Complex Questionnaires

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

Abstract

A principal wishes to ascertain an agent's true profile in order to decide whether or not to accept his request. The principal designs a questionnaire regarding the agent's profile and commits himself to accept the request if the agent's answers satisfy certain conditions. The agent is boundedly rational in that he cannot fully understand the principal's conditions and can only detect certain "regularities" in them. It is shown that the principal can construct a complex enough questionnaire that will allow him to respond optimally to agents who tell the truth and to almost eliminate the probability that an agent will succeed in cheating.

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 combinatorical issues.

1. **Introduction**

In many principal-agent situations, a principal needs to make a decision on the basis of 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 instrument in order to induce the agent to provide full and accurate information. The economic literature has focused on two such instruments: verification and incentives. If the information that the principal wishes to obtain is (at least partially) verifiable, the principal can ask the agent to present hard evidence to support his statement. In cases where hard evidence is not available or is not sufficient, the principal can still induce the agent to be honest by designing an incentive scheme, according to which the agent will be rewarded (or penalized) as a function of the information he provides. However, these tools are often prohibitively expensive or insufficient to incentivize the agent to provide the information needed by the principal.

The purpose of this paper is to suggest another tool that can be used by a principal to reduce the probability that an agent can cheat effectively. Instead of asking the agent direct questions regarding the relevant information, the principal can design a "complex" questionnaire, so that an agent who is boundedly rational and is considering lying, will find it difficult to come up with answers that will induce the principal to take the action desired by the agent.

The analysis is carried out in the context of a simple principal-agent persuasion model. The principal interacts with many agents over time who present him with a request. In each case, the principal needs to 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.

Assume that in order to obtain the information he needs, the principal designs a questionnaire that asks the agent a set of questions regarding his profile and that he will accept the agent's request if and only if his response to the questionnaire (i.e., the set of answers provided by the agent) falls within a certain set of possible answers (hereafter referred to as the acceptable set). Suppose that the principal's goal in designing the questionnaire is twofold: his first priority is to make the right decision regarding the agent's request (from his point of view) if the agent answers the questionnaire honestly. His secondary priority is to minimize the probability that the agent's request will be accepted if he has been dishonest, that is, if he has come up with answers to the questionnaire that are independent of his true profile.

We show that in order to achieve his two goals the principal can design a complex questionnaire, in which the agent is not simply asked to report his profile but whether his profile satisfies certain conditions. Once such a questionnaire has been constructed, the principal's policy (i.e., his decision whether or not to accept the agent's request) will be to treat the agent's answers as if they came from an honest agent. The principal will try to design the questionnaire in such a way that the request of a dishonest agent will be accepted with only a low probability. The ability of the principal to design such a questionnaire will depend on the cognitive procedure used by a dishonest but boundedly rational agent. An optimal questionnaire is one that fully achieves the principal's first goal and at the same time minimizes the probability that a dishonest agent has his request accepted.

The core of our model consists of the assumptions regarding the cognitive 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 agents do not know (or do not fully understand) the principal's policy (i.e., which responses he will accept) although they can detect (or are able to understand) certain "regularities", i.e. interdependencies between the answers to different questions in the set of acceptable responses. An agent in the model is characterized by the "level" of regularities he can detect. The most boundedly rational agent (an agent of level 0) will only be able to conclude whether the answer to a particular question must be positive or negative in order for his request to be accepted. An agent of level 1 will also be able to observe whether in the set of acceptable responses the answer to a particular question uniquely determines the answer to an additional question. An agent of level k will be able to observe whenever an answer to a set of k questions uniquely determines the answer to an additional question.

In order to illustrate the procedure, consider a person staring at a picture of an orchard during fruit picking season. He will be able to discern some patterns. An unsophisticated individual will only be able to observe that the "picture is green". A more observant individual will notice that pixels are ordered in such a way that they form the shapes of trees. Finally, a really astute individual will notice that next to each tree with fruit on it, there is a person with a ladder.

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 in an effective way. We show that, when the principal uses an optimal questionnaire, the larger is the set of profiles which the principal would like to accept, the less likely it becomes that a dishonest agent will have his request accepted. Our main result states that when the set of acceptable profiles is very large the principal can reduce to almost zero the probability of a dishonest agent cheating effectively.

The most closely related previous model is presented in our earlier paper Glazer and Rubinstein (2012). Both papers examine a persuasion situation with a boundedly rational agent. The main difference between the two lies in our assumption concerning the procedure used by the boundedly rational agent in an attempt to come up with a persuasive (but not necessarily truthful) story. In Glazer and Rubinstein (2012), it was assumed that the agent uses his true profile as a basis and modifies it trying to come up with an acceptable profile. How the agent does it is affected by the framing of the acceptance rules announced by the principal. Here we assume that if the agent decides to lie, he ignores his true profile and he instead responds to the questionnaire in a way which is compatible with regularities he has been able to detect in the set of acceptable responses.

The paper is also related to the growing literature on "behavioral mechanism design". Eliaz (2002) investigates the implementation problem when some of the agents are "faulty", in the sense that they fail to act optimally. In Cabrales and Serrano (2011), there must exist a mechanism that induces players’ actions to converge to the desired outcome when they follow best-response dynamics in order for a social choice function to be implementable. De Clippel (2011) extends standard implementation theory by assuming that agents’ decisions are determined by choice functions which are not necessarily rationalizable. In an earlier work, Glazer and Rubinstein (1998) introduced the idea that the mechanism itself can affect agents’ preferences and thus the design of the optimal mechanism. Kamien and Zemel (unpublished,1990) is one of the few papers in Economics to suggest that if cheating is difficult, then this can affect the design of optimal mechanisms.

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 the action r if the profile is in $R = \Omega - A$. Denote the size of A by n. A persuasion problem is the pair (Ω, A).

A questionnaire

A questionnaire is a collection of "questions". Each question is of the form $q_X =$"Does your profile belong to the set X?" where $X \subseteq \Omega$. We use the notation $X(q)$ for the set of profiles that the question q is asking about. The agent responds to each question with a "Yes" or a "No".

Two examples of questionnaires are:

(i) The *one-click* questionnaire which consists of the $|\Omega|$ questions of the form $q_{\{\omega\}}$. 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, that is $X(q_k) = \{\omega \mid \omega_k = 1\}$.

A *response* to *a* questionnaire Q is a truth assignment to all questions in Q. Let $\Omega^*(Q)$ be the set of all possible responses to Q. Given an agent's profile ω, let $Q(\omega)$ be the response to Q given by an honest agent whose profile is ω, that is $Q(\omega)(q) = Y$ iff $\omega \in X(q)$. When we count the questions in Q by $(q_1, \ldots, q_m)$ we identify a response with an m-vector of zeroes and ones, where the value 1 stands for "Yes" and the value 0 for "No".

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

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

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

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

We say that a questionnaire identifies A if, when all agents are honest, the responses of the agents whose profile is in A differ from those of agents whose profile is in R. Formally, Q identifies A if $A^*(Q,A) \cap R^*(Q,A) = \emptyset$. Generally, whether or not Q identifies A may also depend on the elements in R. However, the "one-click" questionnaire identifies any set A since in such a questionnaire any two profiles induce two different responses. The same is true of the simple questionnaire.

We assume that the principal is committed to respond optimally to agents who tell the truth. That is, he must use a questionnaire that identifies A and adhere to a policy of accepting a response if and only if it is in $A^*(Q,A)$. However, the principal is concerned that agents who are not honest will try to come up with a false acceptable response (i.e., a response in $A^*(Q,A)$). The principal's objective is to design a questionnaire that identifies A and at the same time makes it less likely that agents who choose not to tell the truth will come up with an acceptable answer.

The principal does not know the agent's profile and cannot verify any of the answers made given by the agent.

The Bounded Rationality Element

The agent in our model must answer a questionnaire Q, but does not know the principal's policy for accepting and rejecting responses (i.e., the agent does not "know" $A^*(Q,A)$). Neither does he have any prior beliefs about that policy. However, the agent can detect regularities in the principal's policy. A *regularity* is defined as a sentence in the language of propositional logic, with the variables being the names of the questions in Q, which is true in all responses that the principal accepts.

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 d names of questions or their negations and the consequent φ_2 is another name of a question or its negation. We will refer to such a proposition as a d-*implication*. Given a questionnaire Q, an agent of rank d learns all the d-implications that are true for all responses in $A^*(Q,A)$.

For example, 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_2 is Y" (denoted $-q_1 \rightarrow q_2$). The proposition $q_1 \wedge -q_2 \rightarrow q_3$ is an example of a possible regularity of rank 2.

Note that the agent in our model is boundedly rational not only in the sense that he can only detect relatively simple regularities in the principal's policy but also that:

(i) He is not capable of observing regularities in the set of rejected responses. This assumption appears to be reasonable in cases where the agent gets to observe other agents whose request has been accepted (such as job candidates who have been hired or winners of a prize), but not those whose request has been rejected (i.e., who didn't get hired or didn't receive the prize).

(ii) The agent is not capable of identifying inconsistent responses. This assumption appears to be reasonable in cases where there is no obvious logical connection between the different questions in the questionnaire.

(iii) While the agent can detect regularities in the set of accepted responses he cannot identify particular acceptable responses. This assumption is reasonable in situations where it is easier for people to conclude, for example, that all "admitted students are males" than to conclude that "all males who applied were admitted".

(iv) His reasoning is limited to the space of responses and he is unable to relate to the space of profiles. If he were capable of "inferring backwards" from the space of responses to the space of profiles, he could probably learn the set A and come up with an acceptable response to the questionnaire as if he possessed one of the profiles in A. In Section 5 we will further discuss this assumption.

Let $M_d(Q,A)$ be the set of responses that satisfy all the d-implications that are true for all responses in $A^*(Q,A)$. By definition, $M_d(Q,A) \supseteq M_{d+1}(Q,A) \supseteq A^*(Q,A)$ 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 $A^*(Q,A)$, regardless of his true profile), he will choose randomly from among the responses in $M_d(Q,A)$. His probability of success is therefore: $\alpha_d(Q,A) = |A^*(Q,A)|/|M_d(A,Q)|$. Obviously, $\alpha_d(Q,A)$ is weakly increasing in d.

The principal's problem

The principal has two objectives in designing a questionnaire: (i) He would like to treat honest agents "properly" (hence, the questionnaire has to identify A and the principal's policy should be to accept only responses given by honest agents whose profile is in A), and (ii) he wishes to minimize the probability that a dishonest agent will be able to successfully deceive him (i.e., the principal wishes to minimize $\alpha_d(Q,A)$).

An optimal questionnaire solves the principal's problem:

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

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 leader faces a distribution of followers' types and seeks a policy that maximizes the principal's expected payoff.

Example 1: A simple questionnaire:

Let $\Omega = \{0,1\}^3$ and A be the set of profiles in which at least two characteristics receives the value 1.

Let Q_1 be the simple questionnaire $\{q_1,q_2,q_3\}$ where q_i is the question about dimension i.

$A^*(Q_1,A) = \{(1,1,1),\ (1,1,0),\ (0,1,1),\ (1,0,1)\}$. $R^*(Q_1,A)$ consists of all other possible responses.

No 0-implication is true in $A^*(Q_1,A)$ since for each question there is an acceptable response in which the question receives the value Y and there is another acceptable response in which the question receives the value N. Thus, $\alpha_0(Q_1,A) = 1/2$.

The 1-implications that are true in $A^*(Q_1,A)$ are the six propositions $-q_j \rightarrow q_k$ where $j \neq k$ (the answer N to any question determines that the answer to the other two questions should be Y). The set of responses that satisfy these six propositions ($M_1(Q_1,A)$) is exactly $A^*(Q_1,A)$ and thus, $\alpha_1(Q_1,A) = 1$.

However, the principal in this case can do even better when $d = 1$. Let Q_2 be the questionnaire $\{q_{12},q_{13},q_{23}\}$ where q_{ij} asks whether the i'th and the j'th characteristics have the value 1 (formally $X(q_{ij}) = \{\omega|\omega_i = \omega_j = 1\}$). The questionnaire Q_2 identifies A as $A^*(Q_2,A) = \{(1,1,1),\ (1,0,0),\ (0,1,0),\ (0,0,1)\}$ and $R^*(Q_2,A) = \{(0,0,0)\}$. No 1-implication is true in $A^*(Q_2,A)$ (i.e., no answer to one question determines the answer to another). Thus, $\alpha_0(Q_2,A) = \alpha_1(Q_2,A) = 1/2$. However, $\alpha_2(Q_2,A) = 1$ since the agent with $d = 2$ detects 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 the q_{23} and thus $M_2(Q_2,A) = A^*(Q_2,A)$.

Example 2: *The one-click questionnaire*

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

An agent of rank 0 will learn to answer N 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 $\alpha_0(Q,A) = n/2^n$ (recall that $|A|= n$).

An agent of rank 1 will in addition observe the regularities $q_{\{\omega\}} \to -q_{\{\omega'\}}$ where $\omega \in A$ and $\omega \neq \omega'$. For $n > 2$, the agent will not detect any additional regularities and therefore $M_1(oneclick,A)$ consists of set $A^*(Q,A)$ and the "constant No" response. Hence, $\alpha_1(oneclick,A) = n/(n+1)$. For $n \leq 2$, we have in addition $-q_{\{\omega\}} \to q_{\{\omega'\}}$ and therefore $\alpha_1(oneclick,A) = 1$.

Example 3: *An anomaly: Increasing the number of questions may hurt the principal.*

The following example shows that increasing the number of questions may increase the chances that a cheater will succeed.

Let $A = \{\omega_1,\omega_2,\omega_3,\omega_4\}$. We compare the questionnaire $Q_1 = \{q_{\{\omega_1,\omega_2\}},q_{\{\omega_3\}},q_{\{\omega_4\}}\}$ to $Q_2 = Q_1 \cup \{q_{\{\omega_1\}}\}$. The following table presents the honest responses for all possible profiles:

	$q_{\{\omega_1,\omega_2\}}$	$q_{\{\omega_3\}}$	$q_{\{\omega_4\}}$	$q_{\{\omega_1\}}$
ω_1	1	0	0	1
ω_2	1	0	0	0
ω_3	0	1	0	0
ω_4	0	0	1	0
$Other$	0	0	0	0

The next table presents (for both questionnaires) the sets of acceptable responses and the set of responses that satisfy all the 1-implications:

	Q_1	Q_2
$A^*(Q_i,A)$	$\{(1,0,0),(0,1,0),(0,0,1)\}$	$\{(1,0,0,1),(1,0,0,0),(0,1,0,0),(0,0,1,0)\}$
$M_1(Q_i,A)$	$A^*(Q_1,A) \cup \{(0,0,0)\}$	$A^*(Q_2,A) \cup \{(0,0,0,0)\}$

Thus, $\alpha_1(Q_1,A) = 3/4$ while $\alpha_1(Q_2,A) = 4/5$!

3. Some Observations

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

Claim 1:

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

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

(iii) If $q'(\omega) \equiv \mathit{constant}$ for all $\omega \in A$ (that is, if $X(q') \supseteq A$ or $-X(q') \supseteq A$), 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 $-q$ (that is $X(-q) = -X(q)$). Then, Q' identifies A and $\alpha_d(Q,A) = \alpha_d(Q',A)$ for all d.

We say that a questionnaire Q is a cover of A if for all $q \in Q$, $X(q) \subseteq A$ and $\cup_{q \in Q} X(q) = A$. We use this definition in Claim 2 which states that $\beta_d(A)$ depends only on the size of A (and not on the size of Ω).

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 Ω.

Proof:

(i) Let $b \in R$. Since Q identifies A then b's honest response to Q is distinct from the responses of all profiles in A. By Claim 1(iv), we can assume that $b \notin X(q)$ 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 $X(q)$.

Now let Q' be the questionnaire $\{q_{X \cap A}$ | there exists $q_X \in Q\}$. 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 $X(q')$ (for some $q' \in Q'$). The honest response of each profile in A to Q and to Q' is identical and therefore, $\alpha_d(Q,A) = \alpha_d(Q',A)$.

(ii) By (i), we can assume that the optimal questionnaire consists only of questions regarding subsets of A and the size of R is immaterial for any $\alpha_d(Q,A)$. ■

The ability of the principal to prevent dishonest agents from cheating effectively depends on the relation between n and d. Claim 3 states that 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 $A^*(Q,A) = \{z^1,\ldots,z^m\}$, where $m \leq n$. We construct (inductively) a set of $m-1$ questions in Q, such that for any profile in A an honest answer to these questions determines the honest answers to all 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 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 set $Q(t) = Q(t-1)$. 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 will be able to detect any combination of responses to $Q(m-1)$ that never appear in $A^*(Q,A)$. Thus, $\alpha_{n-1}(Q,A) = 1$. ■

Claim 4.

(i) If $n = 1$, then $\beta_1(A) = 1$.

(ii) If $n = 2$, then $\beta_1(A) = 1$.

(iii) If $n = 3$, then $\beta_1(A) = 3/4$.

(iv) If $n = 4$, then $\beta_1(A) = 1/3$.

Proof:

(i) Each question receives a unique truth value in A and thus even an agent with $d = 0$ will know what to say.

(ii) Let $A = \{\omega_1, \omega_2\}$. If $q(\omega_1) \neq q(\omega_2)$ for some $q \in Q$, then an agent with $d = 1$ will fully learn the set $A^*(Q,A)$. Otherwise, all answers are constant for honest agents in A and even an agent with $d = 0$ will know how to respond.

(iii) Let $A = \{\omega_1, \omega_2, \omega_3\}$. Then, $\alpha_1(one - click - questionnaire, A) = 3/4$.

Assume, to the contrary, that there is a Q with $\alpha_1(Q,A) < 3/4$. By Claim 2, we can assume that Q is a cover. By Claim 1(ii), Q contains at least three questions. By Claim 1(iii), we can assume that neither of the questions receives a constant truth value. Since $d = 1$, we can assume that no two questions receive identical or opposing truth values for profiles in A. Thus, without loss of generality, Q is the one-click questionnaire, a contradiction.

(iv) Let $A = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $Q = \{q_{\{\omega_1,\omega_2\}}, q_{\{\omega_1,\omega_3\}}, q_{\{\omega_1,\omega_4\}}, q_{\{\omega_1,\omega_2,\omega_3\}}, q_{\{\omega_1,\omega_2,\omega_4\}}, q_{\{\omega_1,\omega_3,\omega_4\}}, q_{\{\omega_2,\omega_3,\omega_4\}}\}$.

The four accepted responses are

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

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

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

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

The question $q_{A-\{\omega_i\}}$ "identifies ω_i". That is, $-q_{A-\{\omega_i\}} \to q_B$ if $\omega_i \in B$ and $-q_{A-\{\omega_i\}} \to -q_B$ if $\omega_i \notin B$. Thus, $A^*(Q,A)$ consists of the four honest responses given by profiles in A and the eight responses in which the last four questions are answered affirmatively and the first three questions receive an arbitrary combination of truth values. Thus, $\alpha_1(Q,A) = 1/3$.

To prove that $\alpha_1(Q,A) \geq 1/3$ for all Q that identify A, we can again assume that Q is a cover of A. Let $Q_2 = \{q_X \in Q| \ X \subseteq A$ contains 2 elements$\}$ and $Q_3 = \{q_X \in Q| \ X \subseteq A$ contains 1 or 3 elements$\}$. By the argument in part (iii) of this claim, we can assume that all questions are in either Q_2 or Q_3 and that $|Q_2| \le 3$ and $|Q_3| \le 4$. Each $q \in Q_3$ "identifies a profile ω_q" (the unique profile that belongs to $X(q)$ or the unique profile that does not belong to $X(q)$). The response of ω_q to q determines (in $A^*(Q,A)$) the answers to all other questions. Thus, the set $M_1(Q,A)$ contains at most the four responses of members of A and at most $2^{|Q_2|}$ responses in which, for every $q \in Q_3$, the answer to q is the opposite answer to that of ω_q. Thus, $|M_1(Q,A)| \le |Q_3| + 2^{|Q_2|} \le 12$ and $\alpha_1(Q,A) \ge 4/12$. ■

4. Almost No Successful Cheating

Our last and main claim states that whatever d is, $\beta_d(A)$ decreases rapidly with the size of A. In fact, even for a moderate size of A the principal can guarantee that the probability of the agent cheating successfully is very small.

The proof uses a concept in 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 its complement.

For example, a collection C is 2–independent if for every two subsets in 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. That is, the fact that a particular element belongs or does not belong to a certain set in the collection is not by itself evidence that it belongs or does not belong to any other set in the collection. For $A = \{1,2,3,4\}$, the collection $C = \{\{1,2\},\{1,3\},\{1,4\}\}$ is 2-independent. Furthermore, it is a maximal 2-independent collection: For any set Y outside C the collection $C \cup \{Y\}$ is not 2-independent. If $1 \in Y$, then either $Y = \{1\}$ and $Y \cap -\{1,j\} = \emptyset$ or $Y \neq \{1\}$ and $-Y \cap \{1,j\} = \emptyset$ for some j. If $1 \notin Y$ then either $Y = \{2,3,4\}$ and $-\{2,3,4\} \cap -\{1,2\} = \emptyset$ or $Y \cap \{1,j\} = \emptyset$ for some j.

We will 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 "very fast" to 0 when $n \to \infty$.

Proof: By Kleitman and Spencer (1973), there is 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. For large n enough (for example, $logn > d$), there exists such a collection, which is also a cover of A^n. Let $Q^n = \{q_X \mid X \in C^n\}$. No d –implication involving these questions is true in A^n. Thus, $\alpha_d(Q^n,A^n) = \frac{n}{2^{|Q^n|}}$ which diminishes rapidly.

A simpler proof uses the sequence of questionnaires Q^n defined by $Q^n = \{q_{\{w\}} | w \in A^n\} \cup \{q_{\{1,w\}} | w \in A^n\}$. ($Q^n$ is the one-click questionnaire appended with questions about the subsets in the non-maximal d –independent cover of A^n $\{\{1,w\} \mid w \neq 1\}$). Using the same logic as in Claim 4 (iv) we get $\alpha_d(Q^n,A^n) = \frac{n}{2^{n-1}}$. ■

5. Discussion

The main purpose of this paper is to formally present the intuition that complex questionnaires may serve a principal’s purpose in the case where he must rely on non-verifiable information provided to him by agents. A complex questionnaire enables the principal to properly treat honest responders and at the same time to make it difficult for dishonest responders to game the system successfully. Thus, sophisticated questionnaires can complement the extensively studied incentives mechanisms in helping a principal obtain non-verifiable information from a self-interested but boundedly rational agent.

Our model introduces two additional elements into the standard literature:

(a) Information acquisition: agents "see" the set of accepted responses and arrive at some simple propositions (i.e., they observe regularities). The agents conclude that the accepted responses must satisfy all of these propositions.

(b) The principal’s objective is twofold: to treat honest agents optimally and at the same time to lower the probability of success for agents who try to cheat.

Our key assumption regarding the agent’s capabilities states that he cannot figure out the profiles in A by observing regularities in the set of acceptable responses. If he could, the agent would mimic an honest response given by one of the profiles in A. The plausibility of this assumption depends on the way in which the questions are framed. For example, if a question is framed explicitly as "is your profile either ω_1 or

ω_2?" then the agent's attention could be drawn to the space of profiles and he might then consider which profiles the principal would like to accept. However, it is less likely that he will think in terms of profiles if the question is framed without mentioning the profiles explicitly, but rather it asks him about some "equivalent" fact that is true if and only if the agent's true profile is either ω_1 or ω_2. The following two examples help to illustrate this point:

Example: The porter's trap: A hotel manager wants to find out whether a particular porter was present in at least two out of three shifts. The true profile is the list of shifts in which the porter was present. The manager could simply use a questionnaire containing three questions, each asking the porter whether he was present in one particular shift. This questionnaire could lead the porter to think that the manager is interested in knowing when he was present. Suppose, however, that the manager knows that for any two shifts if the porter was present in both shifts he must have noticed the arrival and departure of a particular serviceman. In such a case, by asking the porter whether a certain serviceman had arrived and left while he was on duty, the manager can, in fact, conceal from the porter that he is trying to monitor his attendance.

Example: *The financial expert trap:* A financial expert is looking for a knowledgeable assistant. The expert possesses information about three investments, denoted by 1,2 and 3. Each investment will yield one dollar if it succeeds and nothing if it fails. On the basis of his analysis, the expert believes that either investments 1 or 2 or both will succeed, whereas investment 3 will certainly fail. The expert's objective is to test the candidates and to select only those who share his belief about these three investments. Assume that the candidates are of level $d = 1$.

The expert can present the candidates with the questionnaire $Q_1 = \{q_1, q_2, q_3\}$ where q_i is "will investment i succeed?"

In this case, $A^*(Q_1, A) = \{(1,1,0),(1,0,0),(0,1,0)\}$. A candidate will learn that $q_3 = 0$, $-q_1 \rightarrow q_2$ and $-q_2 \rightarrow q_1$, and thus will conclude that the set of acceptable responses is exactly $A^*(Q_1, A)$. A dishonest agent will succeed with probability 1 in giving one of the right answers.

However, assume that there are three funds in the market F_1, F_2 and F_3 and that each fund’s performance depends on that of the three investments. Each entry in the table below presents the incremental payoff to each fund as a result of the success of each of the three investments:

	1	2	3
F_1	1	0	-1
F_2	0	1	-1
F_3	0.5	0.5	1

Consider a questionnaire, $Q_2,$ which consists of three questions: $q_i =$"Do you believe that the total yield of fund F_i is at least 1 ?" There are three responses to Q_2 that are consistent with the expert’s beliefs: $A^*(Q_2,A) = \{(1,1,1),\,(1,0,0),\,(0,1,0)\}$. The responses $(1,0,0)$ and $(0,1,0)$ are the honest responses of a candidate who believes that either investments 1 or 2 (but not investment 3) will succeed and $(1,1,1)$ is the honest response of a candidate who believes that both investments 1 and 2 (but not 3) will succeed. If a candidate thinks that none of them will succeed then he will answer $(0,0,0)$ and if he thinks that 3 will succeed he will answer $(0,0,1)$. The other three responses $(1,1,0),\ (1,0,1),\ (0,1,1)$ are inconsistent.

A candidate who tries to game the system will learn that $-q_1 \to q_2$, $-q_1 \to -q_3$, $-q_2 \to q_1$, $-q_2 \to -q_3$ and hence $M_1(Q_2,A) = A^* \cup \{(1,1,0)\}$. Thus, using the questionnaire Q_2 instead of Q_1, i.e., asking about the funds rather than the investments, allows the expert to reduce to $3/4$ the probability that a dishonest candidate will succeed in impressing him.

References

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

Cabrales, Antonio and Roberto Serrano. 2011. "Implementation in Adaptive Better-Response Dynamics: Towards a General Theory of Bounded Rationality in Mechanism." *Games and Economic Behavior* 73: 360-374.

de Clippel, Geoffroy. 2011. "Behavioral Implementation", mimeo.

Eliaz, Kfir. 2002. “Fault Tolerant Implementation”. *Review of Economic Studies* 69:

589-610.

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

Glazer, Jacob and Ariel Rubinstein, “A Model of Persuasion with a Boundedly Rational Agent”. The *Journal of Political Economy*, 2012.

Kamien, Morton I. and Eitan Zemel. 1990. "Tangled Webs: A Note on the Complexity of Compound Lying". mimeo.

Kleitman, Daniel J. and Spencer, Joel. 1973. Families of k-independent sets. *Discrete Math*. 6: 255–262.