Complex Questionnaires

Jacob Glazer
Tel Aviv University and
The University of Warwick
and
Ariel Rubinstein
Tel Aviv University and
New York University

Abstract

We study a principal-agent model in which the agent is boundedly rational in his ability to understand the principal's decision rule. The principal wishes to elicit an agent's true profile in order to determine whether or not to grant him a certain request. The principal designs a questionnaire and commits himself to accepting certain responses. In designing such a questionnaire, the principal takes into account the bounded rationality of the agent and wishes to reduce the success probability of a dishonest agent who is trying to game the system. It is shown that the principal can construct a sufficiently complex questionnaire that will allow him to respond optimally to agents who tell the truth and at the same time to almost eliminate the probability that a dishonest agent will succeed in cheating.

Keywords: Bounded Rationality, persuasion games, questionnaires.
JEL: Classification D0

The second author acknowledges financial support in the form of ERC grant 269143. Our thanks to Noga Alon, Gil Kalai, Xiaosheng Mu, Neil Thakral and Tal Rabin for their advice on combinatorial issues.

1. **Introduction**

In many principal-agent situations, a principal makes a decision based on information provided to him by an agent. Since the agent and the principal do not necessarily share the same objectives, the principal cannot simply ask the agent to provide him with the relevant information (hereafter referred to as the agent’s profile). He instead must utilize an additional tool in order to induce the agent to provide accurate information. The economic literature has focused on two such tools: verification (requiring the agent to present hard evidence) and incentives (rewarding or penalizing the agent on the basis of the information he provides). However, these tools are often prohibitively expensive or insufficient to achieve the task.

The purpose of this paper is to analyze a different type of tool that can be used by a principal to reduce the probability of an agent cheating successfully. Instead of asking the agent direct questions to elicit the relevant information, the principal can design a sufficiently "complex" questionnaire such that a boundedly rational agent who is considering lying will find it difficult to come up with consistent answers that will induce the principal to take an action desired by the agent.

The analysis is carried out in the context of a simple persuasion model. A principal interacts on a routine basis with many different agents who present him with requests. In each case, the principal must decide whether or not to accept the request. He would like to accept the request if and only if the agent’s profile meets certain conditions, whereas the agent would like his request to be accepted regardless of his true profile. The agent’s profile is known only to himself and cannot be verified by the principal. In order to obtain the information he needs, the principal designs a questionnaire for the agent which contains a set of yes/no questions regarding his profile. The principal accepts the agent’s request if the agent’s response to the questionnaire (i.e., the list of answers he provides) is included within a set of acceptable responses.

At the core of our model are assumptions regarding the procedure used by a boundedly rational agent who instead of answering the questionnaire honestly attempts to come up with a response that will be accepted. We assume that the agent does not know (or does not fully understand) the principal’s policy (i.e., which responses to the questionnaire will be accepted). However, the agent can detect (or is able to understand or is informed of) certain interdependencies between the answers to the various questions in the set of acceptable responses. We refer to such an interdependency as a "regularity". An agent is characterized by the "level" of regularities he can detect. The most boundedly rational agent (an agent of level 0) is only able to determine whether an answer to a particular question must be positive or negative. An agent of level d will be able to determine whether, within the set of acceptable responses, an answer to a set of d questions uniquely determines the answer to an additional question.

Note that we assume the agents can detect regularities in the set of acceptable responses but cannot imitate any particular acceptable response. What we have in mind is that the agent perceives the set of acceptable responses in an analogous way to how a person views a picture of an orchard during fruit picking season. An unsophisticated observer will only be able to see that the picture is green. A more observant individual will notice that the pixels form the shapes of trees. A really astute individual will notice that next to each tree with fruit on it, there is a person with a ladder. Even the most observant individuals, however, will not be able to draw or recall even a tiny part of the picture later on.

The principal’s goal in designing the questionnaire is twofold: his first priority is to make the right decision (from his point of view) when an agent answers the questionnaire honestly. His second priority is to minimize the acceptance probability of a dishonest agent who has abandoned his true profile and, based on the regularities he detects in the set of acceptable responses, tries to guess an acceptable answer. We demonstrate that a complex questionnaire can serve as a tool for the principal to achieve these two goals. The principal’s optimal questionnaire depends on the agent’s level of bounded rationality. The more boundedly rational the agent is, the lower will be the probability that he will succeed in dishonestly responding to the optimal questionnaire.

Following the construction and discussion of the model, we prove two main results: (i) if the principal uses an optimal questionnaire, a dishonest agent’s ability to come up with an acceptable answer depends only on the size of the set of profiles that the principal wishes to accept and (ii) when the set of acceptable profiles is large the principal can design a questionnaire that will reduce to almost zero the probability of a dishonest agent cheating effectively.

2. **The model**

The principal and the agent

The agent possesses private information, referred to as his true *profile*, in the form of an element ω in a finite set Ω. The principal needs to choose between two actions: a (accept) and r (reject). The agent would like the principal to choose the action a, regardless of his true profile. The principal's desired action depends on the agent's true profile: he wishes to choose a if the agent's profile belongs to a set A, a proper subset of Ω, and to choose r if the profile is in $R = \Omega - A$. Denote the size of A by n. A persuasion problem is a pair (Ω, A).

A questionnaire

A questionnaire is a (multi)set of "questions". Each question is of the form "Does your profile belong to the set q?" where $q \subseteq \Omega$. We will denote the question according the set which the question asks about.

The agent responds to each question with a "Yes" (1) or a "No" (0). The principal does not know the agent's profile and cannot verify any of the answers given by him.

Following are two examples of questionnaires:

(i) The *one-click* questionnaire which consists of $|\Omega|$ questions of the form $\{\omega\}$. That is, each question asks whether the agent has a particular profile.

(ii) Let $\Omega = \{0,1\}^K$. A profile contains information about K relevant binary characteristics. *The simple questionnaire* consists of K questions, each of which asks about a distinct characteristic, i.e. $q_k = \{\omega \mid \omega_k = 1\}$.

A *response* to *a* questionnaire Q is a function that assigns of a value of 1 or 0 to each question in Q. It will sometimes be convenient to order the questions in Q, i.e. $(q_1,\ldots,q_L)$, and to identify a response using an L-vector of zeroes and ones. Let $\Theta(Q)$ be the set of all possible responses to Q. Let $\theta(Q,\omega)$ be the response to Q given by an honest agent whose profile is ω, i.e. the vector of length L whose i'th component is 1 if $\omega \in q_i$ and 0 otherwise.

For every A and Q, define the following three sets:

(i) $\Theta(Q,A) = \{\theta(Q,\omega) |\ \omega \in A\}$ (the set of honest responses given by agents whose profiles are in A);

(ii) $\Theta(Q,R) = \{\theta(Q,\omega) | \ \omega \in \Omega - A\}$ (the set of honest responses given by agents whose profiles are in R); and

(iii) $Inconsistent(Q) = \Theta(Q) - \{\Theta(Q,\omega) | \ \omega \in \Omega\}$ (the set of responses that are not given by any honest agent).

We say that a questionnaire Q identifies A if, when all agents are honest, the responses of the agents whose profiles are in A differ from the responses of the agents whose profiles are in R (that is, $\Theta(Q,A) \cap \Theta(Q,R) = \emptyset$). The "one-click" questionnaire (as well as the simple questionnaire) identifies any set A since any two profiles induce two different responses.

An agent does not know the set of acceptable responses. We assume that he is either: (i) "honest" in the sense that he "automatically" tells the truth or (ii) a "manipulator" who, regardless of his true profile, tries to respond to the questionnaire successfully after learning some properties of the set of acceptable responses.

We assume that the principal's first priority is to accept honest agents whose profile is in A and to reject all others. In other words, he seeks a questionnaire that identifies A and adheres to a policy of accepting a response if and only if it is in $\Theta(Q,A)$. The principal's second priority is to design a questionnaire that makes it less likely for a manipulator to come up with an acceptable answer.

The Bounded Rationality Element

At the core of our model is the element of bounded rationality. Were a manipulative agent fully aware of the set of acceptable responses, $\Theta(Q,A)$, he would always choose an acceptable response and the principal would be helpless. However, we assume that an agent is limited in his ability to figure out the set $\Theta(Q,A)$ and does not have any prior beliefs on it. In the spirit of the set theoretic model of knowledge, we assume that an agent detects certain types of regularities in the set. By *regularity,* we are referring to a sentence (in the language of propositional logic with the variables being the names of the questions in Q) that is true in $\Theta(Q,A)$. The agent detects regularities but is not able to cite any particular acceptable response. This phenomenon is common in real life. For example, the fact that we observe that all papers accepted to *Econometrica* contain formal models does not mean that we are able to cite any of them.

The set of regularities detected by an agent is characterized by a rank, which is an integer $d \geq 0$. An agent of rank d can recognize propositions of the form $\varphi_1 \rightarrow \varphi_2$ where the antecedent φ_1 is a conjunction of at most d clauses, each of which is a question or its negation, and the consequent φ_2 is a question (which does not appear in the antecedent) or its negation. We will refer to such a proposition as a *d-implication*. Given a questionnaire Q, an agent of rank d can figure out all the d-implications that are true for all responses in $\Theta(Q,A)$. Thus, an agent of rank 0 observes only regularities such as: "In all accepted responses, the answer to the question q is N" (denoted $-q$). An agent of rank 1 is also able to identify regularities of the type: "In all accepted responses, if the answer to q_1 is N then the answer to q_3 is Y" (denoted $-q_1 \rightarrow q_3$). The propositions $-q_1 \wedge -q_2 \rightarrow q_3$ constitute an example of a regularity of rank 2.

Let $\Theta_d(Q,A)$ be the set of responses that satisfy all the d-implications that are true for all responses in $\Theta(Q,A)$. By definition, $\Theta_d(Q,A) \supseteq \Theta_{d+1}(Q,A) \supseteq \Theta(Q,A)$ for all d.

We assume that if instead of responding honestly to the questionnaire, an agent of rank d is interested in gaming the system (i.e., coming up with a response in $\Theta(Q,A)$, regardless of his true profile), he will choose randomly from among the responses in $\Theta_d(Q,A)$. His probability of success is therefore: $\alpha_d(Q,A) = |\Theta(Q,A)|/|\Theta_d(A,Q)|$. Obviously, $\alpha_d(Q,A)$ is weakly increasing in d.

The principal's problem

As mentioned , the principal has two objectives in designing a questionnaire: His lexicographically first priority is to accept honest agents whose profile is in A and to reject all others. Hence, the questionnaire needs to identify A and the principal's policy should be to accept only responses given by honest agents whose profile is in A. His second priority is to minimize the probability that a manipulator will be able to successfully deceive him (i.e., the principal wishes to minimize $\alpha_d(Q,A)$). In other words, the principal's problem is:

$$min\{\alpha_d(Q,A) \mid Q \ identifies \ A\}.$$

The value of this optimization is denoted by $\beta_d(A)$.

Note that we are not following the standard mechanism design approach according to which the principal faces a distribution of agents' types and seeks a policy that maximizes the principal’s expected payoff.

Example 1:

Recall that the one-click questionnaire, $oneclick$, contains $|\Omega|$ questions (of the form $\{\omega\}$), one for each profile. The set $\Theta(oneclick,A)$ consists of all responses that assign the value 1 to precisely one question $\{\omega\}$ where $\omega \in A$.

An agent of rank 0 will learn to answer 0 to all the questions related to profiles in R. If A contains at least 2 profiles, the agent will learn nothing about how to respond to questions regarding profiles in A and thus $\alpha_0(Q,A) = n/2^n$ (where $n = |A|$).

An agent of rank 1 will, in addition, observe the regularities $\{\omega\} \to -\{\omega'\}$ where $\omega \in A$ and $\omega \neq \omega'$. For $n > 2$, the agent will not detect any additional regularities and therefore $\Theta_1(oneclick,A)$ consists of the set $\Theta(oneclick,A)$ and the "constant 0" response. Hence, $\alpha_1(oneclick,A) = n/(n+1)$. For $n = 2$, we have in addition $-\{\omega\} \to \{\omega'\}$ and therefore $\alpha_1(oneclick,A) = 1$.

Example 2:

We have in mind that a question is not necessarily phrased directly but rather in an equivalent indirect way as demonstrated in the following example:

A principal would like to identify scholars who are interested in at least two of the following three fields: Law, Economics and History. Thus, a profile can be presented as a triple of zeros and ones, indicating whether or not an agent is interested in each field ($\Omega = \{0,1\}^3$) and A is the set of the four profiles in which at least two characteristics receive the value 1.

The principal can simply ask the agent three questions:

1. Are you interested in Law?
2. Are you interested in Economics?
3. Are you interested in History?

This is formalized as the simple questionnaire $Q = \{q_1,q_2,q_3\}$ where q_i is the question about dimension i. The set of acceptable responses is $\Theta(Q,A) = \{(1,1,1),\ (1,1,0),\ (0,1,1),\ (1,0,1)\}$. The set $\Theta(Q,R)$ consists of all other possible responses.

An agent with $d = 0$ cannot detect any regularity in the set of acceptable responses since interest in any particular field or lack thereof is not a necessary requirement for a response to be accepted. That is, neither q nor $-q$ is true in $\Theta(Q,A)$. Thus, $\alpha_0(Q,A) = 1/2$.

An agent with $d = 1$ realizes that if he says he is not interested in one field then he should say that he is interested in the other two. That is, the 1-implications that are true in $\Theta(Q,A)$ are the six propositions $-q_j \rightarrow q_k$ where $j \neq k$. The set of responses that satisfy these six propositions ($\Theta_1(Q,A)$) is exactly $\Theta(Q,A)$. Thus, an agent with $d = 1$ will fully understand the set of acceptable responses, i.e., $\alpha_1(Q,A) = 1$.

Suppose that instead of asking these three questions, the principal uses the following questionnaire:

1. Are you familiar with the book "Sex and Reason"?
2. Are you familiar with the book "The Book Club Murder"?
3. Are you familiar with the book "Which Road to the Past?"?

The first book was written by Richard Posner, a leading figure in Law and Economics. The second book was written by Lawrence Friedman, a well-known scholar who bridges between Law and History. The author of the third book is the prominent economic historian Robert Fogel. Thus, each book spans two of the three fields. For example, a scholar will be familiar with "Sex and Reason" if and only if he is interested in both Law and Economics.

Notice that the acceptable responses to this questionnaire are either "three yes’s" or "a single yes". An agent with $d = 1$ cannot detect whether an answer of Yes or No to one question implies anything about the other two.

Formally, let Q' be the questionnaire $\{q_{12},\ q_{13},\ q_{23}\}$ where q_{ij} asks whether the i’th and j’th characteristics have the value 1, i.e. $q_{ij} = \{\omega \mid \omega_i = \omega_j = 1\}$. The questionnaire Q' identifies A as $\Theta(Q',A) = \{(1,1,1),\ (1,0,0),\ (0,1,0),\ (0,0,1)\}$ and $\Theta(Q',R) = \{(0,0,0)\}$. No 1-implication is true in $\Theta(Q',A)$ and thus $\Theta_1(Q',A)$ contains all 8 possible responses and $\alpha_1(Q',A) = 1/2$. As we will see later, the principal can do even better and reduce this probability to $1/3$.

Notice that an agent with $d = 2$ realizes that any one of the four combinations of answers to q_{12} and q_{13} in the set of acceptable responses uniquely determines the answer to q_{23} and thus $\Theta_2(Q',A) = \Theta(Q',A)$ and $\alpha_2(Q',A) = 1$.

3. Comments on the Bounded Rationality Element

As always, when one departs from the model of the ultra-rational economic agent special assumptions are necessary. We believe that our model captures some interesting aspects of the situation we have in mind although there are other assumptions that could be made and which would also yield interesting results. In what follows, we discuss the assumptions made regarding the agent's bounded rationality.

a. **What does the agent see**? The agent focuses on the space of responses without being able to relate to the space of profiles. If he was capable of "inferring backwards" from the space of responses to the space of profiles, he could probably determine the set A and come up with an acceptable response to the questionnaire, as if he indeed possessed one of the profiles in A. Furthermore, since the agent does not relate to the space of profiles he is not capable of identifying inconsistent responses.

The question of whether a questionnaire can conceal the interest of the principal in differentiating between profiles in A and profiles in R depends on the language available to the principal when framing the questions. In example 2, the question q_{12} can be framed in two different ways: (i) "Are you interested in both Economics and Law?" and (ii) "Are you familiar with the book Sex and Reason?" The availability of the second option makes the second questionnaire more attractive as a tool to elicit the agent's information without hinting the agent regarding the principal's real interest.

b. **What does the agent notice in the set of acceptable responses**? Our key assumption is that the agent notices only certain regularities in the set of acceptable responses. A regularity of rank d is a dependency (within the set of acceptable responses) of the answer to one question on the answers to some d other questions. An agent with $d \geq 1$ is able to detect the regularity $q_1 \to q_2$ whenever such a regularity is true in the set $\Theta(Q, A)$. Notice that such a regularity is true even if there is no acceptable response to Q with a positive answer to q_1. An alternative assumption would be that the agent discerns such a regularity if an addition to it being logically true, there exists at least one acceptable response with affirmative answers to q_1 and q_2. For example, the regularity "all acceptable economists are theoreticians" is true if the acceptable set does not include any economists. However, under the alternative assumption, the agent would detect this regularity only if there exists one acceptable response containing an affirmative answer to the question "Are you an economist?".

Another plausible assumption would be that the agent can detect statistical correlations such as: "Among the acceptable responses, 80% of those who answer Yes to q_1 answered Yes to q_2 as well".

c. **What does the agent not notice**?

We assume that the regularities are observed in the set of acceptable responses but not in the set of rejected responses. This appears to a be reasonable assumption in cases where the agent notices information about agents whose request has been accepted (such as job candidates who have been hired), but not about those whose request has been rejected (those who didn't get hired).

Furthermore, the agent cannot ascertain whether his request will be accepted if his response satisfies a particular proposition. This is a reasonable assumption in situations where it is easier for people to observe that, for example, all "admitted students are males" than "all males who applied were admitted".

d. **An agent is not able to exactly imitate an acceptable profile**

Possession of information about the set of acceptable responses does not necessarily imply familiarity with any particular acceptable response that can be copied. For example, assume you want to sneak into a party that you were not invited to. If you are an agent with $d = 0$ who thinks that what you are wearing is relevant to getting into the party, you will notice that all guests are wearing military uniforms and therefore you will not arrive at the party in a business suit. If you are an agent with $d = 1$, you will also notice that everyone wearing a white uniform is also wearing a navy emblem and thus you will either not arrive in a white uniform or you will wear a navy emblem if you do. However, this does not mean that you know exactly what combination of uniforms, emblems and insignia will keep you from getting caught and it will be impossible for you to duplicate every detail of what any one of the admitted guests is wearing.

This is captured by our assumption that an agent is unable to exactly imitate an acceptable response even though he knows some regularities about the set of acceptable responses. This assumption is also appropriate in situations where the agent is able to obtain partial information from people who have access to the file of acceptable responses without he himself having access.

e. **Framing our model as a conventional model of knowledge**

The agent's problem can be framed as a standard model of knowledge if we define the set of "feasible states" as the set of all non-empty sets of responses. A state is interpreted as the set of acceptable responses used by the principal. We assume that the agent can only ascertain that certain responses are not acceptable. Thus, for example, he cannot determine that there are three acceptable responses or that in 60% of the acceptable responses to a certain question is Yes. Given this kind of knowledge, an agent of rank d is able to determine that the acceptable set of responses can be any non-empty subset of $\Theta_d(Q,A)$. If his prior does not discriminate between the responses, he will conclude that any response in $\Theta_d(Q,A)$ is equally likely to be accepted and that any response outside this set will be rejected.

4. Some Observations

The following claim embodies some simple observations about $\alpha_d(Q,A)$:

Claim 1:

(i) If a combination of answers to m questions in Q never appears in $\Theta(Q,A)$, then such a combination will not appear in any element of $\Theta_d(Q,A)$ for $d \geq m-1$. (For example, if the response of "yes to all" to the questions q_1, q_2 and q_3 does not appear in $\Theta(Q,A)$, then an agent with $d \geq 2$ will detect the regularity $q_1 \wedge q_2 \rightarrow -q_3$.)

(ii) If Q consists of m questions, then $\alpha_d(Q,A) \equiv 1$ for all $d \geq m-1$ (follows from (i)).

(iii) If the answer to q' is the same for all $\omega \in A$ (that is, if $q' \supseteq A$ or $-q' \supseteq A$), then $\alpha_d(Q,A) = \alpha_d(Q \cup \{q'\},A)$ for all d.

(iv) Suppose that Q is a questionnaire that identifies A. Let Q' be a questionnaire obtained from Q by replacing one of the questions $q \in Q$ with $-q$. Then, Q' identifies A and $\alpha_d(Q,A) = \alpha_d(Q',A)$ for all d.

Claim 2 states that the principal can limit himself to questionnaires which are covers of A (where a questionnaire Q is a cover of A if for all $q \in Q$, $q \subseteq A$ and $\cup_{q \in Q} q = A$) and that $\beta_d(A)$ depends only on the size of A (and not on $|\Omega|$).

Claim 2:

(i) If Q identifies A, then there exists a questionnaire Q', which is a cover of A, that identifies A and $\alpha_d(Q,A) = \alpha_d(Q',A)$ for all d.

(ii) $\beta_d(A)$ is a function of $n = |A|$ and is independent of $|\Omega|$.

Proof:

(i) Consider $b \in R$. Since Q identifies A then b's honest response to Q is different from that of any profile in A. By Claim 1(iv), we can assume that $b \notin q$ for all $q \in Q$, i.e., b's honest response to the questionnaire is a constant 0. Since the questionnaire identifies A, every element in A belongs to at least one $q \in Q$.

Now let Q' be the questionnaire $\{q \cap A$ | there exists $q \in Q\}$. Q' identifies A: a response to Q' by a profile outside of A is a constant 0; a profile in A belongs to at least one $q' \in Q'$ and thus Q' is a cover of A. The honest response of each profile in A to any $q \in Q$ is the same as its honest response to $q \cap A \in Q'$ and therefore, $\alpha_d(Q,A) = \alpha_d(Q',A)$.

(ii) By (i), we can assume that the optimal questionnaire is a cover of A and thus the size of R is immaterial for any $\alpha_d(Q,A)$. ■

Claim 3 states that the ability of the principal to prevent dishonest agents from successfully cheating depends on the relation between n and d. Thus, if $d \geq n-1$ then a dishonest agent will be able to fully game the system.

Claim 3: $\alpha_{n-1}(Q,A) = 1$ for all Q.

Proof: Let $\Theta(Q,A) = \{z^1,\ldots,z^m\}$, where $m \leq n$. The claim is trivial for the case of $m = 1$. Otherwise, we could (inductively) construct a set of $m-1$ questions in Q, such that for any profile in A an honest answer to these questions would determine the honest answers to all the others.

In the first stage, let q be a question for which $z^1(q) \neq z^2(q)$. Define $Q(1) = \{q\}$. In $\{z^1,z^2\}$, the answer to q determines the responses to all other questions in Q.

By the end of the $(t-1)$-th stage we have a set $Q(t-1)$ of at most $t-1$ questions such that in $\{z^1,..,z^t\}$ a response to these questions uniquely determines the responses to all the others.

In the t-th stage, consider z^{t+1}. If for every z^s ($s \leq t$) there is a question $q \in Q(t-1)$ such that $z^{t+1}(q) \neq z^s(q)$ (that is, a "signature" of z^{t+1} appears in the answers to $Q(t-1)$), then $Q(t) = Q(t-1)$. If for some $s \leq t$, $z^{t+1}(q) = z^s(q)$ for all q in $Q(t-1)$, then there must be a question $q \notin Q(t-1)$ for which $z^{t+1}(q) \neq z^s(q)$. Let $Q(t) = Q(t-1) \cup \{q\}$. The answers to the (at most t) questions in $Q(t)$ uniquely determine the responses to all other questions in $\{z^1,..,z^{t+1}\}$.

Finally, we reach the set $Q(m-1)$ of at most $(m-1)$ questions. Given that $d \geq n-1 \geq m-1$, the agent detects all the dependencies of the answer to any question outside $Q(m-1)$ on the response to the questions in $Q(m-1)$. Furthermore, he is able to detect any combination of responses to $Q(m-1)$ that never appear in $\Theta(Q,A)$. Thus, $\alpha_{n-1}(Q,A) = 1$. ■

Comments:

(a) We use the above claims to find an optimal questionnaire and to calculate $\beta_d(A)$ for $d = 1$ and some small values of n:

(i) From claim 3, if $n \leq 2$, then $\beta_1(A) = 1$.

(ii) If $n = 3$, the one-click questionnaire is optimal and $\beta_1(A) = 3/4$. To see this, let Q be an optimal questionnaire. By Claim 1(iii), we can assume that neither of the questions receives a constant truth value. Since $d > 0$, we can assume that no two questions receive identical or opposing truth values for profiles in A and thus Q is a set of singletons. By Claim 1(ii), Q contains at least three questions. Thus, $\alpha_1(Q,A) = \alpha_1(one-click\ questionnaire, A)$.

(iii) If $A = \{a,b,c,d\}$, then $Q^* = (\{a,b\},\{a,c\},\{a,d\},\{a\},\{b\},\{c\},\{d\})$ is an optimal questionnaire and $\beta_1(A) = 1/3$. To see this, note that the four accepted responses to Q are:

$(1,1,1,1,0,0,0)$,

$(1,0,0,0,1,0,0)$,

$(0,1,0,0,0,1,0)$,

$(0,0,1,0,0,0,1)$.

The question $\{\omega\}$ "identifies" ω. That is, for any question q we have $\{\omega\} \to q$ if $\omega \in q$ and $\{\omega\} \to -q$ if $\omega \notin q$. Thus, $\Theta(Q^*,A)$ consists of the four honest responses given by profiles in A and the eight responses that answer the last four questions negatively and the first three questions with an arbitrary combination of truth values. Thus, $\alpha_1(Q^*,A) = 1/3$.

To show that $\alpha_1(Q,A) \geq 1/3$ for all Q that identify A, we can assume that Q is a cover of A. By Claim 1, we can assume that $Q = Q_1 \cup Q_2$ where Q_k consists of sets of size k and that $|Q_1| \leq 4$ and $|Q_2| \leq 3$. Each affirmative response to a question $\{\omega\} \in Q_1$ determines (in $\Theta(Q,A)$) the answers to all other questions. Thus, the set $\Theta_1(Q,A)$ contains at most the four responses of members of A and at most $2^{|Q_2|}$

responses θ for which $\theta(q) = 0$ for all $q \in Q_1$. Thus, $|\Theta_1(Q,A)| \leq |Q_1| + 2^{|Q_2|} \leq 12$ and $\alpha_1(Q,A) \geq 4/12$.

(b) Increasing the number of questions may *increase* the probability that a manipulator will succeed. Consider the case of $A = \{a,\ b,\ c,\ d\}$. Let $Q_1 = \{\{a,b\},\ \{c\},\ \{d\}\}$ and $Q_2 = \{\{a,b\},\ \{c\},\ \{d\}, \{a\}\}$. Then, $\Theta(Q_1,A) = \{(1,0,0),\ (0,1,0),\ (0,0,1)\}$ and $\Theta_1(Q_i,A) = \Theta(Q_1,A) \cup \{(0,0,0)\}$ and thus $\alpha_1(Q_1,A) = 3/4$. However, $\Theta(Q_2,A) = \{(1,0,0,1),\ (1,0,0,0),\ (0,1,0,0),\ (0,0,1,0)\}$, $\Theta_1(Q_2,A) = \Theta(Q_2,A) \cup \{(0,0,0,0)\}$ and thus $\alpha_1(Q_2,A) = 4/5$!

5. Preventing (almost all) Successful Cheating

Our last claim states that whatever the value of d, $\beta_d(A)$ decreases very rapidly with the size of A. The proof uses a concept from Combinatorics: a collection C of subsets of A is said to be k-*independent* if for every k distinct members $Y_1,..,Y_k$ of the collection, all the 2^k intersections $\cap_{j=1}^{k} Z_j$ are nonempty, where Z_j is either Y_j or $-Y_j$.

For example, a collection C is 2–independent if for every two subsets of C, Y_1 and Y_2, the four sets $Y_1 \cap Y_2$, $-Y_1 \cap Y_2$, $Y_1 \cap -Y_2$ and $-Y_1 \cap -Y_2$ are nonempty. In other words, the fact that a particular element either does or does not belong to a certain set in the collection is not by itself evidence that it does or does not belong to any other set in the collection. For $A = \{a,b,c,d\}$, the collection $C = \{\{a,b\},\{a,c\},\{a,d\}\}$ is a maximal 2-independent collection.

We will now use a result due to Kleitman and Spencer (1973) which states that the size of the maximal k-independent collections is exponential in the number of elements in the set A.

Proposition: Let (Ω^n, A^n) be a sequence of problems where $|A^n| = n$. For every d, $\beta_d(A^n)$ converges double exponentially to 0 when $n \to \infty$.

Proof: By Kleitman and Spencer (1973), there exists a sequence C^n of $(d+1)$-independent collections of subsets of A^n such that the size of C^n is exponential in n. Thus, for every n large enough the size of C^n is larger than n and therefore we can assume that C^n is a cover of A^n (if not, then there exists a set Z in the collection such that any of its members also belongs to another set in the collection; by replacing Z with $A^n - Z$ we obtain a new $(d+1)$-independent collection of subsets of A^n which is a cover of A^n). Let $Q^n = \{q \mid q \in C^n\}$. Since C^n is a cover of A^n the questionnaire Q^n identifies A^n. No d –implication involving these questions is true in A^n. Thus, $\beta_d(A^n) \leq \alpha_d(Q^n, A^n) = \frac{n}{2^{|Q^n|}}$. ■

Note that the proposition refers to any fixed d. If d increases with n, then the result would not necessarily hold (by Claim 3 if $d_n = n-1$, then $\beta_{d_n}(A^n) \equiv 1$). Note also that there are many sequences of questionnaires which can ensure that the manipulation probability goes to zero. Thus, the prinicpal does not have to choose an optimal questionnaire in order to make the success of a manipulation very unlikely.

6. Related Literature

The main purpose of this paper is to formally present the intuition that complex questionnaires may assist a principal in eliciting non-verifiable information from agents. In other words, the principal can design a sufficiently complex questionnaire that makes it difficult for dishonest responders to game the system successfully, while treating honest responders fairly.

Kamien and Zemel (unpublished,1990) is an early paper that models the difficulty of cheating successfully. The most closely related paper to ours is Glazer and Rubinstein (2012). Both that paper and the current one, examine a persuasion situation with a boundedly rational agent though they differ in the procedure used by the agent to come up with a persuasive story. In Glazer and Rubinstein (2012), an agent’s profile is a vector of characteristics. The agent is asked to declare a profile after the principal has announced a set of conditions that these characteristics must satisfy in order for the request to be accepted. The principal’s conditions are of the same form as the regularities in the current paper. A crucial assumption in Glazer and Rubinstein (2012) is that the agent’s (boundedly rational) procedure of choice is an algorithm that is initiated from his true profile. The principal’s problem is to design the set of conditions cleverly enough to be able to differentiate between the agents he wishes to accept and those he wishes to reject. In the current paper, the principal chooses a questionnaire and commits himself to accept a particular set of responses. The agent is limited in his ability to understand the set of acceptable responses. If he decides to lie, he will then fully abandon his true profile and randomly choose a response to the questionnaire that is compatible with the regularities he has detected.

The current paper is related to the growing literature on "behavioral mechanism design". Rubinstein (1993) studies a monopolist’s pricing decision where the buyers (modeled using the concept of perceptrons) differ in their ability to process the information contained in a price offer. Glazer and Rubinstein (1998) introduce the idea that the mechanism itself can affect agents’ preferences and a designer can sometimes utilize these additional motives to achieve goals he could not otherwise achieve. Eliaz (2002) investigates an implementation problem in which some of the agents are "faulty", in the sense that they fail to act optimally. Piccione and Rubinstein (2003) demonstrate how a discriminatory monopolist can exploit the correlation between a consumer’s reservation values and his ability to recognize temporal price patterns. Cabrales and Serrano (2011) look for a mechanism that induces players’ actions to converge to the desired outcome when they follow best-response dynamics. Jehiel (2011) shows how an auctioneer, by providing partial information about past bids, can exploit the fact that present bidders see only some of the regularities in the distribution of bids as a function of types. De Clippel (2011) and Korpela (2012) extend standard implementation theory by assuming that agents’ decisions are determined by choice functions that are not necessarily rationalizable.

References

Alon, N. (1986): "Explicit construction of exponential sized families of k-independent sets," *Discrete Math*., 58, 191–193.

Cabrales, A. and R. Serrano (2011): "Implementation in Adaptive Better-Response Dynamics: Towards a General Theory of Bounded Rationality in Mechanisms," *Games and Economic Behavior,* 73, 360-374.

de Clippel, G. (2011): "Behavioral Implementation," mimeo.

Eliaz, K. (2002): “Fault Tolerant Implementation,” *Review of Economic Studies,* 69, 589-610.

Glazer, J. and A. Rubinstein (1998): "Motives and Implementation: On the Design of Mechanisms to Elicit Opinions," *Journal of Economic Theory,* 79, 157-173.

Glazer, J. and A. Rubinstein (2012): “A Model of Persuasion with a Boundedly Rational Agent,” The *Journal of Political Economy,* 120, 1057-1082.

Jehiel, P. (2011): "Manipulative auction design," *Theoretical Economics,* 6, 185–217.

Kamien, M.I. and E. Zemel (1990): "Tangled Webs: A Note on the Complexity of Compound Lying," mimeo.

Kleitman, D.J. and J. Spencer (1973): "Families of k-independent sets," *Discrete Math*., 6, 255–262.

Korpela, V. (2012): "Implementation without Rationality Assumptions," *Theory and Decision*. forthcoming.

Piccione, M. and A. Rubinstein: (2003): "Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns," *Journal of European Economic Association*, 1, 212-223.

Rubinstein, A. (1993): "On Price Recognition and Computational Complexity in a Monopolistic Model," *Journal of Political Economy*, 101, 473-484.