Mathematical Social Sciences ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/mss

# Mechanism design when players' preferences and information coincide☆

Marcelo Caffera [a], Juan Dubra [a,*], Nicolás Figueroa [b]

[a] Universidad de Montevideo, Uruguay
[b] Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

*Article history:*
Received 26 January 2017
Received in revised form 14 June 2017
Accepted 21 August 2017
Available online xxxx

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, in a model with complete information among the players, 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). Moreover, the setup, the mechanism and the proofs are all very simple.

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,

☆ A previous version of this paper was circulated as "Getting Polluters to Tell the Truth". Kim-Sau Chung suggested we change the setting to that of the Travelers' Dilemma, and we are grateful for this suggestion. Hugo Hopenhayn noted a bug in a previous version of this paper, and we are grateful for his help. We also thank Atila Abdulkadiroglu, Ezequiel Aguirre, Anil Arya, Jean-Pierre Benoît, Carlos Chavez, Marcelo Cousillas, Federico Echenique, Jeff Ely, Néstor Gandelman, Jonathan Glover, Ana María Ibáñez, Matt Jackson, Larry Kotlikoff, Carlos Lacurcia, Preston McAfee, Stephen Morris and Francesco Squintani. Figueroa acknowledges support from Complex Engineering Systems Institute (CONICYT – PIA – FB0816).

* Corresponding author.
*E-mail address:* dubraj@um.edu.uy (J. Dubra).

http://dx.doi.org/10.1016/j.mathsocsci.2017.08.002
0165-4896/

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 implementation 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 the 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.

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 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 $\varepsilon_1$ (arbitrarily small) and a probability $\varepsilon_2$ 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 $\varepsilon_2$, 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 following:

- the expected value of the distance between the true state and the one chosen by the regulator;
- 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 $\varepsilon_1$, 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 $\varepsilon_1$.

**Theorem 1.** *For any $\varepsilon > 0$ there is a mechanism that has a total expected cost of less than $\varepsilon$. 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] 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 would not work.