Mechanism Design and Communication Costs

Raymond Deneckere and Sergei Severinov *

First Version: December 3, 2000

Second Version: September 19, 2001

This Version: July 12, 2003

---

*Department of Economics, University of Wisconsin, Madison and Fuqua School of Business, Duke University. We thank Yeon Koo Che, Bart Lipman, Larry Samuelson, Robert Wilson and seminar participants at the University of Wisconsin, Caltech, UCLA, Florida, Stanford GSB, USC, Duke, the Winter 2001 North American Econometric Society Meetings, and the 2001 Canadian Economic Theory Conference for helpful comments and suggestions. We are responsible for all errors. Address for correspondence: Sergei Severinov, Fuqua School of Business, Duke University, 1 Towerview Drive, Durham, NC 27708, email: sseverin@duke.edu

Abstract

Although environments in which individuals can costlessly and effortlessly manipulate their information are standard in economics, they represent an extreme point in the spectrum of various possibilities. In fact, available evidence suggests that at least some individuals have limited abilities to misrepresent their true preferences and imitate the behavior of other types. In this paper we focus on the issue of implementation in such environments. We develop an approach allowing to characterize the set of implementable outcomes, and then apply it to derive optimal mechanisms. The key elements of our approach are the absence of any restrictions on the communication structure in a mechanism and the ability of the principal to screen the agents not only on the basis of their preferences over outcomes, but also on the basis of their communication abilities. We explore the implications of agents' limited abilities to misrepresent their information for monopoly regulation, signaling and screening. In particular, we point out why employers may prefer to screen applicants via multiple rounds of interviews rather than via menus of contracts. Our findings also provide a justification for privacy laws.

JEL Nos: C72, D82

Keywords: mechanism design, implementation, costly communication, bounded rationality, honesty, signaling, screening.

1 Introduction.

This paper focuses on mechanism design and implementation in environments where some participants have limited abilities to misrepresent their information. The premise that economic agents can costlessly and effortlessly manipulate information and misrepresent their preferences is prevalent is economics. Nevertheless, the alternative view that such abilities may be limited is well grounded in common intuition and has been explored in a number of contributions. Several reasons for this can be pointed out.

First, it may be costly for an individual to misrepresent the truth for psychological or ethical reasons. Lying may cause her to incur a disutility because of stress or discomfort that she experiences ("blushing," "feeling wrong").[1] Erard and Feinstein (1994) argue that "some taxpayers appear to be inherently honest, willing to bear their full tax burden even when faced with financial incentives to underreport their income. Evidence for such inherently honest taxpayers derives not just from casual introspection; it is also supported by econometric evidence and survey findings..." Alger and Ma (2003) emphasize that physicians with stronger ethical views may be less prone to exaggerate medical problems of a patient in their communication with an HMO. Experimental evidence confirms that a nonnegligible part of the population chooses not to lie regarding private information, even though lying would increase their monetary payoffs.[2] Chen (2000) argues that individuals have a tendency to keep promises, even if this it not always in their self-interest, and shows that this may cause optimal contracts to be incomplete.

Second, in some environments communication may involve submission of credible or veri-

[1]Behavioral psychologists have extensively studied the physical symptoms associated with the emotional discomfort that people experience when lying (see, e.g., Ekman (1973)). The idea that emotions act as predictable and powerful motivation devices guiding economic behavior has been pursued deeply by Robert H. Frank in his book "Passions within Reason: The Strategic Role of Emotions" (1988).

[2]For example, Gneezy (2002) reports experiments with deception games in which responders were known to largely follow the sender's recommendation. Yet the proportion of informed senders who chose *not* to mislead opponents even though misleading was in the senders' best interests varied from 48% to 83% across experiments. Survey evidence paints a similar picture, with a core group of people having no qualms at all about inflating insurance claims, but an even greater fraction considering it unacceptable to do so (Tennyson 1997).

fiable claims. For example, an agent may be able to produce some hard evidence to support her announcements.[3] Existence of verifiable claims makes imitation harder. Indeed, failure of an agent to produce evidence which is known to be available in a certain state of the world can be taken as a proof that this state of the world has not occurred. Then only individuals who possess skills and technology to manufacture evidence will be able to mimic others, while individuals who lack such technologies will not be able to conceal their private information.[4] A 'credible claims' model has been investigated by Lippman and Seppi (1995). They derive conditions on the verifiability of claims that allow the principal to elicit the truth from a group of symmetrically informed agents with conflicting interests. The model of evidence disclosure in court studied by Bull and Watson (2001) can similarly be seen as a process in which 'credible claims' are submitted. Green and Laffont (1986) consider a situation in which the set of types that an agent can imitate depends on the agent's true type. This situation also admits interpretation in the spirit of the 'credible claims' model. Che and Gale (2000) study an optimal mechanism for selling a good to a buyer who may be budget constrained. In this case the ability of a buyer to misrepresent her willingness to pay is limited if the seller can ask her to post a bond and thus credibly disclose some information about her budget.

Finally, misrepresenting the truth may require costly physical actions. Environments of this kind are studied by Lacker and Weinberg (1989), Maggi and Rodriguez-Clare (1995) and Crocker and Morgan (1998).[5] Lacker and Weinberg (1989) argue that 'there are many instances in which lying about the state of nature requires more than simply sending a false signal regarding one's private information. Often, costly actions must be taken to lend credence

[3]Indeed, individuals are often asked to support their statements with some form of evidence. For example, the government requires taxpayers to justify deductions on their tax return with receipts or other documents, and courts verify contractual performance on the basis of submitted evidence.

[4]Consider a class of environments where the 'binding' direction of imitation is from 'lower' (less accident-prone, not creditworthy) to 'higher' (lower-risk, creditworthy) types. For example, in order to obtain a loan or be chosen as a supplier one may need to be perceived as successful, wealthy, or creditworthy. In such environments, some individuals will not be able to exaggerate their prior performance or risk of default, while others could rely on a network of friends or associates to provide them with references, loan money and 'status' goods to support their exaggerated claims.

[5]Our approach differs from the one used by these authors, as we consider an environment where individuals differ in their aversion to/cost of undertaking such actions.

to the signals being sent.' For example, a sharecropper who misrepresents the crop may have to hide part of the output or borrow some from a third party. A taxpayer may have to reallocate her assets or move to a less affluent neighborhood if her report of low wealth and/or income is to be credible. It is reasonable to believe that there is a cost to such evasive actions.

When the agents' abilities to manipulate information are limited, it is natural to say that they experience communication costs, so we will use this term to describe such environments in the sequel. The existing literature points out that the presence of communication costs affects the nature of optimal mechanisms and expands the set of implementable outcomes. In particular, Green and Laffont (1986) and Alger and Ma (2003) show that mechanisms in which an agent does not always report her type truthfully can attain outcomes that are not implementable via mechanisms with truthful reporting. Similarly, Maggi and Rodriguez-Clare (1995) demonstrate that a principal can elicit an agent's private information more efficiently by inducing her to engage in some misrepresentation.

Yet, the extant literature provides neither a systematic characterization of the set of implementable outcomes, nor a method for designing optimal mechanisms. In this paper we address these shortcomings. We develop a general approach to implementation that can be used to characterize the set of implementable outcomes. This enables us to integrate existing models within a common framework, and to examine whether the mechanisms that have been proposed for these models are indeed optimal. We also apply our approach to study the effects of costly communication, and construct optimal mechanisms in a number of environments that have not been studied before.

We view communication costs as an implementation tool that permits relaxing a number of incentive constraints which otherwise have to be imposed on an allocation profile. To exploit this property to the maximal extent, we avoid any exogenous restrictions on the communication stage of a mechanism. More specifically, we allow for mechanisms in which agents send multiple messages or signals. In contrast, most existing contributions adopt the (often implicit) assumption that an agent can send only one message, or make only one announcement/claim.[6] Although this assumption is without loss of generality in environments with

[6]This applies to Green and Laffont (1986), Alger and Ma (2003), Erard and Feinstein (1994), and Maggi

costless communication, it restricts the set of implementable outcomes when communication is costly.

We consider two classes of environments. The first class includes environments in which an agent is endowed with a set of feasible messages which she can send at zero cost, while any messages outside of this set is infeasible (infinitely costly) for her. In these so called "binary cost structure" environments, the set of feasible messages typically depends on agents' privately known utility parameters, but need not be uniquely determined by the latter. The environments studied by Green and Laffont (1986), Lipman and Seppi (1995), Erard and Feinstein (1994), Alger and Ma (2003) and Alger and Renault (1999) belong to this class.

One interpretation of feasible messages adopted in the literature is that they are "truthful" (or non-false) statements about type that can be credibly verified, or alternatively, that some such messages can be ascertained as false when sent by certain types. This view requires directly linking the content of a message to the physical state of the world, i.e. the agent's preference parameters, and presumably, some concept of the verification/authentication process.[7]

Our model is consistent with this interpretation. It also admits a broader one that does not require to associate specific content with a message, and provides a foundation for the concept of verifiability, or provability. According to this view, both the payoff-relevant preference parameters and the sets of feasible messages are drawn from a known probability distribution. This probability distribution and, in particular, the events that occur with positive probability reflect natural interdependencies between the payoff-relevant parameters and the available message sets. For example, a messages m_1 may only be consistent with some realizations

and Rodriguez-Clare (1995). Participants can send several messages in Lipman and Seppi (1995), but their paper focuses on a different set of issues.

[7]Environments where information is verifiable and so cannot be misrepresented, but can be withheld by interested parties, have been studied by Milgrom (1981), Grossman (1981), and Milgrom and Roberts (1986). In their frameworks, full disclosure of information emerges as a robust prediction due to an unravelling argument. Okuno-Fujiwara, Postlewaite and Suzumura (1990) develop this approach further in a competitive environment and derive the conditions under which information revelation does and does not occur. Milgrom (1981), Shin (1994) and Fishman and Hagerty (1990) also study related environments where the agents' information is imperfect and there are physical limits on the amount of information that can be transmitted to the principal.

of the preference parameters, say, θ_1 and θ_2, but not other realizations, such as θ_3 and θ_4. Then the submission of message m_1 establishes that neither θ_3, nor θ_4 have occurred. Note that we use the terms 'message', 'announcement', 'signal', or 'claim' interchangeably. So a message may actually represent submission of a piece of evidence, or expression of an emotional reaction. Thus, the concept of provability inherent in this view relies on inference. A message can be taken as proof of some event not in a direct sense, but only due to the fact that it would not have been available had an alternative event occurred.[8]

In these environments, the largest set of outcomes can be implemented via a mechanism in which each agent type is requested to report all her feasible messages/make all her feasible claims. By requesting each agent type to send all her feasible messages, such mechanism narrows the set of agent types who can fully imitate the behavior of any given type, and thereby eliminates the maximal number of incentive constraints at the communication stage without imposing them on the allocation profile. Clearly, the set of implementable outcomes is larger than in the standard case where all messages are costless/feasible, or where agents are requested to send only a single message. As a consequence, our approach allows the principal to screen the agents better than the mechanisms exhibited by Green and Laffont (1986) and Alger and Ma (2003) in their respective frameworks.

Mechanisms in which each agent sends all her feasible messages provide an analytical tool for characterizing the set of implementable outcomes. Yet, it may be desirable and more economical to rely on as few messages as possible. Addressing this issue, we demonstrate how each agent's set of feasible messages can be subdivided into a number of equivalence classes in such a way that it becomes sufficient to consider mechanisms in which an agent sends only one message from each equivalence class. This method allows to demonstrate under which conditions some outcomes are implementable only if the agent sends all feasible messages (each equivalence class consists of one message) and, conversely, to exhibit the conditions

[8]For example, Lipman and Seppi (1995) point out that the ability to play a musical instrument is easily provable, but not the inability to do so. Obviously, this is due to the fact that a person unable to play a musical instrument does not have a message which it *not available* to a person with such skills. Similarly, the possession of a European Community passport proves that the holder is a citizen of one of the member-states because such passport is not available to a citizen of a country which is not a member of the Community.

under which an agent needs to send only one message.

Green and Laffont (1986) and Alger and Renault (1999) point out that the Revelation Principle fails when agents can send only one message. Our results allow us to reexamine whether and to what extent the Revelation Principle applies when the communication stage is unrestricted. Note that the Revelation Principle, in fact, embodies two distinct statements. First, each agent reveals her type truthfully, i.e. the principal infers the agent's type correctly after the communication stage. Second, each agent sends only one message, i.e. the amount of communication is the minimal necessary to identify her payoff-relevant type.

Our results imply that the first part of the Revelation Principle (construed broadly), still applies when no restrictions are imposed on the communication stage. Indeed, the set of feasible messages that an agent can send is her private information and is, therefore, a part of her type. Thus, by sending all her feasible messages an agent truthfully demonstrates her communication abilities and completely reveals her type. That the first part of the Revelation Principle applies is of great practical importance, as it greatly simplifies the task of characterizing the set of implementable allocations. At the same time, we show that in general sending a single message is no longer sufficient, so the second part of the Revelation Principle fails. Nevertheless, we identify the conditions under which a single message is sufficient in our setting. Thus, our results stand in contrast to the ones in Green and Laffont (1986), Alger and Ma (2003) and Alger and Renault (1999), who point out the failure of truthfulness and show that mechanisms where agents lie and do not reveal their types may be necessary when the communication stage is restricted to a single message.

The second type of environment considered in the paper relaxes the binary cost structure, and imposes a tighter connection between an agent's payoff-relevant type and her communication abilities. Specifically, each agent type has access to a type-dependent "truthful" message that can be sent at zero cost, whereas sending any other message incurs a finite cost that is increasing in the magnitude of type "misrepresentation."

An agent can be requested to send multiple messages, and we allow the cost of sending any particular message to depend on the other communicated messages. Such dependencies may arise naturally. For example, as the agent proceeds sending more messages to the principal, she may learn how to misrepresent her information more effectively and at a lower cost.

One intuitive interpretation of this model is that an agent undergoes a number of tests or inspections. On every test, each agent type can attain her "natural" level of performance at zero cost, while attaining a different performance level requires costly effort. Thus, one can describe this situation as costly information distortion, costly state falsification or, alternatively, a multidimensional costly signalling problem. Lacker and Weinberg (1989), Maggi and Rodriguez-Clare (1995) and Crocker and Morgan (1998) study environments of this type allowing for a single test/inspection.

The results that we obtain in the multi-signal environment are quite surprising. Any allocation profile becomes implementable at zero communication cost if the communication stage involves a sufficiently large number of messages and the marginal cost of misrepresenting the type (sending a message different from the 'natural' one) does not go to zero too fast in the number of messages. When the marginal cost of misrepresenting one's type is zero, it becomes necessary to induce the agent to incur some communication costs and exert effort to produce 'non-truthful' messages. Still, we establish an approximate implementation result: the principal can come arbitrarily close to any continuously differentiable allocation profile, and at the same time keep the communication cost arbitrarily small, if she requests the agent to send a sufficiently large number of messages. Thus, the availability of a sufficiently large number of costly signals/tests significantly expands the set of implementable allocations.

To allow the kind of perfect screening that we have described, the available message space for each message/signal must be sufficiently rich. In particular, each message space must be larger than the set of possible types. In an applied context, this requirement has important implications for the optimal design of testing and examination procedures. Each testing procedure has to be designed in such a way that some levels of performance would be hard to achieve for all types, even the most productive ones. For example, a test or an interview has to include sufficiently complex question, so that attaining absolute performance level would require considerable amount of effort even from the most able types.

The rest of the paper is organized as follows. In Section 2 we study a principal-agent model with binary communication costs, and establish our general implementation results. (In the

Appendix we explore the multi-agent extension of this model). Section 3 studies environments with a continuous cost structure. Section 4 concludes.

2 Binary Cost Structure.

We start by studying a binary cost environment with two players, a principal and an agent. The principal controls a decision (allocation) $x \in X$, yielding a utility of $w(x, \theta)$ to the principal, and $u(x, \theta)$ to the agent. The preference parameter $\theta \in \Theta$ is private information to the agent. We assume that there exists a 'worst outcome' $\underline{x}$, that minimizes $u(x, \theta)$ for all $\theta \in \Theta$.[9]

Let $\mathcal{C}$ denote the admissible message space, i.e. *the set of statements or messages that can be possibly made by some agent type and understood by the principal.* A typical element of $\mathcal{C}$ is a minimal unit of communication, denoted by m. In the sequel, we will refer to m as "message," "claim," "statement," or "announcement" interchangeably.

The communication abilities of the agent are characterized by her *feasible message set* $\mathcal{M} \subset \mathcal{C}$. An agent with feasible message set $\mathcal{M}$ can make a collection of statements (send a collection of messages) $\{m_1, ..., m_n\}$ if and only if $m_i \in \mathcal{M}$ for all $i \in \{1, ..., n\}$. In other words, she incurs zero cost whenever she sends a message $m \in \mathcal{M}$, and an infinite (or prohibitively large) cost if $m \notin \mathcal{M}$.

The feasible message set is also private information to the agent. Therefore, a full description of her type includes both her preference parameter θ and her feasible message set $\mathcal{M}$. Accordingly, we define an agent's type as $t = (\theta, \mathcal{M})$. Let $T \subset \Theta \times 2^{\#\mathcal{C}}$ be the set of all possible types, where $2^{\#\mathcal{C}}$ denotes the 'power set' (set of all subsets) of $\mathcal{C}$. We assume that the probability distribution $F(.)$ over T is common knowledge. Let $\mathcal{N}$ denote the set of all possible communication sets, i.e. $\mathcal{N} = \{\mathcal{M} | \exists \theta \in \Theta : (\theta, \mathcal{M}) \in T\}$.

Thus, our environment is completely characterized by a 6-tuple $\{X, \mathcal{C}, u(\cdot), w(\cdot), T, F(.)\}$. In general, there need not be any specific relation between θ and $\mathcal{M}$. In particular, letting $I(\mathcal{M}) = \{\theta | (\theta, \mathcal{M}) \in T\}$ and $K(\theta) = \{\mathcal{M} | (\theta, \mathcal{M}) \in T\}$, it may or may not be that $I(\mathcal{M})$ or $K(\theta)$ are singletons.

[9] In an environment with transferable utility this assumption involves no loss of generality.

Example 1 *Green and Laffont (1986): $\mathcal{C} = \Theta$. The preference parameter θ determines $\mathcal{M}$ uniquely, and the mapping from θ to $\mathcal{M}$ is common knowledge. In this case, one can use the notation $M(\theta)$ instead of $\mathcal{M}$.*

For this model, Green and Laffont (1986) show that the Revelation Principle holds if and only if the following ***Nested Range Condition*** *is satisfied: For any three distinct elements $\theta_1, \theta_2, \theta_3 \in \Theta$, if $\theta_2 \in M(\theta_1)$ and $\theta_3 \in M(\theta_2)$, then $\theta_3 \in M(\theta_1)$.*

Example 2 *Alger and Ma (2003): $\mathcal{C} = \Theta = \{\theta_L, \theta_H\}$. The feasible communication set $\mathcal{M}$ of an agent with valuation θ_p, $p \in \{L, H\}$ is either $\{\theta_p\}$ (the agent is 'honest'), or $\{\theta_L, \theta_H\}$ (the agent is 'strategic'). Thus, θ does not determine the feasible communication set $\mathcal{M}$ uniquely. However, there is a correlation between θ and $\mathcal{M}$.*

If the probability α that the agent is 'honest' is sufficiently high, it is optimal for the principal to extract all surplus from the 'honest' type with high valuation θ_H. However, the high-valuation 'strategic' agent then misrepresents her valuation, and in equilibrium imitates the low-valuation one.

Example 3 *The 'credible claims' model of Lipman and Seppi (1995). Several agents with conflicting interests (preferences over actions) observe the state of the world θ. In state θ the same set of credible claims $\mathcal{M}_\theta = \{b_\theta^1, ..., b_\theta^{n_\theta}\}$ is available to each agent. As in Example 1, the correspondence mapping θ to $\mathcal{M}_\theta$ is common knowledge. Lipman and Seppi (1995) explore under what conditions on $\mathcal{M}_\theta$ a principal who does not fully commit to the choice of action, can extract truthful information about the state of the world from the agents.*

Thus, our model encompasses the ones explored by Alger and Ma (2003) and Green and Laffont (1986). There are two main differences between our paper and the rest of the literature. First, our central goal is to characterize the whole set of implementable allocation profiles and design a mechanism to implement them, while most of the literature focuses on finding mechanisms optimal in specific situations. Second, to achieve this goal we consider the largest possible set of mechanisms, without any modeler-imposed restrictions. In particular, we focus on mechanisms in which agents may be requested to send several messages, or make several announcements. In contrast, most related contributions (e.g. Green and Laffont (1986), Alger and Ma (2003), Maggi and Rodriguez-Clare (1995)) assume, sometimes implicitly, that an agent can make only one statement or send only one message.[10]

Mechanisms in which agents send several messages exploit the limits on their abilities to manipulate information to the fullest, and thus allow to screen them to the largest possible extent. Such mechanisms are natural in many contexts. Clearly, an agent may be requested to submit several verifiable claims, documents or pieces of evidence. For example, an applicant for a loan has to confirm her ability to pay by providing evidence about her income, assets, liabilities, etc.

Furthermore, different departments of the same organization or different economic institutions (which one can view in combination as one 'grand mechanism') often collect information about employees, clients, customers or suppliers separately. For example, income has to be reported on income tax forms, loan and mortgage applications, children's college financial aid forms, etc. Then, the agent's skills, tastes and access to technologies for manipulating information determine whether she can communicate different claims to different receivers.

Turning to the issue of implementation, let us define a social choice function $f(\theta, \mathcal{M})$ as a mapping from the set of types into outcomes: $f(.) : T \mapsto X$. To characterize implementable allocation profiles we will use the mechanism $G(.) : 2^{\#\mathcal{C}} \cup \phi \rightrightarrows X$ defined as follows. Let M be the set of reported messages. Then $G(M) = \{f(\theta, M)\}_{\theta \in I(M)}$ and $G(M) = \underline{x}$ if $\theta \notin I(M)$ (recall that $\underline{x}$ is the worst outcome).

In mechanism $G(.)$ the communication stage is followed by a stage in which the agent is offered a menu of allocations. The offered menu depends on the collection of messages M sent by the agent. Specifically, it includes the set of outcomes assigned by the social choice function $f(.,.)$ to all types with feasible communication set M. Clearly, only types whose feasible communication set includes M can access the respective menu. Therefore, we will call $G(.)$ a "password" mechanism.[11]

[10] Lipman and Seppi (1995) allow for multiple messages, but the asymmetry of information and the mechanism design approach distinguish our framework from theirs.

[11] Menus are unnecessary if for all $(\theta, \mathcal{M}) \in T$, we have $\theta \in \mathcal{M}$, i.e. truthful announcements of the preference parameter are always in the feasible message set. Then mechanism $G(.)$ can be replaced with a simpler mechanism $\tilde{G}(.)$ in which the messages are reported in a sequence and the first message is interpreted as the

We will say that mechanism $G(.)$ $\mathcal{M}-$truthfully implements social choice function $f(.)$, if it is optimal for the agent to make all statements/send all messages that are feasible for her, and then choose the allocation corresponding to her true preference parameter. That is, the optimal strategy of type $(\theta, \mathcal{M})$ is to send a collection of messages $\{m_1, ..., m_n\} = \mathcal{M}$ and then choose allocation $f(\theta, \mathcal{M})$ from the menu.[12] Our first result implies that the principal can fully exploit the agent's limited ability to manipulate information by requiring that she send all her feasible messages:

Theorem 1 *Any implementable social choice function is $\mathcal{M}-$truthfully implementable via mechanism $G(\cdot)$.*[13]

Furthermore, there exist social choice functions that are not implementable by mechanisms that do not require the agent to send all of her feasible messages.

Proof: Fix an environment $\{X, \mathcal{C}, u(.), w(.), T, F(.)\}$, and suppose that the social choice function $f(.)$ is implementable via some mechanism $(S, g(.))$ with strategy space S and outcome function $g: S \mapsto X$. Let $s(\theta, M)$ be an equilibrium strategy of type $t = (\theta, M)$ in this mechanism. Since $(S, g(.))$ implements $f(.)$, we have $f(\theta, M) = g(s(\theta, M))$. As in the proof of the Revelation Principle, consider a new mechanism with outcome function $G(.)$ where $G(M) = \{g(s(\theta, M))\}_{\theta \in I(M)}$, and $G(M) = \underline{x}$ otherwise. Note that $G(.)$ offers a menu of outcomes after the communication stage.

Let us show the optimal strategy of an agent with some arbitrary type $(\theta, \mathcal{M})$ in mechanism $G(.)$ is to send all messages in $\mathcal{M}$ and choose $g(s(\theta, \mathcal{M}))$. Suppose to the contrary that there exists a profitable deviation, i.e. the agent of type $(\theta, \mathcal{M})$ gets a higher payoff by sending a collection of messages $\mathcal{M}'$ and choosing the outcome $g(s(\theta', \mathcal{M}'))$. Then $s(\theta', \mathcal{M}')$ cannot be feasible for the agent of type $(\theta, \mathcal{M})$ in mechanism $(S, g(.))$. Therefore, action $s(\theta', \mathcal{M}')$ must

announcement of the true preference parameter. Formally, $\tilde{G}(M) = f(\theta, M)$ if θ is the first message in the collection of messages M sent by the agent and $(\theta, M) \in T$, and $\tilde{G}(M) = \underline{x}$ otherwise.

[12]It is intuitive to call such strategy 'truthful' because the agent fully reveals her *feasible communication set* and chooses the allocation which is designed for a type with her preference parameter.

[13]Note that the existence of a worst punishment $\underline{x}$ is needed only to punish an agent who reports a message set $\mathcal{M}$ that is not feasible for any agent type (i.e. such that $I(\mathcal{M}) = \emptyset$), yet there exist $\mathcal{M}_1, \mathcal{M}_2 \in \mathcal{N}$ s.t. $\mathcal{M} \subset \mathcal{M}_i$ for $i = 1, 2$.

involve sending a message m'' s.t. $m'' \in \mathcal{M}'$, but $m'' \notin \mathcal{M}$. Consequently, type $(\theta, \mathcal{M})$ cannot send the collection of messages $\mathcal{M}'$ in mechanism $G(\cdot)$. This contradiction establishes that $f(.)$ is $\mathcal{M}-$ truthfully implementable via mechanism $G(.)$

To show that mechanism $G(\cdot)$ permits to implement a larger set of social choice functions than mechanisms in which the agent does not have to send all her feasible messages, consider the following example. The decision space is given by $X = \{x_1, x_2, x_3, \underline{x}\}$, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, the admissible message space is $\mathcal{C} = \{m_1, m_2, m_3\}$ and the relation between θ and $\mathcal{M}$ is characterized by the following correspondence $M(\theta)$ which is common knowledge: $M(\theta_1) = \{m_1, m_2\}$, $M(\theta_2) = \{m_2, m_3\}$, $M(\theta_3) = \{m_3, m_1\}$. The outcome $\underline{x}$ is the worst for every type, and the rest of the payoff structure is as follows:

$$u(x_1, \theta_1) < u(x_2, \theta_1) < u(x_3, \theta_1)$$
$$u(x_2, \theta_2) < u(x_3, \theta_2) < u(x_1, \theta_2)$$
$$u(x_3, \theta_3) < u(x_1, \theta_3) < u(x_2, \theta_3)$$

Consider the social choice function $f(\theta_1) = x_1$, $f(\theta_2) = x_2$, $f(\theta_3) = x_3$. Note that $f(.)$ assigns the least preferable alternative to every agent type. Then a mechanism which assigns some allocation x_i, $(i \in \{1, 2, 3\})$ after receiving only one of the messages $\hat{m} \in \{m_1, m_2, m_3\}$ cannot implement $f(.)$. One of the types $j \neq i$ who can send message $\hat{m}$, will find is strictly profitable to do so and obtain allocation x_i rather than x_j.

However, $f(.)$ can be implemented via mechanism $G(.)$: $G(\{m_1, m_2\}) = x_1$, $G(\{m_2, m_3\}) = x_2$, $G(\{m_3, m_1\}) = x_3$, $G(\mathcal{M}) = \underline{x}$, $\forall \mathcal{M} \neq \{m_i, m_j\}$ $(i \neq j)$. *Q.E.D.*

The intuition behind Theorem 1 is easy to understand. When two types have different *feasible communication sets*, one of them (say, type A) cannot send all the messages that are feasible for the other (say, type B). If the mechanism exploits this property, then an implementable allocation profile need not satisfy the standard incentive constraint that type A gets a higher payoff from the allocation designed for her than from the allocation designed for type B. Obviously, the larger is the set of incentive constraints that are eliminated in this way, the less restricted is the allocation profile, and the larger is the set of implementable social choice functions. Because mechanism $G(.)$ requires the agent to send all feasible messages, it allows to eliminate the maximal possible number of incentive constraints at the communication stage. Therefore the set of implementable social choice functions is maximal under $G(.)$.[14]

In mechanism $G(.)$ type $(\theta, \mathcal{M})$ can obtain the allocation designed for type $(\theta', \mathcal{M}')$ if and only if the former can fully mimic the latter i.e., if $\mathcal{M}' \subset \mathcal{M}$. Theorem 1 thus allows to specify which incentive constraints must be imposed on the allocation profile.

Corollary 1 *A social choice function* $f : T \rightarrow X$ *is implementable if and only if for all* $(\theta, \mathcal{M}) \in T$ *the following incentive constraints are satisfied :*

$$u(f(\theta, \mathcal{M}), \theta) \geq u(f(\theta', \mathcal{M}'), \theta) \quad \forall (\theta', \mathcal{M}') \in T \text{ such that } \mathcal{M}' \subset \mathcal{M}.$$

Taken together, Theorem 1 and Corollary 1 provide an analytical tool allowing to characterize the whole set of implementable outcomes via a mechanism in which agents send all their feasible messages. To see how dramatic the effect of sending several messages can be, consider the following condition.

Non-Nested Range Condition (NNRC): Consider any $(\theta, \mathcal{M}), (\theta', \mathcal{M}') \in T$ s.t. $\theta \neq \theta'$. Then $\mathcal{M} \setminus \mathcal{M}' \neq \emptyset$.[15]

[14] Applying our approach to the motivating example studied by Green and Laffont (1986), we can implement a larger set of social choice functions than are implementable using the approach used by these authors. Namely, consider an environment with decision set $X = \{x_1, x_2, x_3\}$, type space and communication space $\Theta = \{\theta_1, \theta_2, \theta_3\}$. The relation between θ and $\mathcal{M}$ is characterized by the following correspondence $M(.)$: $M(\theta_1) = \{\theta_1, \theta_2\}$, $M(\theta_2) = \{\theta_2, \theta_3\}$, $M(\theta_3) = \{\theta_3\}$. The payoff structure is as follows:

$$u(x_1, \theta_i) < u(x_3, \theta_i) < u(x_2, \theta_i) \ \ \forall i \in \{1, 2, 3\}$$

Consider the social choice function $f(\theta_1) = x_1$, $f(\theta_2) = x_2$, $f(\theta_3) = x_3$. Then $f(.)$ is not implementable via any mechanism in which the agent's sends only one message, i.e. announces some $\hat{\theta} \in \Theta$. To see this, suppose otherwise. Then all three agent types must send different messages in equilibrium. Since type θ_3 can only send message "θ_3", it follows that type θ_2 must send message "θ_2." However, type θ_1 would then imitate the message sent by type θ_2.

On the other hand, $f(.)$ is $\mathcal{M}-$ truthfully implementable via the mechanism with the following outcome function $G(.)$: $G(\theta_2, \theta_3) = x_2$, $G(\theta_3) = x_3$, and $G(\mathcal{S}) = x_1$, where $\mathcal{S}$ is any other report.

[15] The analogue of NNRC in the framework studied by Lipman and Seppi (1995) is their *two-way disprovability* condition. Also, see the Appendix for the generalization of NNRC to the multi-agent case.

Under NNRC, an agent's communication abilities uniquely determine her utility parameter θ (i.e. the mapping $I(\mathcal{M})$ from the space of feasible communication sets $\mathcal{N}$ into Θ is a function), and a type with one preference parameter can never fully mimic a type with a different preference parameter and send the same set of messages that the latter is able to send. So, an agent's preference parameter can be inferred perfectly by identifying her communication abilities. We therefore have:

Lemma 1 *Suppose NNRC holds. Then any social choice function $f : \Theta \to X$ is implementable.*[16]

Proof: By Corollary 1, an implementable social choice function need not satisfy any incentive constraints. Therefore, any social choice function is implementable. *Q.E.D.*

A drawback of mechanism $G(.)$ is its potentially lengthy communication stage, as each agent type has to send all her available messages. Thus, it is useful to explore whether the mechanism designer can rely on fewer messages without reducing the set of implementable social choice functions.

Recall that the main reason for requesting that multiple messages be sent is to distinguish between agents with different feasible message sets. From this perspective, message m serves as a proof that the sender's feasible message set is different from feasible message sets that do not contain m. Thus, if two or more messages contain the same proof, i.e. rule out the same collection of feasible message sets, then all but one of them are redundant. We pursue this line of reasoning below to eliminate as many redundant messages as possible.

Let us for simplicity assume that $\mathcal{C}$ is finite, and consider any $\mathcal{M} \equiv \{m_1, m_2, ..., m_n\} \in \mathcal{N}$. Partition $\mathcal{M}$ into $k_{\mathcal{M}} < \infty$ equivalence classes $J_i(\mathcal{M})$ $(i = 1, ..., k_{\mathcal{M}})$, such that messages in the same equivalence class belong to the same collection of elements of $\mathcal{N}$. Formally, m' and m'' belong to the same equivalence class $J_i(\mathcal{M})$ when the following conditions holds: for all $\mathcal{M}' \in \mathcal{N}$ we have $m' \in \mathcal{M}'$ iff $m'' \in \mathcal{M}'$.

[16]Note that the principal can implement any social choice function that depends *only on the agent's preference parameter.* If one wishes to implement all social choice functions, including the ones that also depend on $\mathcal{M}$, then the following stronger version of NNRC is sufficient: Consider any $(\theta, \mathcal{M}), (\theta', \mathcal{M}') \in T$ s.t. either $\theta \neq \theta'$ or $\mathcal{M} \neq \mathcal{M}'$. Then $\mathcal{M} \setminus \mathcal{M}' \neq \emptyset$.

Thus for all $i = 1, ..., k_{\mathcal{M}}$, there exists $A_i(\mathcal{M}) \subset 2^{\mathcal{N}}$ s.t. $m \in J_i(\mathcal{M})$ if and only if $m \in \mathcal{M} \backslash \cup_{M \in A_i(\mathcal{M})} M$. So, by sending any message $m \in J_i(\mathcal{M})$, agent type $(\theta, \mathcal{M}\}$ establishes that her feasible message set does not belong to $A_i(\mathcal{M})$.

Note that each message m belongs to exactly one equivalence class, because the collection of message sets in $\mathcal{N}$ to which m does not belong is well-defined. Since any $\mathcal{M}$ is finite, the number of equivalence classes $k_{\mathcal{M}}$ is also finite.

Let us now eliminate redundant equivalence classes. First, eliminate all $J_i(\mathcal{M})$ s.t. $A_i(\mathcal{M}) \subset A_j(\mathcal{M})$ for some j. Then, re-number all the equivalence classes in an increasing order from $i' = 1$ to $i' = k'_{\mathcal{M}}$, and eliminate all equivalence classes $J_{i'}(\mathcal{M})$ s.t. $A_{i'}(\mathcal{M}) \subset \cup_{j' \geq i'} A_{j'}(\mathcal{M})$. We will refer to the classes remaining after this elimination as basic equivalence classes. Let $k^b_{\mathcal{M}}$ be the total number of basic equivalence classes, and $J^b_i(\mathcal{M})$ denote a representative basic equivalence class.

Next, define mechanism H as follows. For all $\mathcal{M} \in \mathcal{N}$, and all $i = 1, ..., k^b_{\mathcal{M}}$, fix $m^i_{\mathcal{M}} \in J^b_i(\mathcal{M})$. In mechanism H the agent is offered a menu of allocations $\{f(\theta, \mathcal{M})\}_{\theta \in I(\mathcal{M})}$ if she sends a collection of messages $\{m^1_{\mathcal{M}}, ..., m^{k^b_{\mathcal{M}}}_{\mathcal{M}}\}$. If the agent sends a collection of messages that does not correspond to any $\mathcal{M}$ in the described way, then she is assigned the allocation $\underline{x}$. Thus, in mechanism H the agent is requested to send only one message from each basic equivalence class.

Note that mechanism H is well-defined, i.e. any collection $\{m_1, ..., m_k\}$ of messages either defines some $\mathcal{M} \in \mathcal{N}$ uniquely, or does not correspond to any $\mathcal{M} \in \mathcal{N}$ at all. To see this, suppose that collection of messages $\{m_1, ..., m_k\}$ is such that $\exists \mathcal{M} \in \mathcal{N}$ for which $k = k^b_{\mathcal{M}}$ and $m_i \in J^b_i(\mathcal{M})$ $\forall i = 1, ..., k^b_{\mathcal{M}}$. Let us show that such $\mathcal{M}$ is unique. First, consider $\mathcal{M}'$ s.t. $\mathcal{M} \setminus \mathcal{M}' \neq \emptyset$. Then by construction there exists a basic equivalence class $J^b_i(\mathcal{M})$ such that $J^b_i(\mathcal{M}) \setminus \mathcal{M}' \neq \emptyset$. Therefore, the message $m \in \{m_1, ..., m_n\}$ which belongs to $J^b_i(\mathcal{M})$ cannot be sent by an agent with communication abilities $\mathcal{M}'$. Now suppose that $\mathcal{M} \subset \mathcal{M}'$ and the inclusion is strict. Then there exists a basic equivalence class $J^b_i(\mathcal{M}')$ s.t. the agent with communication abilities $\mathcal{M}$ cannot send any $m \in J^b_i(\mathcal{M}')$. Then, we have:

Theorem 2 *A social choice function is implementable via mechanism H if and only if it is implementable via mechanism G. Furthermore, there exist social choice functions imple-*

mentable via mechanism G (and hence H) but not implementable via any mechanism in which some type $(\theta, \mathcal{M})$ sends less than $k^b_{\mathcal{M}}$ messages.

Proof: By construction of H, there are only two types of deviations for an agent of type $(\theta, \mathcal{M})$ that we need to consider: (i) she can send a collection of messages $\{m^1_{\mathcal{M}'}, ..., m^{k^b_{\mathcal{M}'}}_{\mathcal{M}'}\}$ corresponding to the message set $\mathcal{M}' \subset \mathcal{M}$ and choose an allocation $f(\theta', \mathcal{M}')$ for some $\theta' \in I(\mathcal{M}')$, (ii) she can send a collection of messages $\{m^1_{\mathcal{M}}, ..., m^{k^b_{\mathcal{M}}}_{\mathcal{M}}\}$, but choose an allocation $f(\theta'', \mathcal{M})$ $(\theta'' \neq \theta)$ from the menu. Observe that agent of type $(\theta, \mathcal{M})$ can also obtain allocations $f(\theta', \mathcal{M}')$ and $f(\theta'', \mathcal{M})$ in mechanism $G(.)$. Therefore, a social choice function implementable via mechanism H has to satisfy the same set of incentive constraints as a social choice function implementable via mechanism $G(.)$. Hence, a social choice function is implementable via $G(.)$ if it is implementable via H. The"only if" part follows by Theorem 1.

To establish that an agent of type $(\theta, \mathcal{M})$ has to send at least $k^b_{\mathcal{M}}$ messages, consider the example studied in the proof of Theorem 1. We have $k^b_{\mathcal{M}_i} = 2$ for all $i = 1, 2, 3$. Thus, each agent type must send her both feasible messages in mechanism H. In the proof of Theorem 1, we have established the existence of a social choice function that cannot be implemented if any type sends less than 2 messages. *Q.E.D.*

The intuition behind Theorem 2 is as follows. Corollary 1 implies that the communication stage of a mechanism should be constructed in such a way that it would be impossible for an agent to imitate an agent of a different type if the former cannot send all messages available to the latter. In mechanism H, an agent sends exactly one message that is not available to some other type. Therefore, mechanism H minimizes the amount of communication without sacrificing the scope of implementability. Theorem 2 has the following simple corollary.

Corollary 2 *(i) Suppose that $\mathcal{M} \in \mathcal{N}$ is such that there exists $m_{\mathcal{M}} \in \mathcal{M}$ satisfying $m_{\mathcal{M}} \notin \mathcal{M}'$ for all $\mathcal{M}' \in \mathcal{N}$ s.t. $\mathcal{M} \setminus \mathcal{M}' \neq \emptyset$. Then any implementable social choice function is implementable via a mechanism in which type $(\theta, \mathcal{M})$ sends only 'identifier' message $m_{\mathcal{M}}$.*[17]

(ii) If $\mathcal{M} \in \mathcal{N}$ is such that for all $m \in \mathcal{M}$ there exists $\mathcal{M}' \in \mathcal{N}$ s.t. $\mathcal{M} \setminus \mathcal{M}' = m$, then in mechanism H an agent with feasible message set $\mathcal{M}$ has to send all feasible messages.

[17] The analogue of identifier message in the framework studied by Lipman and Seppi (1995) is their *full report* message.

Corollary 2 characterizes the conditions under which the two extreme cases obtain: either one or all messages have to be sent. We do not require that an 'identifier' message $m_{\mathcal{M}}$ could not be imitated by an agent with different communication abilities $\mathcal{M}' \neq \mathcal{M}$. Rather, it must be the case that if $m_{\mathcal{M}} \in \mathcal{M}'$, then $\mathcal{M} \subset \mathcal{M}'$.

Using the results of Theorem 1 and Corollary 2, we can address the issue of the applicability of the Revelation Principle in environments with communication costs. Green and Laffont (1986), Alger and Renault (1999) and Alger and Ma (2003) point out that the Revelation Principle fails when an agent's ability to imitate others is limited, because it may be optimal to induce some types to send non-truthful messages which do not reveal their types. Green and Laffont (1986) show that the Nested Range Condition (see example 1) is necessary and sufficient for the Revelation Principle to hold.

Note that it is natural to view the Revelation Principle as a combination of the following two statements regarding an optimal mechanism: (i) each agent reveals her true type; (ii) each agent sends only one message, i.e. a minimal amount of communication is required. In standard environments these properties hold jointly. However, in environments with communication costs where the applicability of the Revelation Principle is problematic, these two properties have to be considered separately. So, let us let us examine them in our framework where the restriction to a single message is lifted.

Theorem 1 implies that part (i) can be extended to the costly communication case. So, the basic message of the Revelation Principle remains valid. One can focus on mechanisms where the agents report their types truthfully, where truthfulness is construed broadly. Indeed, the set of feasible messages is agent's private information and, therefore, should be regarded as part of her type. By sending all messages in this set the agent truthfully demonstrates her communication abilities. Her choice from the menu also correctly reveals her preference parameter.

However, part (ii) regarding the minimality of communication does not generally hold. In most cases, a mechanism has to rely on more than a minimal amount of communication to ensure implementability.

Still, Corollary 2 identifies a necessary and sufficient condition - the existence of an 'identifier' message $m_{\mathcal{M}}$ for each type of communication abilities $\mathcal{M} \in \mathcal{N}$- under which only a single message is required. In this case, after receiving 'identifier' message $m_{\mathcal{M}}$, the principal offers a menu of allocations $\{f(\theta, \mathcal{M})\}_{\theta \in I(\mathcal{M})}$ to the agent and the agent's choice from the menu correctly identifies the agent's preference parameter. So, the amount of information transmission is the minimal necessary to completely reveal the agent's type $(\theta, \mathcal{M})$, and hence both parts of the Revelation Principle apply. The existence of an 'identifier' message in a more general communication space has the same effect as the *Nested Range Condition (NRC)* in the case where $\mathcal{C} = \Theta$. [18]

To illustrate the applicability and usefulness of our results, we next consider several settings.

Example 4 *A firm with cost function $c(q)$ sells quantity q of a good to a consumer with a quasi-linear utility function $v(q, \theta) - p$, where p is a transfer to the firm and $v(q, \theta)$ is assumed to be concave, increasing in both arguments and possess a positive cross-partial derivative. The parameter θ is privately known by the agent and takes one of n possible values $0 < \theta_1 < ... < \theta_n$. Let $\pi_i > 0$ denote the probability that $\theta = \theta_i$. The firm wishes to maximize the expected value of $p - c(q)$ where p is the transfer paid by the agent. The agent has limited ability to misrepresent her true value: she can do so by only one step down. Specifically, $M(\theta_i) = \{\theta_{i-1}, \theta_i\}$ for $i > 1$, and $M(\theta_1) = \theta_1$.*

According to Corollary 1, the only incentive constraint that has to be imposed is that type θ_2 be unwilling to mimic type θ_1. So, in the optimal mechanism all types other than θ_1 receive efficient quantities, and only type θ_2 obtains a positive surplus. This is so because type θ_2 has a reporting advantage relative to all other types. Specifically, type θ_i $(1 < i < n)$ can claim to be of type θ_{i-1}. Yet, such claim is not available to any higher type θ_k, $k > i$. However, there is no claim available to θ_1 that is not available to θ_2.

[18] In fact, there is a close formal connection between the two, as we can generalize *NRC* as follows:

NRC*: $\forall \mathcal{M} \in \mathcal{N}$ *there exists an identifier message $m_{\mathcal{M}}$ satisfying the condition of part (i) in Corollary 2.*

Thus, **NRC*** allows to restrict attention to mechanisms with only a single message plus a choice from the menu. Also, if $\mathcal{C} = \Theta$ and an agent's communication abilities are a function of her utility parameter, i.e. $\mathcal{M} = M(\theta)$, **NRC*** reduces to the *Nested Range Condition* with message "θ" playing the role of an identifier of type θ.

In implementation, we can rely on the 'password' mechanism. Furthermore, we can exploit the structure of incentive constraints to reduce the number of messages to one per type from two messages per type required in mechanism H. Indeed, the following mechanism would implement the optimal allocation profile. If type θ_i ($i > 1$) is announced, the mechanism should assign the efficient quantity for type θ_{i+1} and a transfer leaving θ_{i+1} at her reservation level. If type θ_1 is announced, the mechanism should offer a self-selecting menu consisting of two quantity/transfer pairs for types θ_2 and θ_1.

This example and the discussion preceding it highlight the importance of an agent's communication abilities in determining the amount of surplus that she obtains. A highly productive agent with limited communication abilities may not earn any surplus, while less productive agents with better communication abilities would be able to obtain a positive surplus.

Finally, let us illustrate our results by applying them in an environment where some agents may be unable to misrepresent their types at all (due to honesty or bounded rationality), while others have unlimited ability to do so. Alger and Ma (2003) (see Example 2) develop this idea in an adverse selection model of health care provision where the patient can either be sick (high cost of treatment) or healthy (low cost of treatment), and the physician treating her is either 'honest' (unable to lie) or 'strategic.' Their mechanism is based on a single round of communication, i.e. one message is sent by the physician. In a companion paper (Severinov and Deneckere 2003), we analyze a non-linear pricing model in which a monopoly faces a consumer with valuation $u(q, \theta)$ for quantity q of the good. The preference parameter θ is drawn from an interval. Its true value as well as the consumer's ability or inability to misrepresent it are her private information. We apply our approach to implementation relying on the "password" mechanism where the principal offers a menu of contracts contingent on the collection of messages sent by the agent in the communication stage.

To highlight the distinction between the two approaches in the most simple way, let us focus on the non-linear pricing problem. Applying the approach of Alger and Ma (2003) in this context, consider a mechanism where an agent makes one announcement $\hat{\theta}$ about her preference parameter and is assigned an allocation $\{q(\hat{\theta}), t(\hat{\theta})\}$ on the basis of her report. Then, the cardinality of the maximal set of allocations is equal to the cardinality of the set of preference parameters. Because 'honest' consumers always report their valuations truthfully, and 'strategic' consumers choose reports to maximize their payoffs, the firm faces a choice between two alternatives. First, it can offer an allocation profile that keeps consumers who report truthfully at their reservation utility levels. This strategy allows to extract full surplus from 'honest' consumers. However, 'strategic' consumers then choose to underreport their valuations, reducing the efficiency of the mechanism and the firm's expected profits. Alternatively, the firm can extract a larger surplus from 'strategic' consumers by offering an allocation profile that makes reporting true valuation incentive compatible. However, incentive compatibility comes at a cost: in this case 'honest' consumers receive the same allocations as 'strategic' consumers with the same valuations, and so end up with positive surpluses.

The results of Alger and Ma (2003) imply that the first strategy is optimal when the agent is very likely to be 'honest,' for then the rent extraction from 'honest' types becomes the dominant motive. Precisely, when the consumer's valuation can take only two values: high θ_H or low θ_L, both low-valuation types and the high-valuation 'strategic' type consume the quantity that is efficient for the low-valuation consumer. They also pay the same transfer, one that gives no surplus to the low-valuation types but, obviously, a positive surplus to the high-valuation 'strategic' type. Thus, the 'strategic' high-valuation agent consumes an inefficiently low quantity. At the same time, the high-valuation 'honest' type is assigned the efficient quantity and zero surplus. In contrast, when the proportion of 'honest' consumers in the population is small, it becomes optimal for the principal to induce 'strategic' consumers to tell the truth and forego full rent extraction from the 'honest' consumers. In this case, a consumer's allocation depends only on her valuation, and is identical to the one in the standard case without 'honest' consumers.

In contrast, Corollary 1 implies that the firm need not face such a tradeoff between leaving surplus to the 'honest' consumers and assigning more efficient quantities to the 'strategic' ones. Using mechanism H or a simpler 'password' mechanism, the firm can implement an allocation profile that consists of as many quantity/transfer pairs as there are types (counting the honest/strategic property as part of the type). It can distinguish 'honest' consumers from 'strategic' without leaving any surplus to the former, while inducing the latter to make self-selecting choices from the menu offered at the second stage of the mechanism. The only incentive constraints that must be imposed are that a 'strategic' consumer does not wish to imitate any other type.

More precisely, in mechanism $H(.)$ the consumer is asked to make two announcements $(\hat{\theta}_1, \hat{\theta}_2)$ regarding her valuation. If the same valuation is reported twice, i.e. $\hat{\theta}_1 = \hat{\theta}_2$, then the mechanism assigns an efficient quantity corresponding to the announced valuation and a transfer that leaves no surplus to this announced type. If, on the other hand, $\hat{\theta}_1 \neq \hat{\theta}_2$, the mechanism offers a self-selecting menu of contracts. Incentive compatibility implies that a consumer with valuation exceeding the lowest one obtains positive surplus when choosing from this menu. The inability of an 'honest' type to lie forces her to report her true valuation both times, and so she does not receive any surplus and her valuation is identified for free. Meanwhile, all 'strategic' consumers send conflicting messages: $\hat{\theta}_1 \neq \hat{\theta}_2$, thereby proving their ability to manipulate information, and gaining access to the screening menu. [19]

This mechanism is more profitable for the principal. It is also more socially efficient, as it generates a smaller quantity distortion than any mechanism in which the agent can make only one statement about her valuation.[20] In Severinov and Deneckere (2003) we characterize the optimal allocation profile for the considerably more complex case in which θ can take any value in some interval $[\underline{\theta}, \overline{\theta}]$. We demonstrate that in the optimal mechanism, (i) 'strategic' consumers are assigned higher quantities than in the standard case without 'honest' consumers,

[19]The required amount of communication can be further reduced by exploiting the special structure of this example. Instead of asking for two messages, the principal can request only one. If the reported valuation $\hat{\theta}$ is strictly above the lowest one $\underline{\theta}$, the modified mechanism assigns the allocation which mechanisms H assigns after the report $(\hat{\theta}, \hat{\theta})$. If the consumer reports $\underline{\theta}$, she is offered the same menu as in H following a pair of conflicting announcements. As a further justification of using the term 'password' to describe this mechanism, note that the message $\underline{\theta}$ can be seen as a 'password' required to gain access to the self-selecting menu allowing to earn a positive surplus.

[20]When θ takes only two values: high θ_H and low θ_L with probabilities π and $1-\pi$ respectively, and a consumer is 'strategic' with probability α, both high-valuation types are assigned an efficient quantity, while both low-valuation types are assigned quantity q_L s.t. $v_q(q_L, \theta_L) - \frac{\alpha\pi(v_q(q_L,\theta_H) - v_q(q_L,\theta_L))}{1-\pi} = c'(q_L)$. The quantity q_L is below the efficient level but the distortion is smaller than in the standard case without 'honest' types, because the 'honest' high-valuation consumer does not earn an informational rent. The only type who obtains a positive surplus is the 'strategic' high-valuation one whose net payoff is equal to $v(q_L, \theta_H) - v(q_L, \theta_L)$.

except the ones with valuations near the top who are assigned standard second-best quantities. (ii) 'Honest' consumers get quantities that are below the first-best, with the exception of those near the 'top' who are assigned first-best quantities. (iii) Most surprisingly, there is no exclusion: all consumers with valuations exceeding the marginal production cost, whether 'honest' or 'strategic,' are assigned positive quantities. Furthermore, all of the 'strategic' consumers obtain larger surpluses than in the absence of 'honest' consumers. Thus, 'strategic' consumers benefit from the presence of 'honest' consumers.

An interesting and possibly surprising implication of our results is that individuals who make conflicting or contradictory statements need not be penalized for such behavior. Such individuals obtain higher payoffs than individuals who do not make contradictory statements and are less suspect of lying. This prevents individuals with low personal cost of lying from imitating someone else's behavior and improves the overall efficiency of the mechanism.

Rewarding individuals who make contradictory statements, rather than punishing them, does not appear to be at odds with reality. For example, there is a wide range of institutions that collect the same information from individuals, but perform their functions separately from each other, without cross-checking information submitted to them. In many cases, sharing such information about individuals is prohibited by privacy laws. Privacy laws then appear to encourage individuals to make contradictory statements, and our results explain why such institutional design may be optimal.

A case in point is statements about income. An individual has to declare her income to the Internal Revenue Service for tax purposes, to banks when applying for a loan, to colleges and Universities when requesting financial aid for children, and so on. These institutions do not cross-check information with each other. This leaves open the possibility that an individual might submit different reports to different institutions. Specifically, people who are more prone to misrepresentation or have weaker ethical norms could overstate their income when applying for a loan, and/or understate it on the financial aid form, but still report the truth to the IRS. From a social welfare point of view, this outcome can dominate the outcome that would arise if these institutions cross-checked individuals' reports and imposed penalties for contradictory statements. For in this case, individuals prone to misrepresentation might find it optimal to also misrepresent their incomes for tax purposes. Conceivably, this could result in a higher social welfare loss. Thus, our results suggests that privacy laws prohibiting information sharing across institutions may be a part of a socially optimal institutional design.

3 Non-binary Costs and Multiple Messages.

In this section we consider a class of environments in which the cost structure is no longer required to be binary, and in which there is a tighter connection between the agent's payoff-relevant type θ and her communication abilities.

More specifically, the cost of a message depends on the magnitude of "misrepresentation" of the payoff-relevant parameter. The principal chooses the number of messages that an agent is requested to send. Each message is characterized by some specific content, i.e. it is sent along a specific dimension or line.

For example, the principal may query the agent's ability by asking her to undergo a number of tests. Each test assesses the agent's ability in a different way. In this case, message i corresponds to the agent's performance on test i. Similarly, in the job-market context the principal (employer) may assign a number of interviewees to evaluate the agent (job-candidate) in separate conversations. Each of the interviewees may use a different method to assess the agent's ability. In fact, the results of this section suggest that a more extensive interviewing of job-applicants, although also costly for the employer, may be preferable to on-the-job screening via incentive schemes.

Alternatively, each message can correspond to the presentation of a different piece of evidence about the state of the world, or the outcome of one of many inspections or audits that the principal subjects the agent to. For example, parent corporations regularly carry out accounting and other audits of their subsidiaries. Firms try to ensure the integrity of their accounts by employing both internal accounting staff and hiring external auditors.

It is natural to assume that in these and similar contexts the agent has to incur a cost in order to misrepresent her payoff-relevant type (ability, cost, creditworthiness, etc.), or undertake an effort in order to conceal it. Thus, the agent can control the signals that the principal receives, but can only do so at a certain type-dependent cost.[21]

Formally, we assume that the type space Θ is a metric space, with metric $||.||_\Theta$. In the communication stage of the mechanism the agent sends n messages. The i-th message m_i belongs to the set M^i which is a metric space with metric $||.||_i$. The agent's communication costs are measured by a nonnegative continuous cost function $C^n(.) : \Pi_{i=1}^n M^i \times \Theta \mapsto R_+$. Thus, when the agent of type θ sends a collection of messages $\{m_1, ..., m_n\}$, she incurs a cost equal to $C^n(m_1, ..., m_n, \theta)$.[22]

The cost function $C^n(m_1, ..., m_n, \theta)$ is assumed to have the following properties. First, there exists an injective mapping $\gamma^i(.) : \Theta \mapsto M^i$ s.t. message m_i is costless for type θ if and only if $m_i = \gamma^i(\theta)$. Specifically, for all n and $(m_1, .., m_{i-1}, .., m_{i+1}, ..., m_n)$, $\gamma^i(\theta) = \arg\min_{\tilde{m}_i} C^n(m_1, ..., \tilde{m}_i, ..., m_n, \theta)$, and $C^n(m_1, ..., m_n, \theta) = 0$ if and only if $m_i = \gamma^i(\theta)$ for all $i = 1, ..., n$. Thus, $C^n(m_1, ..., m_n, \theta) > 0$ if there exists $i \in \{1, ..., n\}$ s.t. $m_i \neq \gamma^i(\theta)$.

Message $\gamma^i(\theta)$ can be regarded as truthful type revelation by type θ. In contrast, sending message $m_i \neq \gamma^i(\theta)$ involves costly type misrepresentation/distortion of information.

Each message space M_i is assumed to be sufficiently large in the following sense. There exists an open set (in the topology generated by the metric $||.||_i$) of messages B_i and $\nu > 0$ s.t. for all $\theta \in \Theta$, (i) $\gamma^i(\theta) \notin B_i$, and (ii) there exists $m'_i(\theta) \in B_i$ s.t. $C^n(m_1, .., m'_i(\theta), ..., m_n, \theta) - C^n(m_1, .., \gamma_i(\theta), ..., m_n, \theta) \geq \nu$.

As in the previous section, an agent of type θ obtains utility $u(x, \theta)$ when the principal selects an allocation $x \in X$. The set X is compact, $u(x, \theta)$ is continuous in x, and there exists a 'worst' outcome $\underline{x}$ which minimizes $u(., \theta)$ for all θ. Thus, when an agent with preference parameter θ sends n messages $m_1, ..., m_n$ and the principal chooses allocation x, the agent's

[21] Milgrom (1981), Fishman and Hagerty (1990), and Glazer and Rubinstein (1998) and (2001) study models in which the agent receives a number of noisy signals about the state of the world. The agent cannot distort these signals, but can disclose only a subset of them to the principal because of the limited capacity of the communication channels. Our framework is related. The difference lies in the fact that in our case the agent can fully (but at a cost) control the noise in each signal, and the capacity of the communication channels is sufficient to transmit all signals.

[22] In a more general model, the communication cost function $C^n(.)$ may also depend on a separate parameter t, so that agents with the same preference parameter θ but different t may have different communication costs.

overall payoff is given by:

$$u(x,\theta) + C^n(m_1, ..., m_n, \theta) \tag{1}$$

Note that we allow the marginal cost of misrepresenting type in the i-the message, i.e. sending message $m_i \neq \gamma^i(\theta)$, to depend on the whole array of messages sent by the agent. This dependence may arise naturally in a variety of economic environments. For example, as the agent sends more messages misrepresenting her type, (s)he may learn how to do so more efficiently and at a lower cost. The only condition that we need for our results to hold is that the marginal cost of sending a non-truthful message $m_i \neq \gamma^i(\theta)$ that involves a small misrepresentation (i.e. m_i lies in some small neighborhood of $\gamma^i(\theta)$) does not go to zero too quickly as the number of messages increases.

Our model is similar to the 'costly state falsification' framework studied by Lacker and Weinberg (1989) and Maggi and Rodriguez-Clare (1995). In their framework the agent can distort the signal received by the principal by expending a costly effort. However, only one signal is sent to the principal in their model, whereas we allow for multiple signals.

We will demonstrate that the availability of multiple signals has powerful implications for implementation in costly distortion environments. Our first result shows that any allocation profile is implementable when the type space is finite.

Theorem 3 *Suppose that the space of utility parameters Θ is finite, and that there exist $\alpha \in [0,1)$ and $\underline{c} > 0$ s.t. $C^n(m_1, ..., \gamma_i(\theta'), ..., m_n, \theta) - C^n(m_1, ..., \gamma^i(\theta), ..., m_n, \theta) \geq \frac{\underline{c}}{n^\alpha}$ for all $\theta, \theta' \in \Theta$, $\theta \neq \theta'$ and $m_j \in M^j$ where $j = 1, ..., n$.*

Then there exists $N < \infty$ such that if the agent has to send at least N messages in the mechanism, every allocation profile $x : \Theta \to X$ is implementable, at zero communication cost.

Proof: Fix an allocation profile $x(\theta)$, and consider a mechanism in which the agent is asked to send n messages. The outcome function $g(.)$ maps the arrays of messages sent by the agent into allocations as follows:

$$g(m_1, ..., m_n) = \begin{cases} x(\theta) \text{ if } m_i = \gamma^i(\theta), \ \ \forall i \in 1, ..., n \\ \underline{x}, \quad \text{otherwise.} \end{cases} \tag{2}$$

Since $C^n(\gamma^1(\theta), ..., \gamma^n(\theta), \theta) = 0$, the agent of type θ cannot improve her payoff by sending a vector of messages such that $m_i \neq \gamma^i(\theta')$ for some i and all $\theta' \in \Theta$.

If the agent of type θ sends a vector of messages $(\gamma^1(\theta'), ..., \gamma^n(\theta'))$ she obtains a payoff $u(x(\theta'), \theta) - C^n(\gamma^1(\theta'), ..., \gamma^n(\theta'), \theta)$. So, we need to show that for all θ and θ'

$$u(x(\theta'), \theta) - C^n(\gamma^1(\theta'), ..., \gamma^n(\theta'), \theta) \leq u(x(\theta), \theta)$$

Let $Z = \max_\theta \max_{x' \neq x} |u(x', \theta) - u(x, \theta)|$. From the continuity of $u(x, \theta)$ in x, the compactness of X, and the finiteness of Θ it follows that Z is finite. Note that $\forall \theta, \theta' \in \Theta$ s.t. $\theta \neq \theta'$ we have: $C^n(\gamma^1(\theta'), ..., \gamma^n(\theta'), \theta) \geq n^{1-\alpha}\underline{c}$. Let N be the smallest integer s.t. $N \geq \left(\frac{Z}{\underline{c}}\right)^{1/1-\alpha}$. Then the above incentive constraint holds if $n \geq N$. No communication costs are incurred because in equilibrium the agent of type θ sends a collection of messages $(m_1, ..., m_n)$ such that $m_i = \gamma^i(\theta)$ for all i. *Q.E.D.*

Note that the cost of misrepresenting one's type by sending message $m_i \neq \gamma^i(\theta)$ could be very small, and decreasing in the number of messages. But provided the speed with which this cost converges to zero as the number of messages increases is not too fast, it becomes prohibitively expensive for the agent to imitate other types when the number of messages (tests) that must be sent (undergone) is sufficiently large.

Finiteness of the type space Θ has played a crucial role in the proof above. From here on, we will assume that the space Θ has infinitely many elements. Theorem 3 demonstrates that it is easy to guarantee that non-local incentive constraints are satisfied, i.e. that an agent of type θ not be willing to imitate a type that is sufficiently different from θ. However, local incentive constraints preventing an agent from imitating types that are 'close' to hers pose a bigger problem.

We establish two kinds of results for this case. When the marginal cost of a small misrepresentation in signal i (slightly distorting signal i away from truth) is non-zero and does not vanish too fast, we show that the principal can still elicit the agent's information at zero cost. When the marginal cost of sending a non-truthful message is zero at $\gamma^i(\theta)$, implementation becomes harder. Nevertheless, when the marginal cost of small misrepresentation in signal i increases in the magnitude of misrepresentation at a rate that is bounded away from zero for each n, and does not go to zero too fast as n becomes large, the principal can still implement almost all allocation profiles. However, the agent has to incur a very small communication cost.

We will henceforth assume that Θ is a compact convex subset of a finite-dimensional Euclidean space. We will also adopt the following mild regularity assumptions. These conditions need to hold only for small misrepresentations, rather than globally. In other words, all that is required is that there be some 'start-up' costs of misrepresentation.

Assumption 1 *There exists* $L < \infty$ *s.t.* $\|x(\theta') - x(\theta)\| \leq L\|\theta' - \theta\|_\Theta$*, for all* $\theta', \theta \in \Theta$.

Assumption 2 *There exists* $r < \infty$ *s.t.* $||\gamma^i(\theta') - \gamma^i(\theta)||_i \geq r\|\theta' - \theta\|_\Theta$*,* $\forall \theta', \theta \in \Theta$*, and* i.

Assumption 3 *There exists* $K < \infty$ *s.t.* $|u(x', \theta) - u(x, \theta)| \leq K\|x' - x\|$*, for all* $\theta \in \Theta$.

Assumption 1 requires the allocation to be Lipschitz continuous with Lipschitz constant L, while Assumption 2 says that "truthful" messages are sufficiently sensitive to the type. Finally, Assumption 3 strengthens the continuity requirement on the agent's utility function to uniform continuity in x. Since $X \times \Theta$ is compact, Assumption 3 holds if $u(.,.)$ is C^1.

Our next theorem shows that all allocation profiles are implementable at zero cost when the marginal cost of misrepresenting type in signal i is positive and does not go to zero too quickly in the number of messages. Note that this condition has to hold only locally, i.e. for small misrepresentations.

Theorem 4 *Suppose that* $C^n(m_1, ..., m_i, ..., m_n, \theta) \geq C^n(m_1, ..., m_i', ..., m_n, \theta)$ *if* $||m_i - \gamma^i(\theta)||_i \geq ||m_i' - \gamma^i(\theta)||_i$*. Suppose furthermore that there exist* $\delta > 0$*,* $\alpha \in [0, 1)$*, and* $\underline{c} > 0$ *s.t.* $C^n(m_1, ..., m_i, ..., m_n, \theta) - C^n(m_1, ..., \gamma^i(\theta), ..., m_n, \theta) \geq \frac{\underline{c}}{n^\alpha}||m_i - \gamma^i(\theta)||_i$ *for all* $\theta \in \Theta$ *and* $(m_1, ..., m_n)$ *satisfying* $\max_{j=1,...,n} ||m_j - \gamma^j(\theta)||_j \leq \delta$.

Then there exists $N < \infty$ *such that any allocation profile* $x(\theta)$ *satisfying Assumption 1 is implementable with zero communication cost via a mechanism in which the agent has to send at least* N *messages.*

Proof: Fix an allocation $x(\theta)$ that satisfies Assumption 1. Consider the same mechanism as in Theorem 3. This mechanism implements allocation profile $x(\theta)$ if the following incentive constraint holds for all θ and θ':

$$u(x(\theta'), \theta) - u(x(\theta), \theta) \leq C^n(\gamma^1(\theta'), ..., \gamma^n(\theta'), \theta) - C^n(\gamma^1(\theta), ..., \gamma^n(\theta), \theta)$$

Using Assumptions 1-3, we obtain that this condition is satisfied if:

$$KL||\theta - \theta'||_\Theta \leq n^{1-\alpha}\underline{c} \min\{\delta, \min_{i=1,...,n} ||\gamma^i(\theta') - \gamma(\theta)||_i\}$$

Since Θ is compact, $\exists A > 0$ s.t. $||\theta - \theta'||_\Theta \leq A$. Using Assumption 2, we conclude that the above inequality holds if $n^{1-\alpha} \geq \frac{KL}{\underline{c}} \max\{\frac{A}{\delta}, \frac{1}{r}\}$. *Q.E.D.*

Intuitively, local incentive constraints are satisfied because the marginal benefit of sending a non-truthful message $m_i \neq \gamma^i(\theta)$ is finite for type θ, while the marginal cost of doing so is bounded from below by $\frac{c}{n^\alpha}$ when m_i is near $\gamma^i(\theta)$. Therefore, when the number of messages is sufficiently large, the cost of a misrepresentation exceeds the benefit of an improved allocation. This ensures that no type is willing to send non-truthful messages, and so no communication costs are incurred.

The above reasoning suggests that the set of implementable allocation profiles may be more restricted when the marginal cost of type misrepresentation is zero near the 'truthful' message $\gamma^i(\theta)$. Indeed, implementation in this case will generally require that the agent send some "non-truthful" messages and incur some misrepresentation costs. Nevertheless, we will demonstrate that by carefully constructing the communication stage of the mechanism, the principal can elicit the agents' private information at a negligible cost.

For technical convenience, we will adopt a number of simplifying assumptions in this case. First, we specialize our model to a one-dimensional type space, i.e. $\Theta = [\overline{\theta}, \underline{\theta}]$. We also partition the outcome x into a production/consumption assignment $q \in Q$, where Q is a compact subset of $\mathbf{R}^l$, and a transfer p. Thus, $x = (q, p)$. Further, the agent's utility function is taken to be quasilinear in transfers, i.e.:

$$u(x, \theta) = v(q, \theta) + p$$

We also assume that $v(q, \theta)$ is twice continuously differentiable, and all messages have the same cost structure i.e. $M_i = M$, the message space M is a compact, convex subset of $\mathbf{R}$, $\gamma^i(\theta) =$

$\gamma^j(\theta) = \gamma(\theta)$, and that the cost function is symmetric in messages, i.e. $C^n(m_1, ..., m_i, ..., m_j, ..., m_n, \theta) = C^n(m_1, ..., m_j, ..., m_i, ..., m_n, \theta)$ for all n and $i, j = 1, ..., n$. Our earlier assumption that the message space is sufficiently large translates into the following one: $\exists \rho > 0$ s.t. $\inf_{\theta\in\Theta} \gamma(\theta) - \rho \in M$ and $\sup_{\theta\in\Theta} \gamma(\theta) + \rho \in M$. Finally, we assume that $\gamma(\theta)$ is continuously differentiable. Then, by Assumption 2, $|\gamma'(\theta)| \geq r \ \forall \theta \in [\underline{\theta}, \overline{\theta}]$.

When focussing on a particular message m_i, we will write the cost function as $C^n(m_i, m_{-i}, \theta)$ where m_{-i} stands for the vector of $n-1$ messages other than message i. We make the following assumptions on the cost function:

Assumption 4 *$C^n(m_1, ..., m_n, \theta)$ is twice continuously differentiable.*
(i) $C_i^n(m_i, m_{-i}, \theta) = 0$ if $m_i = \gamma(\theta)$.
There exist $\delta > 0$, $\overline{\omega}$, $\underline{\omega} > 0$ and $\alpha \in [0, 1)$ s.t. for all θ, and $m_i, m_j \in [\gamma(\theta) - \delta, \gamma(\theta) + \delta]$, we have:
(ii) $\underline{\omega}\frac{1}{n^\alpha} \leq C_{ii}^n(m_i, m_{-i}, \theta) \leq \overline{\omega}\frac{1}{n^\alpha}$.
(iii) $C_{i\theta}^n(m_i, m_{-i}, \theta) < 0$ and $\underline{\omega}\frac{1}{n^\alpha} \leq |C_{i\theta}^n(m_i, m_{-i}, \theta)| \leq \overline{\omega}\frac{1}{n^\alpha}$.
(iv) $C_{\theta\theta}^n(m_i, m_{-i}, \theta) \geq -\kappa$ for some $0 \leq \kappa < \infty$.

We then have the following theorem:

Theorem 5 *For every continuously differentiable allocation profile $\{q(\theta), p(\theta)\}$ and every $\varepsilon > 0$, there exist $N < \infty$ and transfer rule $p^\varepsilon(\theta)$ satisfying $|p^\varepsilon(\theta) - p(\theta)| < \varepsilon$ for all θ such that the allocation profile $\{q(\theta), p^\varepsilon(\theta)\}$ is implementable via a mechanism in which the agent sends at least N messages. Furthermore, the total communication cost incurred by the agent in this mechanism does not exceed ε.*

Proof: see the Appendix.

In the mechanism that we use in the proof an agent with preference parameter θ sends n_1 costless "truthful" messages $\gamma(\theta)$ and n_2 costly messages $s(\theta) \neq \gamma(\theta)$. So, $p^\varepsilon(\theta)$ is obtained by adding the communication cost $C^{n_1+n_2}(s(\theta), ..., s(\theta), \gamma(\theta), ..., \gamma(\theta), \theta)$ to the target transfer $p(\theta)$ plus possibly a small adjustment term. Then, absent the adjustment term, the first-order condition necessary for local incentive compatibility becomes:

$$\frac{dp(\theta)}{d\theta} = -\nabla_q v(q(\theta), \theta)^\tau \nabla q(\theta) + C_\theta^{n_1+n_2}(s(\theta), ..., s(\theta), \gamma(\theta), ..., \gamma(\theta), \theta)) \tag{3}$$

(where the superscript τ denotes the transpose of a vector).

Since $C_i^n(m_1,...,m_i,...,m_n,\theta) = 0$ for all i s.t. $m_i = \gamma(\theta)$ and $C^n(\gamma(\theta),...,\gamma(\theta),\theta) = 0$, we also have $C_\theta^n(\gamma(\theta),...,\gamma(\theta),\theta) = 0$. So, in the absence of costly messages $s(\theta) \neq \gamma(\theta)$ the second term in (3) is zero, in which case this first-order condition is standard. In particular, it imposes a tight link between $q(\theta)$ and $p(\theta)$. Costly messages $s(\theta)$ eliminate the need for any such link. By choosing n_2 and $s(\theta)$ appropriately we can ensure that (3) holds for arbitrary $(q(\theta), p(\theta))$.

Further, the convexity of $C(.)$ under small misrepresentations implies that we can choose n_2 large enough that (3) holds, yet the agent's communication cost $C^{n_1+n_2}(s(\theta),...,s(\theta),\gamma(\theta),...,\gamma(\theta),\theta)$ is less than the chosen ϵ.

Finally, a sufficiently large number n_1 of costless truthful messages guarantees that the agent's payoff is quasi-concave in messages m. So, truthful messages allow the second-order conditions to be satisfied and ensure global incentive compatibility without imposing any restrictions on the set of implementable allocation profiles.

It is instructive to compare Theorem 5 with the results in Maggi and Rodriguez-Clare (1995) (referred to as MR below) who study a similar model under the restriction that the agent can send only one message concerning her type. Although MR focus on finding the optimal (profit-maximizing) mechanism, and we are concerned with implementability, comparing the necessary and sufficient conditions in the two models allows the reader to appreciate the full effect of multiple signals.

The necessary first-order condition for implementation in MR is equivalent to (3) with $n_2 = 1$ (see condition (N) in Lemma 1 of MR). Thus, communication in MR also allows to weaken the link between the allocation $q(\theta)$ and the corresponding transfer, and hence to implement a larger set of social choice functions. But the degree to which this link is weakened is limited, because the magnitude of the required misrepresentation $s(\theta)$ and associated communication costs could be quite large when $n_2 = 1$. In contrast, when n_2 is sufficiently large, co-dependence between $q(\theta)$ and $p(\theta)$ is eliminated at a very small cost.

Finally, consider the second-order conditions for implementation. To satisfy them, MR have to impose the following restrictions: $q'(\theta) \geq 0$ and $s'(\theta) \geq 0$. In contrast, as explained above, in our model second-order conditions hold because the agent also has to send a number of 'truthful' costless messages $m_i = \gamma(\theta)$. Consequently, we are able to implement any continuously differentiable allocation profile at a small (communication) cost.

The results of this section have a number of interesting implications for screening and signaling. In particular, consider a standard job-market screening problem, i.e. an employer hiring an employee of unknown ability. It many real-world situations employers ask job-candidates to undergo a number of different tests or interviews. The tests or interviews can be regarded as multiple signals or messages sent by the job-candidate. Our results indicate that the problem of asymmetric information regarding the employee's ability can be overcome. The employer can learn the employee's ability at a small or negligible cost, if the tests/interviews that the job-candidate goes through have the following properties: (i) Each test identifies the employee's ability accurately, if the employee does not attempt to manipulate the results of the test by expending effort. (ii) The employee incurs some cost of effort when she attempts to misrepresent her type, and the cost of effort is positive, increasing in the magnitude of desired misrepresentation on a test, and does not go to zero too fast in the number of tests, i.e. the learning process is not too fast.

In our framework, the marginal cost of a signal may depend on the content and number of other signals sent by the agent. For example, the amount of effort that an agent with ability θ has to expend in order to perform at a level corresponding to ability $\theta' \neq \theta$ in the n-th test, may depend on how much effort she had put into preparation for other tests and how many other tests she has taken. Each course or test may become shorter and require less effort as their number increases. Still, our results hold when the effect of the true ability θ on the cost of sending signal $m \neq \gamma(\theta)$ does not go to zero too quickly in n.

Our analysis can explain why employers may prefer to test the job-applicants very thoroughly before offering them a job, rather than to present them with self-selecting menus of contracts right away, as suggested by the literature on screening, and why incentives given to the employees on the job are not as strong as predicted by the contracting literature. Indeed, we show that if a job candidate has to go through a large number of tests or interviews, it will be too costly for her to undertake effort sufficient to significantly misrepresent her ability.

So, a testing procedure constructed along the lines suggested in this section will produce an accurate estimate of the candidate's ability, and an employer would not have to perform much additional on-the-job screening by offering high-powered incentives and menus of contracts.

Conceivably, the principal or the agent may also have to incur a fixed cost in association with each test or interview. The presence of such costs would limit the feasible number of interviews/tests, and could make perfect screening too costly. Nevertheless, if the fixed costs per test are not too high, multi-test procedures would still dominate the ones relying on a single test. An optimal policy would then involve some combination of pre-hiring testing and on-the-job screening.

Furthermore, it is plausible that the fixed costs are associated with a particular test, and not a particular job-candidate. Such test-specific fixed costs amortized over all the job-candidates who undergo it, create less of an obstacle for increasing the number of tests. Then, our model will predict that larger firms who interview a large numbers of applicants, will have more thorough testing and evaluation procedures and rely less on the on-the-job- screening than smaller ones. This prediction appears to be broadly consistent with reality.

Indeed, the interviewing process in many professional job-markets appears to be consistent with the idea of requesting multiple messages/signals from a candidate, with each signal being of somewhat different nature. For example, in the context of a departmental visit on the academic job-market a prospective candidate meets with faculty members working in different fields. Each conversation provides a separate signal of the candidate'a ability, because different faculty members, especially if they work in different fields, assess the candidate from different perspectives. Similarly, consulting firms and investment banks have developed rigorous selection procedures involving multiple rounds of interviews. Some interviews involve solving cases, while others involve conversations and discussions with associates, managers, and partners.

4 Conclusions

In this paper we have developed a general approach to implementation in environments where agents have limited abilities to misrepresent information or incur some costs while doing so.

We have demonstrated that in such environments the ability of the principal to offer mechanisms in which an agent sends several messages significantly expands the set of implementable outcomes. In some cases, the principal may be able to uncover the truth at a very low, or negligible, cost. In particular, we show that this is the case when misrepresentation requires distorting multiple aspects of the truth or producing multiple pieces of false evidence, even if each new piece of evidence involves progressively less additional cost. Thus, our research provides a deeper understanding of the conditions under which the problem of asymmetric information can be effectively dealt with.

In instances where the principal cannot completely overcome the information asymmetry, our work also has interesting implications. In particular, we have shown that an individual's communication abilities can play an important role in determining her payoff. Further, we have argued that it may not be optimal to punish individuals who make conflicting statements and, thus, can be easily accused of lying. By not punishing individuals who make such conflicting statements, the mechanism prevents them from imitating others, which enhances the overall social efficiency of the outcome.

5 Appendix

Proof of theorem 5:

Fix a continuously differentiable allocation profile $\{q(\theta), p(\theta)\}$ and some $\varepsilon > 0$.

Given two positive integers n_1 and n_2, a punishment allocation $(\underline{q}, \underline{p})$ which gives any type θ a lower payoff than $v(q(\theta), \theta) + p^{\epsilon}(\theta) - \epsilon$, and a "misrepresentation" message profile $s(\theta) : \Theta \mapsto M$, let us define a mechanism $G_{n_1,n_2}(g(.), s(.))$ as follows. Let $(m_1(\theta), ..., m_{n_1+n_2}(\theta))$ be a vector of messages such that:

$$m_1(\theta) = m_2(\theta) = ... = m_{n_1}(\theta) = s(\theta) \text{ and } m_{n_1+1}(\theta) = ... = m_{n_1+n_2}(\theta) = \gamma(\theta)$$

Mechanism $G_{n_1,n_2}(g(.), s(.))$ maps a vector of $n_1 + n_2$ messages $(m_1, ..., m_{n_1+n_2})$ sent by the agent into an outcome according to the following outcome function $g(.)$:

$$g(m_1, ..., m_n) = \begin{cases} (q(\theta), p^{\epsilon}(\theta)), & \text{if } (m_1, ..., m_{n_1+n_2}) = (m_1(\theta), ..., m_{n_1+n_2}(\theta)) \text{ for some } \theta \in \Theta \\ (\underline{q}, \underline{p}), & \text{otherwise.} \end{cases}$$

We will prove that there exist n_1, n_2, $s(.)$ and a transfer function $p^\epsilon(\theta)$ satisfying $\sup_\theta |p^\epsilon(\theta) - p(\theta)| < \epsilon$ such that $G_{n_1,n_2}(g(.), s(.))$ implements the allocation profile $\{q(\theta), p^\epsilon(\theta)\}$.

First, we set up some simplifying notation. Let $c(x, y, \theta) = C^{n_1+n_2}(m_1, ..., m_{n_1+n_2}, \theta)$, where $m_i = x$ for all $i \in \{1, ..., n_1\}$, and $m_i = y$ for all $i \in \{n_1 + 1, ..., n_1 + n_2\}$. Observe that, for brevity, we have omitted the dependence of $c(x, y, \theta)$ on n_1 and n_2. However, this dependence will always be accounted for. In particular, differentiation yields $c_1(x, y, \theta) = n_1 C_i^{n_1+n_2}(x, ..., x, y, ..., y, \theta)$ for all $i \in \{1, ..., n_1\}$, $c_2(x, y, \theta) = n_2 C_i^{n_1+n_2}(x, ..., x, y, ..., y, \theta)$ for all $i \in \{n_1 + 1, ..., n_1 + n_2\}$, and $c_3(x, y, \theta) = C_\theta^{n_1+n_2}(x, ..., x, y, ..., y, \theta)$.

Next, we construct $s(\theta)$. To this effect, define $U(\theta) = v(q(\theta), \theta) + p(\theta)$. By assumption $U(.)$ is a C^1 function, so we can select $U^\varepsilon(.) \in C^2$ s.t. $\sup_\theta |U^\varepsilon(\theta) - U(\theta)| < \varepsilon/2$. Now fix some $b > 0$. Then there exists $N_s > 0$ s.t. for all $n_1 + n_2 \geq N_s$ with $n_2/n_1 \leq b$ there is a unique $s(\theta) \in [\gamma(\theta) - \delta/2, \gamma(\theta) + \delta/2]$ satisfying:

$$c_3(s(\theta), \gamma(\theta), \theta) = v_\theta(q(\theta), \theta) - U_\theta^\varepsilon(\theta) \tag{4}$$

The existence and uniqueness of such $s(\theta)$ follows from two observations. First, by Assumption 4 (v), $c_3(s, \gamma(\theta), \theta)$ is decreasing in s and is positive (respectively, negative/zero) if $s < \gamma(\theta)$ (respectively, $s > \gamma(\theta)/s = \gamma(\theta)$). Second, for $s \in (\gamma(\theta) - \delta, \gamma(\theta)) \cup (\gamma(\theta), \gamma(\theta) + \delta)$, we have $|c_3(s, \gamma(\theta), \theta)| \geq n_1 |s - \gamma(\theta)| \underline{\omega} \frac{1}{(n_1+n_2)^\alpha} \geq \frac{1}{1+b}(n_1 + n_2)^{1-\alpha} \underline{\omega} |s - \gamma(\theta)|$, which grows without bound as $n_1 + n_2$ gets large.

In particular, let $E = \max_{\theta \in \Theta} |v_\theta(q(\theta), \theta) - U_\theta^\varepsilon(\theta)|$. Since $q(.)$, $U^\varepsilon(.)$ and $v(.,.)$ are C^1 functions, we have $E < \infty$. Then it follows from (4) that $|s(\theta) - \gamma(\theta)| \leq \frac{(1+b)E}{\underline{\omega}(n_1+n_2)^{1-\alpha}}$, which is less than $\delta/2$ when $n_1 + n_2$ is large enough.

Combining this bound with Assumption 4 yields, $c(s(\theta), \gamma(\theta), \theta) \leq \frac{(s(\theta)-\gamma(\theta))^2 \overline{\omega}(n_1+n_2)^{1-\alpha}}{2(1+b)} \leq \frac{E^2 \overline{\omega}}{2(1+b)\underline{\omega}^2 (n_1+n_2)^{1-\alpha}}$. So, there exists $N_\epsilon > 0$ s.t. $c(s(\theta), \gamma(\theta), \theta) < \epsilon/2$ if $n_1 + n_2 \geq N_\epsilon$ and $n_2/n_1 \leq b$.

Since $v(\cdot, \cdot)$ and $U^\varepsilon(\cdot)$ are C^2 functions, $q(.)$ is a C^1 function, and Θ is compact, the right-hand side of (4) is a C^1 function of θ, and the absolute value of its derivative is uniformly bounded from above by some positive constant $\hat{V} < \infty$. Furthermore, since $c(x, y, \theta)$ is C^2 function, $|c_{13}(s, \gamma(\theta), \theta)| \geq \underline{\omega} \frac{n_1}{n^\alpha}$ for all $s \in [\gamma(\theta) - \delta, \gamma(\theta) + \delta]$, and $c_{33}(s, \gamma(\theta), \theta) \geq -\kappa$, it follows that $s'(\theta)$ is a continuous function satisfying:

$$|s'(\theta)| \leq \frac{\hat{V}+\kappa}{\underline{\omega}\frac{n_1}{n^\alpha}} \tag{5}$$

Next, let $p^\epsilon(\theta) = U^\varepsilon(\theta) - v(q(\theta),\theta) + c(s(\theta),\gamma(\theta),\theta)$. In the remainder of the proof we will show that n_1 and n_2, can be chosen in such a way that it is optimal for the agent of type θ to send n_1+n_2 messages satisfying $m_i(\theta) = s(\theta)$ for $i \in \{1,...,n_1\}$ and $m_i(\theta) = \gamma(\theta)$ for $i \in \{n_1+1,...,n_1+n_2\}$, i.e. for all $\theta,\theta' \in \Theta$ we have:

$$U^\varepsilon(\theta) \geq U^\varepsilon(\theta') + v(q(\theta'),\theta) - v(q(\theta'),\theta') + c(s(\theta'),\gamma(\theta'),\theta') - c(s(\theta'),\gamma(\theta'),\theta) \tag{6}$$

At first, suppose that $|\gamma(\theta') - \gamma(\theta)| > \delta$. Then rewrite (6) as follows:

$$c(s(\theta'),\gamma(\theta'),\theta) - c(\gamma(\theta),\gamma(\theta),\theta) + c(\gamma(\theta),\gamma(\theta),\theta) - \epsilon/2 \geq \int_\theta^{\theta'} (U^\varepsilon_\theta(z) - v_\theta(q(\theta'),z))dz \tag{7}$$

Note that $c(s(\theta'),\gamma(\theta'),\theta) - c(\gamma(\theta),\gamma(\theta),\theta) \geq \frac{(\gamma(\theta')-\gamma(\theta))^2\underline{\omega}n_2}{2(n_1+n_2)^\alpha} \geq \frac{\delta^2\underline{\omega}n_2}{2(n_1+n_2)^\alpha}$, while $(U^\varepsilon_\theta(z) - v_\theta(q(\theta'),z)) \leq \sup_{\theta,\theta'\in\Theta\times\Theta}|U^\varepsilon_\theta(z) - v_\theta(q(\theta'),z)| < \infty$.

When $|\gamma(\theta') - \gamma(\theta)| \leq \delta$, use (4) to rewrite (6) as follows:

$$\int_\theta^{\theta'} (v_\theta(q(z),z) - v_\theta(q(\theta'),z) + c_3(s(\theta'),\gamma(\theta'),z) - c_3(s(z),\gamma(z),z))dz \leq 0 \tag{8}$$

Note that when $\theta' > \theta$, we have

$$c_3(s(\theta'),\gamma(\theta'),z) - c_3(s(z),\gamma(z),z) = \int_z^{\theta'} c_{13}(s(x),\gamma(x),z)s'(x) + c_{23}(s(x),\gamma(x),z)\gamma'(z)dx \leq$$
$$\left(n_1\overline{\omega}\frac{\hat{V}+\kappa}{\underline{\omega}\frac{n_1}{(n_1+n_2)^\alpha}} - n_2 r\underline{\omega}\right)\frac{(\theta'-z)}{(n_1+n_2)^\alpha} = -\left(\frac{n_2 r\underline{\omega}}{(n_1+n_2)^\alpha} - (\hat{V}+\kappa)\frac{\overline{\omega}}{\underline{\omega}}\right)(\theta'-z)$$

Let $F = \max_{\theta,\theta'\in\Theta\times\Theta}|v_\theta(q(z),z) - v_\theta(q(\theta'),z)|$. Collecting terms we obtain that (8) holds if:

$$-\left(\frac{n_2 r\underline{\omega}}{(n_1+n_2)^\alpha} - (\hat{V}+\kappa)\frac{\overline{\omega}}{\underline{\omega}}\right) + F \leq 0$$

Observe that for every $h \in (0,1)$ there exists $\hat{N}(h)$ s.t. the latter inequality holds when $n \geq \hat{N}(h)$ and $n_2 \geq hn_1$.

The case of $\theta' < \theta$ is handled in a similar way. Thus, we conclude that the mechanism $G_{n_1,n_2}(.)$ implements the allocation profile $(q(\theta),p^\epsilon(\theta))$ when $b \leq n_1/n_2 \leq \frac{1}{h}$, and $n_1+n_2 \geq \bar{N} \equiv \max\{N_\epsilon, \hat{N}(h), \tilde{N}(h)\}$. For example, we can choose $1/2-\zeta \leq n_1/n_2 \leq 1/2+\zeta$ for some small ζ and then set $\bar{N}$ appropriately. *Q.E.D.*

5.1 Multi-agent Environment.

In this Appendix we extend the binary cost model of Section 2 to a multi-agent environment. So, suppose that, instead of one, there are k agents. To the extent possible, we maintain the same notation as in Section 2. Particularly, X denotes the space of allocations, and $\mathcal{C}$ denotes the admissible message space which is assumed to be finite. Agent i's preference parameter and feasible message set are denoted by θ_i and $\mathcal{M}_i$ respectively, with $\theta_i \in \Theta_i$ and $\mathcal{M}_i \in \mathcal{N}_i \subset 2^{\#\mathcal{C}}$. Thus, agent i's type $t_i = (\theta_i, \mathcal{M}_i)$ lies in $T_i \subset \Theta_i \times 2^{\#\mathcal{C}}$, while her utility function is $u_i : X \times \Theta \to \mathbb{R}$.

Let $t = (t_1, ..., t_n)$ denote the profile (vector) of agents' types, and $T \subset T_1 \times ... \times T_n$ denote the type space. Note that this inclusion could be strict because some type profiles may be infeasible. T is finite because Θ_i and $\mathcal{C}$ are. The prior probability $F(.)$ over T is common knowledge among the agents and the principal. The marginal probability over i's types is denoted by $F_i(\theta_i, \mathcal{M}_i)$. Then $I_i(\mathcal{M}_i) \subset \Theta_i$ is defined as follows: $\theta_i \in I_i(\mathcal{M}_i)$ if and only if $F_i(\theta_i, \mathcal{M}_i) > 0$. To focus on a particular player i, we will also use the notation $(\theta_i, \mathcal{M}_i, \theta_{-i}, \mathcal{M}_{-i})$ for a profile of types where $(\theta_i, \mathcal{M}_i)$ is player i's type, and $(\theta_{-i}, \mathcal{M}_{-i})$ is a profile of types of players other than i.

A social choice function for this environment is a mapping from the set of types into the set of outcomes, denoted by $f : T \to X$. The outcome function of the mechanism $G(.)$ in the multi-agent environment is defined as follows. In the communication stage, each agent is asked to report all her feasible messages. Let M_i be the set of messages sent by agent i. If $F(M_1, .., M_k) = 0$, i.e. the received message profile is inconsistent with the prior distribution of types, then $G(M_1, ..., M_n) = \underline{x}$.

If $F(M_1, .., M_k) > 0$, then the communication stage of the mechanism is followed by a simultaneous move game in which player i has to choose one vector in the set $\{f(\theta_i, M_i, \theta_{-i}, \mathcal{M}_{-i}) | \theta_i \in I_i(M_i), (\theta_i, M_i, \theta_{-i}, \mathcal{M}_{-i}) \in T\}$.[23] Finally, after all agents simultaneously make their choices from their respective menus, the mechanism assigns the allocation

[23]Note that the entries in the vector $\{f(\theta_i, M_i, \theta_{-i}, \mathcal{M}_{-i})\}$ correspond to *all* $(\theta_{-i}, \mathcal{M}_{-i})$ s.t. $(\theta_i, M_i, \theta_{-i}, \mathcal{M}_{-i}) \in T$. So, agent i is not able to infer other agents' reports M_{-i} from the menu that she is offered.

$f(\hat{\theta}_i, M_i, ..., \hat{\theta}_k, M_k)$ where $\hat{\theta}_i$ corresponds to the vector chosen by player i.

We say that mechanism $G(\cdot)$ $\mathcal{M}-$truthfully implements the social choice function $f(.)$ if this mechanism has a Bayesian Nash equilibrium in which each agent i sends all her feasible messages, i.e. reports $M_i = \mathcal{M}_i$ and then chooses the vector corresponding to her true preference parameters θ_i.[24] The following Theorem represents an extension of Theorem 1 to the multi-agent setting:

Theorem 6 *Any implementable social choice function is $\mathcal{M}-$truthfully implementable via mechanism $G(\cdot)$.*

Proof: Suppose that the social choice function $f(.)$ is implementable in Bayesian Nash equilibrium via some mechanism $(S, g(.))$, with strategy space S and outcome function $g: S \mapsto X$. Let $s_i(t_i)$ be an equilibrium strategy of type $t_i = (\theta_i, \mathcal{M}_i)$ in this mechanism, and let $s(t) = (s_1(t_1), ..., s_n(t_n))$. Since the mechanism $(S, g(.))$ implements $f(.)$, we have $f(t) = g(s(t))$.

Let us establish that mechanism $G(.)$ $\mathcal{M}-$truthfully implements $f(.)$. For this, we need to show that in mechanism $G(.)$ it is an optimal strategy for agent i of type $t_i = (\theta_i, \mathcal{M}_i)$ to report $M_i = \mathcal{M}_i$ and choose the row corresponding to θ_i when all other agents follow such $\mathcal{M}-$truthful strategies. Suppose to the contrary that some type t_i can obtain a higher expected payoff in mechanism $G(.)$ by sending a collection of messages $M_i' \subset \mathcal{M}_i$ and choosing a row corresponding to $\theta_i' \in I_i(M_i')$ when all other agents follow $\mathcal{M}-$truthful strategies. Then $s_i(\theta_i', M_i')$ cannot be feasible for the agent of type $(\theta_i, \mathcal{M}_i)$ in the mechanism $(S, g(.))$. Therefore, the strategy $s_i(\theta_i', M_i')$ must involve sending a message m_i s.t. $m_i \in M_i'$, but $m_i \notin \mathcal{M}_i$. But then type $t_i = (\theta_i, \mathcal{M}_i)$ could not send the collection of messages M_i' in the mechanism $G(\cdot)$. Contradiction. *Q.E.D.*

Theorem 6 implies that an implementable allocation profile has to satisfy the following

[24] As in the single-agent case, the menu stage can be dispensed with if $\theta_i \in \mathcal{M}_i$ for all $(\theta_i, \mathcal{M}_i) \in T_i$, i.e. truthful announcements of the preference parameter are always in the message set. Then mechanism $G(.)$ can be replaced with a simpler mechanism $\tilde{G}(.)$ in which the messages are reported in a sequence, and the first message sent by agent i is interpreted as an announcement of her true preference parameter θ_i. Formally, $\tilde{G}(M_1, ..., M_k) = f(\theta_1, M_1, ..., \theta_k, M_k)$ if θ_i is the first message in M_i $\forall i \in 1, .., k$ and $(\theta_1, M_1, ..., \theta_k, M_k) \in T$ and $\tilde{G}(M_1, ..., M_k) = \underline{x}$ otherwise.

incentive constraints for each i and $t_i = (\theta_i, \mathcal{M}_i) \in T_i$:

$$\sum_{t_{-i} \in T_{-i}} u(G(t_i, t_{-i}), \theta) F_i(t_{-i}|t_i) \geq \sum_{t_{-i} \in T_{-i}} u(G(t'_i, t_{-i}), \theta) F_i(t_{-i}|t_i), \text{ for all } t'_i = (\theta'_i, \mathcal{M}'_i) \text{ s.t. } \mathcal{M}'_i \subset \mathcal{M}_i \tag{9}$$

A complete study of implementation in the multiple agent case is beyond the scope of this paper. Instead, we will focus on showing how the principal can exploit the interdependencies between the agents' types to satisfy (9) and extract their private information more effectively. In this regard, we will highlight two issues. First, we derive conditions allowing the principal to detect a misrepresentation of the payoff-relevant information (i.e. θ's) by any agent with a positive probability. Second, we study how correlation between agents' types coupled with limits on communication affects the principal' task of extracting the agents' information.

First, consider the following version of the Non-Nested Range Condition ($NNRC$) for the multi-agent case:

NNRCM: *Suppose that* $(\theta_1, \mathcal{M}_1, ..., \theta_k, \mathcal{M}_k)$ and $(\theta'_1, \mathcal{M}'_1, ..., \theta'_k, \mathcal{M}'_k) \in T$. If $\theta_i \neq \theta'_i$ for some i, then there exists j s.t. $\mathcal{M}_j \setminus \mathcal{M}'_j \neq \emptyset$.

Then we have:

Lemma 2 *If NNRCM holds, then any social choice function* $f(\theta_1, ..., \theta_k) : \Theta \mapsto X$ *is implementable in Bayesian Nash Equilibrium.*

To establish the lemma, suppose that all agents but i follow $\mathcal{M}$-truthful strategies in mechanism $G(.)$. If agent i follows a strategy of misrepresenting her θ, and instead, pretends to have preference parameter θ', then with probability 1 the reported type profile will be inconsistent, i.e. will not be in the set of feasible types T, and the mechanism will assign allocation $\underline{x}$ in this case.

Under NNRCM, an agent's deviation from reporting her preference parameter truthfully is detected with probability 1, because for each profile of types and each agent i there exists a "test" agent j whose truthful reporting identifies that some agent must be lying when agent i actually lies. All agents are punished as a consequence, but the punishment $\underline{x}$ does not have to be too large: it just has to be worse than the equilibrium allocation for any agent type.

We can weaken NNRCM by providing the following condition under which an agent's deviation is detected with some positive probability, but not necessarily for sure:

WNNRCM: Consider some $(\theta_i, \mathcal{M}_i) \in T_i$ and $\theta_i' \neq \theta_i$. Then for all $\mathcal{M}_i' \subseteq \mathcal{M}_i$ there exists $(\theta_{-i}, \mathcal{M}_{-i})$ such that $F(\theta_i, \mathcal{M}_i, \theta_{-i}, \mathcal{M}_{-i}) > 0$ but $F(\theta_i', \mathcal{M}_i', \theta_{-i}, \mathcal{M}_{-i}) = 0$.

Essentially, **WNNRCM** says that when θ_i shifts to θ_i', then in some feasible state of the world some component of $\mathcal{M}_{-i}$ shifts as well.

Under **WNNRCM** the result of Lemma 2 holds provided the principal has sufficiently large punishments at her disposal (for example, when monetary transfers are available and agents do not have limited liability). Large punishments may be required because an agent's deviation is detected with a positive probability, but not necessarily for sure.

Example 5 *A special case in which **WNNRCM** holds is a perfect information environment (i.e. $\theta_1 = .. = \theta_n = \theta$), where one or more agents are 'honest' (i.e. unable to misrepresent θ) with some probability. Then the result of Lemma 2 can be strengthened. It is easy to show that any social choice function $f(\theta)$ is uniquely implementable in Nash equilibrium if two or more agents are 'honest' with a positive probability. If only one agent can be 'honest,' then we can show that for any social choice function $f(\theta)$ there is a social choice function which is arbitrarily close to $f(.)$ and is uniquely implementable in Nash equilibrium.*

Consider another example where **WNNRCM** holds.

Example 6 *Suppose that the profile of preference parameters $(\theta_1, ..., \theta_k)$ (e.g the agents' costs, or abilities) determines/generates the aggregate set of feasible messages (all available evidence), which is dispersed throughout the population in a random way according to $F(.)$. Thus, there exists a function $\mathcal{M}^a(.)$ such that $F(\theta_1, \mathcal{M}_1, ..., \theta_k, \mathcal{M}_k) > 0$ if and only if $\mathcal{M}^a(\theta_1, ..., \theta_k) = \cup_{i=1,...,k} \mathcal{M}_i$.*

*Then, **WNNRCM** reduces to the following condition: for all $(\theta_i, \mathcal{M}_i) \in T_i$ and $\hat{\theta}_i \in \Theta_i$, there exist $(\theta_{-i}', \mathcal{M}_{-i}')$ and $(\theta_{-i}'', \mathcal{M}_{-i}'')$ s.t. $(\theta_i, \mathcal{M}_i, \theta_{-i}', \mathcal{M}_{-i}') \in T$, $(\theta_i, \mathcal{M}_i, \theta_{-i}'', \mathcal{M}_{-i}'') \in T$, and either $\mathcal{M}^a(\hat{\theta}_i, \theta_{-i}') \setminus (\mathcal{M}_{-i}' \cup \mathcal{M}_i) \neq \emptyset$ or $\mathcal{M}^a(\hat{\theta}_i, \theta_{-i}') \setminus \mathcal{M}_{-i}' \neq \mathcal{M}^a(\hat{\theta}_i, \theta_{-i}'') \setminus \mathcal{M}_{-i}''$.*

If this condition holds, then any deviation by agent i which involves reporting some $\hat{\theta}_i$ different from her true θ_i causes the reported profile $(\hat{\theta}_1, \hat{\mathcal{M}}_1, ..., \hat{\theta}_k, \hat{\mathcal{M}}_k)$ to be inconsistent *with a positive probability, i.e.* $\mathcal{M}^a(\hat{\theta}_1, ..., \hat{\theta}_n) \neq \cup_{i=1,...,k}\hat{\mathcal{M}}_i$. *So, the principal can implement any social choice function* $f(.) : \Theta \mapsto X$ *provided that a sufficiently severe punishment is available.*

Finally, let us consider one more aspect of the multiagent model. Even though the principal may not be able to detect a misrepresentation with certainty, she may nevertheless be able to exploit the statistical properties of the distribution of the agents' types. In fact, it is natural for an agent's preference parameter as well as her set of feasible messages to be correlated with the other agents' types. So, the set of feasible messages may be regarded as an additional signal containing statistical information about the types of other agents. For the sake of completeness, let us explore how such statistical dependence affects the principal's ability to extract agents' information. As we demonstrate below, this effect is ambiguous.

To emphasize the link with the existing literature, we will consider the transferable utility case and focus on the issue of surplus extraction. Thus, we assume that any agent i's utility function is given by $v_i(x, \theta) - p_i$ where x is a 'physical' part of the allocation and p_i is agent i's payment to the principal. We assume that $v_i(.)$ has standard properties, including single-crossing, and X is a compact subset of a metric space. The mechanism is said to extract all surplus if the principal implements a social choice function $(x(\theta), p_1(\theta), ..., p_n(\theta))$ s.t. $x(\theta) = \arg\max \sum_{i=1} v_i(x, \theta)$ for all θ, and $E_{\theta_{-i}} v_i(x, \theta) - p_i(\theta) = 0$ for all θ_i and i.

In the standard case without communication costs Crémer and McLean (1985) have shown that the following condition is sufficient for full surplus extraction via a Bayesian mechanism: For all i and $\theta_i \in \Theta_i$ there does not exists a collection of coefficients $\lambda(\theta_i', \theta_i) \geq 0$ s.t.

$$F(\theta_{-i}|\theta_i) = \sum_{\theta_i' \in \Theta_i} \lambda(\theta_i', \theta_i) F(\theta_{-i}|\theta_i') \tag{10}$$

In other words, the vector of probabilities of other agents' type profiles conditional on a particular type of agent i should not be in the positive cone spanned by the vectors of probabilities of other agent's types conditional on other possible types of agent i. This condition is also necessary if full surplus extraction is desired for all possible utility functions.

When there are restrictions on communication, this condition needs to be modified. As a consequence, the difficulty of the task of surplus extraction may change in either direction.

To see the main point, recall the method developed by Crémer and McLean (1985) to extract surplus. Each agent is offered a menu of lotteries the outcome of which depends on the type reports made by other agents. Only the lottery corresponding to the agent's true type gives her a nonnegative (zero) payoff. Neeman (2002) has pointed out that the effectiveness of this method relies on the fact that extracting an agent's beliefs about other agents' types is equivalent to extracting her preference parameter. In our framework, agent i receives two pieces of information - her preference parameter and her feasible message set. Then the principal can also exploit the fact that an agent's preference parameter is correlated not only with the preference parameters of other agents, but also with the realizations of their feasible message sets. So, the lottery offered to agent i should be made contingent on the revealed θ_{-i} and $\mathcal{M}_{-i}$.[25] Yet, it becomes harder to extract full surplus from agent i if $\mathcal{M}_i$ contains more information about θ_{-i} and $\mathcal{M}_{-i}$ than θ_i does.

Formally, the analogue of the sufficient condition of Crémer and McLean (1985) for full surplus extraction in our framework is the following spanning condition: for all $(\theta_i, \mathcal{M}_i) \in T_i$ there does not exists a collection $\lambda(\theta_i', \mathcal{M}_i', \theta_i, \mathcal{M}_i) \geq 0$ s.t.

$$F(\theta_{-i}, \mathcal{M}_{-i}|\hat{\theta}_i, \hat{\mathcal{M}}_i) = \sum_{(\theta_i', \mathcal{M}_i') \in T_i, \mathcal{M}_i' \supseteq \mathcal{M}_i} \lambda(\theta_i, \mathcal{M}_i, \theta_i', \mathcal{M}_i') F(\theta_{-i}, \mathcal{M}_{-i}|\theta_i', \mathcal{M}_i') \tag{11}$$

Note the restriction $\mathcal{M}_i' \supseteq \mathcal{M}_i$ in the summation which comes from the fact that only types $(\theta_i', \mathcal{M}_i')$ such that $\mathcal{M}_i' \supseteq \mathcal{M}_i$ can imitate $(\theta_i, \mathcal{M}_i)$. In particular, if $\mathcal{M}_i \setminus \mathcal{M}_i' \neq \emptyset \ \forall \mathcal{M}_i' \in \mathcal{N}$, in (11) we only need to consider types of player i with the same feasible message set $\mathcal{M}_i$.

Compared to the standard case, the spanning condition (11) becomes easier to satisfy if $\mathcal{M}_{-i}$ generates additional variability of the probability distribution of other agents' types conditional on agent i's type.

In particular, suppose that $F(\theta_{-i}, \mathcal{M}_{-i}|\theta_i, \mathcal{M}_i) = F(\theta_{-i}, \mathcal{M}_{-i}|\theta_i)$ for all i, $(\theta_i, \mathcal{M}_i) \in T_i$, and $(\theta_{-i}, \mathcal{M}_{-i}) \in T_{-i}$. Then (11) holds if (10) holds. However, the converse is false. It is easy to construct examples of probability distributions such that (11) holds while (10) does not. In this case, surplus extraction is facilitated by the fact that θ_i is a sufficient statistic for $(\theta_{-i}, \mathcal{M}_{-i})$ with respect to $(\theta_i, \mathcal{M}_i)$, i.e only θ_i determines i's beliefs about others' types.

[25] For simplicity, assume that $\theta_i \in \mathcal{M}_i$ if agent's type is $(\theta_i, \mathcal{M}_i)$.

When $\mathcal{M}_{-i}$ and θ_i are not statistically independent, $\mathcal{M}_{-i}$ provides an additional signal that can be used to check agent i's report and to extract full surplus in a broader range of cases.

The opposite conclusion obtains in the following case: $F(\theta_{-i}, \mathcal{M}_{-i}|\theta_i, \mathcal{M}_i) = F(\theta_{-i}, \mathcal{M}_{-i}|\mathcal{M}_i)$ for all $(\theta_i, \mathcal{M}_i) \in T_i$ and $(\theta_{-i}, \mathcal{M}_{-i}) \in T_{-i}$. Then, agent i's beliefs regarding the types of other agents depend only on her feasible message set $\mathcal{M}_i$, and so full surplus cannot be extracted. In particular, the principal has to pay informational rents to satisfy incentive constraints between any pair of types with the same feasible message set $\mathcal{M}_i$. The impossibility of surplus extraction in environments where similar conditions hold has been established by Neeman (2002) and Parreiras (2003). It is driven by the fact that extracting agent i's beliefs about the types of other agents is no longer sufficient for the identification of her preference parameter.

References

Alger, I. and C.A. Ma (2003), "Moral Hazard, Insurance and Some Collusion," *Journal of Economic Behavior and Organization*, 50, pp. 225-247.

Alger, I. and R. Renault (1999), "Honest Agents and Equilibrium Lies," mimeo., Boston College.

Armstrong, M. (1996), "Multiproduct Nonlinear Pricing," *Econometrica*, 64, pp. 51-75.

Bull, J and J. Watson. (2001), "Evidence Disclosure and Verifiability," mimeo, UC San Diego.

Cremer, J. and R. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica*, 53, pp. 345-361.

Cremer, J. and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, pp. 1247-1257.

Che, Y.-K. and I. Gale (2000), "The Optimal Mechanism for Selling to a Budget Constrained Buyer," *Journal of Economic Theory*, 92, pp. 198-233.

Chen, Y. (2000), "Promises, Trust, and Contracts," *Journal of Law, Economics and Organization*, 16, pp. 209-231.

Choné, P. and J.-C. Rochet (1998), "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, pp. 783-826.

Crocker, K. and J. Morgan (1998), "Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts," *Journal of Political Economy*, 106, pp. 355-375.

Ekman, P. (1973)"Darwin and Facial Expression : A Century of Research in Review," *Academic Press*, New York.

Feinstein, B. and J. Erard (1994), "Honesty and Evasion in the Tax Compliance Game," *Rand Journal of Economics*, 25(1), pp. 447-456.

Fishman, M. and K. Hagerty (1990), "The Optimal Amount of Discretion to Allow in Disclosure," *Quarterly Journal of Economics*, 105, pp. 427-444.

Frank, R. (1988), "Passions Within Reason: The Strategic Role of the Emotions," Norton.

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

Glazer, J. and A. Rubinstein (2001), "Debates and Decisions: On a Rationale of Argumentation Rules," *Games and Economic Behavior*, 36, pp. 158-173.

Gneezy, U. (2002), "Deception: The Role of Consequences," mimeo, Graduate School of Business, University of Chicago, pp. 1-41.

Green, J. and J.-J. Laffont (1986), "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, 53, pp. 447-456.

Grossman, S. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality, " *Journal of Law and Economics*, 24, pp. 461-483.

Lacker, J.M. and J.A. Weinberg (1989), "Optimal Contracts Under Costly State Falsification," *Journal of Political Economy*, 97(6), pp. 1345-1363.

Lipman, B. and D. Seppi (1995), "Robust Inference in Communication Games with Partial Provability," *Journal of Economic Theory*, 66(2), pp. 370-405.

Maggi, G. and Rodriguez-Clare, A. (1995), "Costly Distortion of Information in Agency Problems," *Rand Journal of Economics*, 26, pp. 675-689.

Milgrom, P. (1981), "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, pp. 380-391.

Milgrom, P. and J. Roberts (1986), "Relying on the Information of Interested Parties," *Rand Journal of Economics*, 17, pp. 18-32.

Neeman, Z. (2002), "The Relevance of Private Information in Mechanism Design," mimeo, Department of Economics, Boston University.

Okuno-Fujiwara, M., A. Postlewaite and K. Suzumura (1990), "Strategic Information Revelation," *Review of Economic Studies*, 57, pp. 25-47.

Parreiras, S. (2003), "Correlated Information, Mechanism Design and Informational Rents," mimeo, Department of Economics, University of North Carolina.

Royden, H.L. (1987), "Real Analysis," Third Edition, *Prentice Hall*

Rudin, W. (1976), "Principles of Mathematical Analysis," Third Edition *McGraw Hill Publishers*

Severinov, S. and R. Deneckere (2003), "Does a Monopolist Need to Exclude?" mimeo, Fuqua School of Business, Duke University.

Shin, H. S. (1994), "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, 64, pp. 253-264.

Tennyson, S. (1997), "Economic Institutions and Individual Ethics: Consumer Attitudes towards Insurance Fraud," *Journal of Economic Behavior and Organization*, February, pp. 247-265.