy, which specifies her distribution on actions given her signal, and θ_{-i} denote the vector of all signals excluding agent i's signal. Formally, a Bayesian Nash equilibrium is a strategy profile $(s_1, ..., s_n)$ where s_i is a *best response* to s_{-i} – the strategy profile of all other participants – for all i. By best response, we mean that:

$$\mathbb{E}[h_i^M(\mathcal{I}, (s_{-i}(\theta_{-i}), s_i(\theta_i))|\theta_i] \geq \mathbb{E}[h_i^M(\mathcal{I}, (s_{-i}(\theta_{-i}), s_i'(\theta_i))|\theta_i]$$

for any $s_i' \neq s_i$ and all $\theta_i \in \Theta_i$. In words, s_i is a best response if it maximizes agent i's payout regardless of what her realized signal is. Furthermore, in a *strict* Bayesian Nash equilibrium, s_i is the unique best response to s_{-i} for all i.

We say a player's strategy s_i in a game (M, I) is *truthful under query* (C, T) if $\gamma(s_i(\theta_i)) \in T_i(\Pr[\omega, \theta], \theta_i)$ for all $\theta_i \in \Theta_i$. Furthermore, an equilibrium $(s_1, ..., s_n)$ is *truthful under query* (C, T) if all s_i in the equilibrium are truthful.

We emphasize that truthfulness is defined with respect to a query, the information sought by the mechanism designer.

Following Waggoner and Chen [16], we relax the notion of a strict and truthful equilibrium to only require strictness with respect to non-truthful strategies. In particular, we call an equilibrium $(s_1, ..., s_n)$ *quasi-strictly truthful under* (C, T) if, in addition to being a truthful equilibrium, it satisfies the property that for any $s_i' \neq s_i$ that is also a best response to s_{-i}, $\gamma(s_i'(\theta_i)) \in T_i(\Pr[\omega, \theta], \theta_i)$ for all $\theta_i \in \Theta_i$.[1] In other words, when the query specifies more than one truthful response, at a quasi-strictly truthful equilibrium, a player can deviate to reporting another response in the range of the query, but cannot deviate to reporting any response outside of the range of the query, without hurting her expected payoff. Said more simply, at a quasi-strictly truthful equilibrium, a player can only deviate to other truthful strategies while still best responding. This concept is weaker than a *strictly truthful* equilibria, where a player cannot deviate to any other strategy without hurting her expected payoff. All strictly truthful equilibria under (C, T) are also quasi-strictly truthful under (C, T). When a query always specifies a unique response, quasi-strictly truthful and strictly truthful equilibria under (C, T) are equivalent.

We now formalize the goals of a mechanism designer with a signal-specific query (C, T):

1. The mechanism induces a game with a quasi-strictly truthful equilibrium under (C, T) over a large set of information structures; the mechanism works in as many information structures as possible.

2. The payoff functions of the mechanism do not depend on the information structure; the mechanism designer does not need to know the information structure to run the mechanism.

Together, these two properties guarantee that the mechanism is effective (truthful) in most settings, and can be employed even in situations where the mechanism designer is unfamiliar with the domain's information structure, such as when the mechanism designer is asking a previously unasked question.

3. THE MAXIMAL QUERYABLE SET OF INFORMATION STRUCTURES

3.1 An Upper Bound

In this section, we provide an upper bound for the set of all information structures in which a mechanism can answer a signal-specific query.

[1] Waggoner and Chen [16] called such equilibria strongly truthful. We use the term "quasi-strictly" to emphasize that this notion is a relaxation of a strictly equilibrium.

We first extend the concept of *stochastic relevance* used in Miller et al. [9] to our setting.

Definition 2. *An information structure $I = \Pr[\omega, \theta]$ satisfies* stochastic relevance *if and only if for any $i \in \{1, ..., n\}$, and any $\theta_i', \theta_i'' \in \Theta_i$ where $\theta_i' \neq \theta_i''$,*

$$\Pr[\theta_{-i} \mid \theta_i'] \neq \Pr[\theta_{-i} \mid \theta_i''].$$

In words, stochastic relevance means that for every agent i, the joint distribution of all the other agents' signals conditional on agent i's signal is different for different realizations of agent i's signal. This definition of stochastic relevance is a weaker requirement than the one used in in Miller et al. [9], which assumed that for different realizations of agent i's signal, the distribution of *any* other agent's signal would be different. Obviously, the two definitions are equivalent when there are only two agents.

For signal-specific queries, stochastic relevance is a necessary criterion for the information structures where an information elicitation mechanism without verification has a quasi-strictly truthful equilibrium.

Theorem 1. *Consider any information structure I that does not satisfy stochastic relevance and query (C, T) that is signal specific for information structure I. There is no mechanism M that induces a game (M, I) with a quasi-strictly truthful equilibrium under (C, T).*

Proof. Consider any I that does not satisfy stochastic relevance, and any mechanism M with payout function h^M. Since I is not stochastically relevant, there must exist some $i \in \{1, ..., n\}$ and some $\theta_i', \theta_i'' \in \Theta_i$, $\theta_i' \neq \theta_i''$, where

$$\Pr[\theta_{-i} \mid \theta_i'] = \Pr[\theta_{-i} \mid \theta_i''].$$

Let $(s_1, ..., s_n)$ be any truthful equilibrium of M under (C, T). Since the equilibrium is truthful and T is signal specific, $\gamma(s_i(\theta_i')) \neq \gamma(s_i(\theta_i''))$. Construct the strategy s_i', where $s_i'(\theta_i') = s_i(\theta_i'')$ and $s_i'(\theta_i) = s_i(\theta_i)$ for all $\theta_i \neq \theta_i' \in \Theta_i$; since T is signal specific and s_i is truthful, $\gamma(s_i'(\theta_i')) = \gamma(s_i(\theta_i'')) \notin T(\Pr[\omega, \theta], \theta_i')$, and thus s_i' is not truthful under T. We claim that s_i' is also a best response to s_{-i}; this follows directly from our assumption that $\Pr[\theta_{-i} \mid \theta_i'] = \Pr[\theta_{-i} \mid \theta_i'']$. To see this, note that for any distribution on actions $p \in \Delta(A_i)$,

$$\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), p)) \mid \theta_i'] = \mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), p)) \mid \theta_i'']$$

because $\mathcal{I}$ and p are both constants, and the distribution of θ_{-i}, and thus the distribution of $s_{-i}(\theta_{-i})$, is the same conditional on θ_i' and conditional on θ_i''. Accordingly, we have that

$$\begin{aligned}
&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i'))) \mid \theta_i'] \\
=&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i'))) \mid \theta_i''] \\
\leq&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i''))) \mid \theta_i''] \\
=&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i''))) \mid \theta_i'] \\
\leq&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i'))) \mid \theta_i']
\end{aligned}$$

where the inequalities hold because s_i is a best response to s_{-i}. This implies that

$$\begin{aligned}
&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i'))) \mid \theta_i'] \\
=&\mathbb{E}[h_i^M(I, (s_{-i}(\theta_{-i}), s_i(\theta_i''))) \mid \theta_i']
\end{aligned}$$

and thus $s'_i(\theta'_i) = s_i(\theta''_i)$ also maximizes expected utility against s_{-i} when agent i receives signal θ'_i, and s'_i is a best response to s_{-i} by construction. Accordingly, $(s_1, ..., s_n)$ is not a quasi-strictly truthful equilibrium. □

The previous theorem demonstrates that no mechanism can be used to answer a signal-specific query when the information structure is not stochastically relevant. While this result may seem intuitive at first glance – the mechanism designer will obviously only be able to distinguish between different signals if those signals have a distinct impact on an agent's belief about the other agents – it is actually somewhat surprising due to the broad nature of signal-specific queries. In particular, as mentioned earlier, the mechanism designer may not be able to distinguish between which of two distinct signals an agent has received with signal-specific query. Nevertheless, the signals need to be distinguishable (by satisfying stochastic relevance) for a signal-specific query to be answered.

In this sense, the maximal set of queryable information structures (for signal-specific queries) is the set of all information structures that satisfy stochastic relevance. Thus, for any signal-specific query, the goal of the mechanism designer (with respect to the first desirable property) can be restricted to designing a mechanism that induces games with quasi-strictly truthful equilibria over all stochastically relevant information structures.

3.2 A Tight Bound for Signal Elicitation

It is natural to ask whether the stochastic relevance restriction on information structures discussed above is tight. In other words, is there a mechanism that induces games with quasi-strictly truthful equilibrium over all stochastically relevant information structures?

In this subsection, we demonstrate that this bound is tight for the signal elicitation query, mentioned in example (a) and studied in the literature [9, 12, 20]. To recap, a signal elicitation query (C, T) has $C_i = \Theta_i$ and $T_i(\Pr[\omega, \theta_i], \theta_i) = \{\theta_i\}$. With respect to this query, we can in fact construct a mechanism that is quasi-strictly truthful over all stochastically relevant information structures.

3.2.1 Strictly Proper Scoring Rules

The mechanism provided below, and the Knowledge-Free Peer Prediction mechanism developed later in this paper, depend on the concept of a *strictly proper scoring rule*. Accordingly, we discuss scoring rules briefly in this section. For a more complete treatment, see Gneiting and Raftery [3].

Strictly proper scoring rules were initially introduced by Brier [1] as a method to calibrate weather forecasts. Then, they were widely studied [17, 14, 8, 10, 11] and developed into a general approach for incentivizing a risk-neutral agent to honestly reveal her subjective probabilistic assessment of an uncertain event. Consider a random variable o that takes value a in outcome space $\mathcal{O}$. Let $\Delta(\mathcal{O})$ be the set of probability distributions over the outcome space $\mathcal{O}$. Scoring rules and strictly proper scoring rules are defined as follows.

DEFINITION 3. *A* regular scoring rule *is a function*

$$R : \Delta(\mathcal{O}) \times \mathcal{O} \to \mathbb{R} \cup \{-\infty\}$$

where $R(p, o)$ can equal $-\infty$ only when $p_o = 0$, where p_o is the probability of outcome o in distribution p.

Informally, a scoring rule is a function that grades a forecast of an event p against the actual outcome of the event o. We are particularly interested in a specific class of scoring rules.

DEFINITION 4. *A regular scoring rule R is* strictly proper *if for any $p, p' \in \Delta(\mathcal{O})$ such that $p \neq p'$,*

$$\mathbb{E}_p[R(p, o)] > \mathbb{E}_p[R(p', o)].$$

Under a strictly proper scoring rule, a rational, risk-neutral agent with belief $p \in \Delta(\mathcal{O})$ who wishes to maximize her expected score would report her true belief p. We will denote a generic strictly proper scoring rule by R_p from this point forward.

3.2.2 Generalized Peer Prediction

We construct a *Generalized Peer Prediction* mechanism that has a strict truthful equilibrium in all of its induced games (M, I) where I satisfies stochastic relevance. In this way, Generalized Peer Prediction is an optimal mechanism for signal elicitation, in the sense that the set of information structures it can truthfully elicit signals from is maximal. As its name suggests, the Generalized Peer Prediction mechanism is a generalization of the original Peer Prediction mechanism presented by Miller et al. [9] to our more general setting.

The Generalized Peer Prediction mechanism works as follows:

1. Every agent i reports their signal x_i to the mechanism.
2. The mechanism calculates $p_{x_i} \equiv Pr[\theta_{-i} \mid x_i]$, or the posterior distribution on all other signals conditional on x_i, and pays agent i a reward of $R_p(p_{x_i}, x_{-i})$.

where R_p is any strictly proper scoring rule.

In the case of two agents, this mechanism reduces to the original Peer Prediction mechanism. In particular, in this case, every agent's report signal is scored based on how well it predicts a reference agent's report (since x_{-i} is only one report), just as in the original Peer Prediction mechanism. Accordingly, Miller et al. [9] had already successfully elicited signals from the maximal set of information structures in their more restricted setting.

THEOREM 2. *Generalized Peer Prediction has a strict truthful Bayesian Nash equilibrium for all information structures that satisfy stochastic relevance.*

PROOF. This proof is very similar to the proof that the original Peer Prediction mechanism is Bayesian Nash incentive compatible, as seen in [9]. In particular, we will demonstrate that truthful reporting of signal is a strict Bayesian Nash equilibrium.

Fix i. Assume all agents $j \neq i$ are reporting their signal truthfully; then $x_{-i} = \theta_{-i}$. To best respond, agent i chooses x_i to maximize her reward

$$\mathbb{E}[R_p(p_{x_i}, x_{-i})] = \mathbb{E}[R_p(p_{x_i}, \theta_{-i})].$$

Since agent i's posterior distribution on θ_{-i} is p_{θ_i} and R_p is proper, playing $x_i = \theta_i$ is a best response. The strictness of this equilibrium follows immediately from stochastic relevance; since $p_{x_i} \neq p_{\theta_i}$ when $x_i \neq \theta_i$, playing any $x_i \neq \theta_i$ results in a lower expected reward because R_p is strictly proper. □

4. KNOWLEDGE-FREE PEER PREDICTION

The Generalized Peer Prediction mechanism achieves truthful signal elicitation for the maximal set of information structures. However, it requires that the mechanism know the information structure. In this section, we still focus on the signal elicitation query (C, T), where $C_i = \Theta_i$ and $T_i(\Pr[\omega, \theta_i], \theta_i) = \{\theta_i\}$, and design a sequential elicitation mechanism that does not need to know the information structure and has a strictly truthful equilibrium for a set of information structures that is slightly smaller than the maximal queryable set. The mechanism requires $n \geq 3$ agents.

Because a sequential mechanism M induces an extensive-form Bayesian game (M, I) for an information structure I, our solution concept in this section is Perfect Bayesian equilibrium, which is a subgame-perfect refinement of Bayesian Nash equilibrium. Informally, a strategy-belief pair is a Perfect Bayesian equilibrium if the agents' strategies are optimal given their beliefs at any time of the game and the agents' beliefs can be derived from other agents' strategies using Bayes' rule whenever possible. In a strict Perfect Bayesian equilibrium, the strategy of each agent is a unique best response to the strategies of other agents, given her beliefs. As before, an agent's strategy is truthful under query (C, T) if it induces a response in $T_i(\Pr[\omega, \theta], \theta_i)$ for all $\theta_i \in \Theta_i$. Similarly, at a truthful Perfect Bayesian equilibrium all agents play a truthful strategy.

We first note that the maximal queryable set of information structures for sequential mechanisms remains to be the set of stochastically relevant information structures. This is because any extensive-form Bayesian game can be converted to a simultaneous-move Bayesian game, by setting the action space of each agent in the simultaneous-move game to be the cross product of the action spaces of the agent at all stages of the extensive-form game. Any Perfect Bayesian equilibrium of the original extensive-form game is a Bayesian Nash equilibrium of the resulting simultaneous-move game. Thus, by considering sequential mechanisms, we cannot truthfully elicit signals for a larger set of information structures than before, but instead focus on removing the mechanism's dependency on the information structure.

Now we introduce the *Knowledge-Free Peer Prediction* mechanism, defined as follows:

Randomly select a permutation of the $n \geq 3$ agents, and relabel them according to this permutation. For every agent i, select two reference agents $h = (i - 1) \mod n$, $j = (i + 1) \mod n$. Now, all n players play the following sequential game:

1. Round 1: Every agent simultaneously reports a signal $x_i \in \Theta_i$, which can be different from his true signal θ_i, to the mechanism; we call this the *information report*.

2. Round 2: Every agent i receives the information report of player h, x_h, from the mechanism, and then reports a joint distribution p^i over the signals of all agents except agents i and h; we call this the *prediction report*.

At the end of the game, agent i receives payoff

$$\underbrace{R_p(p^j, (x_{-(i,j)}))}_{\text{Information Score}} + \underbrace{R_p(p^i, (x_{-(h,i)}))}_{\text{Prediction Score}}$$

where R_p is any strictly proper scoring rule. As in Prelec [12] and Witkowski and Parkes [20], we call the first term the *information score* because it is dependent on agent i's reported signal (and how that information informs agent j's prediction report), and the second term the *prediction score* because it is dependent on agent i's prediction for the distribution of signals.

Consider the following property:

Definition 5. *An information structure $I = \Pr[\omega, \theta]$ satisfies* second order stochastic relevance *if for any $i \in \{1, ..., n\}$, and any $\theta_i', \theta_i'' \in \Theta_i$ where $\theta_i' \neq \theta_i''$, there exists some $j \neq i$ and $\theta_j \in \Theta_j$ where $\Pr[\theta_j \mid \theta_i'] > 0$ and*

- *if $\Pr[\theta_i'' \mid \theta_j] > 0$, then:*
$$\Pr[\theta_{-(i,j)} \mid \theta_i', \theta_j] \neq \Pr[\theta_{-(i,j)} \mid \theta_i'', \theta_j]$$
- *if $\Pr[\theta_i'' \mid \theta_j] = 0$, then:*
$$\Pr[\theta_{-(i,j)} \mid \theta_i', \theta_j] \neq \Pr[\theta_{-(i,j)} \mid \theta_j]$$

where $\theta_{-(i,j)}$ denotes the vector of all signals excluding the signals of agents i and j.

Second-order stochastic relevance is a stronger condition than stochastic relevance, in the sense that the set of information structures that satisfy the former is a strict subset of the set of information structures that satisfy the latter. At the same time, it is only marginally more restrictive.

By construction, Knowledge-Free Peer Prediction does not require knowledge of the information structure. Furthermore, as shown below, it has a strict truthful Perfect Bayesian equilibrium in all of it's induced games (M, I) where I satisfies second order stochastic relevance.

Theorem 3. *Knowledge-Free Peer Prediction has a strict truthful Perfect Bayesian equilibrium for the signal elicitation query for all information structures that satisfy second order stochastic relevance.*

Proof. We claim that the following assessments constitute a strict Perfect Bayesian equilibrium:

1. Strategies: Every agent truthfully reports their signal in round 1, and predicts $p^i = p_{x_h, \theta_i} \equiv \Pr[\theta_{-(h,i)} \mid x_h, \theta_i]$, or the posterior joint distribution on all other signals conditional on signals x_h and θ_i if $\Pr[\theta_h = x_h \mid \theta_i] \neq 0$ and predicts $p^i = p_{\emptyset, \theta_i} \equiv \Pr[\theta_{-(h,i)} \mid \theta_i]$ otherwise.

2. Beliefs: Every agent i believes that agent h received signal x_h, unless $\Pr[\theta_h = x_h \mid \theta_i] = 0$. In this case, agent i believes that the reporting agent's signal is distributed according to $\Pr[\theta_h \mid \theta_i]$.

First, we show that this assessment is consistent. For agent i, the only salient belief in the game is her belief in round 2 about agent h's signal. Every information set where the report x_h she receives has the property that $\Pr[\theta_h = x_h \mid \theta_i] > 0$ is on the equilibrium path, since there is some θ with non-zero probability where she observes signal θ_i and agent h observes signal x_h, and thus there is some non-zero probability that this information set is arrived at under the strategy profile specified in the assessment. Thus, in these infomation sets, her belief that agent h has observed signal x_h is consistent with Bayes rule as applied to the strategy profile in the assessment, since all players report their signal truthfully in round 1. All other information sets, where the report

x_h she receives has the property that $\Pr[\theta_h = x_h \mid \theta_i] = 0$, are not on the equilibrium path, because it is impossible to arrive at these information sets under the strategy profile specified in the assessment. In these information sets, the Perfect Bayesian equilibrium solution concept requires that agent i's beliefs must be derivable using Bayes rule from some strategy profile. Consider the strategy profile where all agents except agent h report their true signal in round 1, and agent h reports a signal from Θ_h randomly, each with equal probability. In this case, agent i's belief that agent h's signal is distributed according to $\Pr[\theta_h \mid \theta_i]$ is consistent; in particular, x_h gives no information about agent h's signal, so her belief about agent h's signal is just her posterior distribution after receiving her own signal.

Next, we show that this assessment is sequentially rational. Since agent i's total payout is equal to the sum of her prediction score and information score, and her actions in round 1 only impact her information score and her actions in round 2 only impact her prediction score, it suffices to show that she uniquely maximizes her expected prediction score in round 2 and expected information score in round 1. In round 2, if agent i receives a report x_h such that $\Pr[\theta_h = x_h \mid \theta_i] \neq 0$, agent i believes that agent h received signal x_h with probability 1, and thus uniquely maximizes her expected prediction score by predicting $p^i = p_{x_h,\theta_i}$ because R_p is strictly proper. If agent i receives a report x_h such that such that $\Pr[\theta_h = x_h \mid \theta_i] = 0$, agent i's joint distribution on $x_{-(i,h)}$ is just $p_{\emptyset,\theta_i}$, and thus uniquely maximizes her total score by predicting $p^i = p_{\emptyset,\theta_i}$ again because R_p is strictly proper. In round 1, agent i uniquely maximizes her information score by reporting $x_i = \theta_i$, because the information structure satisfies second-order stochastic relevance. Since the agent j who is receiving her signal is not known to agent i, for any $x_i \neq \theta_i$, there is a non-zero probability that agent j will predict $p_{x_i,\theta_j} \neq p_{\theta_i,\theta_j}$ or $p_{\emptyset,\theta_j} \neq p_{\theta_i,\theta_j}$ by second-order stochastic relevance; since R_p is strictly proper, this results in a non-maximal expected information score.

Since the stated assessment is both sequentially rational and consistent, it is a Perfect Bayesian equilibrium. Furthermore, since every agent's actions in the assessment uniquely maximizes his expected payout, this equilibrium is strict. □

Knowledge-Free Peer Prediction is a sequential mechanism, in contrast to previous work which has focused on simultaneous-move mechanisms. However, the sequential nature is not essential to the functioning of the mechanism; we can construct an analogous simultaneous-move mechanism by asking every agent i to give their prediction report conditional on every possible signal that agent h could have received. Then, we no longer have to give agent i the actual information report of agent h, and can ask for both the information and prediction reports at the same time, making the mechanism simultaneous. Intuitively, by asking agent i for her response given any information report from agent h, the mechanism can now play the second stage of the game for agent i, thus removing the need for the sequential nature of the game. It is straightforward to show that in this "virtualized" Knowledge-Free Peer Prediction, there is a quasi-strictly truthful Bayesian Nash equilibrium for all information structures that satisfy second-order stochastic relevance. The drawback of this simultaneous-move mechanism is that it is less intuitive, and requires a significantly more complex prediction report from the participating agents.

5. DISCUSSION

This paper is a theoretical investigation toward understanding the limit of truthfully eliciting private information without verification with or without the knowledge of the information structure. In this section, we discuss the implications of our results and their relation to prior mechanisms and suggest some potential directions for future investigation.

5.1 Elicitability

Building on the impossibility result developed by Waggoner and Chen[11], this paper provides an upper bound on the set of information structures in which a signal-specific query can be answered. Furthermore, we go on to prove that this bound is strict for the signal elicitation query in particular, by extending the Peer Prediction mechanism to the more general setting of this paper. Thus, while it may be the case that signal elicitation is impossible in all information structures, it is important to note that it is possible in most information structures, as stochastic relevance is a weak requirement that is likely satisfied in most real-world situations. Furthermore, signal-specificity may be a relatively strong requirement for queries. For more general queries that are not signal specific, it is likely that at least some non-stochastically relevant information structures are elicitable.

5.2 Comparing Knowledge-Free Peer Prediction to Previous Mechanisms

The main innovation that drives Knowledge-Free Peer Prediction is that it "outsources" the process of calculating the posterior conditional on a signal - as is necessary in Generalized Peer Prediction - to the agents themselves; in this way, the mechanism no longer needs to know the prior, since it is not updating the prior itself. It does so by using the same strategy employed by Bayesian Truth Serum [12] and the Robust Bayesian Truth Serum [20]: requesting a prediction report in addition to an information report. In Knowledge-Free Peer Prediction, this prediction report consists of the predicted distribution of the signals of all other agents, conditional on the signal reported by one other agent, which is exactly the posterior that is necessary to calculate the information score of the other agent.

By combining features from these two classes of mechanisms – Peer Prediction and Bayesian Truth Serum – Knowledge-Free Peer Prediction improves on both. Whereas Bayesian Truth Serum [12] can handle any finite number of signals, it is only truthful for a large number of agents (the exact number being unknown to the mechanism designer). The Robust Bayesian Truth Serum [20] is truthful for any number $n \geq 3$ of agents but can only handle binary signals. The Knowledge-Free Peer Prediction mechanism can handle any finite number of signals and any number $n \geq 3$ of agents. Furthermore, while the mechanism developed by Radanovic and Faltings [13] also achieves this, the set of information structures it is applicable for is a strict subset of the set of information structures that satisfy second-order stochastic relevance, even when restricted to information structures where signals are conditionally independent and identically distributed. Finally, unlike Generalized

Peer Prediction, Knowledge-Free Peer Prediction does not need to know the information structure. The weakness of Knowledge-Free Peer Prediction is that it only works for information structures that satisfy second-order stochastic relevance (and thus requires at least 3 agents), whereas Generalized Peer Prediction works for information structures that satisfy stochastic relevance and any number $n \geq 2$ of agents.

5.3 Future Directions

There are several avenues of potential future work to address the limitations of this paper.

1. Develop a mechanism for knowledge-free signal elicitation that is truthful in all stochastically relevant information structures, or prove that not all stochastically relevant information structures can be queried by a mechanism that does not know the information structure.

2. Demonstrate that the provided upper bound for the maximal set of information structures that are elicitable under a signal-specific query is tight for all such queries, rather than for only the signal-elicitation query.

3. Extend this type of general elicitability analysis to settings where there is no common prior.

6. ACKNOWLEDGEMENTS

The authors are grateful to David Parkes, Bo Waggoner and Jens Witkowski for helpful discussions about this work. This work was partially supported by the National Science Foundation, under grant CCF-0953516. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone.

7. REFERENCES

[1] G. W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[2] A. Dasgupta and A. Ghosh. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, pages 319–330. International World Wide Web Conferences Steering Committee, 2013.

[3] T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[4] R. Jurca and B. Faltings. Minimum payments that reward honest reputation feedback. In *Proceedings of the 7th ACM conference on Electronic commerce*, EC '06, pages 190–199. ACM, 2006.

[5] R. Jurca and B. Faltings. Collusion-resistant, incentive-compatible feedback payments. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 200–209. ACM, 2007.

[6] R. Jurca and B. Faltings. Mechanisms for making crowds truthful. *Journal of Artificial Intelligence Research*, 34(1):209, 2009.

[7] N. Lambert, D. M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *EC '08: Proceedings of the 9th ACM conference on Electronic commerce*, pages 129–138, New York, NY, USA, 2008. ACM.

[8] J. E. Matheson and R. L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096, 1976.

[9] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

[10] R. F. Nau. Should scoring rules be 'effective'? *Management Science*, 31:527–535, 1985.

[11] R. G. Nelson and D. A. Bessler. Subjective probabilities and scoring rules: Experimental evidence. *Journal of Agricultural Economics*, 71(2):363–369, 1989.

[12] D. Prelec. A bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

[13] G. Radanovic and B. Faltings. A robust bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, AAAI '13, 2013.

[14] L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[15] L. von Ahn and L. Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[16] B. Waggoner and Y. Chen. Information elicitation sans verification. In *The ACM EC Workshop on Social Computing and User Generated Content*, 2013.

[17] R. L. Winkler. Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association*, 64(327):1073–1078, 1969.

[18] J. Witkowski, Y. Bachrach, P. Key, and D. C. Parkes. Dwelling on the negative: Incentivizing effort in peer prediction. In *Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing*, HCOMP'13, 2013.

[19] J. Witkowski and D. Parkes. Peer prediction without a common prior. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, EC '12, pages 964–981. ACM, 2012.

[20] J. Witkowski and D. Parkes. A robust bayesian truth serum for small populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, AAAI '12, 2012.