Games and Economic Behavior 101 (2017) 34–48

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

A dynamic non-direct implementation mechanism for interdependent value problems[☆]

Richard P. McLean[a], Andrew Postlewaite[b,*]

[a] Rutgers University, United States
[b] University of Pennsylvania, United States

ARTICLE INFO

Article history:
Received 2 March 2015
Available online 10 July 2015

JEL classification:
C70
D44
D60
D82

Keywords:
Privacy
Auctions
Incentive compatibility
Mechanism design
Interdependent values
Ex post incentive compatibility
Informational size

ABSTRACT

Much of the literature on mechanism design and implementation uses the revelation principle to restrict attention to direct mechanisms. We showed in McLean and Postlewaite (2014) that when agents are informationally small, there exist small modifications to VCG mechanisms in interdependent value problems that restore incentive compatibility. We show here how one can construct a two-stage non-direct mechanism that similarly restores incentive compatibility while improving upon the direct one stage mechanism in terms of privacy and the size of messages that must be sent. The first stage that elicits the part of the agents' private information that induces interdependence can be used to transform certain other interdependent value problems into private value problems.

1. Introduction

Much of mechanism design and implementation theory employs direct mechanisms. In these mechanisms, the participants' actions are taken to be their types, which embody all relevant information that is not common knowledge. In many applications, the analysis can be restricted to direct mechanisms without loss of generality since, for many mechanisms, the revelation principle holds: to any Bayes equilibrium of a game defined for a mechanism with given outcome function and message sets, there exists a payoff equivalent direct revelation mechanism in which truthful reporting is an equilibrium.

While it is correct that restricting attention to direct mechanisms is without loss of generality in this sense, there are important problems in which this is not the case. In a typical application of the revelation principle, one solves a principal–agent problem by maximizing the principal's utility function subject to incentive constraints requiring that the agent(s) announce truthfully. This will indeed yield an equilibrium of the revelation game, but there may be additional equilibria. Postlewaite and Schmeidler (1986) provide an example in which, in addition to the truthful-announcement equilibrium, there is a non-truthful equilibrium in which all agents with private information are strictly better off. *Exact implementation*

[☆] This paper is a revised version of our earlier paper "Informational Size and Two Stage Mechanisms". We thank the National Science Foundation for financial support (NSF grant 1260753). We thank Rakesh Vohra for helpful conversations and Zehao Hu for excellent proofreading.

[*] Corresponding author.
E-mail address: apostlew@econ.upenn.edu (A. Postlewaite).

http://dx.doi.org/10.1016/j.geb.2015.06.007
0899-8256/© 2015 Elsevier Inc. All rights reserved.

aims at identifying *sets* of outcomes that can be implemented in asymmetric information problems in the sense that an outcome is an equilibrium outcome if and only if it is in the set. Much of that literature considers mechanisms in which agents announce more than their private type and demonstrates the benefits of going beyond direct mechanisms.[1]

There is a second reason to consider non-direct mechanisms. In a direct mechanism, agent i typically announces $t_i \in T_i$, where T_i is the set of possible types for i. In the implementation literature initiated by Hurwicz (1972) T_i was often taken to be the set of neoclassical preferences, or essentially the set of concave utility functions on $\mathbb{R}^L_+$. If no additional restrictions are placed on T_i, communicating i's type entails sending an infinite dimensional message. It may well be the case that performance functions that can be exactly implemented in revelation games can also be implemented in non-revelation games in which agents need to send dramatically less information than their full type.[2]

A third, and perhaps most significant, reason to consider non-direct mechanisms is a concern for privacy. Participants in a mechanism might be concerned about revealing all of their private information for two reasons. First, third parties may be able to intercept that information and use it to the detriment of the participating agent.[3] Second, even in the absence of a data breach, agents may be uneasy about how the information they transmit might be used in the future.[4] Given this concern, mechanisms that decrease the amount of information revealed relative to that revealed in a direct mechanism are *a priori* preferred. Our aim in this paper is to show how mechanisms that perform well in interdependent value problems can be improved upon with respect to privacy and the size of messages agents send by moving from a one stage direct mechanism to a two stage non-direct mechanism.

The Vickrey–Clarke–Groves mechanism (hereafter VCG) for private values environments is a classic example of a mechanism for which truthful revelation is ex post incentive compatible. It is well-known, however, that truthful revelation is generally no longer incentive compatible when we move from a private values environment to an interdependent values environment. In McLean and Postlewaite (forthcoming, 2014) (henceforth MP (2014)) we showed that, when agents are informationally small in the sense of McLean and Postlewaite (2002), there exists a modification of a generalized VCG mechanism using small additional transfers that restores incentive compatibility. This paper presents an alternative, two-stage non-direct mechanism that accomplishes the same goal – restoring incentive compatibility for interdependent value problems. The advantage of the two stage mechanism relative to a single stage mechanism is that, for typical problems, agents need to transmit substantially less information.

We will explain intuitively the nature of the savings in transmitted information. Consider a problem in which there is uncertainty about the state of nature. An agent's private information consists of a state dependent payoff function and a signal correlated with the state. A single stage mechanism that delivers an efficient outcome for any realization of agents' types must do two things. First, it must elicit the information agents have about the state of nature to determine the posterior probability distribution given that information. Second, it must elicit agents' privately known state dependent payoffs. A two stage mechanism can separate the two tasks. First, elicit the information about the state of nature, but relay to agents the posterior distribution on the state of nature before collecting any additional information.

When agents are induced to reveal their information about the state of nature truthfully, relaying the posterior distribution on the state of nature converts the interdependent value problem into a private value problem. Knowing the probability distribution on the set of states of nature, agents need only report their expected utility for each possible social outcome rather than their utility for every social outcome in each of the states. Essentially, by moving from a one stage mechanism to a two stage mechanism, we can shift the job of computing expected utilities given the posterior from the mechanism to the agents, reducing the information that must be reported to the mechanism. An important side benefit of separating out the first stage elicitation of information is that the method can be used to transform other interdependent value problems into private value problems.

We provide an example of how our mechanism works in the next section, and present the general mechanism after that.

2. Example

The following single object auction example, a modification of the example in McLean and Postlewaite (2004), illustrates our basic idea. An object is to be sold to one of three agents. There are two equally likely states of the world, θ_1 and θ_2, and an agent's value for the object depends on the state of the world. Agent i's state dependent utility function can be written as $v_i = (v_i^1, v_i^2) = (v_i(\theta_1), v_i(\theta_2))$ where v_i^j is his utility of the object in state θ_j. An agent's utility function is private information. In addition, each agent i receives a private signal $s_i \in \{a_1, a_2\}$ correlated with the state. These signals are independent conditional on the state and the conditional probabilities are as shown in the following table.

[1] See, e.g., Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989) and Jackson (1991).

[2] For example, Postlewaite and Wettstein (1989) show that the Walrasian correspondence can essentially be implemented by a mechanism in which agents announce prices and demands in contrast to the direct mechanism that would require infinite dimensional announcements.

The seminal paper on message size requirements of mechanisms is Mount and Reiter (1974); see Ledyard and Palfrey (1994, 2002) for more recent investigations of the message size requirements of mechanisms.

[3] See, e.g., https://corporate.target.com/about/shopping-experience/payment-card-issue-faq.

[4] See Heffetz and Ligett (2014) for examples and a nice discussion on this issue.

state \ **signal**	a_1	a_2
θ_1	ρ	$1-\rho$
θ_2	$1-\rho$	ρ

where $\rho > \frac{1}{2}$. Consequently, an agent's private information, his type, is a pair (s_i, v_i) and we make two assumptions. First, for any type profile $(s_i, v_i)_{i=1}^3$, the conditional distribution on the state space given $(s_i, v_i)_{i=1}^3$ depends only on the signals (s_1, s_2, s_3). Therefore, the agents' utility functions provide no information relevant for predicting the state that is not already contained in the signal profile alone. Second, we assume that for any type (s_i, v_i) of agent i, the conditional distribution on the signals s_{-i} of the other two agents given (s_i, v_i) depends only on i's signal s_i. Note that the conditional distribution on the state space given (s_1, s_2, s_3) and the conditional distribution on the signals s_{-i} given s_i can be computed using the table above.

Suppose the objective is to allocate the object to the agent for whom the expected value, conditional on the agents' true signal profile, is highest. This is a problem with interdependent values. Agent i's conditional expected value for the object depends on the probability distribution on the states, conditional on the signals of all three agents. MP (2014) show how one can design a direct mechanism to allocate the object to the highest value agent. Each agent reports his type (s_i, v_i) and the mechanism uses the reported signals about the state $s = (s_1, s_2, s_3)$ to compute the posterior distribution $(\pi(\theta_1|s), \pi(\theta_2|s))$ on Θ. This posterior is used along with agent i's announced state dependent utilities to compute i's expected utility $\bar{v}_i(s) = v_i^1\pi(\theta_1|s) + v_i^2\pi(\theta_2|s)$. The mechanism then awards the object to an agent i with the highest expected value $\bar{v}_i(s)$ and that agent pays the second highest expected value from the set $\{\bar{v}_j(s)\}_{j\neq i}$ while agents different from i pay nothing.[5]

As stated, this direct mechanism is (not even) Bayesian incentive compatible. To induce agents to truthfully announce their signals, MP (2014) reward every agent j (winners and losers) with a positive payment z_j if his reported signal is in the majority. Since agents receive conditionally independent signals about the state, agent j maximizes the probability that he gets the reward z_j by announcing truthfully if other agents are doing so. Since the maximal possible gain from misreporting is bounded, a sufficiently large value of z_j will make truthful announcement a Bayes–Nash equilibrium. Agent j's reward z_j need not be very large if ρ is close to 1. When ρ is close to 1, it is very likely that all agents received "correct" signals about the state. Therefore, conditional on his own signal, agent j believes that a lie will, with high probability, have only a small effect on the posterior distribution on Θ. But the expected gain from misreporting will be small if the expected change in the posterior is small. Thus, when agents are receiving very accurate signals about θ, small rewards will support truthful announcement as a Bayes–Nash equilibrium.

Our aim in this paper is to provide a non-direct two stage version of this one stage direct mechanism that accomplishes the same goal but requires less information to be transmitted. As an illustration, consider an alternative two stage, non-direct approach to the same simple allocation problem described above in which the mechanism elicits the information about the unknown state θ in stage one, and then uses a Vickrey auction in stage two based on the expected values that are computed using the information about the state revealed in stage one. In stage one, agents report a (not necessarily truthful) signal profile $s = (s_1, s_2, s_3)$ and the mechanism then publicly posts the posterior distribution $(\rho(\theta_1), \rho(\theta_2)) = (\pi(\theta_1|s), \pi(\theta_2|s))$ computed from s. Using the posted distribution, agents can compute the associated expected payoffs $v_i^1\rho(\theta_1) + v_i^2\rho(\theta_2)$. These expected payoffs are then reported to the mechanism and the mechanism awards the object to the agent who reports the highest expected payoff. If agents report their true signals in stage one and if the true signal profile is $s = (s_1, s_2, s_3)$, then by the well known property of Vickrey auctions, it is a dominant strategy in stage 2 for each agent i to truthfully report his expected payoff $\bar{v}_i(s)$. Of course, agents may have an incentive to misreport their signals in order to manipulate the conditional expected valuations that are used in the second stage to determine the winner and the price. For example, if all agents have state dependent values that are lower in state θ_1 than in state θ_2 ($v_i^1 < v_i^2,\ i = 1, 2, 3$), then an agent who has received signal a_2 may have an incentive to report a_1. Such a misreport will increase the probability weight that the posterior assigns to θ_1. Consequently, this will lower all agents' expected values which, in turn, will affect the price paid by the winner of the object. To induce honest reporting in stage one of a perfect Bayesian equilibrium of this two stage game, we use the same sort of reward system as that employed in the one stage direct mechanism of MP (2014).

While the equilibrium outcomes of the one stage direct mechanism and two stage non-direct mechanism are identical, there is a difference in the information that is reported to the mechanism. In the direct mechanism, agent i reports his type $(s_i, (v_i^1, v_i^2))$. In the two stage game, agent i reports s_i and a real number corresponding to his expected payoff computed using the posted posterior from the first stage. Consequently, agents reveal less information in the non-direct mechanism. We note that moving to two stages is necessary for there to be a reduction in the information sent to the mechanism; it cannot be accomplished with a one stage mechanism, even if we allow non-direct mechanisms. More generally, we first show how one can decompose agents' types into "informationally relevant" and "payoff relevant" components as in the example, and then construct the two stages along the lines of the two parts of the mechanism discussed above. We demonstrate how this two stage approach can transform an interdependent value problem into a private value problem for

[5] If there is more than one agent with the highest expected value the object is awarded to each with equal probability.

the VCG mechanism, but our basic approach can be used for other interdependent value implementation and mechanism design problems.

3. Preliminaries

In this section, we review the structure and salient results from MP (2014). If K is a finite set, let $|K|$ denote the cardinality of K and let $\Delta(K)$ denote the set of probability measures on K. Throughout the paper, $||\cdot||_2$ will denote the 2-norm and, for notational simplicity, $||\cdot||$ will denote the 1-norm. The real vector spaces on which these norms are defined will be clear from the context. Let $\Theta = \{\theta_1, .., \theta_m\}$ represent the finite set of states of nature and let T_i denote the finite set of types of player i. Let $\Delta^*(\Theta \times T)$ denote the set of $P \in \Delta(\Theta \times T)$ whose marginals P_Θ on Θ and P_T on T satisfy the following full support assumptions: $P_\Theta(\theta) > 0$ for each $\theta \in \Theta$ and $P_T(t) > 0$ for each $t \in T$. The conditional distribution induced by P on Θ given $t \in T$ (resp., the conditional distribution induced by (the marginal of) P on T_{-i} given $t_i \in T_i$) is denoted $P_\Theta(\cdot|t)$ (resp., $P_{T_{-i}}(\cdot|t_i)$). Let C denote the finite set of social alternatives. Agent $i's$ payoff is represented by a nonnegative valued function $v_i : C \times \Theta \times T_i \to \mathbb{R}_+$ and we assume that for all i, $v_i(\cdot,\cdot,\cdot) \leq M$ for some $M \geq 0$.

A *social choice problem* is a collection $(v_1, .., v_n, P)$ where $P \in \Delta^*(\Theta \times T)$. An *outcome function* is a mapping $q : T \to C$ that specifies an outcome in C for each profile of announced types. A *mechanism* is a collection $(q, x_1, .., x_n)$ (written simply as $(q, (x_i))$ where $q : T \to C$ is an outcome function and each $x_i : T \to \mathbb{R}$ is a transfer function. For any profile of types $t \in T$, let

$$\hat{v}_i(c; t) = \hat{v}_i(c; t_{-i}, t_i) = \sum_{\theta \in \Theta} v_i(c, \theta, t_i) P_\Theta(\theta | t_{-i}, t_i).$$

Although $\hat{v}$ depends on P, we suppress this dependence for notational simplicity as well. Finally, we make the simple but useful observation that the pure private value model is mathematically identical to a model in which $|\Theta| = 1$.

Definition 1. Let $(v_1, .., v_n, P)$ be a social choice problem. A mechanism $(q, (x_i))$ is:

interim incentive compatible if truthful revelation is a *Bayes–Nash equilibrium*: for each $i \in N$ and all $t_i, t_i' \in T_i$,

$$\sum_{t_{-i} \in T_{-i}} \left[\hat{v}_i(q(t_{-i}, t_i); t_{-i}, t_i) + x_i(t_{-i}, t_i)\right] P_{T_{-i}}(t_{-i}|t_i) \geq \sum_{t_{-i} \in T_{-i}} \left[\hat{v}_i(q(t_{-i}, t_i'); t_{-i}, t_i) + x_i(t_{-i}, t_i')\right] P_{T_{-i}}(t_{-i}|t_i).$$

ex post individually rational if

$$\hat{v}_i(q(t); t) + x_i(t) \geq 0 \text{ for all } i \text{ and all } t \in T.$$

feasible if for each $t \in T$,

$$\sum_{j \in N} x_j(t) \leq 0.$$

outcome efficient if for each $t \in T$,

$$q(t) \in \arg\max_{c \in C} \sum_{j \in N} \hat{v}_j(c; t).$$

4. The model

4.1. Information decompositions

In this section, we show how the information structure for general incomplete information problems, even those without a product structure, can be represented in a way that separates out an agent's information about the state θ. This is important because it is this part of his type that affects other agents' valuations for the social alternatives. The example of Section 2 illustrates how we can elicit truthful reporting of agents' signals about the state when they are correlated.

In that example, an agent has beliefs about other agents' signals that depend on his own signal, and it is important that the beliefs are different for different signals the agent may receive. In the example, agent i's type consists of a signal a_i and a state dependent utility function that is independent of his signal. Consequently agent i has multiple types consisting of the same signal but different utility functions, and all of these types will necessarily have the same beliefs regarding other agents' signals.

It isn't necessary to elicit that part of an agent's type that doesn't affect other agents' valuations (e.g., his utility function in the example) to cope with the interdependence, only the part related to the state θ. To formalize this idea, we recall the notion of information decomposition from McLean and Postlewaite (2004).[6]

[6] This definition is equivalent to the partition formulation in McLean and Postlewaite (2004).

Definition 2. Suppose that $P \in \Delta^*(\Theta \times T)$. An information decomposition for P is a collection $\mathbb{D} = ((A_i, f_i)_{i \in N}, Q)$ satisfying the following conditions:

(i) For each i, A_i is a finite set, $f_i : T_i \to A_i$ is a function, and $Q \in \Delta(\Theta \times A_1 \times \cdots \times A_n)$.[7] For each $t \in T$, define $f(t) := (f_1(t_1), .., f_n(t_n))$ and

$$f_{-i}(t_{-i}) := (f_1(t_1), .., f_{i-1}(t_{i-1}), f_{i+1}(t_{i+1}), .., f_n(t_n)).$$

(ii) For each $t \in T$,

$$P_\Theta(\theta|t) = Q_\Theta(\theta|f(t))$$

(iii) For each i, $t_i \in T_i$ and $a \in A$,

$$\sum_{\substack{t_{-i} \in T_{-i} \\ : f_{-i}(t_{-i}) = a_{-i}}} P_{T_{-i}}(t_{-i}|t_i) = Q_{A_{-i}}(a_{-i}|f_i(t_i))$$

If $t_i \in T_i$, we will interpret $f_i(t_i) \in A_i$ as the "informationally relevant component" of t_i and we will refer to A_i as the set of agent i's "signals." Condition (ii) states that a type profile $t \in T$, contains no information beyond that contained in the signal profile $f(t)$ that is useful in predicting the state of nature. Condition (iii) states that a specific type $t_i \in T_i$ contains no information beyond that contained in the signal $f_i(t_i)$ that is useful in predicting the signals of other agents.

Every $P \in \Delta^*(\Theta \times T)$ has at least one information decomposition in which $A_i = T_i$, $f_i = id$, and $Q = P$ which we will refer to as the *trivial decomposition*. However, the trivial decomposition may not be the only one (or the most useful one as we will show below). For example, suppose that each agent's type set has a product structure $T_i = X_i \times Y_i$ and that $P \in \Delta^*(\Theta \times T)$ satisfies

$$P(\theta, x_1, y_1, .., x_n, y_n) = P_1(\theta, x_1, .., x_n) P_2(y_1, .., y_n)$$

for each $(x_1, y_1, .., x_n, y_n)$ where $P_1 \in \Delta(\Theta \times X)$ and $P_2 \in \Delta(Y)$. Then defining the projection map $p_{X_i}(x_i, y_i) = x_i$, it follows that $\mathbb{D} = ((X_i, p_{X_i})_{i \in N}, P_1)$ is an information decomposition for P.

Remark. If $(v_1, .., v_n, P)$ is a social choice problem, then it follows from the definition that any two information decompositions $\mathbb{D} = ((A_i, f_i)_{i \in N}, Q)$ and $\mathbb{D}' = ((A'_i, f'_i)_{i \in N}, Q')$ for P give rise to the same $\hat{v}_i$, i.e., for all $t \in T$, we have

$$\sum_{\theta \in \Theta} v_i(c, \theta, t_i) Q_\Theta(\theta|f(t)) = \sum_{\theta \in \Theta} v_i(c, \theta, t_i) P_\Theta(\theta|t_{-i}, t_i) = \sum_{\theta \in \Theta} v_i(c, \theta, t_i) Q'_\Theta(\theta|f'(t)).$$

4.2. Informational size

In this paper, a fundamental role is played by the notion of informational size. Suppose that $\mathbb{D} = ((A_i, f_i)_{i \in N}, Q)$ is an information decomposition for $P \in \Delta^*(\Theta \times T)$. In a direct mechanism, agent i reports an element of T_i to the mechanism. Consider an alternative scenario in which each agent i reports a signal $a_i \in A_i$ to the mechanism. If i reports a_i and the remaining agents report a_{-i}, it follows that the profile $a = (a_{-i}, a_i) \in A$ will induce a conditional distribution on Θ (computed from Q) and, if agent i's report changes from a_i to a'_i, then this conditional distribution will (in general) change. We consider agent i to be *informationally small* if, for each a_i, agent i ascribes "small" probability to the event that he can effect a "large" change in the induced conditional distribution on Θ by changing his announced type from a_i to some other a'_i. This is formalized in the following definition.

Definition 3. Suppose that $\mathbb{D} = ((A_i, f_i)_{i \in N}, Q)$ is an information decomposition for $P \in \Delta^*(\Theta \times T)$. Let

$$I^i_\varepsilon(a'_i, a_i) = \{a_{-i} \in A_{-i} | \; ||Q_\Theta(\cdot|a_{-i}, a_i) - Q_\Theta(\cdot|a_{-i}, a'_i)|| > \varepsilon\}$$

and

$$\nu^Q_i(a'_i, a_i) = \min\{\varepsilon \geq 0 | \sum_{a_{-i} \in I^i_\varepsilon(a'_i, a_i)} Q_\Theta(a_{-i}|a_i) \leq \varepsilon\}.$$

The *informational size* of agent i is defined as

$$\nu^Q_i = \max_{a_i \in A_i} \max_{a'_i \in A_i} \nu^Q_i(a'_i, a_i).$$

[7] The conditional distribution induced by Q on Θ given $a \in A$ (resp., the conditional distribution induced by (the marginal of) Q on A_{-i} given $a_i \in A_i$) is denoted $Q_\Theta(\cdot|a)$ (resp., $Q_{A_{-i}}(\cdot|a_i)$).

4.3. Variability of beliefs

The example of Section 2 illustrates how one might induce truthful announcement of agents' signals about the state. An agent who receives the signal a_1 believes that the state is more likely to be θ_1 than θ_2. Given that agents' signals are conditionally independent, he believes that each of the other agents is more likely to have received signal a_1 than a_2. Hence, if those agents are announcing truthfully, he maximizes his chance of receiving the reward z by announcing truthfully as well. More generally, the key to constructing rewards for agent i who might receive signal a_i or a_i' is a requirement that agent i's beliefs regarding other agents' signals when he receives signal a_i differ from his beliefs when he receives signal a_i'. Moreover, the magnitude of the difference matters in inducing truthful reporting. We turn next to defining a measure of the variation of an agent's beliefs.

To define formally the measure of variability, we treat each conditional $Q_{A_{-i}}(\cdot|a_i) \in \Delta(A_{-i})$ as a point in an Euclidean space of dimension equal to the cardinality of A_{-i}. Our measure of variability is defined as

$$\Lambda_i^Q = \min_{a_i \in A_i} \min_{a_i' \in A_i \backslash a_i} \left\| \frac{Q_{A_{-i}}(\cdot|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2} - \frac{Q_{A_{-i}}(\cdot|a_i')}{||Q_{A_{-i}}(\cdot|a_i')||_2} \right\|^2 .$$

If $\Lambda_i^Q > 0$, then the agents' signals cannot be stochastically independent with respect to Q. We will exploit this correlation in constructing Bayesian incentive compatible mechanisms. For a discussion of the relationship between this notion of correlation and that found in the full extraction literature, see MP (2014).

It is important to point out that Λ_i^Q and $\Lambda_i^{Q'}$ are generally different for two decompositions $D = ((A_i, f_i)_{i \in N}, Q)$ and $D' = ((A_i', f_i')_{i \in N}, Q')$ for P. When an agent's type set has a product structure $T_i = X_i \times Y_i$ as in the example of Section 4.1 and $D = ((T_i, id)_{i \in N}, P)$ is the trivial decomposition, then $\Lambda_i^Q = \Lambda_i^P = 0$ for all i. However, for the decomposition $D = ((X_i, p_{X_i})_{i \in N}, P_1)$ of that example, it may in fact be the case that $\Lambda_i^{P_1} > 0$. The utility of decompositions will become apparent when we state Theorem B below.

5. The one stage implementation game

5.1. The generalized VCG mechanism

We now adapt some of our previous results on implementation with interdependent values to the model of this paper. In the special case of pure private values, i.e., when $|\Theta| = 1$, it is well known that the classical VCG transfers will implement an outcome efficient social choice function: in the induced direct revelation game, it is a dominant strategy to honestly report one's type. In the general case of interdependent values, the situation is more delicate.

Let $q : T \to C$ be an outcome efficient social choice function for the problem $(v_1, .., v_n, P)$. For each t, define transfers as follows:

$$\alpha_i^q(t) = \sum_{j \in N \backslash i} \hat{v}_j(q(t); t) - \max_{c \in C} \left[\sum_{j \in N \backslash i} \hat{v}_j(c; t) \right]$$

Note that $\alpha_i^q(t) \le 0$ for each i and t. The resulting mechanism $(q, (\alpha_i^q))$ is the *generalized VCG mechanism with interdependent valuations* (GVCG for short) studied in MP(2014). It is straightforward to show that the GVCG mechanism is ex post individually rational and feasible. In the pure private value case where $|\Theta| = 1$, it follows that for an outcome efficient social choice function $q : T \to C$, the GVCG transfers reduce to the classical VCG transfers. Unfortunately, the GVCG mechanism does not inherit the very attractive dominant strategy property of the pure private values special case. It is tempting to conjecture that the GVCG mechanism satisfies ex post incentive compatibility or perhaps the weaker notion of Bayesian incentive compatibility but even the latter need not hold. There are, however, certain positive results. In MP (2014), it is shown that a modification of the GVCG mechanism is individually rational, approximately ex post incentive compatible, exactly Bayesian incentive compatible when agents are informationally small. To state the main result of MP (2014), we need the notion of an augmented mechanism.

Definition 4. Let $(z_i)_{i \in N}$ be an n-tuple of functions $z_i : T \to \mathbb{R}_+$ each of which assigns to each $t \in T$ a nonnegative number, interpreted as a "reward" to agent i. If $(q, x_1, .., x_n)$ is a mechanism, then the associated *augmented* mechanism is defined as $(q, x_1 + z_1, .., x_n + z_n)$ and will be written simply as $(q, (x_i + z_i))$.

Using precisely the same techniques found in MP (2014), we can prove the following result for one stage direct mechanisms.

Theorem A. *Let $(v_1, .., v_n)$ be a collection of payoff functions.*

(i) Suppose that $q : T \to C$ is outcome efficient for the problem $(v_1, .., v_n, P)$. Suppose that $\mathbb{D} = ((A_i, f_i)_{i\in N}, Q)$ is an information decomposition for $(v_1, .., v_n, P)$ satisfying $\Lambda_i^Q > 0$ for each i. Then there exists an augmented GVCG mechanism $(q, (\alpha_i^q + z_i))$ for the social choice problem $(v_1, .., v_n, P)$ satisfying ex post IR and interim IC.

(ii) For every $\varepsilon > 0$, there exists a $\delta > 0$ such that the following holds: whenever $q : T \to C$ is outcome efficient for the problem $(v_1, .., v_n, P)$ and whenever $\mathbb{D} = ((A_i, f_i)_{i\in N}, Q)$ is an information decomposition for $(v_1, .., v_n, P)$ satisfying

$$\max_i \nu_i^Q \le \delta \min_i \Lambda_i^Q,$$

there exists an augmented GVCG mechanism $(q, (\alpha_i^q + z_i))$ with $0 \le z_i(t) \le \varepsilon$ for every i and t satisfying ex post IR and interim IC.[8]

Part (i) of Theorem A states that, as long as $Q_{A_{-i}}(\cdot|a_i) \ne Q_{A_{-i}}(\cdot|a_i')$ whenever $a_i \ne a_i'$, then irrespective of the agents' informational sizes, the augmenting transfers can be chosen so that the augmented mechanism satisfies Bayesian incentive compatibility. However, the required augmenting transfers will be large if the agents have large informational size. Part (ii) states that the augmenting transfers will be small if the agents have informational size that is small enough relative to our measure of variation in the agents' beliefs.

6. The two stage implementation game

6.1. Preliminaries

Suppose that $(v_1, .., v_n, P)$ is a social choice problem and suppose that $\mathbb{D} = ((A_i, f_i)_{i\in N}, Q)$ is an information decomposition for P.

Throughout this section, we will use the following notational convention:

$$\rho(a_{-i}, a_i) = Q_\Theta(\cdot|a_{-i}, a_i) \text{ and } \rho_\theta(a_{-i}, a_i) = Q_\Theta(\theta|a_{-i}, a_i).$$

Let H_i denote the collection of all functions $u_i : C \to \mathbb{R}_+$. Consequently we identify H_i with $\mathbb{R}_+^C$. For each profile $u = (u_1, .., u_n) \in H := H_1 \times \cdots \times H_n$, let

$$\varphi(u) \in \arg\max_{c\in C} \sum_{i\in N} u_i(c)$$

and define

$$\hat{y}_i(u) = \sum_{j\in N\backslash i} u_j(\varphi(u)) - \max_{c\in C}\left[\sum_{j\in N\backslash i} u_j(c)\right].$$

Therefore, $(\varphi, \hat{y}_1, .., \hat{y}_n)$ defines the classic private values VCG mechanism and it follows that

$$u_i \in \arg\max_{u_i'\in H_i} u_i(\varphi(u_{-i}, u_i')) + \hat{y}_i(u_{-i}, u_i')$$

for all $u_i \in H_i$ and all $u_{-i} \in H_{-i}$.

We wish to formulate our implementation problem with interdependent valuations as a two stage problem in which honest reporting of the agents' signals in stage one resolves the "interdependency" problem so that the stage two problem is a simple implementation problem with private values to which the classic VCG mechanism can be immediately applied. We now define an extensive form game that formalizes the two stage process that lies behind this idea. As in the indirect mechanism for the one stage problem, let $Z = (\zeta_i)_{i\in N}$ be an n-tuple of functions $\zeta_i : A \to \mathbb{R}_+$ each of which assigns to each $a \in A$ a nonnegative number $\zeta_i(a)$ again interpreted as a "reward" to agent i. These rewards are designed to induce agents to honestly report their signals in stage 1.

Given an information decomposition $\mathbb{D}$ and a reward system Z, we define an extensive form game $\Gamma(\mathbb{D}, Z)$ that unfolds in the following way.

Stage 1: Each agent i learns his type $t_i \in T_i$ and makes a (not necessarily honest) report $r_i \in A_i$ of his signal to the mechanism designer. If $(r_1, .., r_n)$ is the profile of stage 1 reports, then agent i receives the nonnegative payment $\zeta_i(r_1, .., r_n)$ and the game moves to stage 2.

Stage 2: If $(r_1, .., r_n) = r \in A$ is the reported type profile in stage 1, the mechanism designer publicly posts the conditional distribution $\rho(r) = Q_\Theta(\cdot|r)$. Agents observe this posted distribution (but not the profile r) and make a second (not necessarily honest) report from H_i to the mechanism designer. If $(u_1, .., u_n) = u \in H$ is the second stage profile of reports,

[8] In MP (2014), it is also shown that the augmented mechanism is approximately ex post incentive compatible in the sense defined in that paper.

then the mechanism designer chooses the social alternative $\varphi(u) \in C$, each agent i receives the transfer $\hat{y}_i(u)$, and the game ends.

We wish to design the rewards ζ_i to accomplish two goals. In stage 1, we want to induce agents to report honestly so that the reported stage 1 profile is exactly $f(t) = (f_1(t_1), .., f_n(t_n))$ when the true type profile is t. In stage 2, upon observing the posted posterior distribution $Q_\Theta(\cdot|f(t))$, we want each agent i to report the payoff function

$$u_i^*(\cdot) = \sum_{\theta \in \Theta} v_i(\cdot, \theta, t_i) Q_\Theta(\theta|f(t)).$$

If these twin goals are accomplished in a perfect Bayesian equilibrium, then the social outcome is

$$\varphi(u^*) \in \arg\max_c \sum_{i \in N} \sum_{\theta \in \Theta} v_i(c, \theta, t_i) Q_\Theta(\theta|f(t)),$$

the transfers are

$$\hat{y}_i(u^*) = \sum_{j \in N \backslash i} v_j(\varphi(u^*), \theta, t_j) Q_\Theta(\theta|f(t)) - \max_{c \in C} \left[\sum_{j \in N \backslash i} v_j(c, \theta, t_j) Q_\Theta(\theta|f(t)) \right],$$

and the ex post payoff to agent i of type t_i is

$$\sum_{i \in N} \sum_{\theta \in \Theta} v_i(\varphi(u^*), \theta, t_j) Q_\Theta(\theta|f(t)) + \hat{y}_i(u^*) + \zeta_i(f(t)).$$

Note that these transfers and payoffs are precisely the GVCG transfers and payoffs defined in Section 5 for the one stage implementation problem.

6.2. *Strategies and equilibria in the two stage game*

Define

$$\Pi := \{Q_\Theta(\cdot|a) : a \in A\}.$$

Given the specification of the extensive form, it follows that the second stage information sets of agent i are indexed by the elements of $A_i \times \Pi \times T_i$. A strategy for agent i in this game is a pair (α_i, β_i) where $\alpha_i : T_i \to A_i$ specifies a type dependent report $\alpha_i(t_i) \in A_i$ in stage 1 and $\beta_i : A_i \times \Pi \times T_i \to H_i$ specifies a second stage report $\beta_i(r_i, \pi, t_i) \in H_i$ as a function of i's first stage report $r_i \in A_i$, the posted distribution π, and $i's$ type $t_i \in T_i$.

We are interested in a Perfect Bayesian Equilibrium (PBE) assessment for the two stage implementation game $\Gamma(\mathbb{D}, Z)$ consisting of a strategy profile $(\alpha_i, \beta_i)_{i \in N}$ and a system of second stage beliefs in which players truthfully report their private information at each stage.

Definition 5. A strategy (α_i, β_i) for player i is *truthful* for i if $\alpha_i(t_i) = f_i(t_i)$ for all $t_i \in T_i$ and

$$\beta_i(f_i(t_i), \pi, t_i)(\cdot) = \sum_{\theta \in \Theta} v_i(\cdot, \theta, t_i)\pi(\theta)$$

for all $\pi \in \Pi$ and $t_i \in T_i$. A strategy profile $(\alpha_i, \beta_i)_{i \in N}$ is *truthful* if (α_i, β_i) is truthful for each player i.

Formally, a **system of beliefs** for player i is a collection of probability measures on $\Theta \times T_{-i}$ indexed by $A_i \times \Pi \times T_i$, i.e., a collection of the form

$$\{\mu_i(\cdot|r_i, \pi, t_i) \in \Delta(\Theta \times T_{-i}) : (r_i, \pi, t_i) \in A_i \times \Pi \times T_i\}.$$

with the following interpretation: when player i of type t_i reports r_i in Stage 1 and observes the posted distribution π, then player i assigns probability mass $\mu_i(\theta, t_{-i}|r_i, \pi, t_i)$ to the event that other players have true types t_{-i} and that the state of nature is θ. As usual, an *assessment* is a pair $\{(\alpha_i, \beta_i)_{i \in N}, (\mu_i)_{i \in N}\}$ consisting of a strategy profile and a system of beliefs for each player.

Definition 6. An assessment $\{(\alpha_i, \beta_i)_{i \in N}, (\mu_i)_{i \in N}\}$ is an *incentive compatible Perfect Bayesian equilibrium (ICPBE) assessment* in the game $\Gamma(\mathbb{D}, Z)$ if $(\alpha, \beta, \mu) = \{(\alpha_i, \beta_i)_{i \in N}, (\mu_i)_{i \in N}\}$ is a Perfect Bayesian Equilibrium assessment and the profile $(\alpha_i, \beta_i)_{i \in N}$ is truthful.

6.3. The main result

Theorem B. *Let $(v_1, .., v_n)$ be a collection of payoff functions.*

(a) Suppose that $q : T \to C$ is outcome efficient for the problem $(v_1, .., v_n, P)$. Suppose that $\mathbb{D} = ((A_i, f_i)_{i\in N}, Q)$ is an information decomposition for P satisfying $\Lambda_i^Q > 0$ for each i. Then there exists a reward system $Z = (\zeta_i)_{i\in N}$ such that the two stage game $\Gamma(\mathbb{D}, Z)$ has an ICPBE (α^*, β^*, μ).

(b) For every $\varepsilon > 0$, there exists a $\delta > 0$ such that the following holds: whenever $q : T \to C$ is outcome efficient for the problem $(v_1, .., v_n, P)$ and $\mathbb{D} = ((A_i, f_i)_{i\in N}, Q)$ is an information decomposition for P satisfying

$$\max_i \nu_i^Q \leq \delta \min_i \Lambda_i^Q ,$$

there exists a reward system $Z = (\zeta_i)_{i\in N}$ such that the two stage game $\Gamma(\mathbb{D}, Z)$ has an ICPBE (α^*, β^*, μ). Furthermore, $0 \leq \zeta_i(a) \leq \varepsilon$ for every i and a.

To prove Theorem B, we proceed in several steps which we outline here. Suppose that $\mathbb{D} = ((A_i, f_i)_{i\in N}, Q)$ is an information decomposition for P.

Step 1: Suppose that (α, β) is a strategy profile with $\alpha_i(t_i) = f_i(t_i)$ for all i and t_i. Suppose that player i is of true type t_i, the other players have true type profile t_{-i}, player i reports r_i in stage 1. Given the definition of $(\alpha_{-i}, \beta_{-i})$, it follows that $\alpha_j(t_j) = f_j(t_j)$ for each $j \neq i$. Therefore, player i of type t_i who has submitted report r_i in stage 1 and who observes $\pi \in \Pi$ at stage 2 will assign positive probability

$$\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}),r_i)=\pi} P_{T_{-i}}(\hat{t}_{-i}|t_i) > 0$$

to the event

$$\{\hat{t}_{-i} \in T_{-i} : \rho(f_{-i}(\hat{t}_{-i}), r_i) = \pi\}.$$

Therefore, i's updated beliefs regarding (θ, t_{-i}) consistent with (α, β) are given by

$$\begin{aligned}\mu_i(\theta, t_{-i}|r_i, \pi, t_i) &= \frac{\rho_\theta(f_{-i}(t_{-i}), f_i(t_i))P_{T_{-i}}(t_{-i}|t_i)}{\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}),r_i)=\pi} P_{T_{-i}}(\hat{t}_{-i}|t_i)} \text{ if } \rho(f_{-i}(t_{-i}), r_i) = \pi \\ &= 0 \text{ otherwise.}\end{aligned}$$

Step 2: Let

$$\overline{w}_i(\cdot, \pi, t_i) := \sum_{\theta\in\Theta} v_i(\cdot, \theta, t_i)\pi(\theta)$$

and for each π and t_{-i} let $w^*_{-i}(\pi, t_{-i}) \in H_{-i}$ be defined as

$$w^*_{-i}(\pi, t_{-i}) := (\overline{w}_j(\cdot, \pi, t_j))_{j\in N\backslash i}.$$

Next, we define the following particular second stage component β_i^* of agent i's strategy as follows: for each $(r_i, \pi, t_i) \in A_i \times \Pi \times T_i$, let

$$\beta_i^*(r_i, \pi, t_i) \in \arg\max_{u_i\in H_i} \sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta} \left[v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \mu_i(\theta, t_{-i}|r_i, \pi, t_i)$$

where $\mu_i(\cdot|r_i, \pi, t_i)$ is defined in Step 1.[9] We then show that

$$\beta_i^*(f_i(t_i), \pi, t_i) = \sum_{\theta\in\Theta} v_i(\cdot, \theta, t_i)\pi(\theta) = \overline{w}_i(\cdot, \pi, t_i).$$

Step 3: If $\alpha_i^*(t_i) = f_i(t_i)$ for all i and t_i and β_i^* is defined as in Step 2, then (α^*, β^*) is a truthful strategy profile. The proof is completed by constructing a system of rewards $Z = (\zeta_i)_{i\in N}$ such that (α^*, β^*, μ) is an ICPBE of the two stage game $\Gamma(\mathbb{D}, Z)$. To accomplish this, we define spherical scoring rule transfers

$$\zeta_i(a_{-i}, a_i) = \varepsilon \frac{Q_{A_{-i}}(a_{-i}|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2}$$

[9] Note that $\beta_i^*(r_i, \pi, t_i)$ exists since each v_i takes on only finitely many values for each i and the set Π is also finite.

as in MP (2014). To prove part (a) of Theorem B, we show that one can find $\varepsilon > 0$ so that (i) deviations at second stage information sets are unprofitable given the beliefs μ defined in step 1 and (ii) coordinated deviations across the two stages are unprofitable. This latter argument depends crucially on the special transfers ζ_i. It is these transfers that induce truthful reporting in stage 1, thus reducing the second stage to a simple implementation problem with private values. To prove part (b), we show that ε can be chosen to be small when each agent's informational size is small enough relative to the variation in his beliefs.

Appendix A

A.1. Preparatory lemmas

Lemma A. *Let M be a nonnegative number and let $\{g_i : C \times \Theta \to \mathbb{R}_+ : i \in N\}$ be a collection of functions satisfying $g_i(\cdot.\cdot) \leq M$ for all i. For each $S \subseteq \{1, .., n\}$ and for each $\pi \in \Delta(\Theta)$, let*

$$F_S(\pi) = \max_{c \in C} \sum_{i \in S} \sum_{\theta \in \Theta} g_i(c, \theta)\pi(\theta).$$

Then for each $\pi, \pi' \in \Delta(\Theta)$,

$$|F_S(\pi) - F_S(\pi')| \leq |S|M||\pi - \pi'||.$$

Proof. See MP (2014). □

Lemma B. *Let M be a nonnegative number and let $\{g_i : C \times \Theta \to \mathbb{R}_+ : i \in N\}$ be a collection of functions satisfying $g_i(\cdot, \cdot) \leq M$ for all i. For each $\pi \in \Delta(\Theta)$, let*

$$\xi(\pi) \in \arg\max_{c \in C} \sum_{i \in N} \sum_{\theta \in \Theta} g_i(c, \theta)\pi(\theta),$$

and

$$\eta_i(\pi) = \sum_{j \in N \backslash i} \sum_{\theta \in \Theta} g_j(\xi(\pi), \theta)\pi(\theta) - \max_{c \in C} \left[\sum_{j \in N \backslash i} \sum_{\theta \in \Theta} g_j(c, \theta)\pi(\theta) \right].$$

Then for each $t \in T$ and all $\pi, \pi' \in \Delta(\Theta)$,

$$\left[\sum_{\theta \in \Theta} g_i(\xi(\pi'), \theta)\pi'(\theta) + \eta_i(\pi') \right] - \left[\sum_{\theta \in \Theta} g_i(\xi(\pi), \theta)\pi(\theta) + \eta_i(\pi) \right] \leq (2n-1)M||\pi - \pi'||$$

Proof.

$$\begin{aligned}
&\left[\sum_{\theta \in \Theta} g_j(\xi(\pi'), \theta)\pi'(\theta) + \eta_i(\pi') \right] - \left[\sum_{\theta \in \Theta} g_j(\xi(\pi), \theta)\pi(\theta) + \eta_i(\pi) \right] \\
&= \sum_{k \in N} \sum_{\theta \in \Theta} g_k(\xi(\pi'), \theta)\pi'(\theta) - \sum_{k \in N} \sum_{\theta \in \Theta} g_k(\xi(\pi), \theta)\pi(\theta) \\
&\quad + \max_{c \in C} \left[\sum_{j \in N \backslash i} \sum_{\theta \in \Theta} g_j(c, \theta)\pi(\theta) \right] - \max_{c \in C} \left[\sum_{j \in N \backslash i} \sum_{\theta \in \Theta} g_j(c, \theta)\pi'(\theta) \right] \\
&= \max_{c \in C} \left[\sum_{k \in N} \sum_{\theta \in \Theta} g_k(c, \theta)\pi'(\theta) \right] - \max_{c \in C} \left[\sum_{k \in N} \sum_{\theta \in \Theta} g_k(c, \theta)\pi(\theta) \right] \\
&\quad + \max_{c \in C} \left[\sum_{j \in N \backslash i} \sum_{\theta \in \Theta} g_j(c, \theta)\pi(\theta) \right] - \max_{c \in C} \left[\sum_{j \in N \backslash i} \sum_{\theta \in \Theta} g_j(c, \theta)\pi'(\theta) \right] \\
&\leq nM||\pi - \pi'|| + (n-1)M||\pi - \pi'||
\end{aligned}$$

where the last inequality follows from Lemma A. □

Lemma C. *Suppose that $Q \in \Delta(\Theta \times A)$ and define a system of rewards $Z = (\zeta_i)_{i \in N}$ where*

$$\zeta_i(a_{-i}, a_i) = \varepsilon \frac{Q_{A_{-i}}(a_{-i}|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2}$$

for each $(a_{-i}, a_i) \in A$. Then for each $a_i, a_i' \in A_i$,

$$\sum_{a_{-i}} \left[\zeta_i(a_{-i}, a_i) - \zeta_i(a_{-i}, a_i')\right] Q_{A_{-i}}(a_{-i}|a_i) \geq \frac{\varepsilon}{2\sqrt{|A|}} \Lambda_i^Q .$$

Proof. Since

$$\left\| \frac{Q_{A_{-i}}(\cdot|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2} - \frac{Q_{A_{-i}}(\cdot|a_i')}{||Q_{A_{-i}}(\cdot|a_i')||_2} \right\|^2 = 2\left[1 - \frac{Q_{A_{-i}}(\cdot|a_i') \cdot Q_{A_{-i}}(\cdot|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2 ||Q_{A_{-i}}(\cdot|a_i')||_2}\right]$$

and $||Q_{A_{-i}}(\cdot|a_i)||_2 \geq \frac{1}{\sqrt{|A-i|}} \geq \frac{1}{\sqrt{|A|}}$, we conclude that

$$\begin{aligned}
\sum_{a_{-i}} \left[\zeta_i(a_{-i}, a_i) - \zeta_i(a_{-i}, a_i')\right] Q(a_{-i}|a_i) &= \sum_{a_{-i}} \left[\varepsilon \frac{Q_{A_{-i}}(a_{-i}|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2} - \varepsilon \frac{Q_{A_{-i}}(a_{-i}|a_i')}{||Q_{A_{-i}}(\cdot|a_i')||_2}\right] Q(a_{-i}|a_i) \\
&= \varepsilon \left[\frac{Q_{A_{-i}}(\cdot|a_i) \cdot Q(\cdot|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2} - \frac{Q_{A_{-i}}(\cdot|a_i') \cdot Q(\cdot|a_i)}{||Q_{A_{-i}}(\cdot|a_i')||_2}\right] \\
&= \varepsilon ||Q_{A_{-i}}(\cdot|a_i)||_2 \left[1 - \frac{Q_{A_{-i}}(\cdot|a_i') \cdot Q_{A_{-i}}(\cdot|a_i)}{||Q(\cdot|a_i)||_2 ||Q(\cdot|a_i')||_2}\right] \\
&= \frac{\varepsilon}{2} ||Q(\cdot|a_i)||_2 \left\| \frac{Q_{A_{-i}}(\cdot|a_i)}{||Q_{A_{-i}}(\cdot|a_i)||_2} - \frac{Q_{A_{-i}}(\cdot|a_i')}{||Q_{A_{-i}}(\cdot|a_i')||_2} \right\|^2 \\
&\geq \frac{\varepsilon}{2\sqrt{|A|}} \Lambda_i^Q . \qquad \square
\end{aligned}$$

A.2. Proof of Theorem B

We will prove part (b) of Theorem B first. To begin, define beliefs $\mu_i(\cdot|r_i, \pi, t_i) \in \Delta(\Theta \times T_{-i})$ for agent i at each information set $(r_i, \pi, t_i) \in A_i \times \Pi \times T_i$ as in Section 5. In addition, define $\overline{w}_i(\cdot, \pi, t_i)$ and $w_{-i}^*(\pi, t_{-i})$ as in Section 5. Let $\alpha_i^*(t_i) = f_i(t_i)$ and recall that β_i^* is defined for agent i as follows: for each $(r_i, \pi, t_i) \in A_i \times \Pi \times T_i$, let

$$\beta_i^*(r_i, \pi, t_i) \in \arg\max_{u_i \in H_i} \sum_{t_{-i} \in T_{-i}} \sum_{\theta \in \Theta} \Big[v_i(\varphi(u_i, w_{-i}^*(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(u_i, w_{-i}^*(\pi, t_{-i}))\Big] \mu_i(\theta, t_{-i}|r_i, \pi, t_i).$$

For notational convenience, we will write $Q_{A_{-i}}(\cdot|a_i)$ as $Q(\cdot|a_i)$ and $P_{T_{-i}}(\cdot|t_i)$ as $P(\cdot|t_i)$ throughout this proof. Choose $\varepsilon > 0$ and define a system of rewards $Z = (\zeta_i)_{i \in N}$ where

$$\zeta_i(a_{-i}, a_i) = \varepsilon \frac{Q(a_{-i}|a_i)}{||Q(\cdot|a_i)||_2}.$$

Since $0 \leq \frac{Q(a_{-i}|a_i)}{||Q(\cdot|a_i)||_2} \leq 1$, for all i, a_{-i} and a_i, it follows that

$$0 \leq \zeta_i(a_{-i}, a_i) \leq \varepsilon.$$

Next suppose that

$$0 < \delta < \frac{\varepsilon}{12nM\sqrt{|A|}}.$$

We will show that (α^*, β^*, μ) is an ICPBE in the game $\Gamma(\mathbb{D}, Z)$ whenever $\max_i \nu_i^Q \leq \delta \min_i \Lambda_i^Q$. To accomplish this, we must show that (α^*, β^*) is truthful, that first stage deviations are unprofitable and that coordinated deviations across stages are unprofitable.

Part 1: To show that (α^*, β^*) is truthful, we must show that

$$\beta_i^*(f_i(t_i), \pi, t_i)(\cdot) = \sum_{\theta \in \Theta} v_i(\cdot, \theta, t_i)\pi(\theta) = \overline{w}_i(\cdot, \pi, t_i)$$

for all $\pi \in \Pi$ and $t_i \in T_i$. i.e., that

$$\overline{w}_i(\cdot, \pi, t_i) \in \arg\max_{u_i \in H_i} \sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta} \left[v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i}), \theta, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \mu_i(\theta, t_{-i} | f_i(t_i), \pi, t_i)$$

for each t_i and each $\pi \in \Pi$. To see this, note that for each $u_i \in H_i$,

$$\begin{aligned}
&\sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta} \Big[v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\Big] \mu_i(\theta, t_{-i} | f_i(t_i), \pi, t_i) \\
&= \sum_{\substack{t_{-i}\in T_{-i} \\ :\rho(f_{-i}(t_{-i}), f_i(t_i))=\pi}} \sum_{\theta\in\Theta} \left[v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \left[\frac{\rho_\theta(f_{-i}(t_{-i}), f_i(t_i)) P(t_{-i}|t_i)}{\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}), f_i(t_i))=\pi} P(\hat{t}_{-i}|t_i)} \right] \\
&= \sum_{\substack{t_{-i}\in T_{-i} \\ :\rho(f(t))=\pi}} \left[\sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) \rho_\theta(f(t)) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \left[\frac{P(t_{-i}|t_i)}{\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}), f_i(t_i))=\pi} P(\hat{t}_{-i}|t_i)} \right] \\
&= \sum_{\substack{t_{-i}\in T_{-i} \\ :\rho(f(t))=\pi}} \left[\sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) \pi(\theta) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \left[\frac{P(t_{-i}|t_i)}{\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}), f_i(t_i))=\pi} P(\hat{t}_{-i}|t_i)} \right] \\
&= \sum_{\substack{t_{-i}\in T_{-i} \\ :\rho(f(t))=\pi}} \left[\overline{w}_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \pi, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \left[\frac{P(t_{-i}|t_i)}{\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}), f_i(t_i))=\pi} P(\hat{t}_{-i}|t_i)} \right] \\
&\le \sum_{\substack{t_{-i}\in T_{-i} \\ :\rho(f(t))=\pi}} \left[\overline{w}_i(\varphi(\overline{w}_i(\cdot, \pi, t_i), w^*_{-i}(\pi, t_{-i})), \pi, t_i) + \hat{y}_i(\overline{w}_i(\cdot, \pi, t_i), w^*_{-i}(\pi, t_{-i}))\right] \\
&\quad \times \left[\frac{P(t_{-i}|t_i)}{\sum_{\hat{t}_{-i}:\rho(f_{-i}(\hat{t}_{-i}), f_i(t_i))=\pi} P(\hat{t}_{-i}|t_i)} \right] \\
&= \sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta} \Big[v_i(\varphi(\overline{w}_i(\cdot, \pi, t_i), w^*_{-i}(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(\overline{w}_i(\cdot, \pi, t_i), w^*_{-i}(\pi, t_{-i}))\Big] \mu_i(\theta, t_{-i} | r_i, \pi, t_i).
\end{aligned}$$

Therefore, (α^*, β^*) is truthful.

Part 2: To show that deviations at second stage information sets are unprofitable, suppose that all players use α^*_i in stage 1 and players $j \neq i$ use β^*_j in stage 2. Then, upon observing π and having reported truthfully in stage 1, it follows from the definition of β^*_j that each player $j \neq i$ reports $\beta^*_j(f_j(t_j), \pi, t_j) = \overline{w}_j(\cdot, \pi, t_j)$ in stage 2. Therefore, the second stage expected payoff to player i who reports $u_i \in H_i$ given the beliefs μ_i defined above is

$$\sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta} \left[v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \mu_i(\theta, t_{-i} | f_i(t_i), \pi, t_i)$$

so the definition of β^*_i implies that

$$\overline{w}_i(\cdot, \pi, t_i) \in \arg\max_{u_i \in H_i} \sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta} \left[v_i(\varphi(u_i, w^*_{-i}(\pi, t_{-i})), \theta, t_i) + \hat{y}_i(u_i, w^*_{-i}(\pi, t_{-i}))\right] \mu_i(\theta, t_{-i} | f_i(t_i), \pi, t_i).$$

Part 3: To show that coordinated deviations across stages are unprofitable for player i, we assume that other players use $(\alpha^*_{-i}, \beta^*_{-i})$ and we must show that, for all $t_i \in T_i$ and all $r_i \in A_i$, we have

$$\begin{aligned}
&\sum_{t_{-i}\in T_{-i}} \left[\sum_{\theta\in\Theta} v_i(\varphi(\overline{w}_i(\cdot, \rho(f(t)), t_i)), w^*_{-i}(\rho(f(t), t_{-i})), \theta, t_i) \rho_\theta(f(t)) + \hat{y}_i(\overline{w}_i(\cdot, \rho(f(t)), t_i)), w^*_{-i}(\rho(f(t)), t_{-i})) \right] \\
&\quad \times P(t_{-i}|t_i) + \sum_{t_{-i}\in T_{-i}} \zeta_i(f(t)) P(t_{-i}|t_i) \\
&\ge \sum_{t_{-i}\in T_{-i}} \max_{u_i\in H_i} \left[\sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i) \rho_\theta(f(t)) + \hat{y}_i(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})) \right]
\end{aligned}$$

$$\times P(t_{-i}|t_i) + \sum_{t_{-i}\in T_{-i}} \zeta_i(f_{-i}(t_{-i}), r_i)P(t_{-i}|t_i).$$

Since

$$\begin{aligned}
&\sum_{\theta\in\Theta} v_i(\varphi(\overline{w}_i(\cdot, \rho(f_{-i}(t_{-i}), r_i), t_i), w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f_{-i}(t_{-i}), r_i) \\
&\qquad + \hat{y}_i(\overline{w}_i(\cdot, \rho(f_{-i}(t_{-i}), r_i), t_i), w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})) \\
&\quad \geq \sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f_{-i}(t_{-i}), r_i) \\
&\qquad + \hat{y}_i(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i}))
\end{aligned}$$

for all $u_i \in H_i$, it follows from Lemma C that for each t_{-i} and each u_i, r_i and t_i,

$$\begin{aligned}
&\sum_{\theta\in\Theta} v_i(\varphi(\overline{w}_i(\cdot, \rho(f(t)), t_i)), w^*_{-i}(\rho(f(t)), t_{-i})), \theta, t_i)\rho_\theta(f(t)) + \hat{y}_i(\overline{w}_i(\cdot, \rho(f(t)), t_i)), w^*_{-i}(\rho(f(t)), t_{-i})) \\
&\qquad - \left[\sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f(t)) + \hat{y}_i(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i}))\right] \\
&\quad = \sum_{\theta\in\Theta} v_i(\varphi(\overline{w}_i(\cdot, \rho(f(t)), t_i)), w^*_{-i}(\rho(f(t)), t_{-i})), \theta, t_i)\rho_\theta(f(t)) + \hat{y}_i(\overline{w}_i(\cdot, \rho(f(t)), t_i)), w^*_{-i}(\rho(f(t), t_{-i}))) \\
&\qquad - \left[\sum_{\theta\in\Theta} v_i(\varphi(\overline{w}_i(\cdot, \rho(f_{-i}(t_{-i}), r_i), t_i), w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f_{-i}(t_{-i}), r_i)\right. \\
&\qquad \left. + \hat{y}_i(\overline{w}_i(\cdot, \rho(f_{-i}(t_{-i}), r_i), t_i), w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i}))\right] \\
&\qquad + \left[\sum_{\theta\in\Theta} v_i(\varphi(\overline{w}_i(\cdot, \rho(f_{-i}(t_{-i}), r_i), t_i), w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f_{-i}(t_{-i}), r_i)\right. \\
&\qquad \left. + \hat{y}_i(\overline{w}_i(\cdot, \rho(f_{-i}(t_{-i}), r_i), t_i), w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i}))\right] \\
&\qquad - \left[\sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f_{-i}(t_{-i}), r_i) + \hat{y}_i(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i}))\right] \\
&\qquad + \sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)[\rho_\theta(f_{-i}(t_{-i}), r_i) - \rho_\theta(f(t))] \\
&\quad \geq -(2n-1)M||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i)|| - M||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i)||
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{t_{-i}\in T_{-i}} \sum_{\theta\in\Theta}\left[v_i(\varphi(\overline{w}_i(\cdot, \rho(f(t), t_i)), w^*_{-i}(\rho(f(t), t_{-i})), \theta, t_i)\rho_\theta(f(t)) + \hat{y}_i(\overline{w}_i(\cdot, \rho(f(t), t_i)), w^*_{-i}(\rho(f(t), t_{-i})))\right] \\
&\qquad \times P(t_{-i}|t_i) + \sum_{t_{-i}\in T_{-i}} \zeta_i(f(t))P(t_{-i}|t_i) \\
&\quad \geq \sum_{t_{-i}\in T_{-i}} \max_{u_i\in H_i}\left[\sum_{\theta\in\Theta} v_i(\varphi(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i})), \theta, t_i)\rho_\theta(f(t)) + \hat{y}_i(u_i, w^*_{-i}(\rho(f_{-i}(t_{-i}), r_i), t_{-i}))\right] \\
&\qquad \times P(t_{-i}|t_i) + \sum_{t_{-i}\in T_{-i}} \zeta_i(f_{-i}(t_{-i}), r_i)P(t_{-i}|t_i) \\
&\qquad + \sum_{t_{-i}\in T_{-i}} [\zeta_i(f(t)) - \zeta_i(f_{-i}(t_{-i}), r_i)]\,P(t_{-i}|t_i) - 2nM \sum_{t_{-i}\in T_{-i}} ||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i))||P(t_{-i}|t_i)
\end{aligned}$$

To complete the proof, we must show that

$$\sum_{a_{-i}}[\zeta_i(a_{-i}, f_i(t_i)) - \zeta_i(a_{-i}, r_i)]\, Q\,(a_{-i}|f_i(t_i)) \geq 2nM \sum_{a_{-i}} ||\rho(a_{-i}, f_i(t_i)) - \rho(a_{-i}, r_i))||\, Q\,(a_{-i}|f_i(t_i)).$$

From the definition of ν_i^Q, it follows that

$$\begin{aligned}
&\sum_{a_{-i}} ||\rho(a_{-i}, a_i) - \rho(a_{-i}, r_i)||\, Q\,(a_{-i}|a_i) \\
&= \sum_{\substack{a_{-i} \\ :||\rho(a_{-i},a_i)-\rho(a_{-i},r_i)||>\nu_i^Q}} ||\rho(a_{-i}, a_i) - \rho(a_{-i}, r_i)||\, Q\,(a_{-i}|a_i) \\
&\quad + \sum_{\substack{a_{-i} \\ :||\rho(a_{-i},a_i)-\rho(a_{-i},r_i)||\leq\nu_i^Q}} ||\rho(a_{-i}, a_i) - \rho(a_{-i}, r_i)||\, Q\,(a_{-i}|a_i) \\
&\leq 2\nu_i^Q + \nu_i^Q \\
&= 3\nu_i^Q.
\end{aligned}$$

Consequently, we can apply Lemma C and the definition of information decomposition to conclude that

$$\begin{aligned}
&\sum_{t_{-i}\in T_{-i}} [\zeta_i(f(t)) - \zeta_i(f_{-i}(t_{-i}), r_i)]\, P(t_{-i}|t_i) - 2nM \sum_{t_{-i}\in T_{-i}} ||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i)||P(t_{-i}|t_i) \\
&= \sum_{a_{-i}} \sum_{\substack{t_{-i}\in T_{-i} \\ :f_{-i}(t_{-i})=a_{-i}}} [\zeta_i(f(t)) - \zeta_i(f_{-i}(t_{-i}), r_i)]\, P(t_{-i}|t_i) \\
&\quad - 2nM \sum_{a_{-i}} \sum_{\substack{t_{-i}\in T_{-i} \\ :f_{-i}(t_{-i})=a_{-i}}} ||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i))||P(t_{-i}|t_i) \\
&= \sum_{a_{-i}} [\zeta_i(a_{-i}, f_i(t_i)) - \zeta_i(a_{-i}, r_i)] \left[\sum_{\substack{t_{-i}\in T_{-i} \\ :f_{-i}(t_{-i})=a_{-i}}} P(t_{-i}|t_i) \right] \\
&\quad - 2nM \sum_{a_{-i}} ||\rho(a_{-i}, f_i(t_i)) - \rho(a_{-i}, r_i)|| \left[\sum_{\substack{t_{-i}\in T_{-i} \\ :f_{-i}(t_{-i})=a_{-i}}} P(t_{-i}|t_i) \right] \\
&= \sum_{a_{-i}} [\zeta_i(a_{-i}, f_i(t_i)) - \zeta_i(a_{-i}, r_i)]\, Q\,(a_{-i}|f_i(t_i)) - 2nM \sum_{a_{-i}} ||\rho(a_{-i}, f_i(t_i)) - \rho(a_{-i}, r_i))||\, Q\,(a_{-i}|f_i(t_i)) \\
&\geq \frac{\varepsilon}{2\sqrt{|A|}} \Lambda_i^Q - (2nM)(3\nu_i^Q) \\
&\geq 0.
\end{aligned}$$

To prove part (a) of the theorem, note that part 3 above shows that

$$\begin{aligned}
&\sum_{t_{-i}\in T_{-i}} [\zeta_i(f(t)) - \zeta_i(f_{-i}(t_{-i}), r_i)]\, P(t_{-i}|t_i) - 2nM \sum_{t_{-i}\in T_{-i}} ||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i)||P(t_{-i}|t_i) \\
&\geq \frac{\varepsilon}{2\sqrt{|A|}} \Lambda_i^Q - 2nM \sum_{t_{-i}\in T_{-i}} ||\rho(f(t)) - \rho(f_{-i}(t_{-i}), r_i)||P(t_{-i}|t_i).
\end{aligned}$$

If $\Lambda_i^Q > 0$, then choosing $\varepsilon > 0$ sufficiently large proves the result.

References

Heffetz, O., Ligett, K., 2014. Privacy and data-based research. J. Econ. Perspect. 28, 75–98.

Hurwicz, L., 1972. On informationally decentralized systems. In: McGuire, C.B., Radner, R. (Eds.), Decision and Organization. Amsterdam. Ch. 14.

Jackson, M., 1991. Bayesian implementation. Econometrica 59, 461–477.

Ledyard, J.O., Palfrey, T.R., 1994. Voting and lottery drafts as efficient public goods mechanisms. Rev. Econ. Stud. 61, 327–355.

Ledyard, J.O., Palfrey, T.R., 2002. The approximation of efficient public good mechanisms by simple voting schemes. Econometrica 67, 435–448.

McLean, R., Postlewaite, A., 2002. Informational size and incentive compatibility. Econometrica 70, 2421–2454.
McLean, R., Postlewaite, A., 2004. Informational size and efficient auctions. Rev. Econ. Stud. 71, 809–827.
McLean, R., Postlewaite, A., 2014. Informational size and two stage mechanisms. Mimeo.
McLean, R., Postlewaite, A., forthcoming. Implementation with interdependent valuations. Theor. Econ.
Mount, K., Reiter, S., 1974. The informational size of message spaces. J. Econ. Theory 8, 161–192.
Palfrey, T., Srivastava, S., 1989. Implementation with incomplete information in exchange economies. Econometrica 57, 115–134.
Postlewaite, A., Schmeidler, D., 1986. Implementation in differential information economies. J. Econ. Theory 39, 14–33.
Postlewaite, A., Wettstein, D., 1989. Implementing constrained walrasian equilibria continuously. Rev. Econ. Stud. 1989 (56), 603–612.