Eliciting Truthful Answers to Multiple-Choice Questions

Preliminary Report

Nicolas Lambert *
Computer Science Department
Stanford University
Stanford, CA 94305

Yoav Shoham
Computer Science Department
Stanford University
Stanford, CA 94305

ABSTRACT

Motivated by the prevalence of online questionnaires in electronic commerce, and of multiple-choice questions in such questionnaires, we consider the problem of eliciting truthful answers to multiple-choice questions from a knowledgeable respondent. Specifically, each question is a statement regarding an uncertain future event, and is multiple-choice – the responder must select exactly one of the given answers. The principal offers a payment, whose amount is a function of the answer selected and the true outcome (which the principal will eventually observe). This problem significantly generalizes recent work on truthful elicitation of distribution properties, which itself generalized a long line of work in elicitation of complete distributions. We provide necessary and sufficient conditions for the existence of payments that induce truthful answers, and give a characterization of those payments. We also study in greater details the common case of questions with ordinal answers, and illustrate our results with several examples of practical interest.

Categories and Subject Descriptors

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

General Terms

Economics, Theory

Keywords

Multiple-choice questions, elicitation, probability, forecasting

1. INTRODUCTION

Assume a principal asks an expert to respond to a questionnaire about a random future event. The questionnaire is composed of multiple-choice questions: for each question, the expert must choose one true answer among a collection of possible answers. For example, the principal might ask *"What horse is the most likely to win the race?"*, or *"What pair of stocks is the most correlated?"*, with, as possible answers, a list of horses or pairs of stocks. Assuming the expert knows the information of interest, the principal must incentivize an honest behavior, by rewarding the expert in a nontrivial way, as a function of both the selected answers and the observed outcome of the event.

Naturally, one can always infer answers for any questionnaire from the full probability distribution of the event. It is well-known that one can elicit the full distribution by rewarding the expert with a *proper* or *strictly proper scoring rule*. (For a survey of the literature, see for example Winkler et al. [11].) However, as the number of outcomes grows, it becomes increasingly difficult to ask for the probability of each outcome. For example, assuming a stock takes as few as 100 different values, and considering only 10 stocks, getting the full distribution—to infer, for example, the most correlated pair—requires an estimate of 10^{100} different parameters. Additionally, any approximation in the elicited distribution could lead to wrong inferences. Besides, the expert may simply not know or not be able to provide the full distribution. Consequently, in practice, asking for the full distribution becomes rapidly non-feasible and it is often more convenient to ask directly for the exact information of interest.

And indeed, researchers have addressed the problem of acquiring more specific information. Savage [9] shows how to elicit the mean of a random variable, Cervera and Munoz [3] devise a particular scoring rule that elicits the median. In a recent research, Lambert et al. [7] investigate the numerical parameters—or properties—of distributions that can be elicited *directly*; that is, for which one can incentivize an expert to provide the true value of the parameters, without requesting any further information. The authors consider the class of properties that are both continuous and nowhere locally constant. They show that many parameters that measure a feature of the distribution belong to the class, for example, the probability of a binary event, the mean, variance or the entropy.

Moving from full distributions to distribution properties enhances greatly our ability to elicit information. However, upon reflection, even this may be too restrictive a setting. Consider for example the question *"Which of the following investments is the most profitable, on expectation: A, B or C?"*, or *"Is horse X at least twice as likely to win the race as any other contestant?"*. These aren't naturally viewed

*Part of this work was done while the author was visiting Yahoo! Research.

EC'09, July 6–10, 2009, Stanford, California, USA.

as properties. And indeed, categorical information is not accounted for in the work of Lambert et al.

In other circumstances, we might be interested in ranges of parameter values. For example, instead of asking *"What is the exact probability of precipitation tomorrow?"*, it might be enough to ask *"Is the probability of precipitation tomorrow: not likely (less than 25%), likely (between 25% and 75%), very likely (greater than 75%) ?"*. This is useful when specific values of the parameters are not needed—for instance, one may not want to fund a project that has more than 50% chances of failure, no matter what the exact likelihood is—or simply because it is sometimes convenient and natural to offer a limited choice of answers.

Finally, one might want to retrieve estimates of discrete parameters, asking *"What is the median and the mode of (discrete) variable X?"*. Not only these parameters are not continuous—and so are excluded from the class considered by Lambert et al.—but they cannot even be modeled as distribution properties. Indeed properties associate a single value to any given distribution, whereas there exist several possible modes or medians for some distributions. We will see later that the ability to allow multiple parameter values for one single distribution is critical to elicit certain information.

All these are examples of multiple-choice questions, which are the topic of this paper. At this point we call for a new set of techniques, as the methods used to elicit full distributions [11], and distribution properties [7], are simply inapplicable. Interestingly, the results regarding the elicitation of distribution properties are mostly negative, while in our extended model, we will show that a variety of multiple-choice questions of practical interest can be elicited.

The paper is organized as follows. We introduce the model and notation in Sections 2 and 3. In Section 4, we provide necessary and sufficient conditions for questionnaires to admit truthful payoffs, and give the general form of those payoffs. In Section 5, we investigate the common case of ordinal answers, and give simpler and more precise characterizations. Often the ability to answer some questionnaire makes it possible to answer another one; this is useful to elicit certain kind of partial information and is discussed in Section 6. Finally we illustrate our results through a series of practical examples in Section 7.

2. MODEL

We consider some random future event with a finite set of possible outcomes Ω. We denote by $\Delta(\Omega)$ the set of distributions over Ω, and we assume that the outcome of the event is drawn according to some probability function $P \in \Delta(\Omega)$. Before the event occurs, a principal asks an expert to fill out a questionnaire about the event, and, after observing its outcome, rewards the expert as a function of his responses and the realized outcome. We assume the expert is able to answer the questionnaire (but need not know the full distribution). A questionnaire is composed of multiple-choice questions (MCQ), such as *"Which horse is most likely to win the race? (a) Horse A, (b) Horse B, (c) Horse C"*, or *"Is the mean sales volume of software X for the month of July: (a) less than 1K copies, (b) between 1K and 3K, (c) greater than 3K copies"*. A question is formally defined by a *question function*.

Definition 1. A *question function*, or *MCQ function*, with a set of possible answers $\mathcal{A}$ and a set of outcomes Ω is a function $\Upsilon : \mathcal{A} \mapsto \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$, $m \geq 1$, with $\mathcal{P}_i \subset \Delta(\Omega)$ for all i and such that $\cup_i \mathcal{P}_i = \Delta(\Omega)$ and, for all $a \neq b$, $\Upsilon(a) \not\subseteq \Upsilon(b)$.

A *questionnaire* $\mathcal{Q}$ is formally defined as a set of question functions $\mathcal{Q} = \{\Upsilon_1, \ldots, \Upsilon_k\}$. Question functions describe multiple-choice questions. The domain $\mathcal{A}$ of a question function Υ stipulates the set of admissible answers for the question. It can be any finite set. The image of Υ (given by the sets $\mathcal{P}_i$) determines the true answers for given distributions of the event: if the true distribution is included in $\Upsilon(a)$, then a is correct, but is not otherwise. In their most general form, multiple-choice questions are lists of true/false statements regarding the distribution of outcomes. With $\mathcal{A} = \{a_1, \ldots, a_m\}$, the question function Υ corresponds to the following MCQ: *"Choose one true statement among the following, regarding the distribution P of the future event under consideration: answer (a_1): $P \in \Upsilon(a_1)$, ..., answer (a_m): $P \in \Upsilon(a_m)$"*. The formalism of question functions will prove to be extremely useful to obtain general results on questionnaires. However, most questions of interest can be formulated in a simple and intuitive way.

For instance, consider our above horse race MCQ. We can take as set of outcomes the possible winners $\{A, B, C\}$; with a distribution P, $P(X)$ is the probability that horse X wins the race. We can use a set of answers $\mathcal{A} = \{a, b, c\}$. Our MCQ is fully described by the question function Υ, where $\Upsilon(a)$ is the set of distributions under which horse A is more likely to win than horse B or C; formally $\Upsilon(a) = \{P \in \Delta(\{A, B, C\}) \mid P(A) \geq P(B), P(A) \geq P(C)\}$, with a similar form for $\Upsilon(b)$ and $\Upsilon(c)$.

To be a valid question function, Υ must satisfy two conditions. The first condition, $\cup_a \Upsilon(a) = \Delta(\Omega)$, states that the multiple-choice question must contain at least one correct answer in all cases (i.e., for all possible distributions of outcomes). The second condition, $\Upsilon(a) \not\subseteq \Upsilon(b)$ for all $a \neq b$ means that we exclude redundant answers. Indeed if $\Upsilon(a) \subseteq \Upsilon(b)$, then b is true whenever a is true, and an expert could simply respond b whenever answer a or b is correct.

In exchange for his service, the principal provides a monetary reward to the expert. Let us first consider the case of single multiple-choice questions (as argued below, this is without loss of generality). The payoff to the expert can depend on his own responses and on the true outcome of the event, that the principal subsequently observes. Such payoffs are defined by a *payoff function*.

Definition 2. A *payoff function* for a question function Υ with domain $\mathcal{A}$ is a function $\Pi : \mathcal{A} \times \Omega \mapsto \Re$.

For an MCQ with set of answers $\mathcal{A}$, a payoff function Π specifies an amount $\Pi(a, \omega)$ that an expert would get when answering a while the true outcome is ω.

Not all rewarding schemes lead to truth telling. To incentivize a risk-neutral expert to answer correctly, the principal should use *(strictly) proper payoffs*, by analogy with the (strictly) proper scoring rules used in the forecasting literature. With proper payoffs, the expert gets a maximum expected reward by responding correctly, so that experts never benefit from lying. If payoffs are strictly proper, the maximum expected reward is *only* obtained by providing a correct answer, any other response yields lower expected payoffs, so that experts always benefit from telling the truth.

Definition 3. A payoff function Π is *proper* for an MCQ Υ if, for all possible answers a,

$$\Upsilon(a) \subseteq \operatorname*{arg\,max}_{P \in \Delta(\Omega)} \mathop{\mathbb{E}}_{\omega \sim P}[\Pi(a,\omega)] \ .$$

If, in addition, for all answers a,

$$\Upsilon(a) \supseteq \operatorname*{arg\,max}_{P \in \Delta(\Omega)} \mathop{\mathbb{E}}_{\omega \sim P}[\Pi(a,\omega)] \ ,$$

the payoff Π is *strictly proper.*

Risk-neutrality is always implicitly assumed with the notion of (strict) properness, as it is with proper scoring rules in the forecasting literature. However, our results easily extend to general utility maximizers. Consider a utility for money $u : \mathcal{R} \mapsto \mathcal{R}$, increasing and bijective. Let Π be a (strictly) proper payoff. Then the payoff $u^{-1} \circ \Pi$ yields the same expected utility as the expected value of Π. So, the optimal strategies of a risk-neutral expert with payoff Π are the same as for an expert with utility u and payoffs $u^{-1} \circ \Pi$. And conversely, for any payoff $\tilde{\Pi}$, the optimal strategies of an expert with utility u are the same as for a risk-neutral expert with payoff $u \circ \tilde{\Pi}$. This means that the type of information we can elicit does not depend on the utility function. Additionally, when the utility function is known, the truth-inducing payoffs can be obtained by a simple transformation of the proper and strictly proper payoffs. If unkown, one can use lottery-based methods (see Savage [9] and Smith [10]).

In general, questionnaires may contain any number of questions. When such a questionnaire consists of questions that admit a (strictly) proper payoff, we can construct a (strictly) proper payoff of the full questionnaire by summation of the payoffs for each question. It is easily shown that this global payoff exhibits a *monotone* property: the amount of money awarded to an expert who switches from an incorrect to a correct answer (for any question) is nondecreasing (increasing with strict properness). However the individual questions need not admit such truthful payoffs for the questionnaire as a whole to admit a truthful payoff.

This seems to suggest that one should also consider (strict) properness for sets of questions. But this is in fact not needed. Indeed, we observe that the truthful elicitation of answers for a full questionnaire reduces to that of a single multiple-choice question, as we can easily encode any finite number of questions into one single MCQ.

LEMMA 2.1. *Any questionnaire $\mathcal{Q} = \{\Upsilon_1, \ldots, \Upsilon_k\}$ is equivalent to some MCQ Υ.*

By equivalence we mean the existence of a bijection between the possible responses of the full questionnaire and the answers of the MCQ, so that answering the former is exactly the same as answering the latter. The number of answers of an equivalent MCQ can be very large, generally exponential in the number of questions of the original questionnaire. However this is inconsequent here, since we focus on the strategic analysis – in practice, it is of course preferable to present questionnaires in their original format, and not in their equivalent reduced form. *In the sequel of this paper, and without loss of generality, we restrict ourselves to the study of individual multiple-choice questions.*

Note that the questions of our questionnaires are single-response: we ask the expert to pick exactly one answer among a set of alternatives. Indeed, our purpose is to get only *one* valid answer, not the complete set – even if for some questions, several answers may be correct. Single-response is often desirable; for example, if one is interested in a 90% confidence interval for a random variable, asking for only one instance of those intervals is often sufficient, and much easier in practice than asking for all of them. Our restriction is without loss of generality, as we will see in the next section that any question/questionnaire for which one is interested in the complete set of answers is equivalent to a single-response MCQ. However, we will also argue that those multiple-response questionnaires never admit a strictly proper payoff.

Naturally, all multiple-choice questions admit a proper payoff, for example a constant payoff rule is proper. However those payoffs do not always punish false experts. In general, it is preferable to achieve strict properness, leading to the notion of direct implementability.

Definition 4. A multiple-choice question Υ is *directly implementable with strictly proper payoffs*, or simply *directly implementable*, when there exists a strictly proper payoff for Υ.

Note that with direct implementability, the expert provides truthfully *exactly* the information of interest. Sometimes, one may ask for *more* information to infer the information of interest—which may not admit a truth-inducing payoff. For example one may ask for the full distribution, or for answers of more detailed questions. The latter case is discussed in Section 6.

3. NOTATION

We denote by $\mathfrak{R}^\Omega$ the set of random variables over Ω. Any density function over Ω is a random variable, and to simplify notation we identify a probability function with its density and write $P(\{\omega\}) = P(\omega)$. Hence $\Delta(\Omega)$ also denotes the set of densities over Ω. We consider $\mathfrak{R}^\Omega$ as a vector space, with the inner product $\langle X, Y \rangle = \sum_{\omega \in \Omega} X(\omega)Y(\omega)$ and the distance $d(X,Y) = \|X - Y\| = \sqrt{\langle X-Y, X-Y \rangle}$. For a subset $\mathcal{S}$ of a vector space $\mathcal{E}$, we note $\operatorname{Span} \mathcal{S}$ the linear span of $\mathcal{S}$, which is the smallest subspace that contains $\mathcal{S}$. We recall that the dimension of a subset of a vector space is the dimension of its linear span. An hyperplane of $\Delta(\Omega)$ is defined as the intersection of $\Delta(\Omega)$ and an hyperplane of $\mathfrak{R}^\Omega$. A real-valued function f on a subset $\mathcal{S}$ of a vector space is *linear* when $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for any scalars α, β and vectors x, y of $\mathcal{S}$.

Given a set of answers $\mathcal{A}$, we identify a payoff function $\Pi(a, \omega)$ with a collection of vectors $\{\Pi(a), a \in \mathcal{A}\}$ of $\mathfrak{R}^\Omega$, where we denote by $\Pi(a)$ the function $\Pi(a, \cdot)$. For a distribution P, we write $\Pi(a, P)$ the value $\mathbb{E}_{\omega \sim P}[\Pi(a, \omega)] = \langle \Pi(a), P \rangle$.

4. CHARACTERIZATION RESULTS

In this section we consider a set Ω of n outcomes and a multiple-choice question Υ with a set of m answers $\mathcal{A} = \{a_1, \ldots, a_m\}$.

4.1 Directly implementable questionnaires

Our primary objective is to identify the questions that are directly implementable, for which it is possible to elicit truthful answers by setting appropriate payoffs. We start with the main conditions that such an MCQ must satisfy:

the set of distributions associated with a correct answer must be a "thick" (of non-empty interior), closed convex polyhedron, and the intersections between the sets of distributions of two different answers must be "thin" (of empty interior, included in an hyperplane).

PROPOSITION 4.1. If Υ *is directly implementable,* then

- *For all* $a \in \mathcal{A}$, $\Upsilon(a)$ *is a nondegenerate closed convex polyhedron of the simplex* $\Delta(\Omega)$.
- *For all* $a, b \in \mathcal{A}$, $a \neq b$, *if* $\Upsilon(a) \cap \Upsilon(b)$ *is not empty then it is a degenerate closed convex polyhedron.*

We recall that a closed convex polyhedron of $\Delta(\Omega)$ is any subset $\mathcal{S}$ of distributions such that $\mathcal{S}$ is the convex envelope of a finite collection of probability functions. A convex polyhedron is said to be *degenerate* when its dimension is strictly less than n (i.e., the polyhedron is "flat").

Each MCQ Υ corresponds graphically to some configuration in the simplex of distributions, that is formed by the shapes of the sets $\Upsilon(a)$ for all answers a. Informally, our result implies that if an MCQ is directly implementable, then it has a convex configuration, in the sense that each cell $\Upsilon(a)$ should be convex. Moreover there should always exist cases for which several answers are simultaneously correct, but those cases should be extremely rare, while those situations for which only one answer is correct should occur quite often. The number of simultaneously correct answers depends on the question, and there are directly implementable questions for which all answers are valid in some rare circumstances.

Our proposition provides conditions that can easily be applied to rule out most MCQ of practical interest that are not directly implementable. We illustrate its use in the following few paragraphs.

Convexity condition. This is the most commonly violated condition. For example, consider the case of two random variables X and Y taking values in a set $\mathcal{X}$ containing at least two elements. Let's define the following MCQ Υ_1: *"Which of the following is true: (a)* $\mathrm{Var}(X) \geq \mathrm{Var}(Y)$, *or (b)* $\mathrm{Var}(Y) \geq \mathrm{Var}(X)$*"*. This question is not directly implementable as it does not satisfy the convexity condition. Indeed take the case of $\mathcal{X} = \{0, 1\}$, and let P and Q be distributions such that X and Y are independent under both P and Q, with $P(X = 0) = 1$, $Q(X = 0) = 0$, and $P(Y = 0) = Q(Y = 0) = 1/3$. Let $R = (P + Q)/2$. Then $\mathrm{Var}_P(X) = \mathrm{Var}_Q(X) = 0 \leq 2/9 = \mathrm{Var}_P(Y) = \mathrm{Var}_Q(Y)$. Yet, $\mathrm{Var}_R(X) = 1/4 \geq \mathrm{Var}_R(Y) = 2/9$. This means that $\Upsilon_1(a)$ is not convex. The reasoning trivially extends to arbitrary sets $\mathcal{X}$.

Closeness condition. Besides being convex, the cells should also be closed. For example, let A be a binary event of Ω. The MCQ Υ_2: *"The probability* p *of* A *satisfies (a)* $p < 0.1$, *(b)* $0.1 \leq p \leq 0.9$, *or (c)* $p > 0.9$*"* is not directly implementable, as $\Upsilon_2(a)$ and $\Upsilon_2(c)$ are not closed.

Nondegeneracy condition. We illustrate the nondegeneracy condition with a similar question Υ_3: *"The probability of A is (a) less than 1/2, (b) equal to 1/2, or (c) greater than 1/2"*. This MCQ is not directly implementable as the set $\Upsilon_3(b)$ is "flat", its dimension is less than n. In fact, it can be shown that for any valid MCQ, whenever an answer is degenerate, some other fails the closeness condition. However, degeneracy is sometimes easier to detect.

Thin intersection condition. Even if the answers correspond to nondegenerate closed convex polyhedra, a multiple-choice question is not always directly implementable. For example, consider the case of three random variables X, Y and Z taking values in $\mathcal{X}$, and let's define the MCQ Υ_4: *"Which of the following statements is true about the mean of* X, Y *and* Z*: (a)* $\mathbb{E}[X] \geq \mathbb{E}[Y]$, *(b)* $\mathbb{E}[Y] \geq \mathbb{E}[Z]$, *(c)* $\mathbb{E}[Z] \geq \mathbb{E}[X]$*"*. This is indeed a valid question as no answer implies another, since no two answers concern the same random variables, and there is always a valid answer. (Otherwise, by contradiction, $\mathbb{E}[X] > \mathbb{E}[Y] > \mathbb{E}[Z] > \mathbb{E}[X]$ under some probability.) Each set $\Upsilon_4(\alpha)$ is a convex polyhedron for any answer α. However, from Proposition 4.1 the question is not directly implementable because $\Upsilon_4(\alpha) \cap \Upsilon_4(\beta)$ is of dimension n for any two answers $\alpha \neq \beta$. Indeed, for any probability function P, either $\mathbb{E}_P[X] \geq \mathbb{E}_P[Y] \geq \mathbb{E}_P[Z]$, or $\mathbb{E}_P[X] \geq \mathbb{E}_P[Z] \geq \mathbb{E}_P[Y]$, or $\mathbb{E}_P[Y] \geq \mathbb{E}_P[X] \geq \mathbb{E}_P[Z]$, or $\mathbb{E}_P[Y] \geq \mathbb{E}_P[Z] \geq \mathbb{E}_P[X]$, or $\mathbb{E}_P[Z] \geq \mathbb{E}_P[X] \geq \mathbb{E}_P[Y]$, or $\mathbb{E}_P[Z] \geq \mathbb{E}_P[Y] \geq \mathbb{E}_P[X]$. By symmetry, for any of the 6 conditions, the set of probabilities satisfying it has the same dimension, and since those sets cover the entire simplex $\Delta(\Omega)$, they are nondegenerate. As each of $\Upsilon_4(\alpha) \cap \Upsilon_4(\beta)$ for $\alpha \neq \beta$ equals one of these sets, they are of dimension n.

An interesting consequence of Proposition 4.1 is the impossibility to use another type of questionnaires for which, instead of only *one* correct answer, the expert is asked to provide *all* the correct answers, when there are more than one. A payoff is then called strictly proper when the expected payoff is maximized when and only when the expert identify the full collection of correct answers for each question. As we did for single-response questions, *multiple-response* questions can be modeled as functions $\Phi : \Delta(\Omega) \mapsto 2^{\mathcal{A}}$. Here $\mathcal{A}$ is the set of possible answers and $\Phi(P)$ gives the full set of correct answers when the outcomes are distributed according to P. The multiple-response MCQ is equivalent to a (single-response) MCQ Υ with a set of possible answers $\Phi(\Delta(\Omega))$ and defined by $\Upsilon(a) = \Phi^{-1}(a)$. As this MCQ corresponds to a partition in the space of probability functions, two answers may never occur simultaneously. Since there are only a finite number of answers, the closeness condition is violated (the elements of a finite partition of a closed, non-degenerate convex polyhedron cannot be all closed). Hence no multiple-response MCQ/questionnaire admits a strictly proper payoff.

If the conditions enunciated in Proposition 4.1 allow to detect non-implementability in most cases, they are not sufficient. For example, take $\Omega = \{1, 2, 3\}$ and consider the following MCQ: *"Which statement is true: (a) Outcome 1 occurs at most half the time, (b) Outcome 1 occurs at least half the time and outcome 2 is more likely than or equally likely to outcome 3, and (c) Outcome 1 occurs at least half the time and outcome 2 is less likely than or equally likely to outcome 3"*. The question satisfies all the conditions of Proposition 4.1, yet is not directly implementable. Let's use the vector notation $(P(\{1\}), P(\{2\}), P(\{3\}))$ to represent a distribution P. Let $P_0 = (1, 0, 0)$, $P_1 = (\frac{1}{2}, \frac{1}{2}, 0)$, $P_2 = (\frac{1}{2}, 0, \frac{1}{2})$, $P_3 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. Let Π be a payoff function. Since a and b are valid answers under P_1, $\Pi(a, P_1) = \Pi(b, P_1)$. And similarly, $\Pi(a, P_2) = \Pi(c, P_2)$, $\Pi(a, P_3) = \Pi(b, P_3) = \Pi(c, P_3)$, $\Pi(b, P_0) = \Pi(c, P_0)$. By linearity, $2\Pi(a, P_3) = \Pi(a, P_1) + \Pi(a, P_2)$, so $2\Pi(c, P_3) = \Pi(b, P_1) + \Pi(c, P_2)$ implying $\Pi(b, P_1) = \Pi(c, P_1)$. Also, since the vectors P_0, P_1, P_2 are independent, $\Pi(b)$ is entirely specified by $\Pi(b, P_0), \Pi(b, P_1), \Pi(b, P_3)$, and $\Pi(c)$ is entirely specified by $\Pi(c, P_0), \Pi(c, P_1), \Pi(c, P_3)$. However, we have shown that $\Pi(b, P_0) = \Pi(c, P_0)$, $\Pi(b, P_3) = \Pi(c, P_3)$, and $\Pi(b, P_1) = \Pi(c, P_1)$. Hence $\Pi(a) = \Pi(b)$ and Π cannot be strictly proper.

Fortunately we are able to obtain an elegant description of the directly implementable multiple-choice questions by using a well-known concept of computational geometry called *power diagram*. Let $\mathcal{C}$ be a convex subset of a real finite-dimensional vector space, with the Euclidian distance d.

Consider a set $\mathcal{S} = \{x_1, \ldots, x_m\}$ of m points or sites of $\mathcal{C}$. Each point x_i has an associated weight $w_i \in \Re$. The weight expresses the power that sites have on their neighboring points, the larger the weight, the more power. More precisely, the measure of power of a site x_i on a point y is given by the power function

$$pow(x_i, y) = d(x_i, y)^2 - w_i \ .$$

The power function plays the same role as a distance: the lower the value $pow(x_i, y)$, the greater the power. The *power cell* $\mathcal{R}_i$ of a site x_i is the region of the space that contains all the points under the influence of x_i: it contains the points y such that $pow(x_i, y) \leq pow(x_j, y)$ for all $j \neq i$. The complete collection of regions $\{\mathcal{R}_1, \ldots, \mathcal{R}_m\}$ is called the *power diagram* of the family of weighted sites $\{(x_i, w_i), 1 \leq i \leq m\}$.

Intuitively, power diagrams represent the regions of the space that are closest to a sphere, where the "distance" between a point and sphere is specially defined and has a simple geometric interpretation. Indeed, assume positive weights (without loss of generality, since power diagrams are invariant under a constant addition to the weights). Then each weighted site (x_i, w_i) corresponds to a sphere of center x_i and radius $\sqrt{w_i}$. Then, if $\sqrt{pow(x_i, y)} \geq \sqrt{w_i}$ the value $\sqrt{pow(x_i, y)}$ corresponds to the distance between y and the point on the sphere centered on x_i that is on a tangent line to the sphere that goes through y.

Power diagrams are closely related to the more commonly known *Voronoi diagrams*, and often interpreted as a generalization of the latter. Given a set of Voronoi sites, Voronoi diagrams consist of regions of the space, each region being the set of points that are closer to a site than to any other site. When all weights are the same, the power diagram is identical to the Voronoi diagram. Introduced by Aurenhammer [1] and Imai et al. [6], power diagrams have been well studied and are commonly used both in computer science [4] and more recently in mathematics [8]. The reader may refer to Aurenhammer [2] for a summary of the literature.

Strict properness means that an expert should maximize her expected payoff when and only when answering correctly. Given a payoff function Π, the expected payoff of a truthful expert is given by

$$P \mapsto \max_{a \in \mathcal{A}} \mathop{\mathbb{E}}_{\omega \sim P}[\Pi(a, \omega)]$$

which, since $\mathbb{E}_{\omega \sim P}[\Pi(a, \omega)] = \langle \Pi(a), P \rangle$, describes the upper envelope of (non-vertical) hyperplanes given by $P \mapsto \langle \Pi(a), P \rangle$ for each answer a. The region $\Upsilon(a)$ of the space of distributions that give answer a true is the piece of the hyperplane given by $P \mapsto \langle \Pi(a), P \rangle$ that is on the envelope. This means that the regions $\Upsilon(a)$ coincide to the projection of an envelope of (non-vertical) hyperplanes. It turns out that those projections correspond exactly to power diagrams. Intuitively, this is due to the fact that, even though the power function is quadratic, the power diagram is only defined by differences of power functions, which are linear. So directly implementable questions correspond to geometric configurations that represent power diagrams in the simplex of distributions.

THEOREM 4.1. *The multiple-choice question Υ is directly implementable* if and only if $\{\Upsilon(a_1), \ldots, \Upsilon(a_m)\}$ *is a power diagram of $\Delta(\Omega)$ for some set of weighted sites.*

PROOF. This can be seen as a consequence of the connection established notably by Aurenhammer [1] between power diagrams and convex polyhedra in one dimension higher.

<u>IF PART.</u> Let's assume that $\{\Upsilon(a_1), \ldots, \Upsilon(a_m)\}$ is a power diagram for the collection of weighted vectors $\{(P_a, w_a)\}_{a \in \mathcal{A}}$ of $\Delta(\Omega)$. Then for all $a, b \in \mathcal{A}$, and all $P \in \Upsilon(a)$,

$$d(P_a, P)^2 - w_a \leq d(P_b, P)^2 - w_b \ .$$

Let $Q \in \Delta(\Omega)$. For $P \in \Delta(\Omega)$, we define

$$\begin{aligned} L_a(P) &= d(P_a, P)^2 - w_a - d(Q, P)^2 \\ &= \|P_a\|^2 - 2\langle P_a, P \rangle - w_a - \|Q\|^2 + 2\langle Q, P \rangle \\ &= \langle P, V_a \rangle + c_a \end{aligned}$$

with $V_a = 2(Q - P_a)$ and $c_a = \|P_a\|^2 - \|Q\|^2 - w_a$. In particular, since $\langle P, 1 \rangle = 1$, $L_a(P) = -\langle P, S_a \rangle$ with $S_a = -V_a - c_a$.

Consider the payoff function $\Pi(a, \omega) = S_a(\omega)$. Let $a, b \in \mathcal{A}$. If $P \in \Upsilon(a)$ and $P \notin \Upsilon(b)$, then by definition of the power diagram, $d(P_a, P)^2 - w_a < d(P_b, P)^2 - w_b$, hence $L_a(P) < L_b(P)$, implying $\mathbb{E}_{\omega \sim P}[\Pi(a, \omega)] > \mathbb{E}_{\omega \sim P}[\Pi(b, \omega)]$. Similarly, if $P \in \Upsilon(a)$ and $P \in \Upsilon(b)$, $d(P_a, P)^2 - w_a = d(P_b, P)^2 - w_b$, and $\mathbb{E}_{\omega \sim P}[\Pi(a, \omega)] = \mathbb{E}_{\omega \sim P}[\Pi(b, \omega)]$. Therefore Π is strictly proper.

<u>ONLY IF PART.</u> Let Π be a strictly proper payoff, and S_a the random variable $S_a(\omega) = \Pi(a, \omega)$. For $P \in \Delta(\Omega)$, let $L_a(P) = -\langle P, S_a \rangle$. Take $c_a = \langle 1, S_a \rangle$ and $V_a = -S_a - c_a$. Let $Q = 1/|\Omega|$. Then Q lies in the interior of $\Delta(\Omega)$, and there exists $\alpha > 0$ small enough such that, if $P_a = Q - (\alpha/2)V_a$, P_a is a well-defined probability for all a. We get

$$\begin{aligned} L_a(P) &= -\langle P, S_a \rangle \\ &= \langle P, V_a \rangle + c_a \\ &= \frac{1}{\alpha}\langle P, 2(Q - P_a) \rangle + c_a \\ &= \frac{1}{\alpha}(d(P, P_a)^2 - d(P, Q)^2 + \alpha c_a - \|P_a\|^2 - \|Q\|^2) \\ &= \frac{1}{\alpha}(d(P, P_a)^2 - d(P, Q)^2 - w_a) \end{aligned}$$

with $w_a = \|Q\|^2 - \|P_a\|^2 - \alpha c_a$. By Definition 3, strict properness imply that $L_a(P) \leq L_b(P)$ for all $a, b \in \mathcal{A}$ and all $P \in \Upsilon(a)$, which means that $d(P, P_a)^2 - w_a \leq d(P, P_b)^2 - w_b$. Therefore $\{\Upsilon(a)\}_{a \in \mathcal{A}}$ is a power diagram for the weighted points $\{(P_a, w_a)\}_{a \in \mathcal{A}}$. $\square$

Since Voronoi diagrams are special instances of power diagrams, the next corollary follows.

COROLLARY 4.1. If $\{\Upsilon(a_1), \ldots, \Upsilon(a_m)\}$ *is a Voronoi diagram for a set of sites,* then Υ *is directly implementable.*

4.2 Payoff functions

We now describe the general shape of the payoffs. We show that the set of proper/strictly proper payoffs form a cone in the space of payoff functions. More precisely, the (strictly) proper payoffs are mixtures of a fixed, finite number of payoff functions, modulo a constant payment. The coefficients of those "base" payoffs are constrained to be nonnegative in proper payoffs, and positive in strictly proper payoffs. Thus, all is needed to design the truth-inducing payoffs are those base payoffs. Base payoffs depend on the information requested and are not subject to particular constraints. However, our next section investigates a special case of questions for which the base payoffs can be easily constructed. Together with our characterization of direct implementability, the next theorem is the second major result of our paper.

THEOREM 4.2. *If the multiple-choice question Υ is directly implementable, then there exists a set of payoff functions $\mathcal{B} = \{\Pi_1, \ldots, \Pi_\ell\}$ for some $\ell \geq 1$, called* a base *of Υ, such that Π is proper (resp. strictly proper)* if and only if*, for all $a \in \mathcal{A}, \omega \in \Omega$,*

$$\Pi(a, \omega) = \Pi_0(\omega) + \lambda_1 \Pi_1(a, \omega) + \cdots + \lambda_\ell \Pi_\ell(a, \omega)$$

for any function $\Pi_0 : \Omega \mapsto \Re$ and any reals $\lambda_1, \ldots, \lambda_\ell \geq 0$ (resp. any $\lambda_1, \ldots, \lambda_\ell > 0$).

PROOF SKETCH. Let $\mathcal{P}$ be the space of payoff functions, i.e., the set of functions $\Pi : \mathcal{A} \times \Omega \mapsto \Re$, considered as a finite dimensional inner product space, the inner product being defined by $\langle \Pi_1, \Pi_2 \rangle = \sum_{a \in \mathcal{A}, \omega \in \Omega} \Pi_1(a, \omega) \times \Pi_2(a, \omega)$.

Assume Υ is directly implementable. A payoff Π if proper when reporting truthfully maximizes the expected payoff, which is equivalent to :

$$\langle \Pi(a) - \Pi(b), P \rangle \geq 0 \quad \forall a, b \in \mathcal{A}, \forall P \in \Upsilon(a) \ . \tag{1}$$

By Proposition 4.1, for all $a \in \mathcal{A}$ the set $\Upsilon(a)$ is a bounded convex polyhedron of $\Re^\Omega$, and so is the convex hull of a finite set of vertices [5]. Let $\mathcal{V}_a$ be the set of vertices of $\Upsilon(a)$. Since $\mathcal{V}_a$ is included in $\Upsilon(a)$, Equation 1 implies that

$$\langle \Pi(a) - \Pi(b), P \rangle \geq 0 \quad \forall a, b \in \mathcal{A}, \forall P \in \mathcal{V}_a \ . \tag{2}$$

However, since all vectors of the convex polyhedron $\Upsilon(a)$ are nonnegative mixtures of the extreme vertices $\mathcal{V}_a$, by linearity of the inner product, Equation 2 implies Equation 1, and so the two sets of inequalities are equivalent.

Equation 2 defines a finite homogeneous system of linear inequalities on the space of payoff functions. Its set of solutions has the form $\mathcal{S} = \mathcal{K} + \mathcal{C}$ (see Eremin [5]). Here $\mathcal{C}$ is a cone formed by a finite set of directrices $\{\Pi_1, \ldots, \Pi_\ell\}$ of the edges of the cone of solutions. That is, $\mathcal{C}$ is the set of vectors $\lambda_1 \Pi_1 + \cdots + \lambda_\ell \Pi_\ell$ of $\mathcal{P}$ with $\lambda_1, \ldots, \lambda_\ell \geq 0$. $\mathcal{K}$ is the kernel of the system of inequalities, the solution of the system of equalities

$$\langle \Pi(a) - \Pi(b), P \rangle = 0 \qquad \forall a, b \in \mathcal{A}, \forall P \in \mathcal{V}_a$$

in the space $\mathcal{P}$. We note that any Π such that $\Pi(a) = \Pi(b)$ for all $a, b \in \mathcal{A}$ is solution. To show that these are the only solutions, let $a, b \in \mathcal{A}$. By Proposition 4.1, $\Upsilon(a)$ has dimension n, thus the linear span of its vertices $\mathcal{V}_a$ is the full space $\Re^\Omega$. If Π is solution of the system, then for all $P \in \mathcal{V}_a$, $\langle \Pi(a) - \Pi(b), P \rangle = 0$, and by linearity the equality remains true for all P in the linear span of $\mathcal{V}_a$. This implies $\Pi(a) = \Pi(b)$. Hence $\mathcal{K} = \{\Pi \in \mathcal{P} \mid \Pi(a, \omega) = \Pi(b, \omega) \, \forall a \neq b\}$.

Therefore, Π is proper if and only if $\Pi \in \mathcal{S}$, or equivalently, if and only if $\Pi = \Pi_0 + \sum_{i=1}^{\ell} \lambda_i \Pi_i$, with $\lambda_1, \ldots, \lambda_\ell \geq 0$ and Π_0 a payoff that does not depend on the answer.

This proves the result for proper payoffs. Intuitively, when considering strict properness, some of the inequalities become strict. The solutions now take the form of the topologic interior of $\mathcal{C}$, modulo constant payments. The interior of the cone $\mathcal{C}$ is described by the positive mixtures of the directrices, which converts into positiveness of the scalars $\lambda_1, \ldots, \lambda_\ell$. The complete proof involves some nontrivial algebraic manipulations and is omitted for brevity. □

5. ORDINAL ANSWERS

Sometimes, the answers of a multiple-choice question are not related in any particular way. This is true, for example, when asking an expert to choose among a list of portfolios for the one with the highest payoff. Such MCQ are *categorical*. But it is not uncommon to find questions whose answers can be compared to one another. For example, consider the MCQ *"The payoff for portfolio X is expected to be: (a) less than \$1K, (b) between \$1K and \$5K, (c) between \$5K and \$10K, (d) greater than \$10K"*. Here the answers are related: for answer *(b)* the profit is larger than for answer *(a)*, but smaller that for answer *(c)*, and even smaller than for answer *(d)*.

In this section we deal with the special—but common—case of multiple-choice questions with ordinal answers, that is, with a set of answers that can be ordered in a meaningful way. Ordinal answers are typical with questions on discrete parameters, such as mode and median, or questions on ranges of continuous parameters, such as mean and variance. (We detail some of those cases in Section 7.) Let Ω be the set of outcomes and let Υ be an MCQ with a set of answers $\mathcal{A}$. Further let's assume that $\prec$ is a strict total order over $\mathcal{A}$, so that we can write $\mathcal{A} = \{a_1, \ldots, a_m\}$, with $a_i \prec a_j$ if $i < j$.

For such a question, we might require more than strict properness. In our above portfolio example, suppose for instance that the true mean payoff is \$4K. Then an expert who responds *(c)* is of better use that an expert who responds *(d)*, even if both are wrong, as his answer is closer to the truth. It is therefore natural to give that expert higher rewards. Such rewarding schemes are called *accuracy-rewarding* (see Lambert et al. [7]).

In the remaining of this section we denote by $\Upsilon^{-1}(P)$ the set of all the correct answers when the true distribution of outcomes is P. For a set of answers $\mathcal{S}$, we write $a \prec \mathcal{S}$ (resp. $\mathcal{S} \prec a$) when $a \prec b$ (resp. $b \prec a$) for all $b \in \mathcal{S}$.

Definition 5. A payoff Π is *accuracy-rewarding* with respect to the strict total order $\prec$ if it is strictly proper and when, for all $P \in \Delta(\Omega)$, all $a, b \in \mathcal{A}$, if either $a \prec b \prec \Upsilon^{-1}(P)$ or $\Upsilon^{-1}(P) \prec b \prec a$, then $\mathbb{E}_{\omega \sim P}[\Pi(a, \omega)] < \mathbb{E}_{\omega \sim P}[\Pi(b, \omega)]$.

Two main reasons motivate the study of MCQ with ordinal answers. First, the principal might want to use payoffs that are accuracy-rewarding, and our primary objective will be to determine the questions that admit such payoffs, and to provide a simple description of those payoffs. Second, we will show that whenever an accuracy-rewarding payoff exists for an MCQ, it is possible to express more precisely than in Theorem 4.2 the general form of the (strictly) proper payoffs. This will prove to be useful in practice (see Section 7). Our next result characterizes the questions that admit accuracy-rewarding payoffs.

THEOREM 5.1. *The multiple-choice question Υ is directly implementable with accuracy-rewarding payoffs* if and only if $\Upsilon(a_i) \cap \Upsilon(a_{i+1})$ *is an hyperplane of* $\Delta(\Omega)$*, for all* $1 \leq i \leq n-1$.

PROOF. <u>IF PART.</u> The construction of an accuracy-rewarding payoff is done in Theorems 5.2 and 5.3.

<u>ONLY IF PART.</u> Let Π be an accuracy-rewarding payoff.

<u>STEP 1.</u> We first show that for all i and $j > i+1$, if $P \in \Upsilon(a_i)$ and $P \in \Upsilon(a_j)$ then $P \in \Upsilon(a_{i+1})$. Suppose by contradiction that there exists i and $P \in \Upsilon(a_i)$, $P \notin \Upsilon(a_{i+1})$, and $P \in \Upsilon(a_j)$ for some $j > i+1$. By Proposition 4.1, $\Upsilon(a_i)$ is a convex polyhedron of nonempty interior. Since $P \in \Upsilon(a_i)$, there exists a sequence of vectors $\{P_k\}_{k\geq 1}$ of $\mathrm{Int}(\Upsilon(a_i))$ that converges towards P. As, for all a, $\Pi(a,\cdot)$ is linear, it is continuous and $\lim_{k\to+\infty} \Pi(a,P_k) \to \Pi(a,P)$. Let $\delta_k = \Pi(a_i,P_k) - \Pi(a_{i+1},P_k)$. Since P_k and P belong to $\Upsilon(a_i)$, but not to $\Upsilon(a_{i+1})$, $\delta_k > 0$, and δ_k converges toward $\delta = \Pi(a_i,P) - \Pi(a_{i+1},P) > 0$. Therefore $\inf\{\delta_k\}_{k\geq1} > 0$. Let $\epsilon = \inf\{\delta_k/2\}_{k\geq1}$. By continuity, there exists K such that

$$|\Pi(a_i,P) - \Pi(a_i,P_K)| \leq \epsilon/2 \ ,$$

and

$$|\Pi(a_j,P) - \Pi(a_j,P_K)| \leq \epsilon/2 \ ,$$

so that, since a_i and a_j are both valid answers under P, $\Pi(a_i,P) = \Pi(a_j,P)$ and

$$|\Pi(a_i,P_K) - \Pi(a_j,P_K)| \leq \epsilon \ .$$

Hence, $\Pi(a_j,P_K) > \Pi(a_i,P_K) - \epsilon = \Pi(a_{i+1},P_K) + \delta_K - \epsilon > \Pi(a_{i+1},P_K)$. However, P_K is in the interior of $\Upsilon(a_i)$, which means according to Proposition 4.1 that a_i is the only valid answer under P_K. But, since $a_i < a_{i+1} < a_j$, and Π is accuracy rewarding, we should have $\Pi(a_{i+1},P_K) > \Pi(a_j,P_K)$.

<u>STEP 2.</u> Now let $1 \leq j \leq m-1$. Let $A_j = \Upsilon(a_1) \cup \cdots \cup \Upsilon(a_j)$, and $B_j = \Upsilon(a_{j+1}) \cup \cdots \cup \Upsilon(a_m)$. By Proposition 4.1, A_j and B_j are polyhedra of dimension n and nonempty interior in $\Delta(\Omega)$, with $A_j \cup B_j = \Delta(\Omega)$. Let $i \leq j < j+1 \leq k$. If $P \in \Upsilon(a_i)$ and $P \in \Upsilon(a_k)$, by a recursive application of our claim above, we find that $P \in \Upsilon(a_i), \Upsilon(a_{i+1}), \ldots, \Upsilon(a_k)$. In particular, $P \in \Upsilon(a_j) \cap \Upsilon(a_{j+1})$. Therefore $A_j \cap B_j = \Upsilon(a_j) \cap \Upsilon(a_{j+1})$. By Proposition 4.1, the dimension of $\Upsilon(a_j) \cap \Upsilon(a_{j+1})$ is at most $n-1$, so that there is an hyperplane $\mathcal{H}$ that contains $A_j \cap B_j$. Suppose that there exists a vector P of $\mathcal{H}$ that does not belong to $A_j \cap B_j$. Since $A_j \cup B_j = \Delta(\Omega)$, $P \in A_j$ or $P \in B_j$, for example $P \in A_j$. Then there exists a vector Q in the interior of B_j with $Q \notin \mathcal{H}$. Note that the segment $]P,Q]$ contains only vectors of A_j or B_j. Since both sets are closed, the segment intersects $A_j \cap B_j$, which is impossible since the $]P,Q]$ does not intersect $\mathcal{H}$. This means that $\mathcal{H} = A_j \cap B_j = \Upsilon(a_j) \cap \Upsilon(a_{j+1})$. $\square$

When the question Υ admits an accuracy-rewarding payoff (with respect to a strict total order), we can get a more precise description of the (strictly) proper payoffs. So let's assume that Υ is directly implementable with an accuracy-rewarding payoff function. Let n_i be a normal to the hyperplane $\Upsilon(a_i) \cap \Upsilon(a_{i+1})$ that is positively oriented, that is, such that $\langle n_i, P\rangle = 0$ for all $P \in \Upsilon(a_i) \cap \Upsilon(a_{i+1})$, and $\langle n_i, P\rangle \geq 0$ for all $P \in \Upsilon(a_{i+1})$. For $1 \leq i < m$, let's define the payoffs

$$\hat{\Pi}_i(a_j,\omega) = \begin{cases} 0 & \text{if } j \leq i \ , \\ n_i(\omega) & \text{if } j > i \ . \end{cases}$$

We prove the following:

THEOREM 5.2. *The set of payoffs* $\{\hat{\Pi}_1, \ldots, \hat{\Pi}_{m-1}\}$ *is a base of* Υ.

PROOF. <u>STEP 1.</u> Let's define, for all i, $\Pi(a_i) = \Pi_0 + \sum_{1\leq j<i} \lambda_j n_j$, with $\lambda_1, \ldots, \lambda_{m-1} \geq 0$.

Since n_i is oriented positively, $\langle n_i, P\rangle \geq 0$ for all $P \in \Upsilon(a_{i+1}), \ldots, \Upsilon(a_m)$ with a strict inequality if $P \notin \Upsilon(a_i)$, and $\langle n_i, P\rangle \leq 0$ for all $P \in \Upsilon(a_1), \ldots, \Upsilon(a_i)$ with a strict inequality if $P \notin \Upsilon(a_{i+1})$.

Let $P \in \Upsilon(a_i)$. If $j < i$,

$$\langle \Pi(a_i), P\rangle - \langle \Pi(a_j), P\rangle = \sum_{j\leq k<i} \lambda_k \langle n_k, P\rangle \geq 0 \ ,$$

and, if $j > i$,

$$\langle \Pi(a_i), P\rangle - \langle \Pi(a_j), P\rangle = -\sum_{i\leq k<j} \lambda_k \langle n_k, P\rangle \geq 0 \ ,$$

the inequalities being strict when $P \notin \Upsilon(a_j)$ and the scalars $\lambda_1, \ldots, \lambda_{m-1} > 0$. Therefore the payoff is incentive compatible, and strictly proper if the scalars λ_k are strictly positive.

<u>STEP 2.</u> Now assume Π is a proper payoff. Then, for all $P \in \Upsilon(a_i) \cap \Upsilon(a_{i+1})$, $1 \leq i < m$, $\langle \Pi(a_i), P\rangle = \langle \Pi(a_{i+1}), P\rangle$, and so $\langle \Pi(a_{i+1}) - \Pi(a_i), P\rangle = 0$. Since by Theorem 5.1 $\Upsilon(a_i) \cap \Upsilon(a_{i+1})$ is an hyperplane of $\Delta(\Omega)$, its linear span is an hyperplane $\mathcal{H}_i$ of $\mathfrak{R}^\Omega$. Therefore, $\Pi(a_{i+1}) - \Pi(a_i) = \lambda_i n_i$, where n_i is a normal to $\mathcal{H}_i$ oriented positively, i.e., such that for all $P \in \mathcal{H}_i$, $\langle n_i, P\rangle = 0$, and $\langle n_i, P\rangle \geq 0$ if $P \in \Upsilon(a_{i+1})$.

Let $P \in \Upsilon(a_{i+1})$, $P \notin \Upsilon(a_i)$, then $\langle \Pi(a_{i+1}), P\rangle \geq \langle \Pi(a_i), P\rangle$, and so $\lambda_i \langle n_i, P\rangle \geq 0$. Since $P \notin \mathcal{H}_i$ and n_i is positively oriented, $\langle n_i, P\rangle > 0$, which implies $\lambda_i \geq 0$ ($\lambda_i > 0$ with strict properness). Therefore $\Pi(a_i) = \Pi(a_1) + \sum_{1\leq j<i} (\Pi(a_{j+1}) - \Pi(a_j)) = \Pi_0 + \sum_{1\leq j<i} \lambda_j n_j$ with $\Pi_0 = \Pi(a_1)$, which concludes the proof. $\square$

Naturally we also wish to obtain accuracy-rewarding payoffs. This is easily done, as they are exactly the strictly proper payoffs as stated in our next theorem, easily proved using the derivation in Step 1 of our above proof.

THEOREM 5.3. If *the payoff* Π *is strictly proper,* then *it is accuracy-rewarding.*

PROOF. Assume Π is strictly proper. Then, by Theorem 5.2,

$$\Pi(a_i) = \Pi_0 + \sum_{1\leq k<i} \lambda_k n_k \ ,$$

with $\lambda_1, \ldots, \lambda_{m-1} > 0$. Let $P \in \Delta(\Omega)$. Since the normals are positively oriented, $\langle n_i, P\rangle > 0$ if $a_i \prec \Upsilon^{-1}(P)$, and $\langle n_i, P\rangle < 0$ if $\Upsilon^{-1}(P) \prec a_i$.

Similarly to the derivation of the proof of Theorem 5.2, if $a_j \prec a_i \prec \Upsilon^{-1}(P)$, then

$$\mathop{\mathbb{E}}_{\omega\sim P}[\Pi(a_i,\omega)] - \mathop{\mathbb{E}}_{\omega\sim P}[\Pi(a_j,\omega)] = \sum_{j\leq k<i} \lambda_k \langle n_k, P\rangle > 0 \ .$$

Similarly, if $\Upsilon^{-1}(P) \prec a_i \prec a_j$, then

$$\mathop{\mathbb{E}}_{\omega\sim P}[\Pi(a_i,\omega)] - \mathop{\mathbb{E}}_{\omega\sim P}[\Pi(a_j,\omega)] = -\sum_{i\leq k<j} \lambda_k \langle n_k, P\rangle > 0 \ .$$

Therefore Π is accuracy-rewarding. $\square$

One interesting application of our results concerns the elicitation of *distribution properties*. Introduced by Lambert et al. [7], distribution properties represent numerical parameters of the distribution of outcomes. They are defined as functions that associate a real value to each distribution. The authors impose two important restrictions on distribution properties: they must be *continuous* and *nowhere locally constant*. The authors characterize the properties that can be elicited truthfully with appropriate payoffs, such properties are said to be *directly elicitable*. As questionnaires with multiple-choice questions only allow a finite number of possible responses, they cannot be used to elicit distribution properties. However, they can be used to elicit ranges of property values, and thus give approximate values of distribution properties.

Formally, let $\Gamma : \Delta(\Omega) \mapsto \mathfrak{R}$ be a distribution property that is continuous and nowhere locally constant, with a range of admissible values $(\underline{\alpha}, \overline{\alpha})$. To divide the set of admissible values into subintervals, let $\alpha_1, \ldots, \alpha_{k+1}$ be any sequence of reals satisfying $\underline{\alpha} = \alpha_1 < \alpha_2 < \cdots < \alpha_{k+1} = \overline{\alpha}$. Let $I_i = [\alpha_i, \alpha_{i+1}]$. We consider the MCQ Υ_p:*"Which interval among $I_1, \ldots, I_k$ contains the true value of the distribution property Γ?"*. Here the set of possible answers is the set of intervals $\{I_i, 1 \leq i \leq k\}$, and an interval I is a valid answer for an event with distribution P when $\Gamma(P) \in I$.

The boundaries, sizes, and number of subintervals can be chosen freely and depend on the purpose of the question. It is often convenient to choose a non-uniform division, and have a larger number of small subintervals in the regions of property values of particular interest. We recall, by Theorem 2 of Lambert et al., that if Γ is directly elicitable then $\Gamma^{-1}(x)$ is an hyperplane in the simplex of distributions for all admissible values x. The following result gives the appropriate payoffs to be used with Υ_p.

THEOREM 5.4. If Γ *is directly elicitable,* then Υ_p *is directly implementable, and* Π *is a proper (resp. strictly proper) payoff* if and only if

$$\Pi(I_i, \omega) = \Pi_0(\omega) + \sum_{j<i} \lambda_j n_j(\omega)$$

for any $\Pi_0 : \Omega \mapsto \mathfrak{R}$ *and any* $\lambda_1, \ldots, \lambda_{k-1} \geq 0$ *(resp. any* $\lambda_1, \ldots, \lambda_{k-1} > 0$*), where for all* j, n_j *is a positively oriented normal to the hyperplane* $\Gamma^{-1}(\alpha_j)$. [1]

PROOF. This is a direct application of Theorems 5.1 and 5.2. We observe that, if Γ is directly elicitable then, by Lemma 3 and Theorem 2 of Lambert et al.[7], $\Gamma^{-1}(\alpha_{i+1}) = \Upsilon(I_i) \cap \Upsilon(I_{i+1})$ is an hyperplane of $\Delta(\Omega)$. □

6. IMPLIED QUESTIONNAIRES

When a multiple-choice question is not directly implementable, no payoff yields a truthful answer. Fortunately it is sometimes possible to answer correctly an MCQ from the solution of another one. This creates a particular relation between multiple-choice questions (and more broadly between questionnaires) that we call *implication*. Intuitively, an MCQ implies another when one can provide a valid answer to the latter if one can answer correctly the former. Recall that the notion of implication and the results of this section naturally generalize to questionnaires with multiple MCQ by Lemma 2.1.

[1]Such that $\langle P, n_j \rangle \geq 0$ for all P with $\Gamma(P) > \alpha_j$.

Definition 6. An MCQ Φ *implies* an MCQ Υ if, for all possible answer b of Φ, there exists an answer a of Υ such that $\Phi(b) \subseteq \Upsilon(a)$.

If a multiple-choice question Φ implies a question Υ, we write $\Phi \Longmapsto \Upsilon$. It is easily shown that implication defines a partial order relation on question functions.

The concept of implication becomes useful when an MCQ that does not admit a strictly proper payoff is implied by an MCQ that admits one. In such circumstances, one can simply use the directly implementable question to infer the answer of the non-directly implementable one. In the two theorems that follow, we show that those situations do exist, and we give a characterization of the multiple-choice questions that can be implied by a directly implementable question.

THEOREM 6.1. If *an MCQ* Υ *is such that, for all possible answers* a *of* Υ, $\Upsilon(a)$ *is a closed polyhedron,* then *there exists a directly implementable MCQ that implies* Υ*.*

PROOF. Let $\{\mathcal{H}_1, \ldots, \mathcal{H}_\ell\}$ the set of hyperplanes that define the facets of the polyhedra $\Upsilon(a)$, $a \in \mathcal{A}$. Let n_i be a normal to $\mathcal{H}_i$. The hyperplanes divide $\Delta(\Omega)$ into cells $\{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$. Each cell $\mathcal{C}_j$ is a convex polyhedron that can be defined by the inequalities

$$s_i^j \langle n_i, P \rangle \geq 0 \ ,$$

for $i = 1, \ldots, \ell$, where for all i, either $s_i^j = 1$ or $s_i^j = -1$. Let Φ be the MCQ with a set of answers $\{1, \ldots, M\}$, defined by $\Phi(k) = \mathcal{C}_k$. We observe that $\Phi(k)$ is included in one of the polyhedra whose facets define the hyperplanes $\mathcal{H}_1, \ldots, \mathcal{H}_\ell$ and so Φ implies Υ.

We construct a strictly proper payoff Π as follows:

$$\Pi(k) = \sum_i s_i^k n_i \ ,$$

for all k. Let $k \neq k'$, and $P \in \mathcal{C}_k$. Then,

$$\langle \Pi(k), P \rangle - \langle \Pi(k'), P \rangle = 2 \sum_{i / s_i^k \neq s_i^{k'}} s_i^k \langle n_i, P \rangle \geq 0 \ ,$$

which implies properness. Furthermore, if $P \notin \mathcal{C}_{k'}$, then there exists i_0 with $s_{i_0}^{k'} \langle n_{i_0}, P \rangle < 0$. Then,

$$\begin{aligned} \langle \Pi(k), P \rangle - \langle \Pi(k'), P \rangle &= s_{i_0}^k \langle n_{i_0}, P \rangle - s_{i_0}^{k'} \langle n_{i_0}, P \rangle \\ &\quad + 2 \sum_{i \neq i_0 / s_i^k \neq s_i^{k'}} s_i^k \langle n_i, P \rangle \ , \\ &\geq s_{i_0}^k \langle n_{i_0}, P \rangle - s_{i_0}^{k'} \langle n_{i_0}, P \rangle \ , \\ &\geq -s_{i_0}^{k'} \langle n_{i_0}, P \rangle \ , \\ &> 0 \ , \end{aligned}$$

which gives strict properness. □

THEOREM 6.2. *There exists a directly implementable MCQ that implies* Υ if and only if Υ *is implied by some MCQ* Φ *that is such that for all possible answers* b, $\Phi(b)$ *is a closed polyhedron.*

PROOF. IF PART. If an MCQ Φ implies Υ, and is such that, for all answers b, $\Phi(b)$ is a closed polyhedron, then by Theorem 6.1, there exists a directly implementable MCQ that implies Φ, and so that implies Υ by transitivity.

ONLY IF PART. If there exists a directly implementable MCQ Φ that implies Υ, then by Proposition 4.1, $\Phi(b)$ is a closed polyhedron for all possible answers b of Φ. $\square$

It is interesting to observe that not every question can be implied by a directly implementable one. In fact, those situations are quite rare: any multiple-choice question whose geometric configuration, in the simplex of distributions, is such that the intersection between the cells of two answers forms a curvy hypersurface can never be implied by a directly implementable question. For example, the MCQ *"Is it true that the variance of X is no less than the variance of Y? (a) Yes or (b) No?"* for given random variables X and Y is not implied by any directly implementable MCQ, since the intersection between the cells of the two answers is given by the equation $\mathrm{Var}_P(X) = \mathrm{Var}_P(Y)$, quadratic in P.

7. EXAMPLES

In this section we apply our results to concrete examples of simple multiple-choice questions. Many of our questions concern the values taken by one or more random variables, represented as X (single random variable) or a vector $\mathbf{X}$ (multiple random variables). In such cases, it is generally assumed that one does not observe the true outcome but only the values taken by the random variables, so that we can assume without loss of generality that each outcome corresponds exactly to one value of X (or $\mathbf{X}$), and for simplicity we write payoffs as $\Pi(a, X)$ (or $\Pi(a, \mathbf{X})$).

7.1 Finding the median/quantiles

Let X be a random variable with a finite range of values $\mathcal{X}$, that contains at least 2 elements. We consider an MCQ to elicit a median of X: *"What is a median of X?"*, taking as possible answers the possible medians, that is, the full range of X. Formally, the question function Υ is defined by $\Upsilon(x) = \{P \in \Delta(\Omega) \mid P(X \le x) \ge 1/2, P(X \ge x) \ge 1/2\}$, for all $x \in \mathcal{X}$.

PROPOSITION 7.1. *This MCQ is directly implementable, and Π is a proper (resp. strictly proper) payoff* if and only if

$$\Pi(x, X) = \Pi_0(X) + \frac{1}{2} f(x) + \begin{cases} f(X) - f(x) & \text{if } X \le x \\ 0 & \text{if } X > x \end{cases}$$

with any function $\Pi_0 : \mathcal{X} \mapsto \Re$ and any nondecreasing (resp. strictly increasing) function $f : \mathcal{X} \mapsto \Re$. Moreover, if f is strictly increasing, then Π is also accuracy-rewarding with respect to the strict total order $<$ on the real line.

Our result can be obtained by application of Theorems 4.2, 5.1, 5.2 and 5.3.

We can easily generalize the above theorem to elicit an α-quantile of X. An α-quantile is any $x \in \mathcal{X}$ satisfying $P(X \le x) \ge \alpha$ and $P(X \ge x) \ge 1 - \alpha$. The MCQ *"What is an α-quantile of X?"* is directly implementable, and Π is a proper (resp. strictly proper) payoff if and only if

$$\Pi(x, X) = \Pi_0(X) + \alpha f(x) + \begin{cases} f(X) - f(x) & \text{if } X \le x \\ 0 & \text{otherwise} \end{cases}$$

with $\Pi_0 : \mathcal{X} \mapsto \Re$ and $f : \mathcal{X} \mapsto \Re$ nondecreasing (resp. strictly increasing).

7.2 Finding the mode

Under the same setting as above, we consider the question *"What is a mode for X?"*, where an answer can be any value of X. We recall that the mode of a random variable is its more frequent value. The corresponding MCQ function is defined by $\Upsilon(x) = \{P \in \Delta(\Omega) \mid P(X = x) \ge P(X = y)\, \forall y \in \mathcal{X}\}$, for all $x \in \mathcal{X}$.

PROPOSITION 7.2. *This MCQ is directly implementable, and Π is a proper (resp. strictly proper) payoff* if and only if

$$\Pi(x, X) = \Pi_0(X) + \lambda \begin{cases} 1 & \text{if } X = x \\ 0 & \text{if } X \ne x \end{cases}$$

for any function $\Pi_0 : \mathcal{X} \mapsto \Re$ and any $\lambda \ge 0$ (resp. $\lambda > 0$).

Note that, by applying Theorem 5.1, one can easily show that the MCQ does *not* admit accuracy-rewarding payoffs.

7.3 Estimating the probability of a binary event

Let $A \subset \Omega$ be a binary event. We divide the interval of possible probabilities $[0, 1]$ into intervals $[\alpha_i, \alpha_{i+1}]$, with $\alpha_i < \alpha_j$ when $i < j$. Let $I_i = [\alpha_i, \alpha_{i+1}]$, and m be the total number of intervals. We consider the multiple-choice question *"What is the range of A's probability?"*, with as set of possible answers the intervals I_i. Any interval I that contains the true probability of A is considered valid.

PROPOSITION 7.3. *This MCQ is directly implementable, and Π is a proper (resp. strictly proper) payoff* if and only if

$$\Pi(I_i, \omega) = \Pi_0(\omega) + \sum_{j<i} \lambda_j \begin{cases} 1 - \alpha_j & \text{if } \omega \in E \\ -\alpha_j & \text{if } \omega \notin E \end{cases}$$

for any function $\Pi_0 : \Omega \mapsto \Re$ and any $\lambda_1, \ldots, \lambda_{m-1} \ge 0$ (resp. any $\lambda_1, \ldots, \lambda_{m-1} > 0$).

7.4 Finding the variable of largest mean

Let $X_1, \ldots, X_k$, $k > 1$, taking values in a set $\mathcal{X}$ that contains at least 2 elements. We consider the MCQ *"What is the random variable whose mean is the largest?"*. Here answers consist of the indices of the random variables, $\{1, \ldots, k\}$.

PROPOSITION 7.4. *This MCQ is directly implementable, and Π is a proper (resp. strictly proper) payoff* if and only if

$$\Pi(i, \mathbf{X}) = \Pi_0(\mathbf{X}) + \lambda X_i$$

for any function $\Pi_0 : \mathcal{X}^k \mapsto \Re$ and any $\lambda \ge 0$ (resp. $\lambda > 0$).

7.5 Ranking the likelihoods of binary events

Let $A_1, \ldots, A_k$ be k binary events (not necessarily independent). We now consider the MCQ *"Rank the events $A_1, \ldots, A_k$ from the most to the least likely"*. In this case the set of answers is the set of permutations of the indices $\{1, \ldots, k\}$ of the events, each answer σ being interpreted as an ordering of the events, where $\sigma(i)$ is the index of the i-th most likely event. An answer σ is correct when, for all $i \le j$, $P(A_{\sigma(i)}) \ge P(A_{\sigma(j)})$. Let X_i be the indicator of A_i.[2]

[2] I.e., $X_i(\omega) = 1$ if $\omega \in A_i$ and $X_i(\omega) = 0$ if $\omega \notin A_i$.

PROPOSITION 7.5. *This MCQ is directly implementable, and the following payoff function*

$$\Pi(\sigma, \mathbf{X}) = \lambda_0 + \sum_i \lambda_i X_{\sigma(i)}(\omega)$$

is strictly proper for any $\lambda_0 \in \Re$, *and any reals* $\lambda_1 > \lambda_2 > \cdots > \lambda_k > 0$.

7.6 Eliciting confidence intervals

Given a random variable X taking values in a set $\mathcal{X}$ with at least 2 elements, we consider the multiple-choice question *"What is a 90% confidence interval for X?"*. The set of answers is the set of intervals $[a, b]$, where $a \leq b$ are two possible values for X. $[a, b]$ is a 90% confidence interval when $P(a \leq X \leq b) \geq 0.9$, $P(a < X \leq b) \leq 0.9$ and $P(a \leq X < b) \leq 0.9$. Unfortunately, this MCQ does not satisfy the thin intersection condition of Proposition 4.1, and so is not directly implementable.

PROPOSITION 7.6. *The MCQ on general confidence intervals is not directly implementable.*

We now restrict ourselves to symmetric confidence intervals with the MCQ *"What is a 90% symmetric confidence interval for X?"*. A 90% symmetric confidence interval for X is any interval $[a, b]$ that contains X with at least 90% probability, and such that the likelihood of X being at least b, or at most a, is at least 5%.

Observing that $[a, b]$ is a valid answer if and only if a is a 0.05-th quantile and b is a 0.95-th quantile, we easily get a family of (strictly) proper payoff functions as a sum of the two (strictly) proper payoffs corresponding to each quantile.

PROPOSITION 7.7. *The MCQ on symmetric confidence intervals is directly implementable, and the payoff functions*

$$\Pi([a,b], X) = \Pi_0(X) + \frac{5}{100} f(a) + \frac{95}{100} g(b) + \begin{cases} f(X) + g(X) - f(a) - g(b) & \text{if } X \leq a \\ g(X) - g(b) & \text{if } a < X \leq b \\ 0 & \text{if } X > b \end{cases}$$

are proper (resp. strictly proper) for any $\Pi_0 : \mathcal{X} \mapsto \Re$ *and any nondecreasing (resp. strictly increasing) functions* $f : \mathcal{X} \mapsto \Re$ *and* $g : \mathcal{X} \mapsto \Re$.

8. CONCLUSION AND FUTURE WORK

We studied the problem of incentivizing an expert to truthfully answer questionnaires with multiple-choice questions. Observing that the study of complete questionnaires reduces to that of a single multiple-choice question, we gave simple necessary conditions for questions to be directly implementable, and showed that directly implementable questions correspond to power diagrams in the simplex of distributions. We also showed that the proper (resp. strictly proper) payoffs are nonnegative (resp. positive) mixtures of a fixed, finite number of particular payoff functions. We then considered the special case of questions with ordered answers. We proposed a simple characterization of questions directly implementable with accuracy-rewarding payoffs and described those payoffs. Next we introduced the concept of implication between questionnaires; we gave necessary and sufficient conditions for a questionnaire to be implied by a directly implementable questionnaire. Finally we illustrated our results with several examples of practical interest. Also note that our analysis can be applied to a prediction market setting, our results appear in the full version of the paper.

It would be interesting, for future work, to validate experimentally our results. Such exercise is facilitated by the growth of online survey systems that can implement complex payoff functions, such as Amazon Mechanical Turk. [3] From a more theoretical perspective, it would be interesting to explore other types of questionnaires that can be created with online web services; for example, *dynamic* questionnaires, where questions displayed may depend on answers of previous questions, or *smooth* questionnaires in which, instead of selecting exactly one answer and rejecting all others, experts specify a degree of correctness for each answer, interpreted as the perceived probability for each answer being correct, etc.

Acknowledgements

The authors would like to thank Mohammad Mahdian and David Pennock for helpful discussions and suggestions, and the anonymous reviewers for their useful comments.

9. REFERENCES

[1] F. Aurenhammer. Power Diagrams: Properties, Algorithms and Applications. *SIAM Journal on Computing*, 16:78, 1987.

[2] F. Aurenhammer. Voronoi Diagrams: a Survey of a Fundamental Geometric Data Structure. *ACM Computing Surveys*, 23(3), 1991.

[3] J. Cervera and J. Munoz. Proper Scoring Rules for Fractiles. *Bayesian Statistics*, 5:513–519, 1996.

[4] M. De Berg, O. Cheong, and M. van Kreveld. *Computational Geometry: Algorithms and Applications*. Springer, 2008.

[5] I. Eremin. *Theory of Linear Optimization*. VSP, 2002.

[6] H. Imai, M. Iri, and K. Murota. Voronoi Diagram in the Laguerre Geometry and Its Applications. *SIAM Journal on Computing*, 14:93, 1985.

[7] N. Lambert, D. Pennock, and Y. Shoham. Eliciting Properties of Probability Distributions. In *Proceedings of the ninth ACM Conference on Electronic Commerce*, pages 129–138, 2008.

[8] J. Richter-Gebert, B. Sturmfels, and T. Theobald. First Steps in Tropical Geometry. *Arxiv preprint math.AG/0306366*, 2003.

[9] L. Savage. Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[10] C. Smith. Personal Probability and Statistical Analysis. *Journal of the Royal Statistical Society. Series A (General)*, 128(4):469–499, 1965.

[11] R. Winkler, J. Muñoz, J. Cervera, J. Bernardo, G. Blattenberger, J. Kadane, D. Lindley, A. Murphy, R. Oliver, and D. Ríos-Insua. Scoring Rules and the Evaluation of Probabilities. *TEST*, 5(1):1–60, 1996.

[3] `http://aws.amazon.com/mturk/`