Coalition-Proof Full Efficient Implementation

Mikhail Safronov[*]

University of Cambridge

July 12, 2018

Abstract

The Vickrey–Clarke–Groves and d'Aspremont–Gerard-Varet mechanisms implement efficient social choice by compensating each agent for the externalities that his report imposes on all other agents. Instead of aggregate compensations, which may lead to profitable coalitional deviations, this paper provides an alternative mechanism, in which each pair of agents directly compensate each other for the pairwise externalities they impose. Under the assumption of independent private values, any agent is guaranteed to receive his ex ante efficient payoff by reporting truthfully, regardless of others' strategies. This absence of ex ante externalities makes the mechanism coalition-proof, and makes all equilibria efficient.

1   Introduction

The problem of externalities which cause economic inefficiency can be solved if there exists a procedure for internalizing the externalities. This paper develops such a procedure in a benevolent social planner's problem in which agents have independent private values and quasilinear preferences. The social planner (she) asks each agent (he) to report his preferences, and then she implements the social outcome which maximizes the total payoff of the agents. Since any agent's report affects the social outcome, the agents impose externalities on each other and may benefit from misreporting their types (preferences). In order to induce truthful reports, the agents should be required to compensate each other for these externalities.

The idea of internalizing the externalities has been used in the classic Vickrey–Clarke–Groves (VCG) and d'Aspremont–Gerard-Varet (AGV) mechanisms, though in these mechanisms agents do not directly compensate each other. In the VCG mechanism, it is the social planner who compensates the agents for the externalities. In the AGV mechanism, the compensation is unfair: if agent i's report imposes externalities on agent j and no externalities on agent k, agent k still has to partially compensate agent i for the former externalities. As a result, both of these mechanisms internalize the aggregate—not the pairwise—externalities and are not resistant to group deviation. In these mechanisms, each agent individually prefers to report truthfully, but a group of agents can coordinate on a misreport and jointly benefit.

The current paper presents an alternative mechanism, which improves upon the VCG and AGV mechanisms by being resistant to coalitional deviations. The mechanism is built assuming *independent private values* - the environment of the AGV mechanism. There are two equivalent versions of the mechanism: the direct mechanism and the sequential mechanism. This paper mainly focuses on the direct mechanism. Agents simultaneously report their types. Then the social planner orders the agents in an

arbitrary sequence, and she treats the agents' reports as if they were arriving one by one, according to that sequence. When the report of agent i "arrives", the social planner updates her beliefs over the efficient social outcome she will choose at the end, and she updates the expected payoffs of the agents from that outcome. The mechanism prescribes any other agent $j \neq i$ to pay agent i the change in j's expected payoff which occurs as a result of i's report. These payments are made for the report of each agent, that is, each pair of agents i, j compensate each other for the *pairwise* externalities of their reports. The sequential (version of the) mechanism is equivalent to the direct version, except that the agents report their types sequentially and publicly.

In the new (direct) mechanism, each pair of agents directly compensate each other for the pairwise marginal externalities caused by their reports. As a result, all externalities are removed at the ex ante level. If any agent i, before learning his type, commits to reporting truthfully, he is guaranteed to get his ex ante efficient payoff, regardless of others' strategies. This result follows from the way the payments are made. First, agent i receives a payment from every other agent j, equal to the change in j's expected payoff caused by i's report. Since agent i reports truthfully, in expectation over i's report that change is zero, and so is j's payment to i. Second, agent i makes a payment to j, equal to the change in i's expected payoff caused by j's report. Effectively, the utility of agent i (his payoff from social choice plus payments received in the mechanism) does not change with j's report. Therefore, i's utility does not change with reports of other agents and is equal to its ex ante value, that is, to i's ex ante efficient payoff.

The idea behind this mechanism is similar to that of property rights in the Coase theorem. Before the mechanism is announced, the social planner expects each agent i to obtain his ex ante efficient payoff. She guarantees that agent i will receive that payoff if he reports truthfully: when reports of other agents change i's expected payoff,

3

he is compensated for these changes. The ex ante utility of agent i does not depend on others' strategies. This guarantee makes the mechanism attractive to any risk- or ambiguity-averse agent, or to any agent who is struggling to predict others' strategies.

The property of no ex ante externalities on any individual agent guarantees the social outcome to be efficient in any Bayesian Nash equilibrium (i.e., full efficient implementation). Since truthful reporting is always an option, in any equilibrium the ex ante utility of each agent is at least as great as his ex ante efficient payoff. The total utility of all the agents is at least as great as the total ex ante efficient payoff. Since the mechanism is ex post budget-balanced, the total payoff of all the agents is ex ante efficient, and so is the social outcome.

The property of no ex ante externalities on any individual agent also ensures that the mechanism is coalition-proof: it is not profitable for any coalition to misreport. Since agents outside the coalition report truthfully, each of those agents is guaranteed to receive his ex ante utility. Thus, the coalition is the residual claimant of the total payoff, which is maximized at truthful reporting.

The social planner treats the agents' reports sequentially, and the agents know that sequence in advance. The payment which agent i receives from reporting his type, is equivalent to the expected payoff of all the other agents; that payoff is estimated conditional on the reports of agents located before i in the sequence and assuming that agents after i report truthfully. The incentives to report truthfully thus lie between those of the VCG and AGV mechanisms. In some environments, the solution concept for the truthful equilibrium in the current mechanism similarly lies between the weak dominance of VCG and the Bayesian Nash equilibrium of AGV. Assuming that any misreport causes inefficiency in the social choice, truthful reporting becomes a uniquely interim correlated rationalizable strategy (as in Dekel, Fudenberg and Morris (2007), and Battigalli et al. (2011)). The last agent strictly prefers to report truthfully regardless of his beliefs about others' reports. Knowing that, the next-to-

4

last agent strictly prefers to report truthfully as well. By induction, all the agents have truthful reporting as their uniquely rationalizable strategy.

The mechanism has other features. The planner treats agents' reports in a certain order, that order determines the monetary transfers to each agent from the mechanism; however, the interim utility of each agent does not depend on the order. The mechanism can be made symmetric by uniformly choosing an order in which the agents' reports are revealed. Indeed, since the mechanism works for any arbitrary deterministic ordering of the agents, it works for random ordering as well. In the resulting symmetric mechanism, each agent pays the externalities that other agents impose on him, and gets paid the Shapley value of the externalities that his report imposes on others. In addition, under certain assumptions, the mechanism can be adjusted to satisfy interim participation constraints.

The paper is organized as follows. Section 2 discusses the relevant literature. Section 3 builds the direct mechanism and shows that truthful reporting is incentive compatible and is the uniquely rationalizable strategy. Section 4 shows the main properties of the mechanism: ex ante removal of externalities, coalition proofness and full implementation. Section 5 discusses the issues related to agents' ordering in the mechanism. Section 6 discusses the sequential version of the mechanism, and points out limitations of the mechanism. In the Appendix the problem of interim participation constraints is covered.

2   Literature review

The idea of internalizing the externalities in an efficient mechanism has given rise to the classic Vickrey–Clarke–Groves (VCG) and d'Aspremont–Gerard-Varet (AGV) mechanisms. In the VCG mechanism, which was introduced by Vickrey (1961), Clarke (1971), and Groves (1973), each agent is paid the externalities that his re-

port imposes on other agents. As a result, truthful reporting is a weakly dominant strategy. The AGV mechanism from the paper by d'Aspremont and Gerard-Varet (1979) uses a similar approach: each agent is paid the expected externalities that his report imposes on other agents. The payment is made budget-balanced by taking it from the other agents with equal shares. As a result, the AGV mechanism is ex post budget-balanced, though the solution concept is weaker: truthful reporting is Bayesian incentive-compatible, rather than weakly dominant.

Crémer and Riordan (1985) design a mechanism in which agents report their types sequentially. The first agent reports his type publicly, and all other agents can condition their reports on his report. Crémer and Riordan show the existence of budget-balanced monetary transfers, which make truthful reporting a weakly dominant strategy for all agents except the first one, and a Bayesian incentive-compatible strategy for the first agent. However, the mechanism by Crémer and Riordan is not coalition-proof. Moulin (1999) designs a sequential mechanism for sharing the production cost of a certain commodity among several agents. In the mechanism the agents sequentially report their preferences for the commodity, and then each agent is asked to pay the incremental cost of production, which has occurred due to his report. The mechanism by Moulin is coalition-proof, although it may not be efficient. In comparison, the mechanism described in the current paper achieves all the properties of coalition-proof full efficient implementation, and it works regardless of agents reporting their types sequentially or simultaneously.

The mechanisms in Samuelson (1985) and Cramton, Gibbons, and Klemperer (1987) perform similarly to the Coase theorem. These works consider environments where the agents have property rights to an asset and trade these rights through efficient mechanisms. The fact that each agent owns a share of the asset imposes participation constraints and makes it impossible to always reach the efficient allocation of property rights. The authors find the conditions on the initial shares under which efficiency is

6

achieved. In contrast, my paper builds a mechanism in which each agent is guaranteed to get his ex ante efficient payoff from the social outcome, which is similar to owning an initial share of an asset. When reporting their types, agents change their efficient payoffs and compensate each other for those changes. Since there is no participation constraint, the mechanism always achieves efficiency.

Another series of papers studies the problem of collusion in mechanism design. Laffont and Martimort (1997, 1998, 2000) consider the environment with two agents and show the optimal outcome to be collusion-proof in the case of independent types. The paper by Che and Kim (2006) extends the model to an arbitrary number of agents and a more general environment with object allocation. Che and Kim show that any incentive-compatible, individually rational mechanism can be adjusted to be collusion-proof in the case where the grand coalition is formed. With an additional requirement of ex post incentive compatibility, the same result holds if a subgroup of agents can form a coalition and the principal knows at least two agents in the subgroup. In another paper on auctions, Che and Kim (2009) show that with passive beliefs and the assumption of impossibility of forming the grand coalition, the seller can achieve the same revenue as in the case of no collusion. In comparison, I consider the problem of achieving efficiency, rather than profit maximization, and do not impose participation constraints. Another difference is that I construct a mechanism where agents directly compensate each other for the pairwise, rather than aggregate, externalities. This mechanism is resistant to *any* coalition, despite the entire coalition behaving as a single player. It is essential for the mechanism in this paper (as well as in the papers by Laffont and Martimort, and Che and Kim), that the social planner has correct beliefs about the agents' type distribution.

The problem of different aspects of the mechanism with collusion has been studied more extensively in auctions. McAfee and McMillan (1992) show that the inability of the cartel members to pay each other reduces their payoffs. Che, Condorelli and

Kim (2013) show that in this case the seller is not hurt by the possibility of collusion. Erdil and Klemperer (2011) propose a new class of payment rules to make the agents less willing to submit non-truthful bids if they are colluding. Biran and Forges (2011) consider the stability of a collusion in auctions with respect to externalities that the each bidder who gets the object may impose on others. Chen and Micali (2012) allow the agents to report not only their value but also the coalition to which they belong. If several agents consistently report being in the same coalition, and one of them wins the good and has to pay, the bids of other coalition members do not increase the payment; this feature induces the agents to reveal that they belong to a coalition.

An independent branch of literature is devoted to *full implementation*: it considers mechanism design in which all equilibria achieve the desired social outcome. In the environment with observable types, the Maskin monotonicity condition (described in Maskin (1998)) is necessary and essentially sufficient for full implementation. This condition is extended to environments with incomplete information in Postlewaite and Schmeidler (1986); and then extended in environments with agents having exclusive information to the Bayesian monotonicity condition in Palfrey and Srivastava (1989). The condition of Bayesian monotonicity is generalized to environments with externalities in Jackson (1991). The idea is that for any undesirable outcome, there is an agent who can credibly inform the designer if this outcome is being played and get rewarded. However, a non-direct mechanism is needed for this communication to be possible. Matsushima (1993) shows that with quasilinear preferences and side payments the Bayesian monotonicity can be replaced with much weaker condition, which is satisfied for a generic class of social outcomes. This result is further developed by Chen, Kunimoto and Sun (2015) where only small transfers are needed for full implementation. A recent paper by Ollár and Penta (2017) shows that full implementation is achieved by using transfer schemes which only elicit payoff-relevant information. In their paper, the mechanism designer uses moment conditions, commonly known to both the designer and the agents, and makes truthful reporting the

uniquely rationalizable strategy. The results by Ollár and Penta hold under both private and interdependent values, in addition, their result on unique rationalizable implementation still holds even if one weakens the restrictions of the common prior assumption, or the designer's knowledge of the distribution of types, or of agents' beliefs.

3   Mechanism

I consider a setup with n agents, denoted as $i \in \{1, ..., n\}$, with $-i$ standing for the set of all agents other than i. Each agent i has a privately observed type $\theta_i \in \Theta_i$, with overall type profile denoted by θ. Each set Θ_i lies in a Euclidean space $\mathbb{R}^{n_i}$ with finite n_i; each space $\mathbb{R}^{n_i}$ is endowed with a Borel sigma-algebra, and each set Θ_i is a measurable compact set in $\mathbb{R}^{n_i}$. Types θ_i are independently distributed across the agents, and the ex ante distribution of types is publicly known. A set of social outcomes S is a Borel-measurable compact set in a Euclidean space $\mathbb{R}^S$. Each social outcome $s \in S$ gives agent i a payoff of $u_i(\theta_i, s)$, the function $u_i(\theta_i, s)$ is continuous in both variables. That is, the setup is characterized by *independent private values*. There is an *efficient* social choice function $s^*(\theta)$ that maximizes the sum of agents' payoffs $\sum_i u_i(\theta_i, s)$ given θ, and $s^*(\theta)$ is assumed to be measurable. The payoff of agent i at the efficient outcome $s^*(\theta)$ is denoted by $u_i(\theta)$.

I allow for monetary transfers and assume that the agents have quasilinear preferences: if agent i receives monetary transfer of size x_i, his total utility is equal to $u_i(\theta_i, s) + x_i$. Later in the paper, I will refer to the "payoff" as the payoff $u_i(\theta_i, s)$ from the social outcome, and to the "utility" as the payoff plus monetary transfers. A social planner commits to an efficient mechanism: each agent i reports his type $\hat{\theta}_i$, and then the social planner chooses $s^*(\hat{\theta})$ as a function of the total report profile $\hat{\theta}$. The goal of the social planner is to find transfers $x_i(\hat{\theta})$, to induce agents to report truthfully. The

VCG and AGV mechanisms achieve this goal by making all agents report their types simultaneously, and compensating each agent for the aggregate externality that his report imposes on others.

DEFINITION 1 *In the VCG mechanism, the monetary transfer to each agent i is*

$$x_i^{VCG}(\hat{\theta}) = \sum_{k \neq i} u_k(\hat{\theta}) - \max_{s \in S} \sum_{k \neq i} u_k(\hat{\theta}_k, s)$$

That is, agent i receives a (non-positive) monetary transfer equal to the externality that i imposes on all other agents.

DEFINITION 2 *In the AGV mechanism, the net monetary transfer to agent i is:*

$$x_i^{AGV}(\hat{\theta}) = E_{\theta_{-i}} \sum_{k \neq i} u_k(\hat{\theta}_i, \theta_{-i}) - \sum_{j \neq i} \frac{1}{n-1} E_{\theta_{-j}} \sum_{k \neq j} u_k(\hat{\theta}_j, \theta_{-j})$$

Each agent i gets paid the expected externality his report imposes on other agents. The AGV mechanism is ex post budget-balanced.

The VCG and AGV mechanisms are incentive compatible, however, they are both susceptible to coalitional deviations (Section 3.1). I will now introduce new transfers, which will achieve coalition-proofness. First, the agents are *arbitrarily* ordered into a sequence $1, 2, 3, ..., n$, which is publicly known before the agents start reporting their types. Second, the agents *simultaneously* submit their reports.

DEFINITION 3 *Agent i's strategy is a Borel-measurable function σ_i: $\Theta_i \longrightarrow \Delta(\Theta_i)$ that determines his report $\hat{\theta}_i$ as dependent on his true type θ_i. Agent i reports truthfully, if $\sigma_i(\theta_i) \equiv \theta_i$, $\forall \theta_i$.*

For any agent j, one can estimate j's expected payoff, given the reports of agents $1, ..., i-1$, and taking expectation over the reports of agents $i, ..., n$:

10

$E_{\theta_i,...,\theta_n}[u_j(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, ..., \theta_n)]$. Respectively, the mechanism prescribes every other agent $j \neq i$ to pay i the marginal change in j's expected payoff which is caused by i's report:[1]

DEFINITION 4 *Given the total submitted report to be $\hat{\theta}$, each agent $j \neq i$ pays agent i the change in expectation of j's payoff, caused by report $\hat{\theta}_i$:*

$$x_{ij}(\hat{\theta}) \equiv E_{\theta_{i+1},...,\theta_n}[u_j(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)] - E_{\theta_i,\theta_{i+1},...,\theta_n}[u_j(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, \theta_{i+1}, ..., \theta_n)]$$

The net monetary transfer to agent i is

$$x_i(\hat{\theta}) = \sum_{j \neq i} \left(x_{ij}(\hat{\theta}) - x_{ji}(\hat{\theta}) \right)$$

Since the payment $x_{ij}(\hat{\theta})$ depends only on reports of agents $1, ..., i$, I will sometimes use either the notation $x_{ij}(\hat{\theta}) = x_{ij}(\hat{\theta}_1, ..., \hat{\theta}_i)$, or simply x_{ij} when there would be no confusion. This payment x_{ij} can be negative (i.e., agent j receives a positive transfer from agent i) if i's report causes negative changes in the expectation of payoff u_j. Such pairwise transfers x_{ij} are made for the report of each agent i, and from each agent $j \neq i$. Thus, any two agents exchange monetary transfers between themselves according to the pairwise marginal externalities they impose on each other.

LEMMA 1 *The mechanism is ex-post budget balanced: for any $\hat{\theta}$, $\sum_i x_i(\hat{\theta}) = 0$.*

Proof.

$$\sum_i x_i(\hat{\theta}) = \sum_i \sum_{j \neq i} \left(x_{ij}(\hat{\theta}) - x_{ji}(\hat{\theta}) \right) = \sum_{i,j,i \neq j} \left(x_{ij}(\hat{\theta}) - x_{ji}(\hat{\theta}) - x_{ij}(\hat{\theta}) + x_{ji}(\hat{\theta}) \right) = 0$$

---

[1]These transfers can be made either immediately after agent i's report, or at the end, after all the agents have submitted their reports.

3.1 Example

In this section I demonstrate how the new mechanism works and that it achieves coalition-proofness, unlike the VCG and AGV mechanisms. Assume there are three agents $\{1, 2, 3\}$ that live in a city, with agents $1, 3$ living close to each other. The social planner can build a new park for either agents $1, 3$, or for agent 2. She can build only one park, and at a zero cost. Each agent i has one of two types L or H, with $Prob(\theta_i = H) = Prob(\theta_i = L) = \frac{1}{2}$, for all i. The payoff from having the park for each agent equals 8 if the agent's type is L, and 20 if the agent's type is H; the payoff from not having a park is zero.

The efficient social choice is to build the park for agent 2 if agents' type profile is $(\theta_1, \theta_2, \theta_3) = (L, H, L)$, and to build the park for agents $1, 3$ otherwise. The vector of agents' efficient payoffs (u_1, u_2, u_3) is represented in Table 1 below as dependent on type profile θ:

$(\theta_1, \theta_3) \backslash \theta_2$	L	H
L,L	8,0,8	0,20,0
L,H	8,0,20	8,0,20
H,L	20,0,8	20,0,8
H,H	20,0,20	20,0,20

Table 1. Efficient payoffs.

In this example neither the VCG nor AGV mechanism is resistant to group deviation: agents $1, 3$ can form a coalition and their total utility increases if they misreport their types - that is, it is possible for them to pay each other so that each benefits from misreporting. In the VCG mechanism, if agents $1, 3$ have type L each, they can both report H: their chance of having the park increases from $1/2$ to 1, and none of them will have to pay any monetary transfers since none of them is pivotal.[2] Similarly,

---

[2]In fact, such reports constitute an inefficient equilibrium: agents $1, 3$ report H, while agent 2

in the AGV mechanism (Definition 2), if agents $1, 3$ have types L each, and agent 2 reports truthfully, if agents $1, 3$ submit reports $\hat{\theta}_1, \hat{\theta}_3 = L, H$, rather than truthful reporting, their expected (over 2's report) total utility increases from -1 to 5.5.

Now let's find transfers in the new mechanism. Let the planner order the agents in the following sequence: $\{1, 2, 3\}$, and consider all agents to report low types. Before the reports, the social planner estimates the agents' ex ante efficient payoffs by taking the average across all eight cells from Table 1: $E_\theta[u_1, u_2, u_3](\theta) = (13, 2.5, 13)$. After agent 1 reports type $\hat{\theta}_1 = L$, the updated expected payoffs of agents are average across the first two rows of Table 1: $E_{\theta_{-1}}[u_1, u_2, u_3](\theta_1 = L, \theta_{-1}) = (6, 5, 12)$. By Definition 4, the pairwise transfer to agent 1 from each of agents $j = 2, 3$ is the change in j's expected payoff: $x_{12}(\hat{\theta}_1 = L) = 5 - (2.5) = 2.5$, $x_{13}(\hat{\theta}_1 = L) = 12 - 13 = -1$. After agent 2 reports $\hat{\theta}_2 = L$, the updated expected payoffs of agents will be: $E_{\theta_3}[u_1, u_2, u_3](\theta_1 = L, \theta_2 = L, \theta_3) = (8, 0, 14)$. The pairwise transfers to agent 2 are the marginal changes in expected payoffs of agents $1, 3$: $x_{21}(\hat{\theta}_1 = L, \hat{\theta}_2 = L) = 8 - 6 = 2$, $x_{23}(\hat{\theta}_1 = L, \hat{\theta}_2 = L) = 14 - 12 = 2$. Afterwards, agent 3's report does not affect the expected payoffs of other agents, and $x_{31} = x_{32} = 0$.

The pairwise transfers $x_{12}, x_{13}, x_{21}, x_{23}, x_{31}, x_{32}$ can be thus calculated for all reports $\hat{\theta}$:

$(\hat{\theta}_1, \hat{\theta}_3) \backslash \hat{\theta}_2$	L	H
L,L	2.5,-1,2,2,0,0	2.5,-1,-2,-2,-4,10
L,H	2.5,-1,2,2,0,0	2.5,-1,-2,-2,4,-10
H,L	-2.5,1,0,0,0,0	-2.5,1,0,0,0,0
H,H	-2.5,1,0,0,0,0	-2.5,1,0,0,0,0

Table 3. Pairwise transfers in the new mechanism.

With such monetary transfers, if agents $1, 3$ have types $\theta_1 = \theta_3 = L$, and *assuming*

---

reports truthfully.

that agent 2 *reports truthfully,* then the total expected utility of $1, 3$ is maximized if they both report truthfully (unlike in VCG or AGV). The total utility of agents $1, 3$ is $U_{1,3} \equiv u_1 + u_3 + x_{12} - x_{21} - x_{23} + x_{32}$. If agents $1, 3$ report truthfully, their total expected utility (over 2's report) is 15.5. If agent 1 reports $\hat{\theta}_1 = H$, then $U_{1,3} = u_1 + u_3 + x_{12} = 13.5$. Similarly, let agent 1 report $\hat{\theta}_1 = L$, and agent 3 report $\hat{\theta}_3 = H$. Then the expected utility of agents $1, 3$ equals 13.5, which is still smaller than the utility of truthtelling.

There is a general way to show that *any* misreporting by agents $1, 3$ is not beneficial to them as a group. *Regardless of the joint strategy of agents* $1, 3$, the utility of truthful agent 2, $U_2 \equiv u_2 - x_{12} + x_{21} + x_{23} - x_{32}$, in expectation over θ_2, equals 2's ex ante efficient payoff of 2.5. For example, if agents $1, 3$ report their types to be $\hat{\theta}_1 = \hat{\theta}_3 = L$, then with probability $1/2$ agent 2 has type $\theta_2 = L$ ($\theta_2 = H$), and 2's total utility is $U_2 = 1.5$ ($U_2 = 3.5$). The average is 2.5. Since agent 2 has his utility equal to his ex ante efficient payoff (when reporting truthfully), and the mechanism is budget-balanced, agents $1, 3$ as a group become the residual claimants of the total payoff and suffer from misreporting. Moreover, each of agents $1, 3$ can also guarantee his ex ante efficient payoff by reporting truthfully (Section 4), thus leaving no profitable deviations for any group.

Note that the current mechanism has similarities with the *sequential* mechanism by Cremer and Riordan (1985): in both mechanisms, each agent gets compensated for either the expected externalities, or the ex post externalities its report imposes on others. However, the mechanism of Cremer and Riordan considers aggregate rather than pairwise externalities, and does not achieve coalition-proofness: in the example, the transfer to agent 2 is $u_1(\hat{\theta}) + u_3(\hat{\theta}) - E_{\theta_2,\theta_3}[u_1 + u_2 + u_3](\hat{\theta}_1, \theta_2, \theta_3)$, the transfer to agent 3 is $u_1(\hat{\theta}) + u_2(\hat{\theta}) - E_{\theta_2,\theta_3}[u_1 + u_2 + u_3](\hat{\theta}_1, \theta_2, \theta_3)$. Let agent 1 report type L. If agents $2, 3$ have types $\theta_2, \theta_3 = H, L$, then their total utility will strictly increase if they misreport both their types to be high; and the social choice will change. Such a

deviation is not profitable under the new mechanism.

3.2   Incentive compatibility and rationalizability

The transfers given by Definition 4 have the following property:

LEMMA 2 *For any agent i and reports $\hat{\theta}_1, ..., \hat{\theta}_{i-1}$, if agent i reports truthfully, every agent $j \neq i$ expects to pay zero to agent i, over i's report: $E_{\theta_i}[x_{ij}(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i)] = 0$.*

Proof.

The amount x_{ij} that agent $j \neq i$ pays agent i is

$$E_{\theta_{i+1}, ..., \theta_n}[u_j(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)] - E_{\theta_i, \theta_{i+1}, ..., \theta_n}[u_j(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, \theta_{i+1}, ...\theta_n)]$$

If one substitutes i's true type θ_i for his report $\hat{\theta}_i$ and takes the expectation over i's type, the value of j's payment to agent i becomes zero. In other words, agent j pays i the change in j's expected payoff caused by i's report, and that change has to be zero in expectation by the law of iterated expectations.   *Q.E.D.*

The mechanism with transfers given by Definition 4 is incentive compatible: if all agents but i report truthfully, then agent i prefers to report truthfully as well.

PROPOSITION 1 *For any report $\hat{\theta}_1, ..., \hat{\theta}_{i-1}$, any pair of types $\theta_i, \hat{\theta}_i \in \Theta_i$ and given that all agents $j > i$ report truthfully, one has:*

$$E_{\theta_{i+1}, ..., \theta_n}\left[u_i(\theta_i, s^*(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, \theta_{i+1}, ..., \theta_n)) + x_i(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, \theta_{i+1}, ..., \theta_n)\right] \geq$$

$$E_{\theta_{i+1}, ..., \theta_n}\left[u_i(\theta_i, s^*(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)) + x_i(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)\right]$$

Proof.

Since all agents $j > i$ report truthfully, due to Lemma 2, agent i expects to pay zero to each of them. Thus, agent i's report affects only pairwise transfers x_{ik} made to him, and his expected payoff from the social outcome, which sum up to

$$\sum_{k \neq i} x_{ik}(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i) + E_{\theta_{i+1}, ..., \theta_n}[u_i(\theta_i, s^*(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n))] =$$

$$= \sum_{k \neq i} E_{\theta_{i+1}, ..., \theta_n}[u_k(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)] - \sum_{k \neq i} E_{\theta_i, \theta_{i+1}, ..., \theta_n}[u_k(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, \theta_{i+1}, ... \theta_n)] +$$

$$+ E_{\theta_{i+1}, ..., \theta_n}[u_i(\theta_i, s^*(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i, \theta_{i+1}, ..., \theta_n))]$$

The first and the third terms sum up to the total expected payoff across all the agents, which is maximized if i reports truthfully. The second term does not depend on i's report, because an expectation over θ_i is taken. $Q.E.D.$

Under the following assumption, which requires that any individual misreport decreases the total payoff, it will be shown that truthful reporting is a uniquely rationalizable strategy in this mechanism:

ASSUMPTION 1 *For any agent i, any type profile θ_{-i}, and any two types $\theta_i \neq \theta_i'$:*

$$\sum_{j \neq i} u_j(\theta_j, s^*(\theta_{-i}, \theta_i)) + u_i(\theta_i, s^*(\theta_{-i}, \theta_i)) > \sum_{j \neq i} u_j(\theta_j, s^*(\theta_{-i}, \theta_i')) + u_i(\theta_i, s^*(\theta_{-i}, \theta_i'))$$

I use Assumption 1 only for the remainder of this section, and do not require it anymore in the next sections. While the Assumption is restrictive, it holds in some applications, for example, in some problems of efficient choice of quantity of public good. More generally, Assumption 1 holds in the problems that satisfy "strict" version of "single-crossing condition with public concavity" (SCC-PC), described in the paper by Ollár and Penta (2017): for each i, the type set Θ_i is a compact subset on a real line $\mathbb{R}$, the set of social outcomes S is a compact set in $\mathbb{R}$, the payoff function $u_i(\theta_i, s)$ is twice continuously differentiable with $\frac{\partial^2 u_i(\theta_i, s)}{\partial \theta_i \partial s} > 0$, $\frac{\partial^2 u_i}{\partial^2 s} < 0$, and the allocation rule

is linear: $s^*(\theta) = \sum_i \alpha_i \theta_i$ with $\alpha_i > 0$ for all i. An example of such an environment is when agents have quadratic valuations.

I use the concept of interim correlated rationalizability (ICR) from Dekel, Fudenberg and Morris (2007), and Battigalli et al. (2011). Agents' ex ante type distribution is a common knowledge, therefore, I restrict analysis to the payoff types of agents (that is, θ_i for each agent i). Each type θ_i forms a set of measurable conjectures $\sigma_{\theta_i}(\theta_{-i})[\hat{\theta}_{-i}]$ over the report $\hat{\theta}_{-i}$ of the other agents, dependent on their type profile θ_{-i}. The ex ante distribution of type profile θ_{-i} is denoted by a measurable function $f_{-i}(\theta_{-i})$. For each θ_i and $\sigma_{\theta_i}(\theta_{-i})$, i's best-response correspondence consists of reports $\hat{\theta}_i$ that maximize i's expected utility:

$$BR_{\theta_i}(\sigma_{\theta_i}(\theta_{-i})) \equiv$$

$$\operatorname*{argmax}_{\hat{\theta}_i \in \Theta_i} \int_{\Theta_{-i} \times \Theta_{-i}} \left[u_i(\theta_i, s^*(\hat{\theta}_i, \hat{\theta}_{-i})) + x_i(\hat{\theta}_i, \hat{\theta}_{-i}) \right] d\sigma_{\theta_i}(\theta_{-i})[\hat{\theta}_{-i}] df_{-i}(\theta_{-i})$$

where the integral is taken over all possible pairs $(\theta_{-i}, \hat{\theta}_{-i}) \in \Theta_{-i} \times \Theta_{-i}$.

For each set $T_{-i} \in \Theta_{-i}$, $\sigma_{\theta_i}(\theta_{-i})$ is called *consistent* with T_{-i}, if $\sigma_{\theta_i}(\theta_{-i})$ assigns probability 1 to the event $\hat{\theta}_{-i} \in T_{-i}$ (if T_{-i} is not measurable, then instead, $\sigma_{\theta_i}(\theta_{-i})$ would assign probability 1 to the event $\hat{\theta}_{-i} \in T'_{-i}$ for some measurable $T'_{-i} \subseteq T_{-i}$). ICR consists of an iterated deletion procedure in which, for each type θ_i, a report $\hat{\theta}_i^*$ survives the k-th round of deletion if and only if it can be justified by a conjecture that is consistent with the previous round of deletion:

DEFINITION 5 *(**ICR-Rationalizability**) For each i and θ_i, let $R_i^0(\theta_i) = \Theta_i$, and for each $k = 1, 2, ...$, let $R_{-i}^{k-1}(\{\theta_j\}_{j \neq i}) = \times_{j \neq i} R_j^{k-1}(\theta_j)$, and let $R_i^k(\theta_i) =$*

$$\left\{ \hat{\theta}_i^* : \begin{array}{l} \textit{there exists a measurable } \sigma_{\theta_i} : \Theta_{-i} \to \Delta(\Theta_{-i}) \textit{ such that} \\ (1) \forall \theta_{-i}, \ \sigma_{\theta_i}(\theta_{-i}) \textit{ is consistent with } R_{-i}^{k-1}(\theta_{-i}) : supp(\sigma_{\theta_i}(\theta_{-i})) \subseteq R_{-i}^{k-1}(\theta_{-i}) \\ (2) \ \hat{\theta}_i^* \in BR_{\theta_i}(\sigma_{\theta_i}(\theta_{-i})) \end{array} \right\}$$

and let $R_i(\theta_i) \equiv \bigcap_{k \geq 0} R_i^k(\theta_i)$

That is, agent i allows for correlation between reports of other agents, however, agent i keeps the ex ante belief in the (independent) distribution of others' types.

PROPOSITION 2 *In the mechanism with transfers given by Definition 4, under Assumption 1, for any i and θ_i, $R_i(\theta_i) = \{\theta_i\}$.*

Proof.

The social planner orders the agents in a sequence $1, ..., n$. The total amount of the pairwise transfers paid to agent n when n's report is revealed, is:

$$\sum_{k \neq n} [u_k(\hat{\theta}_1, ..., \hat{\theta}_{n-1}, \hat{\theta}_n) - E_{\theta_n} u_k(\hat{\theta}_1, ..., \hat{\theta}_{n-1}, \theta_n)]$$

The second term does not depend on agent n's report, $\hat{\theta}_n$. The first term together with n's payoff from the social choice, causes agent n's utility to be equal to the total payoff of all the agents from the social choice. Take any two types $\theta_n \neq \theta'_n$, and consider agent n of type θ_n misreporting to be θ'_n. By Assumption 1, for each report $\hat{\theta}_{-n}$ of other agents, the misreport of θ'_n causes the total payoff of all agents to decrease by a strictly positive amount $y(\hat{\theta}_{-n}) > 0$, dependent on $\hat{\theta}_{-n}$. Since $s^*(\theta)$ is measurable, so is $y(\hat{\theta}_{-n})$, and thus, for any conjecture μ_n of agent n over others' report $\hat{\theta}_{-n}$, the expected value of y is positive, and so is the expected loss of agent n from misreport. Thus, agent n does not want to misreport for each of his type.

Now let's look at agent $n-1$. Since agent n reports truthfully, and agent $n-1$ believes θ_n to have its ex ante distribution, agent $n-1$ expects to pay zero to agent n (Lemma 2). Thus, report $\hat{\theta}_{n-1}$ affects only the transfers made to agent $n-1$ at the stage when the planner reveals $n-1$'s report. These transfers equal to:

$$\sum_{k \neq n-1} [E_{\theta_n}[u_k(\hat{\theta}_1, ..., \hat{\theta}_{n-2}, \hat{\theta}_{n-1}, \theta_n)] - E_{\theta_{n-1}, \theta_n}[u_k(\hat{\theta}_1, ..., \hat{\theta}_{n-2}, \theta_{n-1}, \theta_n)]]$$

Report $\hat{\theta}_{n-1}$ affects only the first term, which, together with agent $n-1$'s payoff from the social choice, makes his utility equal to the total expected payoff of all the

18

agents. Similar to the case of agent n, one can show that for each type θ_{n-1} and each conjecture μ_{n-1}, if agent $n-1$ misreports his type to be $\theta'_{n-1} \neq \theta_{n-1}$, he will expect to suffer a strictly positive loss. Thus, agent $n-1$ has a unique best strategy to report truthfully.

Similarly, all agents $1,...,n-2$ can each be shown to have truthful reporting as the only interim correlated rationalizable strategy as well.   Q.E.D.

Note that one needs the assumption of independent private values for Proposition 2: private values ensure that the payoff u_i of each agent i does not depend on θ_{-i} and does not depend on whether the reports of agents $1,...,i-1$ were truthful or not; while independence allows i's expected payoff to be independent from truthful reports of agents $i+1,...,n$. On the contrary, with correlated types, agents' beliefs about each other's types would differ from those of the social planner, and the transfers in the mechanism, that are built according to the planner's beliefs, would not induce truthtelling. In particular, the mechanism is not ex post incentive compatible, as shown in the example below.

Consider a public good problem with two agents 1,2 with quadratic valuations: agents' types θ_1, θ_2 are i.i.d., and are uniformly distributed on $[0, 1]$, and the payoff function of agent i is $u_i(\theta_i, s) = -(\theta_i - s)^2$. The efficient social choice is $s^*(\theta) = \frac{\theta_1+\theta_2}{2}$, with payoffs $u_1(\theta) = u_2(\theta) = -\frac{(\theta_1-\theta_2)^2}{4}$. The ex ante payoff of each agent $E_\theta u_i(\theta) = E_\theta[-\frac{(\theta_1-\theta_2)^2}{4}] = -\frac{1}{24}$, while after the report $\hat{\theta}_1$ of agent 1, the expected payoff of each agent $E_{\theta_2} u_i(\hat{\theta}_1, \theta_2) = \frac{1}{4}(-\hat{\theta}_1^2 + \hat{\theta}_1 - \frac{1}{3})$. Respectively, the monetary transfer from agent 2 to agent 1 is $x_{12}(\hat{\theta}_1) = E_{\theta_2} u_2(\hat{\theta}_1, \theta_2) - E_\theta u_2(\theta) = \frac{1}{4}(-\hat{\theta}_1^2 + \hat{\theta}_1 - \frac{1}{6})$, while the monetary transfer from 1 to 2 is $x_{21}(\hat{\theta}) = u_1(\hat{\theta}) - E_{\theta_2} u_1(\hat{\theta}_1, \theta_2) = \frac{1}{2}\hat{\theta}_1\hat{\theta}_2 - \frac{1}{4}\hat{\theta}_1 - \frac{1}{4}\hat{\theta}_2^2 + \frac{1}{12}$.

The payoff from the social choice for agent 1 is equal to $u_1(\theta_1, s^*(\hat{\theta})) = -(\theta_1 - \frac{\hat{\theta}_1+\hat{\theta}_2}{2})^2$. The total utility of agent 1 from report $\hat{\theta}$ is $u_1(\theta_1, s^*(\hat{\theta})) + x_{12}(\hat{\theta}_1) - x_{21}(\hat{\theta})$, and if one considers terms that contain $\hat{\theta}_1$, one gets in total $-\frac{1}{2}\hat{\theta}_1^2 + \frac{1}{2}\hat{\theta}_1 + \theta_1\hat{\theta}_1 - \hat{\theta}_1\hat{\theta}_2$. If agent 1 does not observe 2's report $\hat{\theta}_2$, and believes 2 to report truthfully, in expectation

$\hat{\theta}_2 = \frac{1}{2}$, which reduces the expression above to $-\frac{1}{2}\hat{\theta}_1^2 + \theta_1\hat{\theta}_1$, with truthful report $\hat{\theta}_1^* = \theta_1$ being optimal for agent 1. However, if 1 knows 2's report $\hat{\theta}_2$, 1's optimal report becomes $\hat{\theta}_1^* = \theta_1 + (\frac{1}{2} - \hat{\theta}_2)$. The additional term of $\frac{1}{2} - \hat{\theta}_2$ makes agent 1 misreport his type in the "opposite" direction compared to 2's report: this would make the planner to overestimate the negative externalities 2's report imposes on 1, thus reducing the transfer x_{21} agent 1 has to pay.

Since the mechanism is not ex post incentive compatible, the result in Proposition 2 cannot be extended to belief-free rationalizability (Battigalli et al., 2011). In addition, and as discussed in Section 6, the reporting in the mechanism can be made sequential. In that case, the truthful reporting would be Bayesian incentive compatible, but not uniquely interim sequentially rationalizable (Penta, 2012) nor uniquely (belief-free) backwards rationalizable (Penta, 2015).

4   Properties of the mechanism

In this section I show the main property of the mechanism, later referred to as *ex ante removal of externalities* (here and later I do not require Assumption 1 that was used to show Proposition 2), for each agent i:

THEOREM 1 *In the efficient mechanism with transfers given by Definition 4, the truthful strategy of agent i: $\sigma_i(\theta_i) \equiv \theta_i$ guarantees him his ex ante efficient payoff, in expectation over type θ_i, regardless of others' reports:*
$$\forall \hat{\theta}_{-i},\ E_{\theta_i}[u_i(\theta_i, s^*(\theta_i, \hat{\theta}_{-i})) + x_i(\theta_i, \hat{\theta}_{-i})] = E_\theta u_i(\theta).$$

Note that Theorem 1 allows all agents except for i to coordinate their reports, still they cannot affect i's expected utility.

Proof.

The social planner "reveals" agents' reports one by one, in the order $1, ..., n$. After the reports of agents $1, ..., j$ are revealed, the *current* expected payoff of agent i is $E_{\theta_{j+1}, ..., \theta_n}[u_i(\hat{\theta}_1, ..., \hat{\theta}_j, \theta_{j+1}, ..., \theta_n)]$, while the *current* transfer to agent i (that is, a sum of transfers to/from agent i as determined by the reports of agents $1, ..., j$) is $-\sum_{k \leq j} x_{ki}$ if $j < i$, or is $-\sum_{k \leq j, k \neq i} x_{ki} + \sum_{k \neq i} x_{ik}$ if $j \geq i$. Let's show that the total *current* utility of agent i—sum of i's current expected payoff, and i's current transfer—does not change with the report of any agent $m \neq i$, and does not change in expectation over the truthful report of agent i.

Indeed, when the report of agent $m \neq i$, $\hat{\theta}_m$, is revealed, i's current expected payoff changes from $E_{\theta_m, ..., \theta_n}[u_i(\hat{\theta}_1, ..., \hat{\theta}_{m-1}, \theta_m, ..., \theta_n)]$ to $E_{\theta_{m+1}, ..., \theta_n}[u_i(\hat{\theta}_1, ..., \hat{\theta}_m, \theta_{m+1}, ..., \theta_n)]$, however, i pays back that change to m through x_{mi}. In other words, i always compensates the other agents for the change in his payoff, and thus, their reports (joint or not) cannot affect i's total current utility. When i's report is revealed, due to Lemma 2, every agent $j \neq i$ pays zero in expectation to i; plus, the change in i's current expected payoff, $E_{\theta_{i+1}, ..., \theta_n}[u_i(\hat{\theta}_1, ..., \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)] - E_{\theta_i, ..., \theta_n}[u_i(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i, ..., \theta_n)]$, caused by i's report, is zero in expectation over i's truthful report. Thus, i's current utility does not change in expectation over θ_i.

Before agents start reporting their types, i's current utility is $E_\theta u_i(\theta)$, and so is i's current utility after all reports, in expectation over θ_i: $E_{\theta_i}[u_i(\theta_i, s^*(\theta_i, \hat{\theta}_{-i})) + x_i(\theta_i, \hat{\theta}_{-i})] = E_\theta u_i(\theta)$. $\hfill Q.E.D.$

The assumption of independent private values is essential for Theorem 1. With correlated types, there may be agent j who reports before agent i, and who can condition his report $\hat{\theta}_j$ on i's (expected truthful) report, affecting i's expected utility. With interdependent values i's ex post payoff from social choice directly depends on others' types, thus causing pairwise transfers not to properly compensate agent i for the externalities others impose on him.

4.1 Coalition-proofness

Theorem 1 shows that any agent can guarantee himself his ex ante efficient payoff, regardless of others' strategies. This allows to extend incentive compatibility result in Proposition 1 to groups: any group C of agents cannot increase their total expected utility by misreporting. A joint strategy of agents in C is defined as a Borel-measurable function $\sigma^C : \prod_{i \in C} \Theta_i \to \Delta \left(\prod_{i \in C} \Theta_i \right)$, allowing the coalition to condition reports of all its agents $i \in C$ on its entire type profile θ_C.

DEFINITION 6 *A mechanism is coalition-proof, if for any coalition $C \subset \{1, ..., n\}$, and any joint type profile of its members θ_C,*

$$E_{\theta_{-C}} \sum_{i \in C} [u_i(\theta_i, s^*(\theta_C, \theta_{-C})) + x_i(\theta_C, \theta_{-C})] \geq E_{\theta_{-C}} \sum_{i \in C} \left[u_i(\theta_i, s^*(\hat{\theta}_C, \theta_{-C})) + x_i(\hat{\theta}_C, \theta_{-C}) \right]$$

Definition 6 implies that the members of coalition C are able to transfer money to each other; thus a profitable coalitional deviation improves the *sum* of its members' utilities, rather than each of its members utilities separately. Coalition C can be thought of as a single player: the types of its members are a common knowledge within C; and there is no threat of a subcoalitional deviation. The property of coalition-proofness from Definition 6 resembles a concept of *Strong Nash equilibrium* in Aumann (1959), and Bernheim, Peleg, and Whinston (1987), in which agents in a coalition coordinate their reports without a risk of a further subcoalitional deviation. The current paper differs in the presence of privately observed types and monetary transfers within the coalition.

THEOREM 2 *The efficient mechanism with transfers given by Definition 4 is coalition-proof.*

Proof.

Given any type realisation θ_{-C} of agents outside coalition C, and any report $\hat{\theta}_C$ of coalition C, one gets the following expression for the total utility of all agents:

$$\sum_{i \in C}(u_i(\theta_i, s^*(\hat{\theta}_C, \theta_{-C})) + x_i(\hat{\theta}_C, \theta_{-C})) + \sum_{j \notin C}(u_j(\theta_j, s^*(\hat{\theta}_C, \theta_{-C})) + x_j(\hat{\theta}_C, \theta_{-C})) =$$

$$= \sum_{i \in C} u_i(\theta_i, s^*(\hat{\theta}_C, \theta_{-C})) + \sum_{j \notin C} u_j(\theta_j, s^*(\hat{\theta}_C, \theta_{-C}))$$

which holds due to ex post budget balance (Lemma 1). Since each agent $j \notin C$ reports truthfully, if one takes expectation over θ_{-C} of the above equation, due to Theorem 1, j's expected utility equals his ex ante efficient payoff $E_\theta u_j(\theta)$.

$$E_{\theta_{-C}} \left[\sum_{i \in C}(u_i(\theta_i, s^*(\hat{\theta}_C, \theta_{-C})) + x_i(\hat{\theta}_C, \theta_{-C})) \right] + \sum_{j \notin C} E_\theta u_j(\theta) =$$

$$= E_{\theta_{-C}} \left[\sum_{i \in C} u_i(\theta_i, s^*(\hat{\theta}_C, \theta_{-C})) + \sum_{j \notin C} u_j(\theta_j, s^*(\hat{\theta}_C, \theta_{-C})) \right]$$

The total expected utility of agents in C (first term on left-hand side) is equal to a total payoff of all the agents (right-hand side), up to a constant. The total payoff is maximized if coalition C reports truthfully. [3]   *Q.E.D.*

4.2   Full efficient implementation

Theorem 1 guarantees that each agent can get his ex ante efficient payoff. Respectively, in any Bayesian-Nash equilibrium, in which agents' total report $\hat{\theta}$ is a random variable dependent on θ, the total payoff of agents is ex ante efficient:

THEOREM 3 *In any Bayesian-Nash equilibrium:* $E_\theta \left[\sum_i u_i(\theta_i, s^*(\hat{\theta})) \right] = \sum_i E_{\theta_i} u_i(\theta)$

---

[3]In other words, given any type θ_C, if agents in C misreport, then this misreport will cause the expected (over reports of non-collusive agents) total payoff to decrease by X, and, respectively, the expected utility of C will decrease by X.

Proof.

In any Bayesian-Nash equilibrium,[4] the overall report $\hat{\theta}$ is a measurable function that depends on θ, according to the total equilibrium strategy profile of all agents, σ. The ex post total utility of any agent i is equal to $u_i(\theta_i, s^*(\hat{\theta})) + x_i(\hat{\theta})$. Agent i's equilibrium strategy gives him at least as much as his truthful report. Thus, in expectation over θ_i, i gets at least his ex ante efficient payoff: $E_{\theta_i}[u_i(\theta_i, s^*(\hat{\theta})) + x_i(\hat{\theta})] \geq E_\theta u_i(\theta)$. Since any agent can guarantee to get his ex ante efficient payoff, by taking expectation over the total profile θ, and taking sum across all agents, one gets

$$E_\theta \left[\sum_i (u_i(\theta_i, s^*(\hat{\theta})) + x_i(\hat{\theta})) \right] \geq E_\theta \sum_i u_i(\theta)$$

Since the mechanism is budget-balanced, $\sum_i x_i(\hat{\theta}) = 0$, and

$$E_\theta \left[\sum_i u_i(\theta_i, s^*(\hat{\theta})) \right] \geq E_\theta \sum_i u_i(\theta)$$

The total payoff cannot exceed its efficient value, thus the expression above holds with equality.   *Q.E.D.*

Theorem 3 implies that the total payoff in equilibrium, $\sum_i u_i(\theta_i, s^*)$ reaches its efficient value with probability 1. However, the total payoff can be inefficient conditional on zero probability events. For example, assume a single-good auction environment with two agents 1,2. Agent 1 has valuations of either 5 or 10 with probability 1/2-1/2 each, while agent 2 has valuation 15 with probability 1, and valuation 7 with probability 0. If agent 1 with valuation 5 misreports his valuation to be 10, there would be an inefficiency in case of agent 2 having a valuation of 7. However, this is a zero probability event, thus this misreport does not reduce the expected monetary transfer of agent 1, making the misreport an equilibrium choice.

---

[4]Truthful reporting is a Bayesian-Nash equilibrium, thus the set of equilibria is non-empty.

5 Ordering of agents

The ordering of the agents' reports determines the ex post monetary transfers (together with their utilities). However, the total interim utility of agent i—estimated after i learnt his type but before the agents announce their types in the mechanism—does not depend on the ordering:

PROPOSITION 3 *In the efficient mechanism with transfers given by Definition 4, for any agent i and any type θ_i, under truthful reporting, agent i's interim utility does not depend on the ordering.*

Proof.

The total utility of agent i is equal to:

$$u_i(\theta_i, s^*(\theta)) - \sum_{j \neq i} x_{ji}(\theta) + \sum_{j \neq i} x_{ij}(\theta)$$

The social choice $s^*(\theta)$ does not depend on the ordering, and neither does i's payoff u_i. Due to Lemma 2, for any j, $E_{\theta_j} x_{ji}(\theta) = 0$. Finally, for each $j \neq i$, transfer $x_{ij}(\theta)$ is equal to:

$$E_{\theta_{i+1},\ldots,\theta_n}[u_j(\theta_1,\ldots,\theta_{i-1},\theta_i,\theta_{i+1},\ldots,\theta_n)] - E_{\theta_i,\theta_{i+1},\ldots,\theta_n}[u_j(\theta_1,\ldots,\theta_{i-1},\theta_i,\theta_{i+1},\ldots\theta_n)]$$

Taking expectation over truthful reporting of types $\theta_1,\ldots,\theta_{i-1}$, one gets

$$E_{\theta_{-i}} u_j(\theta_i, \theta_{-i}) - E_\theta u_j(\theta)$$

that depends only on θ_i.   *Q.E.D.*

It is possible to make the mechanism symmetric by randomizing over the agents' orderings:

COROLLARY 1 *The statements of Theorems 1, 2, 3 hold for the symmetric mechanism: Take transfers given by Definition 4 for each ordering of the agents, and take the average across all orderings.*

In the resulting symmetric mechanism, each agent i receives the following transfer:

$$x_i(\hat{\theta}) = E_\theta u_i(\theta) - u_i(\hat{\theta}) + \quad (1)$$

$$+ \sum_{m=0}^{m=n-1} \frac{m!(n-m-1)!}{n!} \sum_{j_1,j_2,...,j_m \neq i} [E_{\theta_{j_1},...,\theta_{j_m}} a(\{\hat{\theta}_k\}_{k \neq i,j_1,j_2,...,j_m}, \hat{\theta}_i, \theta_{j_1}, ..., \theta_{j_m}) -$$

$$- E_{\theta_i,\theta_{j_1},...,\theta_{j_m}} a(\{\hat{\theta}_k\}_{k \neq i,j_1,j_2,...,j_m}, \theta_i, \theta_{j_1}, ..., \theta_{j_m})]$$

where $\hat{\theta}$ is the total report submitted, and $a(\theta) = \sum_i u_i(\theta)$ is the efficient total payoff.

The symmetric payment scheme from Corollary 1 has similarities with coalitional games and the Shapley value (Shapley (1953)). The Shapley value is defined as follows: In an n-player coalitional game each group $S \subset n$ of players is assigned a value $v(S)$ - the total sum of payoffs that members of S are able to obtain by cooperation. The Shapley value for player i is:

$$\phi_i(v) = \sum_{S \subset n/\{i\}} \frac{|S|!(|n| - |S| - 1)!}{|n|!}(v(S \cup \{i\}) - v(S))$$

In other words, the Shapley value gives i his marginal contribution to a random subcoalition S. The Shapley value describes a way to distribute the total payoff $v(n)$ between players, and it has nice properties of efficiency, linearity and symmetry.

In this paper, for each group S, one can define the effect of type θ_S on the total payoff, that is, one estimates the total expected payoff of all agents, given that types in S are known: $v(S) \equiv E_{\theta_{-S}} \left[\sum_{j=1}^n u_j(\theta_j, s^*(\theta_S, \theta_{-S})) \right]$. With such a value function $v(S)$, the corresponding Shapley value gives each agent i the marginal externality his report imposes on the total payoff, conditional on knowing reports of a random group $S \subset n/\{i\}$. The scheme from Corollary 1 makes agent i pay the externality that other agents' reports impose on his payoff (first line in expression(1)); and get paid the Shapley value of his report - a marginal externality caused by i's report $\hat{\theta}_i$ on the total payoff, conditional on knowing the types in a random group $S = n/\{i, j_1, ..., j_m\}$.

6 Conclusion

This paper provides an efficient mechanism with agents having independent private values and quasilinear preferences. In the mechanism, each pair of agents directly exchange monetary transfers according to the pairwise externalities they impose on each other. The transfers internalize the externalities, thereby achieving coalition-proofness and full efficient implementation.

The reporting is simultaneous in the mechanism, although the reports are treated in a sequential manner. In fact, the reporting can be made sequential, so that agents could observe the previous reports before reporting themselves. The sequential reporting would change strategies of agents and coalitions, since agents (coalitions) could condition their reports on previous reports. Nevertheless, it can be shown that the statement of Theorem 1 holds even if agents submit their reports sequentially and publicly, and can condition their reports on all previous reports (Safronov, 2016). Indeed, in the proof of Theorem 1, even if agents that report after i can condition their reports on i's report $\hat{\theta}_i$, i would still guarantee to get his ex ante efficient payoff if reporting truthfully. Moreover, with an appropriate definition of strategies and a concept of equilibrium, the proofs of Theorems 2, 3 can be readjusted so that statements of Theorems hold in the sequential case.

At the same time, the unique rationalizability result from Proposition 2 is not directly extended to the sequential case. For dynamic games, interim correlated rationalizability is extended to the concept of interim sequential rationalizability (Penta, 2012), and for truthful reporting to be a uniquely sequentially rationalizable strategy, the mechanism would have to be ex post incentive compatible, which is not true. In order to achieve unique rationalizablity, one would need a new concept, according to which agent i keeps ex ante beliefs about the types of agents that have not reported yet.

The mechanism requires the type distribution to be common knowledge, which is a

widely used assumption in the mechanism design literature. This assumption may seem to be too demanding, according to the Wilson doctrine. However, Crémer (1996) shows that is generally impossible to have truthful reporting as a weakly dominant strategy for coalitions, that is, the social planner needs knowledge of types distribution to achieve coalition proofness. The mechanism in this paper achieves coalition proofness assuming independent private values; an extension to correlated types is not straightforward. Crémer and McLean (1988) have studied auctions with agents having correlated types, they have shown that the auctioneer can use the correlation to extract full surplus from the agents: each agent would face a lottery with weights dependent on others' reports, and the lottery can be adjusted to ensure truthful reporting. The method developed by Crémer and McLean's was later used in other mechanism design papers with correlated types. However, this method requires that for each agent i, the conditional distribution of others' types varies enough with i's type θ_i. This requirement is likely to fail when one considers coalitions as players: the number of possible types for a large coalition C is greater than the number of type profiles of agents outside the coalition, thus the type profile θ_{-C} outside coalition C would not vary enough with θ_C. A new method has to be developed to achieve coalition-proofness in the case of correlated types.

References

ARROW, K. (1979): "The Property Rights Doctrine and Demand Revelation Under Incomplete Information," in *Economics and Human Welfare*, ed. by M. Boskin. New York: Academic Press.

AUMANN, R. (1959): "Acceptable points in general cooperative n-person games," in *Contributions to the Theory of Games IV*, Princeton Univ. Press, Princeton, N.J..

BATTIGALLI, P., DI TILLIO , A., GRILLO, E., AND A. PENTA (2011): "Interactive Epistomology and Solution Concepts for Games with Asymmetric Information," *The B.E. Journal of Theoretical Economics*, Volume 11 (Advances), Article 6.

BIRAN, O., AND F. FORGES (2011): "Core-Stable Rings in Auctions with Independent Private Values," *Games and Economic Behavior*, 73, 52–64.

BERNHEIM, D., PELEG, B., AND M. WHINSTON (1987): "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, 42, 1–12.

CHE, Y., AND J. KIM (2006): "Robustly Collusion-Proof Implementation," *Econometrica*, 74, 1063–1107.

———(2009): "Optimal Collusion-Proof Auctions," *Journal of Economic Theory*, 144, 565–603.

CHE, Y., CONDORELLI, D., AND J. KIM (2013): "Weak Cartels and Collusion-Proof Auctions," working paper.

CHEN, J., AND S. MICALI (2012): "Collusive Dominant-Strategy Truthfulness," *Journal of Economic Theory*, 147, 1300–1312.

CHEN, Y.-C., KUNIMOTO, T., AND Y. SUN (2015): "Implementation with Transfers," *Discussion Paper No, 2015-04*.

CLARKE, E. (1971): "Multipart Pricing of Public Goods," *Public Choice*, 11, 19–33.

CRAMTON, P., GIBBONS, R., AND P. KLEMPERER (1987): "Dissolving a Partnership Efficiently," *Econometrica*, 55, 615–632.

CRÉMER, J., AND M. RIORDAN (1985): "A Sequential Solution to the Public Goods Problem," *Econometrica*, 53, 77–84.

CRÉMER, J., AND R.P. MCLEAN (1988): 'Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257.

CRÉMER, J. (1996): "Manipulations by Coalitions under Asymmetric Information: The Case of Groves Mechanism," *Games and Economic Behavior*, 13, 39–73.

D'ASPREMONT, C., AND L. GERARD-VARET. (1979): "Incentives and Incomplete Information," *Journal of Public Economics*, 11, 25–45.

DEKEL, E., FUDENBERG, D., AND S. MORRIS. (2007): "Interim Correlated Rationalizability," *Theoretical Economics*, 2, 15–40.

ERDIL, A., AND P. KLEMPERER (2011): "A New Payment Rule for Core-Selecting Package Auctions," *Journal of the European Economic Association*, 8, 537–547.

GROVES, T. (1973): "Incentives in Teams," *Econometrica*, 41, 617–631.

JACKSON, M. (1991): "Bayesian Implementation," *Econometrica*, 59, 461–477.

KRISHNA, V., AND M. PERRY. (2000): "Efficient Mechanism Design," unpublished.

LAFFONT, J., AND D. MARTIMORT (1997): "Collusion under Asymmetric Information," *Econometrica*, 65, 875–911.

———(1998): "Collusion and Delegation," *The Rand Journal of Economics*, 29, 280–305.

———(2000): "Mechanism Design with Collusion and Correlation," *Econometrica*, 68, 309–342.

MASKIN, E. (1998): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

MATSUSHIMA, H. (1993): "Bayesian Monotonicity with Side Payments," *Journal of Economic Theory*, 59, 107–121.

MCAFEE, R. AND J. MCMILLAN (1992): "Bidding Rings," *American Economic Review*, 82, 579–599.

MOULIN, H. (1999): "Incremental Cost Sharing: Characterization by Coalition-Strategy Proofness," *Social Choice and Welfare*, 16, 279–320.

MYERSON, R. (2007): "Virtual Utility and the Core for Games with Incomplete Information," *Journal of Economic Theory*, 136, 260–285.

MYERSON, R. AND M. SATTERTHWAITE (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281.

OLLÁR, M. AND A. PENTA (2017): "Full Implementation and Belief Restrictions," *American Economic Review*, 107, 2243–2277

PALFREY, T. AND S. SRIVASTAVA (1989): "Implementation with Incomplete Information in Exchange Economies," *Econometrica*, 57, 115–134.

——— (1989): "Mechanism Design with Incomplete Information: A Solution to the Implementation Problem," *Journal of Political Economy*, 97, 668–691.

PENTA, A. (2012): "Higher Order Uncertainty and Information: Static and Dynamic Games," *Econometrica*, 80, 631–660.

PENTA, A. (2013): "On the Structure of Rationalizability on Arbitrary Spaces of Uncertainty," *Theoretical Economics*, 8, 405–430.

PENTA, A. (2015): "Robust Dynamic Implementation," *Journal of Economic Theory*, 160, 280–316.

POSTLEWAITE, A. AND D. SCHMEIDLER (1986): "Implementation in Differential Information Economies," *Journal of Economic Theory*, 39, 14–33.

SAFRONOV, M. (2016): "A Coasian Approach to Efficient Mechanism Design," Cambridge-INET working paper 1619.

SAMUELSON, W. (1985): "A Comment on the Coase Theorem," in *Game-Theoretic Models of Bargaining*, ed. by A. Roth. Cambridge University Press.

SHAPLEY, L. (1953): "A value for n-person Games," in *Contributions to the Theory of Games*, ed. by H. Kuhn and A. Tucker, Ann. Math. Studies 28, Princeton University Press.

VICKREY, W. (1961): "Counterspeculation, Auctions and Competitive Sealed Tenders," *Journal of Finance*, 16, 8–37.

WILLIAMS, S. (1999): "A Characterization of Efficient, Bayesian Incentive Compatible Mechanisms" *Economic Theory*, 14, 155–180.

WILSON, R. (1987): "Game-Theoretic Analysis of Trading Processes", *Advances in Economic Theory*, ed. by Bewley. Cambridge University Press.

A   Individual rationality

The mechanism in this paper is coalition-proof, efficient and budget-balanced, however, it is assumed that agents cannot quit the mechanism. Adding participation constraint as another requirement is impossible in general, due to Myerson–Satterwaite impossibility theorem: if the agents could decide on participation after learning their types, they might choose to quit.[5] Despite the impossibility result, a different question may be asked. The current paper extends the property of incentive compatibility to coalition-proofness. Assuming there exists an incentive-compatible mechanism M that also satisfies individual rationality, is it possible to repeat the exercise of extending the incentive compatibility to coalition proofness while preserving individual rationality? The answer is positive if there is "enough structure" on agents' types, as shown below.

I consider the setting by Krishna and Perry (2000). The set of possible social choices S

---

[5]By Theorem 1, if the agents decided on participation at the ex ante stage, the mechanism would be as attractive to each agent as in the first-best case.

is finite and has k elements. The agents $i \in \{1, ..., n\}$ have independently distributed private values, with each type θ_i having a continuous density function, with full support on a compact and convex subset Θ_i of Euclidean space $\mathbb{R}^k$. A direct efficient mechanism M is a pair (s^*, x) of a measurable efficient social choice rule $s^*(\hat{\theta})$ and a transfer vector $x(\hat{\theta})$, as dependent on the report profile $\hat{\theta}$. Mechanism M is incentive compatible, if truthtelling is a Bayesian Nash equilibrium in the resulting game.

PROPOSITION 4 *If a direct efficient mechanism M is incentive compatible, budget-balanced, and interim individually rational, then there exists an alternative direct efficient mechanism M', that satisfies all the three properties above, and in addition satisfies the statements of Theorems 2, 3.*

Proof.

Given the efficient rule $s^*(\hat{\theta})$ of mechanism M, one can construct a coalition-proof mechanism $\hat{M}$ using transfers from Definition 4, the latter transfers are denoted as $x^{\hat{M}}(\hat{\theta})$. Since both mechanisms M and $\hat{M}$ are incentive compatible, by Lemma 1 in Krishna and Perry, there exists a set of type-independent constants $\{y_i\}_{i=1}^n$, such that for each type θ_i of agent i, i's total interim utility in mechanism M differs from i's total interim utility in mechanism $\hat{M}$ by y_i. Let's show that a mechanism M', that has the same efficient rule $s^*(\hat{\theta})$ and transfers $x_i^{M'}(\hat{\theta}) = x_i^{\hat{M}}(\hat{\theta}) + y_i$, satisfies the claim of Proposition 4. First, by construction, both M and M' give the same interim utility to each agent, thus M' is interim individually rational. Next, i's ex ante total utility, $E_\theta[u_i(\theta_i, s^*(\theta)) + x_i^M(\theta)]$, in mechanism M differs from i's ex ante total utility in mechanism $\hat{M}$ by y_i. Since both $M, \hat{M}$ have the same allocation rule, it has to be that the ex ante monetary transfer to i, $E_\theta x_i^M(\theta)$, in mechanism M differs from the one in $\hat{M}$ by y_i. Since both M and $\hat{M}$ are budget-balanced, this means that $\sum_i y_i = 0$, and that mechanism M' is (ex post) budget balanced. Finally, adding a constant transfer y_i into mechanism $\hat{M}$ did not change incentives, thus mechanism M' is incentive compatible, and it satisfies Theorems 2, 3. $\qquad$ *Q.E.D.*