Appendix B: Examples.

In this Appendix, we pursue two different types of applications of our model. First, we consider cases in which agents can misrepresent their private information only locally. Second, we analyze models in which some fraction of agents are unable to misrepresent their private information, while the remainder are able to misrepresent it fully.

Local Misrepresentation

In many instances, agents may be able to credibly misrepresent their private information only locally. For example, a taxpayer may be able to underreport her or his income only to a limited extent. Similarly, a buyer may only be able to partially hide the value of a transaction to him. Consider the following variation of the usual screening model.

Example 7 Screening with limited ability to misrepresent.

A risk-neutral firm with cost function \( c(q) \) sells quantity \( q \) of the 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 concave, increasing in both arguments and possesses a positive cross-partial derivative. The parameter \( \theta \) is privately known by the agent and takes one of \( n \) possible values \( 0 < \theta_1 < \ldots < \theta_n \). Let \( \pi_i > 0 \) denote the probability that \( \theta = \theta_i \). The agent can only misrepresent her true value by one step down, i.e. \( M(\theta_i) = \{\theta_i, \theta_{i-1}\} \) for \( i > 1 \), and \( M(\theta_1) = \theta_1 \).

In the standard case, where \( M(\theta_i) = \Theta_i \), the structure of the optimal solution is entirely driven by local (downward) incentive compatibility considerations. Since the agent is still able to claim to be the next lower type, one might naively assume that the nature of the optimal solution will remain unchanged. But according to Corollary 1, the only incentive constraint that needs to be imposed is that type \( \theta_2 \) be unwilling to mimic type \( \theta_1 \). The reason for this is simple: any type \( \theta_i \) \( (i > 2) \) is unable to tell a “lie” that is available to type \( \theta_{i-1} \). Type \( \theta_2 \) has a reporting advantage relative to all other types, because only \( \theta_2 \) can fully imitate another type. So, in the profit maximizing mechanism only type \( \theta_2 \) obtains positive surplus.

The solution has the following properties. Any type \( i = 2, \ldots, n \) consumes its efficient quantity \( q_i^* \) maximizing \( u(q, \theta_i) - c(q) \), \( \theta_1 \) consumes \( q_1^* \) solving \( \frac{\partial u(q, \theta_1)}{\partial q} - \frac{\pi_2}{\pi_1 + \pi_2} \frac{\partial u(q, \theta_2)}{\partial q} = \frac{\pi_1 c'(q)}{\pi_1 + \pi_2} \). The transfers to the firm are given by: \( p^*(\theta_i) = u(q_i^*, \theta_i) \) for \( i \neq 2 \), and \( p^*(\theta_2) = u(q_2^*, \theta_2) - \frac{\pi_2}{\pi_1 + \pi_2} \frac{\partial u(q, \theta_2)}{\partial q} \).
\[ u(q^*_1, \theta_2) + u(q^*_4, \theta_1), \]

This example highlights 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.

An even stronger result emerges when we extend the previous example to a multi-agent setting and consider the following version of the Myerson-Satterthwaite bilateral bargaining model:

**Example 8** Bilateral bargaining with limited ability to misrepresent.

A buyer with valuation \( b \) and a seller with cost \( c \) bargain over the sale of a single good. The traders’ types are drawn from some joint probability distribution \( F_{bc}(\cdot) \) and can take on one of \( n \) possible values respectively denoted by \( b_1 < b_2 < \ldots < b_n \) and \( c_1 < c_2 < \ldots c_n \) s.t. \( b_{i-1} < c_{i-1} < b_i \leq c_i \) for all \( i \). If trade takes place at price \( p \), then the buyer’s and seller’s surplus are \( b_i - p \) and \( (p - c_j) \), respectively. The buyer and the seller can misreport their types by at most one grid point, i.e.

\[
M^b(b_i) = \{b_{i-1}, b_i, b_{i+1}\} \quad \text{for} \quad i = 2, \ldots, n - 1, \quad M^b(b_1) = \{b_1, b_2\}, \quad M^b(b_n) = \{b_{n-1}, b_n\},
\]

\[
M^s(c_j) = \{c_{j-1}, c_j, c_{j+1}\} \quad \text{for} \quad j = 2, \ldots, n - 1, \quad M^s(c_1) = \{c_1, c_2\} \quad \text{and} \quad M^s(c_n) = \{c_{n-1}, c_n\}.
\]

These restrictions on players’ communication abilities imply that there exists an ex-post budget balanced, ex-post individually rational incentive compatible mechanism with efficient trade (i.e. the good gets transferred from a seller type \( c_j \) to a buyer type \( b_i \) iff \( i > j \)). Indeed, define the following outcome functions: the good gets transferred from seller \( c_j \) to buyer \( b_i \) iff \( i > j \), at a price \( p(b_i, c_j) \in [c_j, b_i] \), which is non-decreasing in \( b_i \) and \( c_j \). By construction,

---

\(^{26}\)Let us contrast our approach to that of Green and Laffont (1986) where the principal is restricted to select from the set of direct mechanisms, in which the agent sends a single message. Determining the optimal mechanism over this restricted class becomes a rather unwieldy exercise, because one has to optimize over the set of all direct mechanisms, not just the truth-telling ones. In the latter class, the principal faces a tradeoff between extracting full surplus from high valuation agents, and having to pool lower-valuation agents. For example, with three types, if type \( \theta_2 \) is left with an incentive to lie by announcing that its type is \( \theta_1 \), the principal can extract full surplus from type \( \theta_3 \), but then must pool types \( \theta_2 \) and \( \theta_1 \) (or else price type \( \theta_1 \) out of the market completely).
the outcome is ex-post efficient, ex-post individually rational, and ex-post budget balanced. According to Corollary 1, the only incentive constraints that have to be imposed are that buyer type $b_2 (b_{n-1})$ be unwilling to mimic buyer type $b_1 (b_n)$, and that seller type $c_2 (c_{n-1})$ be unwilling to imitate seller type $c_1 (c_n)$. Observe that buyer type $b_2$ has no incentive to mimic type $b_1$, as $b_1$ never gets to trade. Similarly, seller type $c_{n-1}$ has no incentive to mimic seller type $c_n$. Further, buyer type $b_{n-1}$ has no incentive to mimic buyer type $b_n$ as he trades at a lower price with all seller types $c_j$ with $j < n - 1$ and cannot profit from trading with type $c_{n-1}$. A similar statement holds for seller type $c_2$.

**Honest agents**

Next, we illustrate our results by applying them to 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. Two papers in the literature have analyzed models of this kind. Alger and Ma (2003) (see Example 2) develop 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 misrepresent the patient’s condition to the HMO) or ‘strategic.’ Similarly, Erard and Feinstein (1994) present a signalling model of income taxation in which some fraction of taxpayers is unable to underreport their incomes, while the others can fully misrepresent their incomes. Neither of these models, however, fully exploit the limits on agents’ communication abilities.

In this section, we will first provide a general overview of the differences between the approach followed by these authors and the one developed in this paper, and then focus on the income taxation problem. So, consider an adverse selection problem in which an agent’s preference parameter $\theta$ can vary between high (high income or productivity, good health, low cost) and low (low income or productivity, poor health, high cost). Additionally, an agent can either be ‘strategic’ or ‘honest.’ In line with the approach of Alger and Ma (2003) and Erard and Feinstein (1994), construct a mechanism in which the agent makes one announcement $\hat{\theta}$ about her preference parameter, and is assigned an allocation $x(\hat{\theta})$ on the basis of her report. Then the dimension of the set of implementable allocations is equal to the cardinality of the set of preference parameters. Because ‘honest’ agents always report their valuations...
truthfully, and ‘strategic’ agents choose reports to maximize their payoffs, the principal faces a choice between two alternatives. First, it can offer an allocation profile that keeps the agents who report truthfully at their reservation utility levels. This strategy extracts full surplus from ‘honest’ types. However, ‘strategic’ high types then choose to underreport, reducing the efficiency of the mechanism and the principal’s expected payoff. Alternatively, the principal can extract a larger surplus from ‘strategic’ agents by offering an allocation profile that makes reporting truthfully incentive compatible. However, incentive compatibility comes at a cost: in this case ‘honest’ agents receive the same allocations as ‘strategic’ agents with the same preference parameters, and so end up with positive surpluses.

The results of Alger and Ma (2003) imply that the first strategy is optimal when the fraction of ‘honest’ agents is large, for then the rent extraction from ‘honest’ types becomes the dominant motive. However, when the fraction of ‘honest’ agents is small, it becomes optimal for the principal to induce ‘strategic’ agents to tell the truth and forego full rent extraction from the ‘honest’ agents. In this case, an agent’s allocation depends only on her valuation, and is identical to the one in the standard case without ‘honest’ agents.

In contrast, Corollary 1 implies that the principal need not face such a trade-off between leaving surplus to the ‘honest’ agents and assigning more efficient allocations to the ‘strategic’ ones. Using mechanism $H(\cdot)$ in which the agent is asked to make two announcements $(\hat{\theta}_1, \hat{\theta}_2)$ regarding her preference parameter, or a simpler and equivalent ‘password’ mechanism in which a single report $\hat{\theta}$ is followed by a self-selecting menu of contracts, the principal can implement an allocation profile that, in principle, 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’ agents from ‘strategic’ ones 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’ agent does not wish to imitate any other type. This mechanism is more profitable for the principal and is more efficient, as it generates smaller allocative distortions than any mechanism in which an agent can make only one statement about her valuation.\footnote{In a companion paper (Severinov and Deneckere 2003), we analyze a version of the non-linear pricing model in which a fraction of the population is unable to misrepresent the marginal valuation of quantity.} To illustrate this discussion, we consider...
the following example in detail.

**Example 9** (Optimal income taxation with a fraction of honest taxpayers). Consider a taxpayer who has private information regarding his income $I$. There are two possible levels of income, $I_L$ and $I_H$, with $I_L < I_H$. The probability that $I = I_L$ equals $\pi$. The taxpayer is risk averse, with a strictly concave utility function $u(\cdot)$ over after-tax income, satisfying $u(0) = 0$ and $u'(0) = \infty$. A fraction $\gamma$ of the population consists of taxpayers who are unable to misrepresent their income.

The tax authority (principal) selects a tax $t(x)$ and an audit probability $\alpha(x)$ for every possible income report $x$. An audit, which costs $c$ to perform, allows the tax authority to uncover the taxpayer’s true income $I$. So, the tax authority also chooses the tax schedule $t(x, I)$ applied when a taxpayer reporting income $x$ is audited and her true income is found to be $I$. There is limited liability, so the constraints $t(x) \leq x$ and $t(x, I) \leq I$ have to hold.

To simplify the notation, set $t_L = t(I_L)$, $t_H = t(I_H)$, $t_{LL} = t(I_L, I_L)$, $t_{LH} = t(I_L, I_H)$, $t_{HH} = t(I_H, I_H)$, $t_{HL} = t(I_H, I_L)$, $\alpha_L = \alpha(I_L)$ and $\alpha_H = \alpha(I_H)$. It is straightforward to see that (regardless of the approach) it is optimal for the principal to: (i) impose maximal punishment if an audit reveals misreporting, (ii) induce ‘strategic’ low income type to report her income truthfully extracting all her rents, and (iii) never to audit high income reports. Consequently we have $t_L = I_L$, $t_{LL} = t_{HL} = I_L$, $t_{LH} = t_{HH} = I_H$, and $\alpha_H = 0$.

**Proposition 2** In the optimal ‘password’ mechanism a taxpayer reporting high income is taxed the full amount $I_H$. If $(1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{w(I_H - I_L)} > \pi c$, then a taxpayer reporting low income is given a menu of two tax/audit probability pairs: $(I_H - u^{-1}((1 - \alpha_L)u(I_H - I_L)), 0)$, and $(I_L, \alpha^*_L)$, where $\alpha^*_L$ solves

$$(1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(u^{-1}((1 - \alpha_L)u(I_H - I_L)))} = \pi c$$

We apply our approach to implementation relying on a “password” mechanism and characterize the optimal allocation profile for the considerably more complex case in which $\theta$ can take any value in some interval $[\underline{\theta}, \bar{\theta}]$. A surprising finding of our analysis is that 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.
In equilibrium, ‘strategic’ high income types report $I_L$ and choose the first element from the menu, while low income types report $I_L$ and choose the second element from the menu.

If \((1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(I_H - I_L)} \leq \pi c\), then any taxpayer reporting low income (which includes all low income types and the strategic high income type) pays a tax $t_L = I_L$ and does not get audited, i.e. $\alpha_L = 0$.

In the ‘password’ mechanism, the mechanism designer separates ‘honest’ high income types from ‘strategic’ ones without leaving any surplus to the former. This is accomplished by requiring any agent who wishes to access the menu designed for ‘strategic’ high income types to send a low income report.

In contrast, in the approach used by Erard and Feinstein (1994), the tax agency is limited to ‘direct mechanisms’ in which a taxpayer reports her income once. She then either pays tax right away, or is audited before her final tax amount is determined. The revenue-maximizing mechanism designed by using this approach is characterized below:

**Proposition 3** The optimal direct taxation mechanism (with a single income report from the taxpayer) is as follows. There exists $\hat{c} > 0$ s.t. when $c < \hat{c}$, the tax authority sets $t_H = t_{LH} = I_H - u^{-1}((1 - \alpha_L)u(I_H - I_L)), t_L = I_L, \alpha_H = 0$ and chooses $\alpha_L$ solving:

$$u'(u^{-1}((1 - \alpha_L)u(I_H - I_L)))\pi c = (1 - \pi)u(I_H - I_L)$$

In this mechanism, the ‘strategic’ high income type reports his income truthfully.

When $c > \hat{c}$, the tax authority sets $t_H = t_{LH} = I_H, t_L = I_L, \alpha_H = \alpha_L = 0$, the strategic high income type reports $I_L$, and the honest high income type is held at her reservation level.

Comparing the allocations from Propositions 2 and 3, we see that unless audit costs are so high that the tax authority does not audit low income report in either solution, the tax authority raises a higher expected revenue from the optimal password mechanism.

An interesting and possibly surprising implication of our model is that individuals who provide conflicting or contradictory information need not be penalized for such behavior. Indeed, in the password mechanism, the strategic high income type first reports a low income, but then repents by selecting the tax designed for the strategic high income type from the menu. It is
as if the tax agency, upon receiving a report of low income, asks the taxpayer to confirm her report and allows him/her to revise it upward in exchange for a smaller penalty. In practice, this can be interpreted as an initial interview with the IRS to assert if the taxpayer made no ‘mistake’. At this interview a taxpayer is offered a menu of choices. Only insistence on a low income then leads to a full-blown audit. More broadly, individuals who provide contradictory statements may receive 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.

Proofs of Propositions 2 and 3.

Proof of Proposition 2:

Let us construct the optimal mechanism \( G(\cdot) \) where each taxpayer can be asked to send more than one message, or, alternatively, is given a menu to choose from after sending a single message. We start by discussing the game form of the mechanism, and then derive the optimal allocation profile.

In mechanism \( G(\cdot) \) the taxpayer is asked to make two announcements \((\hat{I}_1, \hat{I}_2)\) regarding her income. If high income is reported twice, i.e. \( \hat{I}_1 = \hat{I}_2 = I_H \), then the mechanism should assign a tax that leaves no surplus to this announced type. The inability of an ‘honest’ high-income type to lie forces her to report her true valuation both times, so she does not receive any surplus and her valuation is identified for free. If both messages are \( I_L \), then the mechanism should treat the taxpayer as a low-income one (either ‘strategic’ or ‘honest’). If, on the other hand, \( \hat{I}_1 = I_H \) and \( \hat{I}_2 = I_L \), then the mechanism should assign a tax and audit probability which makes it incentive compatible for high income strategic type to choose this combination rather than imitate low-income types by sending messages \( \hat{I}_1 = \hat{I}_2 = I_L \). Incentive compatibility implies that a high income strategic type obtains a positive surplus.

A simpler way to implement the same allocation profile is via the password mechanism relying on just one message. Specifically, the principal should offer a tax rate \( t_{HH} = I_H \) to any taxpayer reporting income level \( I_H \) since only an ‘honest’ high income type will report such income. A taxpayer who reports income level \( I_L \) is given a menu consisting of two elements: (a) paying a tax \( t_H = I_H - u^{-1}(1 - \alpha_L)u(I_H - I_L) \) and not being audited, and (b) paying
a tax $I_L$ and being audited with probability $\alpha_L$, resulting in a tax of $t_{LH} = I_H$ if the audit reveals income $I_H$. In equilibrium, the strategic high income type will first report income $I_L$, and then select the option $(t_H, \alpha_H)$ from the menu. Note that her incentive constraint holds. Since the principal extracts full rent from the honest high income type, the objective becomes

$$V_3(\alpha_L) = \pi(I_L - \alpha_L c) + (1 - \pi)[I_H - (1 - \gamma)u^{-1}((1 - \alpha_L)u(I_H - I_L))]$$

Differentiating, we get:

$$V_3'(\alpha_L) = -\pi c + (1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(u^{-1}((1 - \alpha_L)u(I_H - I_L)))}$$  (3)

It is easy to show that $V_3'(\alpha_L)$. Then $\alpha_L^*$ solves $V_3'(\alpha_L) = 0$ if $(1 - \pi)(1 - \gamma)\frac{\pi}{\pi'}(I_H - I_L) > \pi c$, and $\alpha_L^* = 0$ otherwise. \textit{Q.E.D.}

**Proof of Proposition 3:**

The expected payoffs to a taxpayer with income $I$ who reports $x$ and the corresponding payoff to the principal are given by respectively:

$$U(x, I) = (1 - \alpha(x))u(I - t(x)) + \alpha(x)u(I - t(x, I))$$

$$v(x, I) = (1 - \alpha(x))t(x) + \alpha(x)(t(x, I) - c)$$

As we argued in the main text, the optimal mechanism belongs to either of the two categories. First, it could be truthful so that every type reports her true income. Second, it could be non-truthful where the strategic high-income type reports a low income. To prove the proposition, we will derive the best truthful and non-truthful mechanisms and then compare their profitability.

Start by deriving the best non-truthful mechanism in which the strategic high-income type reports income $I_L$. Clearly, in such mechanism it is optimal to set $t_H = I_H$, $\alpha_H = 0$, $t_L = I_L$ and $t_{LH} = I_H$. Then the tax authority should select the audit probability $\alpha_L$ to maximize her expected payoff given by:

$$V_1(\alpha_L) = \pi(I_L - \alpha_L c) + (1 - \pi)[\gamma I_H + (1 - \gamma)((1 - \alpha_L)I_L + \alpha_L(I_H - c))]$$

Differentiating $V_1(\alpha_L)$ with respect to $\alpha_L$ we get: $V_1'(\alpha_L) = -\pi c + (1 - \pi)(1 - \gamma)(I_H - I_L - c)$. So, it is optimal to set $\alpha_L = 1$ if $(1 - \pi)(1 - \gamma)(I_H - I_L - c) > \pi c$ and to set $\alpha_L = 0$ otherwise.
Next, we derive the best truthful mechanism in which the strategic high income reports his income truthfully. In this mechanism \( t_H \) must be set to satisfy the incentive constraint \( U(I_H, I_H) \geq U(I_L, I_H) \). Optimality requires making this constraint binding, which is done by setting \( t_H = I_H - u^{-1}((1 - \alpha_L)u(I_H - I_L)) \). Consequently, the principal must select \( \alpha_L \) to maximize her expected payoff:

\[
V_2(\alpha_L) = \pi(I_L - \alpha_L c) + (1 - \pi)[I_H - u^{-1}((1 - \alpha_L)u(I_H - I_L))]
\]

This expression implies that the best mechanism coincides with the one in which \( \gamma = 0 \). Differentiating \( V_2 \) produces

\[
V_2'(\alpha_L) = -\pi c + (1 - \pi)\frac{u(I_H - I_L)}{u'(u^{-1}((1 - \alpha_L)u(I_H - I_L)))}.
\]

If \((1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(I_H - I_L)} > \pi c\), then \( \alpha_L^* \) solves \( V_2'(\alpha_L^*) = 0 \). If \((1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(I_H - I_L)} \leq \pi c\), then \( \alpha_L^* = 0 \).

Finally, we need to determine which of the two mechanisms - the best truthful one or the best non-truthful one - yields the highest expected tax revenue. To compare the two, observe first that \( u(x) > xu'(x) \) for all \( x > 0 \) since \( u \) is strictly concave. So,

\[
(1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(I_H - I_L)} > (1 - \pi)(1 - \gamma)(I_H - I_L).
\]

If \((1 - \pi)(1 - \gamma)(I_H - I_L - c) > \pi c\), then \( \alpha_L = 1 \) in the optimal non-truthful mechanism. If we also set \( \alpha_L = 1 \) in the optimal truthful mechanism, then the two mechanisms will have the same expected payoff for the tax authority. But optimal \( \alpha_L \) in the optimal truthful mechanism is strictly less than 1. Thus \( \max_\alpha V_2(\alpha) > V_2(1) = V_1(1) \). Consequently, over this range the best truthful mechanism dominates.

On the other hand, if \((1 - \pi)(1 - \gamma)\frac{u(I_H - I_L)}{u'(I_H - I_L)}(I_H - I_L) \leq \pi c\), then \( \alpha_L = 0 \) in the best truthful and best non-truthful mechanisms. But \( t_H = I_L \) in the former and \( t_H = I_H \) in the latter. Obviously, setting \( t_H = I_H \) is better for the tax authority since it allows to extract rent from the ‘honest’ high income type. Consequently, over this range, the best non-truthful mechanism dominates.
Finally, if $(1 - \pi)(1 - \gamma)(I_H - I_L - c) \leq \pi c < (1 - \pi)(1 - \gamma) \frac{u(I_H - I_L)}{w(I_H - I_L)}$, we need to compare the corresponding values of the objective functions $V_1(0)$ to $V_2(\alpha^*_L)$. Note that $V_1(0) = \pi I_L + (1 - \pi)[\gamma I_H + (1 - \gamma)I_L]$, which is independent of $c$. The envelope theorem implies that $\frac{dV_2(\alpha^*_L)}{dc} = -\pi \alpha^*_L < 0$. Furthermore, as shown above, $V_2(\alpha^*_L) > V_1(0)$ if $c = \frac{1 - \pi}{\pi}(1 - \gamma)(I_H - I_L)$, and $V_1(0) > V_2(\alpha^*_L)$ if $c = \frac{1 - \pi}{\pi} \frac{u(I_H - I_L)}{w(I_H - I_L)}$. Hence there exists a unique $\tilde{c} \in [\frac{1 - \pi}{\pi}(1 - \gamma)(I_H - I_L), \frac{1 - \pi}{\pi} \frac{u(I_H - I_L)}{w(I_H - I_L)}]$ such that $V_1(0) = V_2(\alpha^*_L).$ Q.E.D.