Mechanism Design when players' Preferences and information coincide

Dubra, Juan and Caffera, Marcelo and Figueroa, Nicolás

Universidad de Montevideo, Universidad de Montevideo, Pontificia Universidad Católica de Chile

2 December 2016

Mechanism Design when players' preferences and information coincide.*

Marcelo Caffera
Universidad de Montevideo

Juan Dubra[†]
Universidad de Montevideo

Nicolás Figueroa
Pontificia Universidad Católica de Chile

December 2, 2016

Abstract

It is well known that when players have private information, vis a vis the designer, and their preferences coincide it is hard to implement the socially desirable outcome. We show that with arbitrarily small fines and arbitrarily noisy inspections, the social choice correspondence can be fully implemented (truth telling is the unique Nash equilibrium).

In a model where types are correlated, we know since Crémer and McLean (1985) that a mechanism designer can get the parties to reveal their types, but that might involve arbitrarily large ex post losses for the participants. Alternatively, with costly state verification, as in Townsend (1979), one can introduce potentially large fines and get players to report their types in equilibrium.

In this note we show that a combination where the mechanism designer can:

- inspect the truthfulness of a player's report about the state of nature with an arbitrarily small probability, with an inspection technology that actually finds something about the state with a very small probability (i.e. a noisy technology);

- impose arbitrarily small fines if a report is discovered to be false;

allows for a mechanism that yields truth telling as the unique equilibrium. The arbitrarily small fines are never paid in equilibrium, and the expected cost of the mechanism is virtually 0: the inspection is materialized with an arbitrarily small probability, so the cost of the state verification is almost never paid. In addition we show that there is no mechanism that can achieve a cost of exactly 0.

The main contribution of our note is to show that a combination of "a little" of noisy (cheap) state verification, and arbitrarily small fines can yield uniqueness of truth telling where full costly state verification or large losses might be needed to obtain truth telling (and not uniqueness).

1

The main assumptions we make are that all players agree on what constitutes a better state of nature, and that it is common knowledge that they agree on what state has occurred. We believe that these assumptions are a good first approximation of several situations of interest. Our initial interest in this problem was on how to get firms to reveal the true cost of abating pollution; if firms have access to the same abatement technologies, our results show that the regulator can achieve the first best pollution level as the unique outcome. Similarly, a regulator setting the price of utilities that share the same production technology could elicit their true marginal cost and again achieve first best production levels. The assumption of common costs has been used often in different literatures, the most prominent one being common value auctions where the common value arises because the product is the same for (say) oil firms, and they share a production technology. Another application could be Grameen Banking, where a small group of related potential debtors knows what is the largest amount of money that all debtors could repay with certainty, and the banker is trying to find out what that number is. Finally, a well known example where those features of our model are met exactly is the travelers' dilemma of Basu (1994), in which an airline lost the luggage of two travelers who had purchased the same antique: it is common knowledge that the cost is the same for both travelers, and the airline is trying to elicit what that number is.

To understand what our assumptions imply in relation to the standard design problem, we note that usually, in order to get truthful revelation, the designer lets the agents self-select an outcome from a well defined menu. In order to construct such a menu, the mechanism designer must be able to exploit players' state dependent utilities. In our setup, however, players' preferences are uniform (they are the same, irrespective of the state): they all want a larger reimbursement by the airline, a larger estimate of the cost of production or of abatement. The uniform preferences make the implementation problem hard. Nevertheless, in our setup the planner knows that the information is common across agents (as is standard in the original mechanism design literature, where players, with knowledge of the true state, have to report a preference profile specifying every player's preference). This knowledge enables the planner to incentivize the agents to agree on the common announcement, and the use of (small) fines can be used to break ties so that players are pitched against each other. The knowledgeable reader will notice the connection between this mechanism and the recent and growing literature about implementation with evidence, or with costs of lying. We discuss the connection to this literature after presenting our main results; it suffices here to say that results are not nested, and that our main theorem precedes most of these results (see Caffera and Dubra, 2005; to the best of our knowledge, the only earlier papers include Green and Laffont (1986), Lipman and Seppi (1995) and Bull and Watson, 2004).

Although we develop our results in a context where the state is known by all players in order to show that the result is not a consequence of this assumption, we extend our results to a simple example in which the state is not known exactly.

More generally our paper, and part of the literature on implementation, can be thought in the context of Becker and Stigler (1974): what is the cost of implementing the desired social outcome, and what is an efficient way to enforce the outcome? Becker and Stigler were the first to ask how can we pay "an enforcer" in the most effective way. In our paper the players are the enforcers themselves, as most of the enforcing is done by the players of the (optimal) mechanism, where they are pitched against each other. By designing an appropriate mechanism we have made the cost of enforcement negligible. The assumptions of our paper seem suited to cases where the players have commited a crime jointly, so they share the same information and preferences.

We now present our model and results, and then their discussion in the context of the relevant literature.

1 Model and Results

There are $m \geq 2$ players, and the state space is S. The mechanism designer will have to choose an action, and we assume that for each state there is a distinct optimal action, so in order to simplify the analysis, we assume that the space of actions the mechanism designer can choose is also S (the designer wants to match each state $s \in S$ with its optimal action which we identify with s). For example, in terms of the travelers' dilemma, if the cost of the lost items is \$4, the designer wants to pay each traveler \$4. As another example, suppose the designer is a regulator choosing a price $p(c)$ (depending on the marginal cost of abatement) that polluters must pay per ton of CO2. If the marginal cost of abatement is $c = \$5$, the regulator would like to choose $c = 5$ and announce that the price to be paid is $p(5)$.

As discussed above, we assume that all players agree on what a "better state" or "better action" is; the assumption is that all players have a common utility function over S. For action s and transfer t to the designer, the utility or profit is $u(s) - t$, for $u : S \to \mathbf{R}$. The state is drawn according to some probability measure H on S, and we assume that the support of the random variable $u(s)$ is an interval (if all players are declaring state s, this assumption allows a player to "slightly" undercut all players' announcements).

In a direct mechanism, the designer can investigate a claim made by a player, with a cost of k. In the example of the travelers' dilemma it could try to contact the seller abroad, or by having access to the luggage, they could find a receipt; in the case of the regulator trying to abate pollution, it could hire engineers to check the claims made about the cost of abatement; in the case of a regulator regulating utilities (with the same cost), it could also hire engineers to check the player's claim. The inspection technology is such that for some fixed, exogenous (arbitrarily small) probability $\varepsilon_2 > 0$ the inspection tells whether the report was truthful or not (in case the report was not truthful, the inspection does *not* say what is the true cost of the object, just that the report was not truthful). In addition, the airline (or the regulator) has the ability to impose an arbitrarily small fine if the report was found to be false; this is in line with the penalty imposed in Basu's original mechanism, but our fine can be arbitrarily small.

The cost k of the inspection *must* be interpreted as a small number. Since the probability of the inspection yields information only with an arbitrarily small probability, the inspection could just mean checking another time to see whether the luggage (of one of the players) was not in fact lying beside the carrousel. The point is that the inspection could be the simplest of actions by the regulator; even the slightest effort in trying to find out the state (so long as it has *some* chance of yielding any information) could be interpreted as an inspection.[1] In this regard, the assumption that the regulator must sometimes carry out an inspection is not as "harmful" as in the case of Abreu and Matsushima (1992) where the planner must commit to choosing an outcome which might be far from a socially desirable one.[2]

The technology for making inspections is given by a cost k, a maximum fine ε_1 (arbitrarily small) and a probability ε_2 of finding if the inspected party cheated or not. When an individual is sampled, the inspection yields the "answer" "uncertain, u" with probability $1 - \varepsilon_2$ for an arbitrarily small $\varepsilon_2 > 0$. With probability ε_2, the inspection tells whether the report was truthful or not (in case the report was not truthful, the inspection does *not* say what is the true state).

The regulator must design a mechanism to minimize the sum of:

- the expected value of the distance between the true state and the one chosen by the regulator;

---

[1] We make the cost k explicit, but others have introduced forms of inspections where the information comes "for free" to the mechanism designer (see for example Midjord, 2013; in that paper the designer may learn the true state with some probability; here the true state is never revealed to the designer).

[2] The cost k being small also helps in the sense that if the planner did not have the commitment power to implement the inspection if the cost was high, and players knew this, the mechanism wouldn't work. We thank the referee for pointing out the relevance of the size of k.

- the expected cost of making inspections;

In the equilibrium of the mechanism that minimizes those costs there will be no fines. Still, we do not include the fines in the objective function in order to clarify that the cost of the mechanism is not low because of the existence of fines.

A direct mechanism is given by a triplet (l, t, f).

a. $l : S^m \to \Delta \equiv \left\{ x \in \mathbf{R}_+^m : \sum x_i = p \right\}$ for some $p \in (0, 1)$ specifies with what probability each player will be sampled for inspection, as a function of all players' reports; the probability that one player will be inspected is p.

b. $t : S^m \to S$ specifies the action to be taken by the designer as a function of players' announcements.

c. $f : \{g, i, u\} \to \{0, \varepsilon_1\}$ represents the fines to be paid by the only individual who is inspected, which must be lower than ε_1, depending on the outcome of the inspection: $g$uilty of lying, $i$nnocent, or $u$ncertain. In order to make the mechanism simpler, and to avoid giving the mechanism designer more degrees of freedom, we fix fines at either 0 or ε_1.

Theorem 1. *For any $\varepsilon > 0$ there is a mechanism that has a total expected cost of less than ε. In that mechanism, the unique equilibrium is truth telling: all players report the true state. The only cost of the mechanism is the (small) chance that the regulator will inspect the report of one of the players, and in that case the total cost will be k.*

Our mechanism is as follows: the regulator selects a $p < \varepsilon/k$ then

1. Players simultaneously announce a state;

2. The regulator decides whether to inspect a player, or not; he inspects with probability p. If no inspection will be carried out, one of the states with the smallest profit is chosen at random. If an inspection will happen, step 3 is carried out.

3. If the profits of all announcements coincide, the regulator chooses one player randomly (uniformly) and inspects the report. If the profits of the announcements do not all coincide, the regulator: identifies the player or players who announced the largest-profit state – let $n < m$ be the number of players in that group; randomly (uniformly) selects one of them and inspects either his report (with probability $\pi \in \left(\frac{1}{2}, 1 \right)$) or any other player's with probability $1 - \pi$ (this yields a chance of inspection of $p \frac{1 - \pi}{m - 1}$ for players who don't declare the highest profit state). The idea is to monitor only one player, and to monitor more intensely those players who are most likely lying. A player is fined if and only if: his report is false and it is inspected and the inspection discovers that the report was false.

4. The designer selects an action corresponding to one of the smallest-profit announcements.

A strategy for a player in the game that this mechanism defines is a function $\sigma : S \to S$ that announces a state for each possible real state. We now prove our main result.

Proof of Theorem 1. Truth Telling is an Equilibrium. We first show that truth telling is an equilibrium. Without loss of generality, consider the situation of player 1 when all other players are reporting the true state $(s, s, ..., s)$. Notice that declaring the true s leads to the implementation of s, consistent with all the declarations of players 2 through m. If player 1 reports $\widehat{s}_1 \neq s$, two things could happen, depending on the profile of states announced by players $2, ..., m$:

4

- $u(\widehat{s}_1) \geq u(s)$. In this case player 1 does not change the utility of the regulator's action (if $u(\widehat{s}_1) > u(s)$, the regulator chooses s; if $u(\widehat{s}_1) = u(s)$, he could choose either s or $\widehat{s}_1$, but the utility is still $u(s)$ as if player 1 declared s), and player 1 could be fined (if sampled and the inspection showed a lie).

- $\widehat{s}_1 < s$. In this case, the player lowers the utility of the action chosen by the regulator, and in addition he could be fined.

Then, player 1 is worse off deviating, and hence, declaring the truth is better than declaring anything else, proving that truth telling is an equilibrium.

There is no other equilibrium. Suppose there is a profile $\boldsymbol{\sigma}$ of strategies such that for some player j, $\sigma_j(s_{lie}) \neq s_{lie}$ for some $s_{lie} \in S$ and suppose that this profile of strategies is an equilibrium. That is, suppose there is an equilibrium without truth telling. Without loss of generality, suppose $j = 1$. Then, for state s_{lie} *all* announcement must be lies, since if one individual were telling the truth all other players would be strictly better off telling the truth, since (relative to lying) they would weakly increase u, and strictly reduce the chance of being fined (to 0). Since one player is inspected with probability p, the average chance of a player being inspected is p/m.

Note first that in s_{lie} no player j can declare a $\sigma_j(s_{lie})$ that yields lower profit than s_{lie} (he would be strictly better off declaring s_{lie} by reducing the fine); this implies that in this proposed equilibrium players can still undercut each other's announcements.

Take any player that in state s_{lie} declares a state with the highest profit; suppose it is player 1.

- If he is the only one declaring such a profit, he can strictly reduce the chance of inspection by matching one of the lowest announcements, and he doesn't change the outcome.

- If any other player (say player 2) is also declaring a state with maximum profit, player 1 still has a probability q of being inspected that is (weakly) larger than p/m. We will show that he is strictly better off slightly undercutting the announcement of player 2: if he does, player 1 will never be one of the ones announcing the state with the largest utility (the ones inspected with higher probability) and therefore will strictly reduce his chance of being fined, because his chances of being inspected fall. In particular, if he is not announcing the highest state, his chance of being inspected is

$$p\frac{1-\pi}{m-1} < \frac{p}{m} \leq q.$$

  If the amount by which it undercuts the announcement of player 2 is small, this deviation is profitable. This shows that the profile $\boldsymbol{\sigma}$ is not an equilibrium. ∎

We note that the assumption that the support of $u(s)$ is an interval cannot be dropped, since in the proof we use that players can "slightly undercut" other players' announcements. If the support had a hole, that would not be possible: if all players were declaring a false state s for which there was no s' with $u(s')$ close to $u(s)$ it would not pay for any player to undercut s.

The mechanism achieves a cost of virtually 0, and there are no fines in equilibrium. In particular, there is a sequence of mechanisms, the costs of which converge to 0, but none achieves the 0 cost. To complement this analysis, we now show that there is no mechanism that achieves a cost of exactly 0.

Theorem 2. *There is no mechanism in which an equilibrium with truth telling yields a cost of exactly* 0; *that is, some inspections are necessary for truth telling. Moreover, for any mechanism and any equilibrium*

with truth telling of that mechanism, there must be a positive probability of inspection for every s with $u(s) > \underline{u} \equiv \min \text{Support} u(S)$. That is, inspections occur with positive probability for most realizations of s.

Proof of Theorem 2. If for some $u(s^*) > \underline{u}$ there are no inspections, for any s with $u(s) < u(s^*)$ it is an equilibrium for all to report s^*, which contradicts that there is truth telling. ∎

As for $m \geq 2$, the assumption is also necessary: with only one player, he can always claim the highest-profit state, and the fine would not force him to tell the truth (since the fine is small, and the chance of being discovered is also small). Also regarding the number of players, an additional feature of our mechanism, is that it is robust to collusion by less than m agents.

2   An example with imperfect correlation.

The fact that all players agree exactly on the state seems like a reasonable assumption in the cases that we mentioned in the introduction. Moreover, it is a feature of two of the problems that motivated this paper (the original travelers' dilemma and the minimum cost of abating pollutants). Still, one may wonder whether our result is a knife edge result that would disappear if players do not observe the same state. In order to show that it is not, we now provide a simple example in the context of the travelers's dilemma where the types of players are imperfectly correlated, and players don't know each other's type, and show that there is also a unique equilibrium with truth telling. The point of the example is not to show that the mechanism would work with small noise in general, which would be beyond the scope of this note, but rather to show that uniqueness is not the result of the perfect correlation.

Formally, we now assume that there are $m = 2$ players, and that for each i, individual i bought an object with a cost of $c_i = c + \mu_i$ for c drawn from $[a, A]$ for some $a > 0$ according to a distribution H, and μ_i distributed, for some $h < \frac{1}{2}$, as

$$p(\mu) = \left\{ \begin{array}{cc} h & \mu = \pm g \\ 1 - 2h & \mu = 0 \end{array} \right. .$$

Here g is the size of the error, and it measures the "opposite" of affiliation. In order to claim that there is some "continuity" in our game, we now show that for sufficiently low noise (g) the unique equilibrium of the game in the mechanism described above is still truthful reporting of c_i.

The original travelers' dilemma is that the airline lost the objects of two travelers but neither the airline nor the regulator know the true cost. In the case without noise, it is common knowledge that c is drawn from a non-degenerate distribution H. The dilemma is that a reasonably looking mechanism yields as the unique outcome that both players declare the lowest price: if both announcements coincide, the airline pays the individuals their claim; if one is lower than the other, both get the low price, and the announcer of the higher price must pay \$2 to the other traveler.

Consider the following adaptation of our mechanism above to cope with noise. Given a maximum expected cost of ε, the airline selects a $p < \varepsilon/k$ then

1. players announce their types, call σ_i the announcements.

2. The airline inspects the announcement of the highest σ with probability $\frac{2}{3}p$, and the smallest with probability $\frac{1}{3}p$ (and samples each with probability $\frac{p}{2}$ in case of a tie). If $|\sigma_1 - \sigma_2| > 2g$, both get $\min\{\sigma_1, \sigma_2\}$; this is also the case if an announcement is false, in which case a fine is applied to the player who reported the false claim. If no announcement is false and $|\sigma_1 - \sigma_2| \leq 2g$, the airline pays each player his announcement.

This mechanism allows for players to differ in their announcements, if they are not too far apart. The mechanism designer knows the structure of the game, so he knows g. Also, the two travelers have actually paid c_i (not c, and they don't remember exactly).

Theorem 3. *For any ε, there exists a $\overline{g} > 0$ such that for all $g < \overline{g}$ the unique equilibrium of the mechanism described in 1-2 above is truth telling, and the total cost is less than ε.*

The proof is simple, but tedious and long, so it is included in an online appendix.

3   Discussion

There is a vast literature about how to obtain truth telling as an equilibrium, and at the same time obtain the first best outcome. For instance, Kwerel (1977) obtains truth telling as one of potentially many equilibria when the regulator sells pollution licenses (which are assumed to be traded in a perfectly competitive market) and subsidizes firms which buy them in excess of their needs.[3] Also, Dasgupta, Hammond and Maskin (1980) use the Groves-Clarke mechanism to obtain dominant strategy truth telling with an unbalanced budget. Spulber (1988) presents a mechanism that, contrary to what happens with ours, does not attain the first best outcomes.

These prior papers are concerned with pollution control, and our paper is also relevant to this problem. In particular, our mechanism can be applied to situations where the regulator wants to attain the optimal pollution levels, and firms have access to the same abatement technologies.[4] In this context, we note that the mechanisms proposed in the literature are seldom used. This may happen because they possess two undesirable features that are not present in our mechanism.

The first is that they are complicated, a complaint often directed to the mechanism design literature. They are based on taxes, subsidies, or tradeable permits and these types of instruments have several implementation problems as compared to classic "command and control" instruments. Although they have been used recently, they have applied only in very specific contexts, and their implementation has been slow. For example, regulators in some countries are not educated in environmental economics and do not see the advantages of these instruments in terms of cost-effectiveness and efficiency; they see "command-and-control" instruments as stronger statements of support for environmental protection. Moreover, other regulators may think that it is immoral to let firms pollute just because they paid some taxes, or because they purchased pollution permits. Policymakers may also be reluctant to impose further costs on firms because of the impact on employment. Also, incentive-based instruments shift control decisions from regulatory staff to polluting firms, possibly affecting the regulator's job security and prestige.[5] The second is that they focus on whether truth telling is a Nash equilibrium of the revelation game, and not on whether truth telling is the unique equilibrium. If declaring large abatement costs is an equilibrium that yields higher profits for all firms, one will not observe firms telling the truth, but rather overestimating their costs.

Our mechanism is free from both problems. On the one hand, it is extremely simple, which makes it "likely" that players will understand their incentives.[6] In particular, we do not use some of the standard

---

[3] See Montero (2007) who fixes a problem with Kwerel's mechanism, but still retains the undesirable feature of unbalanced budget.

[4] Our mechanism can also be applied to finding out the optimal fisheries exploitation, when firms know the amount of fish available to all.

[5] These and other arguments are well documented in the literature. See for example Bohm and Russell (1985), Russell and Powell (1996), Lewis (1996), Keohane, Revesz and Stavins (1998).

[6] We thank Matt Jackson for many of the references in this Section, and for his comments regarding the importance of the simplicity of the mechanism and the proof.

techniques, like cross reporting, used in the literature on implementation with complete information. In addition, its unique equilibrium is truth telling.[7]

Moreover, our mechanism works with an arbitrarily imprecise inspection technology, and for arbitrarily small fines. Although inspections and fines have been used in the past and it is "known" that they help in the implementation problem, our assumptions are weaker and different than the ones that have been used before. For example, the important works of Mookherjee and P'ng (1989) and Ortuño-Ortin and Roemer (1993) used costly but perfectly informative inspections and sizeable fines. Our inspections can be as uninformative as one wants, and the fines can be arbitrarily small. Arya and Glover (2005) use a public signal that may be only slightly correlated with the player's reports to implement truth telling (to the owner of a firm) by a manager and his auditor. In their model, however, fines for lying can be large.

And of course, our results do not follow from any of the existing theorems in the literature. That is, there is no theorem that ensures that the social choice correspondence of the regulator, or any selection from it, is fully implementable in Nash equilibrium. The results in Jackson, Palfrey and Srivastava (1994) do not apply to our setting. Most importantly, their theorems are for implementation in undominated Nash, and our results are full Nash implementation (we get uniqueness without requiring that the strategies be undominated). Moreover, their Theorem 1 is for three or more firms, and their Theorem 3 requires the existence of a "worse outcome" that is not present in our setup.[8]

Finally, our results are not subject to the criticisms to full implementation in complete information that have been raised by Chung and Ely (2003), since our setup is, in their terminology, one of "private values".

There are three other strands of the literature that are related to our paper. The first is the literature on costly state verification, but none of the papers in that literature implies our results. Some of the closer papers in that literature are Diamond (1984), Gale and Hellwig (1985), Williamson (1986), Border and Sobel (1987), and most importantly due to the multilateral nature of our problem, Krasa and Villamil (1994) and Winton (1995)

Another branch of the literature to which our paper relates, especially the example with noise, is that on "informational size". In our basic setup, where all agents have the same valuation, agents have "nonexclusive information", as in Postlewaite and Schmeidler (1986). Basically, any information an agent might have is redundant given the information of other players.

In our example, where information is noisy, we have that agents are informationally small, as in McLean and Postlewaite (2002). Here, an agent's impact on posterior is positive but small, given other agents' types. McLean and Postlewaite (2004) argues that with informationally small agents, it is possible to obtain efficiency, even in the presence of interdependent valuations, where generic results indicate that this is impossible with independent types (Jehiel and Moldovanu, 2001). The result is accomplished with a version of the lotteries introduced in Crémer and McLean (1988), which might involve arbitrarily high payments.[9] In our model, by introducing an arbitrary small cost and an arbitrary low probability of inspection, we also achieve the first best, with an arbitrarily low budget. In addition to the standard analysis, we rule out other undesirable equilibria, where all agents inflate their reports.

---

[7]Dasgupta et al. also criticize Kwerel for the assumption that permits are traded in perfectly competitive markets and because of the weak "implementation" concept: that truth telling is a Bayesian Nash equilibrium. An additional problem of Kwerel is that his regulator has an unbalanced budget. In Dasgupta et al., if one requires a balanced budget one only obtains that truth telling is a Bayes Nash equilibrium (and neither uniqueness, nor dominant strategy implementation).

[8]A worse outcome in that setting would be to give each player $\min \text{Support} u(S)$ and a lottery which yields the fine with probabilty ε. We do not need to include such an outcome in our space of allocations for our mechanism to work. Our mechanism inspects only one player.

[9]A more exact comparison with Crémer and McLean is impossible because their mechanism cannot be adapted exactly to this setting. The reason is that in their mechanism there is only one good to be allocated among the agents who are competing for it. In this setting, agents objectives are more or less aligned.

There is, finally, a third strand of the literature which is related to our work. Some recent papers (Ben-Porath and Lipman (2012), Dutta and Sen (2012), Kartik and Tercieux (2012), and Kartik, Tercieux and Holden (2013) among others) consider a problem which is very similar to ours in that players know the state and there is a tie breaking rule in favor of honesty. In these interesting papers, players know the state of the world and can either provide "evidence" or have a lexicographic preference for telling the truth. As noted by Midjord (2013), whose planner could learn the true state with some (possibly small probability), the way all these mechanisms work is similar: the planner uses his/her small edge to break ties, and pitch players against each other. In every case, a player has a slight incentive to deviate from a "lying" equilibrium by providing different evidence (which may not change the outcome), or telling the truth (again without changing the outcome), but still breaking the proposed equilibrium. In our case we break undesired equilibria by letting players undercut each others' announcements and therefore avoiding the inspection (and potential fine). Our mechanism can be modified to avoid inspections if we let players also declare when an announcement is false.

To illustrate, suppose there are two travelers, and the mechanism chooses the lowest cost announcement and sets a fine of ε if in a tie the other player confesses. Players only like to confess if the announcement is false, and it doesn't change the outcome. In equilibrium, announcements can't be different (the player with the lowest announcement would want to increase it slightly), and can't be equal and false (both would like to confess, but in that case, each would have an incentive to declare a slightly larger cost).

The point of the previous example is to establish a simple formal connection between the two kinds of arguments. The results in this paper and in the prior literature are not nested however. In the honesty literature, players "are born" with a taste for honesty which we don't have available. But in favor of that literature, our planner can make inspections, which the planners in that literature cannot. Similar arguments apply to the availability of "evidence" (it is available in that literature, it is not in our context). In that sense we view our results as complementary, as different sets of assumptions may be more convenient in different setups. An important "technical" remark is that our model cannot be nested within the evidence or preference for honesty models. To illustrate, in the evidence models the set of messages that can be sent is different in different states of nature; that is what makes things implementable in that literature. One way to interpret that is as in Green and Laffont (1986) who argue that it is as if the designer has some information and can inflict "severe punishment" whenever an individual makes a false statement. What makes our problem harder (in this dimension) is that players can name any state in every state (and we have small fines, and not "severe punishment"). A similar comment applies to Lipman and Seppi (1995) where individuals have access to a technology where claims can be refuted; the requirement of refutability can be weakened when players preferences are not completely aligned. In our paper we have neither of those conditions.

We must stress that our results were circulated prior to most results in the literatures on evidence and honesty (see for example, Caffera and Dubra, 2005).

Finally, some of our results extend to situations where preferences over the state do not coincide; if anything, that makes the implementation problem harder. We have tried to keep the setup simple, and still applicable to a set of interesting and important problems. That extension is beyond the scope of this work.

References

[1] Arya, A. and J. Glover, (2005), "Maintaining Auditor Independence," mimeo.

[2] Basu, K. (1994), "The player's Dilemma: Paradoxes of Rationality in Game Theory," *American Economic Review P&P* **84(2)**, 391-95.

[3] Becker, G. S. and G. J. Stigler (1974), "Law Enforcement, Malfeasance and Compensation of Enforcers," *Journal of Legal Studies* **3(1)**, 1-18.

[4] Ben-Porath, E. and B. Lipman (2012), "Implementation with Partial Provability," *Journal of Economic Theory* **147**, pp. 1680-1724.

[5] Böhm, P. and C. S. Russell, (1985), "Comparative analysis of alternative policy instruments," in *Handbook of Natural Resource and Energy Economics*, Vol. I, A.V. Kneese and J.L. Sweeney. (Eds.), Elsevier.

[6] Border, K. and J. Sobel (1987), "Samurai accountant: a theory of auditing and plunder", *Review of Economic Studies*, **54** (4), 525-40.

[7] Bull, J. and J. Watson (2004), "Evidence Disclosure and Verifiability," *Journal of Economic Theory* **118**, 1-31.

[8] Caffera, M. and J. Dubra (2005), "Getting Polluters to Tell the Truth," SSRN wp 710346; also EconWPA of April 2005.

[9] Chung, K. and J. Ely, (2003), "Implementation with Near Complete Information," *Econometrica*, 71, 3, 857-871.

[10] Crémer, J. and R. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent," *Econometrica*, **53(2)**, 345-61.

[11] Crémer, J. and R. McLean (1988), "Full extraction of the surplus in bayesian and dominant strategy auctions," *Econometrica*, **56(6)**, 1247-57.

[12] Dasgupta, P., P. Hammond and E. Maskin, (1980), "On Imperfect Information and Optimal Pollution Control," *Review of Economic Studies*, **47** (5), 857-60.

[13] Diamond, D.W. (1984), "Financial Intermediation and Delegted Monitoring," *Review of Economic Studies*, **51** (3), 393-414.

[14] Dutta, B. and A. Sen (2012), "Nash Implementation with Partially Honest Individuals," *Games and Economic Behavior* **74(1)**, pp. 154-69

[15] Gale, D. and M. Hellwig (1985) "Incentive-Compatible Debt Contracts: The One-Period Problem," *Review of Economic Studies*, **52** (4), 647-63.

[16] Green, J. R. and J.-J. Laffont (1986), "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies* **53**, 447-56.

[17] Jackson, M. O., T. R. Palfrey, and S. Srivastava (1994): "Undominated Nash Implementation in Bounded Mechanisms," *Games and Economic Behavior*, 6, 474–501.

[18] Jehiel, P. and B. Moldovanu (2001), "Efficient design with interdependent values," *Econometrica* **69**, 1237–59.

[19] Kartik, N. and O. Tercieux (2012), "Implementation with Evidence," *Theoretical Economics* **7**, pp. 323-55.

[20] Kartik, N., O. Tercieux and R. Holden (2014), "Simple Mechanisms and Preferences for Honesty," *Games and Economic Behavior* **83**, pp. 284-90.

[21] Keohane, N. O., R. L. Revesz and R. N. Stavins, (1998), "The Choice of Regulatory Instruments in Environmental Policy," *Harvard Environmental Law Review*, 22 (2), 313-367.

[22] Krasa, S. and A. Villamil (1994), "Optimal multilateral contracts", *Economic Theory* 4, 167-87.

[23] Kwerel, E. (1977), "To Tell the Truth: Imperfect Information and Optimal Pollution Control," *Review of Economic Studies*, **44** (3), 595-601.

[24] Lewis, T. (1996), "Protecting the Environment when Costs and Benefits are Privately Known," *RAND Journal of Economics*, **27**, 819-47.

[25] Lipman, B. and D.J. Seppi, (1995), "Robust Inferences in Communication Games with Partial Provability," *Journal of Economic Theory* **66**, 370-405.

[26] Midjord, R. (2013), "Full implementation of rank dependent prizes," *Economics Letters* **119**, 261-3.

[27] McLean, R. and A. Postlewaite (2002), "Informational size and incentive compatibility," *Econometrica* **70(6)**, 2421–53.

[28] McLean, R. and A. Postlewaite (2004), "Informational size and efficient auctions," *Review of Economic Studies*, **71(3)**, 809–27.

[29] Montero, J.P. (2007) "A simple auction mechanism for the optimal allocation of the commons," mimeo.

[30] Mookherjee, D. and I. Png, (1989), "Optimal Auditing, Insurance, and Redistribution," *the Quarterly Journal of Economics,* Vol. 104, 2, pp. 399-415.

[31] Ortuño-Ortin, I. and J. Roemer (1993), "Politically realistic implementation with inspection: The equity-honesty-welfare trade-off," *Economics and Politics* 5, 255-270.

[32] Postlewaite, A. and D. Schmeidler (1986), "Implementation in differential information economies," *Journal of Economic Theory*, **39(1)**,14–33.

[33] Russell, C. S. and P. T. Powell, (1996), "Choosing Environmental Policy Tools, Theoretical Cautions and Practical Considerations," IADB, Washington D.C., - No. ENV-102.

[34] Spulber, D. (1988), "Optimal environmental regulation under asymmetric information," Journal of Public Economics **35(2)**, pp. 163-81.

[35] Townsend, R. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* **21**, 265-93.

[36] Williamson, S. (1986), "Costly monitoring, financial intermediation, and equilibrium credit rationing", *Journal of Monetary Economics* **18**, 159-79.

[37] Winton, A. (1995), "Costly State Verification and Multiple Investors: The Role of Seniority," *The Review of Financial Studies* **8(1)**, 91-123.

4   Online Appendix, not for publication.

Proof of Theorem 3. The number $\varepsilon_1\varepsilon_2$ is the expected value of the fine, if an individual is lying: the probability that the inspection will be effective times the value of the fine; for ease of notation let $q \equiv p\varepsilon_1\varepsilon_2$, since those always appear multiplied.

Truth telling is an equilibrium. Suppose every player is playing truthtelling. Then, marginal changes are not worth it (due to the penalty: changing the report to c' increases the penalty in a discrete fashion, times a discrete positive probability), and big changes won't change your payoff substantially because others are playing truthfully, and the payoff will be the min of others' reports.

There is no other equilibrium. Suppose there is an equilibrium profile of strategies (σ_1, σ_2) and a c_l such that for some player j, $\sigma_j(c_l) \neq c_l$. That is, suppose there is an equilibrium without truth telling. Without loss of generality, suppose $j = 1$, and let P denote 1's beliefs over σ_2 given c_l.

  Case 1: $P(\sigma_2 = \sigma_1(c_l)) > 0$. In this case player 1 is strictly better off playing $\sigma_1(c_l) - v$ for very small v : it strictly and discreetly decreases the expected fine by lowering the probability of inspection, while decreasing only marginally his payment across states. Formally, (omit (c_l) in $\sigma_1(c_l)$):

$$u(\sigma_1(c_l)) = P(\sigma_2 > \sigma_1)\left(\sigma_1 - \frac{q}{3}\right) + P(\sigma_2 = \sigma_1)\left(\sigma_1 - \frac{q}{2}\right) + P(\sigma_2 + 2g \geq \sigma_1 > \sigma_2)\left(\sigma_1 - \frac{2}{3}q\right)$$
$$+ P(\sigma_1 > \sigma_2 + 2g) E\left(\sigma_2 - \frac{2}{3}q \mid \sigma_1 > \sigma_2 + 2g\right)$$

To calculate the payoff in a small deviation, let $\overline{v}$ be sufficiently small that $\sigma_1 - \overline{v}$ does not fall below any other σ_2, or any other $\sigma_2 + 2g$ (this last part would be good for 1, but would complicate the comparison of payoffs):

$$\sigma_1(c_l) - \overline{v} > \max\left\{\{\sigma_2 : P(\sigma_2) > 0 \text{ and } \sigma_2 < \sigma_1\} \cup \{\sigma_2 + 2g : P(\sigma_2) > 0 \text{ and } \sigma_2 + 2g < \sigma_1\}\right\}. \quad (1)$$

This can be done because the set in (1) is finite. The payoff to $\sigma_1(c_l) - v$ for $v < \overline{v}$ is

$$u(\sigma_1(c_l) - v) = P(\sigma_2 > \sigma_1)\left(\sigma_1 - v - \frac{q}{3}\right) + P(\sigma_2 = \sigma_1)\left(\sigma_1 - v - \frac{q}{3}\right) + P(\sigma_2 + 2g \geq \sigma_1 > \sigma_2)\left(\sigma_1 - v - \frac{2}{3}q\right)$$
$$+ P(\sigma_1 > \sigma_2 + 2g) E\left(\sigma_2 - \frac{2}{3}q \mid \sigma_1 > \sigma_2 + 2g\right)$$

  We obtain

$$u(\sigma_1(c_l) - v) - u(\sigma_1(c_l)) = P(\sigma_2 > \sigma_1)(-v) + P(\sigma_2 = \sigma_1)\left(-v - \frac{q}{3} + \frac{q}{2}\right) + P(\sigma_2 + 2g \geq \sigma_1 > \sigma_2)(-v)$$
$$= P(\sigma_2 = \sigma_1)\frac{q}{6} - v(1 - P(\sigma_1 > \sigma_2 + 2g))$$

so for sufficiently small v, individual 1 is better off undercutting his equilibrium strategy, which is a contradiction.

  Case 2: $P(\sigma_2 \leq \sigma_1(c_l)) = 0$. In this case, player 1 is strictly better off increasing his σ_1 to any value $\sigma_1(c_l) < \sigma_1 < \min\{\sigma_2 : P(\sigma_2) > 0\}$. The refund from the airline would increase in every case, without changing the chances of being inspected.

  For the following two cases, let $\overline{\sigma}_2 \equiv \max\{\sigma_2 : P(\sigma_2) > 0 \text{ and } \sigma_2 < \sigma_1\}$.

Case 3: $P\left(\sigma_2 = \sigma_1\left(c_l\right)\right) = 0$ and $\overline{\sigma}_2 + 2g \geq \sigma_1$. In this case there is a positive probability σ_2 close by from the intended play by player 1. In that case, player 1 is better off undercutting $\overline{\sigma}_2$. In order to keep the proof simple, instead of undercutting, player 1 will tie $\overline{\sigma}_2$. The payoff of playing σ_1 is

$$
\begin{aligned}
u\left(\sigma_1\left(c_l\right)\right) &= P\left(\sigma_2 > \sigma_1\right)\left(\sigma_1 - \frac{q}{3}\right) + P\left(\overline{\sigma}_2\right)\left(\sigma_1 - \frac{2}{3}q\right) + P\left(\overline{\sigma}_2 > \sigma_2 \geq \sigma_1 - 2g\right)\left(\sigma_1 - \frac{2}{3}q\right) \\
&\quad + P\left(\sigma_1 > \sigma_2 + 2g\right)E\left(\sigma_2 - \frac{2}{3}q \mid \sigma_1 > \sigma_2 + 2g\right)
\end{aligned}
$$

and that of playing $\sigma_1 = \overline{\sigma}_2$ is

$$
\begin{aligned}
u\left(\overline{\sigma}_2\right) &\geq P\left(\sigma_2 > \sigma_1\right)\left(\overline{\sigma}_2 - \frac{q}{3}\right) + P\left(\overline{\sigma}_2\right)\left(\overline{\sigma}_2 - \frac{1}{2}q\right) + P\left(\overline{\sigma}_2 > \sigma_2 \geq \sigma_1 - 2g\right)\left(\overline{\sigma}_2 - \frac{2}{3}q\right) \\
&\quad + P\left(\sigma_1 > \sigma_2 + 2g\right)E\left(\sigma_2 - \frac{2}{3}q \mid \sigma_1 > \sigma_2 + 2g\right)
\end{aligned}
$$

(the weak inequality is because the payoff in the last term may be $\overline{\sigma}_2 \geq \sigma_1 - 2g$ instead of $\sigma_2 < \sigma_1 - 2g$ when $\sigma_2 + 2g \geq \overline{\sigma}_2$). Then,

$$
\begin{aligned}
u\left(\overline{\sigma}_2\right) - u\left(\sigma_1\left(c_l\right)\right) &\geq P\left(\sigma_2 > \sigma_1\right)\left(\overline{\sigma}_2 - \sigma_1\right) + P\left(\overline{\sigma}_2\right)\left(\overline{\sigma}_2 - \overline{\sigma}_1 - \frac{1}{2}q + \frac{2}{3}q\right) + P\left(\overline{\sigma}_2 > \sigma_2 \geq \sigma_1 - 2g\right)\left(\overline{\sigma}_2 - \sigma_1\right) \\
&= P\left(\overline{\sigma}_2\right)\frac{1}{6}q + \left(1 - P\left(\sigma_1 > \sigma_2 + 2g\right)\right)\left(\overline{\sigma}_2 - \sigma_1\right) \geq P\left(\overline{\sigma}_2\right)\frac{1}{6}q - \left(1 - P\left(\sigma_1 > \sigma_2 + 2g\right)\right)2g
\end{aligned}
$$

so if $g < P\left(\overline{\sigma}_2\right)\frac{1}{6}q$ the deviation is profitable. Note that $P\left(\overline{\sigma}_2\right)$ is independent of g: for any g, and any c_l there are only 5 possible values of σ_2, and their probabilities do not depend on the value of g. Moreover, since H is bounded away from 0, $P\left(\overline{\sigma}_2\right)$ is also bounded away from 0 and the same g works for every possible c_l.

Case 4: $P\left(\sigma_2 \geq \sigma_1\left(c_l\right)\right) = 0$ and $\overline{\sigma}_2 + 2g < \sigma_1$. In this case, player 1 is better off playing $\sigma_1 = \overline{\sigma}_2 + 2g$: the payoff strictly improves when $\sigma_2 = \overline{\sigma}_2$ (he was being paid $\overline{\sigma}_2$, because his claim was far away from $\overline{\sigma}_2$, and in the deviation he'll get paid $\overline{\sigma}_2 + 2g$) and doesn't change otherwise. Formally, we have

$$
u\left(\sigma_1\left(c_l\right)\right) = P\left(\sigma_2 > \sigma_1\right)\left(\sigma_1 - \frac{q}{3}\right) + P\left(\overline{\sigma}_2\right)\left(\overline{\sigma}_2 - \frac{2}{3}q\right) + P\left(\overline{\sigma}_2 > \sigma_2\right)E\left(\sigma_2 - \frac{2}{3}q \mid \overline{\sigma}_2 > \sigma_2\right)
$$

and the payoff of playing $\overline{\sigma}_2 + 2g$ is

$$
u\left(\overline{\sigma}_2 + 2g\right) = P\left(\sigma_2 > \sigma_1\right)\left(\overline{\sigma}_2 + 2g - \frac{q}{3}\right) + P\left(\overline{\sigma}_2\right)\left(\overline{\sigma}_2 + 2g - \frac{2}{3}q\right) + P\left(\overline{\sigma}_2 > \sigma_2\right)E\left(\sigma_2 - \frac{2}{3}q \mid \overline{\sigma}_2 > \sigma_2\right).
$$

The difference in payoffs is then

$$
u\left(\overline{\sigma}_2 + 2g\right) - u\left(\sigma_1\left(c_1\right)\right) = P\left(\overline{\sigma}_2\right)\left(\overline{\sigma}_2 + 2g - \frac{2}{3}q - \overline{\sigma}_2 + \frac{2}{3}q\right) = P\left(\overline{\sigma}_2\right)2g > 0
$$

so that the deviation is profitable.

Case 5: $P\left(\sigma_2 = \sigma_1\left(c_l\right)\right) = 0$ and $\overline{\sigma}_2 + 2g < \sigma_1$ and $P\left(\sigma_2 > \sigma_1\left(c_l\right)\right) > 0$. In this case, player 1 is better off slightly increasing his claim: for $\sigma_2 < \overline{\sigma}_2$ it changes nothing (since he will be paid $\sigma_2 9$, and for $\sigma_2 > \overline{\sigma}_2$ (and there are some with positive probability), it strictly increases player 1's payoff.

∎