Coalition-Proof Full Efficient Implementation

Mikhail Safronov*

University of Cambridge

January 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 types, 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.

*E-mail: mikhailsafronov2014@u.northwestern.edu. This project was started during my PhD study at Northwestern University. I thank the editor, the associate editor, and anonymous referees for comments and suggestions which helped to vastly improve the paper. I am grateful to Matt Elliott, Aytek Erdil, Robert Evans, Ben Golub, Qingmin Liu, Roger Myerson, Mariann Ollár, Alessandro Pavan, Antonio Penta, Doron Ravid, Soenje Reiche, Ludovic Renou, Hamid Sabourian, Bruno Strulovici, Juuso Toikka, Asher Wolinsky, and participants at the 2016 Annual Conference of the Royal Economic Society, the 21st Coalition Theory Network workshop, the 2016 North American Summer Meeting of the Econometric Society, the Stony Brook 27th International Conference on Game Theory, and the 5th World Congress of the Game Theory Society for fruitful discussion. I am thankful for the financial support provided by the Center for Economic Theory and the Graduate School at Northwestern University, and by the Cambridge-INET Institute. All errors are mine.

1 Introduction

The problem of externalities which cause economic inefficiency can be solved if there exists a procedure for internalizing of the externalities. This paper develops such a procedure in a benevolent social planner's problem in which agents have independent private types and quasilinear preferences. The social planner (she) asks each agent (he) to report his private type, 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. In order to induce truthful reports, the agents should be required to compensate each other for these externalities.

The idea of internalizing of 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. In the new mechanism each pair of agents directly compensate each other for the pairwise externalities. There are two equivalent versions of the mechanism: the direct mechanism and the sequential mechanism. In the sequential mechanism the agents

are ordered in an *arbitrary* sequence and report their types sequentially according to that ordering. Each report is publicly observed, including by the agents who have yet to report. When any agent i reports his type, 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. The direct (version of the) mechanism is equivalent to the sequential version, except that the agents report their types simultaneously. Each pair of agents exchange the same monetary transfers *as if* the reports were submitted sequentially.

In the new mechanism (sequential or direct), 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,

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 weak Perfect Bayesian 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. Full efficient implementation holds regardless of whether agents act individually or in coalitions. In the latter case, any agent is guaranteed his ex ante efficient payoff by refusing to join a coalition and reporting truthfully.

The set of equilibria in the mechanism always contains the truthful equilibrium. 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.

In the sequential version of the mechanism, in the truthful equilibrium 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 who report before i and assuming that the agents who report after i, will 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

sequentially rationalizable strategy. The last agent strictly prefers to report truthfully regardless of others' reports. Knowing that, the next-to-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 order in which the agents report their types determines the monetary transfers to each agent from the mechanism; however, the interim utility of each agent does not depend on the order. In the direct (simultaneous) version of the mechanism, symmetry could be imposed 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 sequential 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 considers the direct version of the mechanism. Section 6 discusses setting with imposed participation constraints. Section 7 concludes.

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 achieved. In comparison, 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 private 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, even in the case where the entire coalition behaves as a single player.

The assumption of passive beliefs, which is used widely in the models of collusion, was motivated in Myerson (2007). The agents report their types to the social planner, though they are not yet committed to them. A third party proposes a collusion, and if successful, the agents involved resubmit their reports to the social planner. If the collusion fails, the reports are unchanged. I do not require this assumption: with an

endogenous process of coalition formation, any agent can refuse to join the coalition and report truthfully, in which case he is guaranteed to get his ex ante efficient payoff, since this refusal will not make others' strategies depend on the actual type of that agent. This possibility causes the mechanism to yield the efficient outcome in any equilibrium.

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 (2015) shows that full implementation is achieved by using a direct mechanism. 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.

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. Types are independently distributed across the agents, and the ex ante distribution of types is publicly known. There is a set of social outcomes S, each outcome $s \in S$ gives agent i a payoff of $u_i(\theta_i, s)$. That is, the setup is characterized by *independent private values*. I assume that for any type profile θ, there exists an *efficient* social outcome $s^*(\theta) \equiv \arg\max_{s \in S} \sum_i u_i(\theta_i, s)$, which maximizes the sum of the agents' payoffs,[1] given θ. The payoff of agent i at the efficient outcome $s^*(\theta)$ is denoted by $u_i(\theta)$. The sets Θ_i and S are measurable; in addition, for any group of agents C and their type profile θ_C, and any agent i, the expectation $E_{\theta_{-C}} u_i(\theta_C, \theta_{-C})$ is assumed to exist.

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 $u_i(\theta_i, s) + x_i$.

[1] In case of several outcomes that maximize the total payoff, one of them is arbitrarily chosen to be $s^*(\theta)$.

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 private 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. The resulting mechanism will have two equivalent versions, that prescribe the same net transfer $x_i(\hat{\theta})$ to any agent i: the agents report their types either simultaneously (that is, direct mechanism), or sequentially and publicly. I will postpone discussion of the direct mechanism to Section 5 and

focus on the sequential version. The agents are *arbitrarily* ordered into a sequence $1, 2, 3, ..., n$, which is publicly known before the agents start reporting their types. The agents report according to that sequence, and each report is publicly observed. In the mechanism, each agent i can condition his report on all the previous reports:

DEFINITION 3 *Agent i's strategy σ_i: $\Pi_{k=1}^{i-1}\Theta_k \times \Theta_i \longrightarrow \Delta(\Theta_i)$ determines his report $\hat{\theta}_i$ as dependent on his true type θ_i and the reports of agents before him $\hat{\theta}_1, ..., \hat{\theta}_{i-1}$. Agent i reports truthfully, if $\sigma_i(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i) \equiv \theta_i$, $\forall(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i)$.*

Given any set of reports $\hat{\theta}_1, ..., \hat{\theta}_i$ of agents $1, ..., i$, the social planner can estimate, for any agent j, j's expected payoff from the social choice: $E_{\theta_{i+1},...,\theta_n} u_j(\hat{\theta}_1, ..., \hat{\theta}_i, \theta_{i+1}, ..., \theta_n)$. After agent i has submitted report $\hat{\theta}_i$, 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:[2]

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

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

caused by report $\hat{\theta}_i$. The net monetary transfer to agent i is

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

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

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

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.

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

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. 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: if agents $1, 3$ have types L each, 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 agents $1, 3$ can both report H: they will have the park and none of them will have to pay any monetary transfers.[3] Similarly, 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 agents report their types 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 $\hat{\theta}$:

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

$(\hat{\theta}_1, \hat{\theta}_3) \backslash \hat{\theta}_2$	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 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 $\hat{\theta}_3 = H$ if $\hat{\theta}_2 = H$.[4] 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.

[4] Agent 3's report does not affect the social choice if $\hat{\theta}_2 = L$.

REMARK 1 *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:* $E_{\theta_i} x_{ij}(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i) = 0$*.*

Indeed, 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:

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

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

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. *Q.E.D.*

In the mechanism each agent prefers to report truthfully, given any previous reports and anticipating future reports to be truthful. Moreover, truthful reporting is a uniquely rationalizable strategy under the assumption that any individual misreport by an agent decreases the total payoff.

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—in particular, it requires the cardinality of the set of social choices to be at least as large as the cardinality of any set Θ_i—it is satisfied in certain environments. For example, in a Hotelling model with agents having single-peaked preferences and identical quadratic losses, the efficient location is the average of agents' preferences; any misreport will cause inefficiency.

I use the concept of interim sequential rationalizability from Penta (2012); which, in the environment of independent private values and each player reporting once, is as follows. Any agent i, after observing previous reports $\hat{\theta}_1, ..., \hat{\theta}_{i-1}$, has set M_i of conjectures $\mu_i[\hat{\theta}_1, ..., \hat{\theta}_{i-1}]$, each conjecture being a probability distribution over pure strategies of players $j > i$, $\sigma_j(\theta_j, \hat{\theta}_1, ...\hat{\theta}_{j-1}) : \Theta_j \times \Pi_{k=1}^{j-1}\Theta_k \to \Theta_j$. A pure strategy $\sigma_i(\theta_i, \hat{\theta}_1, ...\hat{\theta}_{i-1})$ is a sequential best response for player i of type θ_i to conjecture μ_i, if it maximizes the expected payoff of player i with respect to μ_i. Interim sequential rationalizability (ISR) consists of sequential deletion procedure for each type of each player: the set of pure strategies for player i that survive the k-th step of iterated deletion, is denoted as ISR_i^k, with $ISR^k \equiv \{ISR_i^k\}_{i=1}^n$. The procedure of iterated deletion works as follows: ISR^0 consists of all pure strategies for all players, while for each $k > 0$, ISR_i^k consists of all pure strategies of player i each of which is a best response to some belief μ_i, that has a support over pure strategies of players $j > i$ in ISR^{k-1}. Denote $ISR \equiv \bigcap_{k\geq 0} ISR_k$ as the set of interim sequentially rationalizable strategies.

Proposition 2 *In the sequential mechanism with transfers given by Definition 4, under Assumption 1, ISR consists of a unique strategy of truthful reporting.*[5]

Proof.

[5]With Assumption 1, the statement of Proposition 2 also holds for the sequential mechanism in Crémer and Riordan (1985).

Let's number the agents in the order in which they submit their reports as $1, ..., n$. The amount of the transfers paid to agent n 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. By Assumption 1, n's utility is *uniquely maximized* at the truthful report $\hat{\theta}_n = \theta_n$.

Now let's look at agent $n-1$. Since agent n reports truthfully, agent $n-1$ expects to pay zero to agent n, regardless of report $\hat{\theta}_{n-1}$. Thus, report $\hat{\theta}_{n-1}$ affects only the transfers made to agent $n-1$ at the stage when he reports. 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 agents. By Assumption 1, 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 sequentially rationalizable strategy as well. *Q.E.D.*

Note that one needs the assumption of independent private values for Proposition 2: private values allow the agents not to care about others' types and whether the reports of previous agents were truthful or not, while independence allows the agents' expected payoffs to be independent from future truthful reports.

4 Properties of the mechanism

In this section I show the main property of the mechanism, referred later 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: $\sigma_i(\hat{\theta}_1, ..., \hat{\theta}_{i-1}, \theta_i) \equiv \theta_i$ guarantees him his ex ante efficient payoff, in expectation over type θ_i, regardless of others' strategies:*
$\forall \hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_{i+1}(\hat{\theta}_i), ..., \hat{\theta}_n(\hat{\theta}_i)$, $E_{\theta_i} u_i(\theta_i, s^*(\theta_i, \hat{\theta}_{-i})) + E_{\theta_i} x_i(\theta_i, \hat{\theta}_{-i}) = E_\theta u_i(\theta)$.

Note that Theorem 1 allows all agents except for i to coordinate their reports, and agents who report after i can condition on his report $\hat{\theta}_i$. In other words, even if all agents but i unite together, they cannot affect i's expected utility.

Proof.

After agents $1, ..., j$ have submitted their reports, 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[6] 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 agent $m \neq i$ submits his report $\hat{\theta}_m$, 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 change in his payoff, and thus, their reports (joint or not) cannot affect i's total current utility. When i submits his report, due to Lemma 2, each other agent $j \neq i$ pays zero in expectation to i; plus, the change in i's current expected

[6] That is, a sum of transfers to/from agent i as determined by the reports of agents $1, ..., j$.

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: $E_{\theta_i}(u_i(\theta_i, s^*(\theta_i, \hat{\theta}_{-i})) + x_i(\theta_i, \hat{\theta}_{-i})) = E_\theta u_i(\theta)$. *Q.E.D.*

The assumption of independent private values is essential for Theorem 1. With interdependent 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 non-private 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.

There exist efficient mechanisms with alternative transfers that satisfy the statement of Theorem 1, for example with transfers given below:

DEFINITION 5 *Given the total submitted report to be $\hat{\theta}$, agent j pays agent i the change in expectation of j's payoff:*

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

caused by report $\hat{\theta}_i$.

In other words, the new transfers are calculated as if the social planner has ignored j's report. Using the same technique, one can show that

PROPOSITION 3 *The conclusion of Theorem 1 holds with transfers given by Definition 5.*

4.1 Full efficient implementation and 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. Let's define a joint strategy of agents in C as the collection of strategies of its members $\sigma^C = \{\sigma_i^C\}_{i \in C}$, with $\sigma_i^C : \prod_{j<i} \Theta_j \times \prod_{k \in C} \Theta_k \to \Theta_i$. That is, the agents in C can coordinate their reports, and each agent i in C can condition his report on all previous reports of agents $j < i$, and on the entire type profile in 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, any joint strategy of agents in C will give them weakly less total expected utility as compared to truthful reporting, assuming all other agents outside C report truthfully.*

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 private types of its members are a common knowledge within C; and there is no threat of a subcoalitional deviation. Note that the coalition decides on its strategy after learning its type θ_C, and estimates the total utility before the mechanism. 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 private types and monetary transfers within the coalition.

THEOREM 2 *The efficient mechanism with transfers given by Definition 4 is coalition-proof.*[7]

[7]One should note that even in case of agents partitioned into several coalitions, each of which

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. 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. [8]

Q.E.D.

Definition 6 can be interpreted as a coalition being formed before the mechanism is announced: the members of a coalition act as a single player. An alternative concept of coalition formation, is related to the property of *strong collusion proofness* (intro-

maximizing the sum of its members' utilities, any equilibrium among those coalitions is efficient. Indeed, Theorem 1 guarantees any agent (coalition) ex ante efficient payoff.

[8] 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) decrease X in the total payoff, and the expected utility of C will decrease by X.

duced by Laffont and Martimort (1997)).[9] There is an *explicit process* of coalition formation: a third party proposes a side contract to a subgroup of agents, and if they all agree, they submit a joint report to the original mechanism, as controlled by the third party. A mechanism is strongly collusion-proof, if in *any* weak Perfect Bayesian equilibrium the outcome is efficient. This property does not follow directly from Definition 6: even if no group of agents benefits from any deviation, they may still be stuck in a bad equilibrium, - if an agent refuses to join a coalition, others may unfavorably update their beliefs about his type.

With an explicit process of coalition formation, a third party proposes the following side contract to a subgroup $C \subset \{1, ..., n\}$ of agents:

- Each agent i in C simultaneously reports his private type to the third party. Let's denote this report as θ'_i, with the total report profile as θ'_C;
- The third party randomizes the joint report $\hat{\theta}_C$ of agents in C to the social planner, according to a probability measure $\nu(\theta'_C)$;[10]
- Any agent $i \in C$ receives a side monetary transfer $y_i(\theta'_C)$, in addition to the transfer $x_i(\hat{\theta})$ from the original mechanism. Side transfers are weakly budget-balanced: $\sum_{i \in C} y_i(\theta'_C) \leq 0$.

The timing is as follows. Each agent learns his private type, and the social planner announces the mechanism. Then the third party offers a side contract $\nu(\theta'_C)$, $\{y_i(\theta'_C)\}_{i \in C}$ to the agents in C. The agents in C simultaneously state their decisions on whether to join the coalition. This profile of acceptance decisions is observable to all the agents in C. If all the agents in C accept the side contract, they simultaneously

[9] In the current paper, agents do not get to choose whether to participate in the mechanism, unlike in Laffont and Martimort (1997).

[10] Since the agents report sequentially in the original mechanism, the third party can condition the report $\hat{\theta}_i$ of any collusive agent $i \in C$ on the previous reports.

report their private types to the third party, which enforces the implementation of the side contract. Otherwise the coalition is not formed and the agents proceed to participate in the original mechanism individually.[11]It is assumed that the non-collusive agents do not observe whether the coalition is formed or not, however, in equilibrium they predict correctly the report distribution of the agents in C.[12]

The strategy of any non-collusive agent $j \notin C$ σ_j: $\Pi_{k=1}^{j-1}\Theta_k \times \Theta_j \longrightarrow \Delta(\Theta_j)$ determines his report $\hat{\theta}_j$ as dependent on his true type θ_j and the reports of agents before him $\hat{\theta}_1, ..., \hat{\theta}_{j-1}$. The strategy of agent $i \in C$ is described by a) his vote decision $D_i(\theta_i) \in \Delta(\{0,1\})$ as to whether to agree to join the coalition, with realized *vote* $d_i = 1$ ($d_i = 0$) meaning that i agrees (does not agree); b) report $\theta_i'(\theta_i)$ to submit to the third party if the coalition is formed; and c) report $\hat{\theta}_i(\{d_l\}_{l\in C} \neq 1^{|C|}, \hat{\theta}_1, ...\hat{\theta}_{i-1}, \theta_i)$ to submit to the social planner if the coalition is not formed, as dependent on votes to join the coalition. I consider the concept of *weak Perfect Bayesian Nash equilibrium*. In equilibrium each non-collusive agent j observes previous reports $\{\hat{\theta}_l\}_{l<j}$, and has *correct* beliefs about the (joint) distribution of future reports. Each collusive agent i updates his beliefs about types of the other agents in C, based on their votes, $\{d_l\}_{i\neq l\in C}$, this update is Bayesian whenever possible and arbitrary otherwise. For all agents that report after him, agent i keeps ex ante beliefs about types of non-collusive agents, and updated beliefs for collusive agents. All the agents submit their reports (to the third party or to the planner) to maximize their expected utility each, with respect to their beliefs.

DEFINITION 7 *An efficient mechanism satisfies strong collusion-proofness, if for any subset $C \in \{1, ..., n\}$, and any side contract $\nu(\theta_C')$, $\{y_i(\theta_C')\}_{i\in C}$, in any weak Perfect*

[11]The agents in C may update their beliefs about each other's type from the fact of rejection to form the coalition, and adjust their reports accordingly.

[12]For example, in McAfee, McMillan (1992) non-collusive bidders choose best responds to bidding strategies of collusive bidders, while Che and Kim (2006) make non-collusive agents incentivized to participate and report truthfully.

Bayesian Nash equilibrium, the social outcome is efficient.

THEOREM 3 *The efficient mechanism with transfers given by Definition 4 satisfies strong collusion-proofness.*

Proof.

Consider any weak Perfect Bayesian equilibrium, in which, given type profile θ, the (random) report to the social planner is $\hat{\theta}(\theta)$, and each agent k gets expected net monetary transfer $m_k(\theta)$ (from the mechanism and potentially side contract). Let's show that any agent k gets at least his ex ante efficient payoff in expectation over θ_k:

$$E_{\theta_k}(u_k(\theta_k, s^*(\hat{\theta}(\theta))) + m_k(\theta)) \geq E_\theta u_k(\theta) \tag{1}$$

If any non-collusive agent $j \notin C$ reports truthfully, then, by Theorem 1, j gets $E_\theta u_j(\theta)$ as total utility, in expectation over θ_j. For each report $\hat{\theta}_1, ..., \hat{\theta}_{j-1}$, agent j knows the joint distribution of reports $\mu_j[\hat{\theta}_1, ..., \hat{\theta}_{j-1}, \hat{\theta}_j](\{\hat{\theta}_m\}_{m>j})$[13] of agents who report after him, as dependent on his report $\hat{\theta}_j$. The equilibrium report of j, $\hat{\theta}_j(\theta_j, \hat{\theta}_1, ..., \hat{\theta}_{j-1})$ gives j higher utility as compared to reporting truthfully, in expectation over μ_j: $E_{\mu_j}(u_j(\theta_j, s^*(\hat{\theta}_{-j}, \hat{\theta}_j)) + x_j(\hat{\theta}_{-j}, \hat{\theta}_j)) \geq E_{\mu_j}(u_j(\theta_j, s^*(\hat{\theta}_{-j}, \theta_j) + x_j(\hat{\theta}_{-j}, \theta_j))$. Taking expectation over the joint distribution of $\hat{\theta}_1, ..., \hat{\theta}_j$, and over θ_j, one gets (1) for $k = j$.

For a collusive agent $i \in C$, if i refuses to join a coalition and reports truthfully, the statement of Theorem 1 holds, and i gets $E_\theta u_i(\theta)$ as total utility, in expectation over θ_i. After i refuses to join a coalition, and before agents start reporting, i has beliefs about the joint distribution of reports before him, $\widetilde{\mu}_i[\{d_l\}_{i \neq l \in C}](\hat{\theta}_1, ..., \hat{\theta}_{i-1})$, and about the joint distribution of reports after him, $\mu_i[[\{d_l\}_{i \neq l \in C}], \hat{\theta}_1, ..., \hat{\theta}_{i-1}, \hat{\theta}_i](\{\hat{\theta}_m\}_{m>i})$, both dependent on the profile of votes $\{d_l\}_{i \neq l \in C}$ to join a coalition.[14] After i refuses to join a coalition, and once it is his turn to report, i's equilibrium report, $\hat{\theta}_i(\{d_l\}_{l \in C} \neq$

[13]The notation $\mu[x](y)$ means the distribution of y, conditional on x.

[14]Those beliefs may not coincide with actual equilibrium behavior, if there is an off-path vote, including by i.

$1^{|C|}, \hat{\theta}_1, ...\hat{\theta}_{i-1}, \theta_i)$ gives i higher utility as compared to reporting truthfully, over μ_i: $E_{\mu_i}(u_i(\theta_i, s^*(\hat{\theta}_{-i}, \hat{\theta}_i)) + x_i(\hat{\theta}_{-i}, \hat{\theta}_i)) \geq E_{\mu_i}(u_i(\theta_i, s^*(\hat{\theta}_{-i}, \theta_i)) + x_i((\hat{\theta}_{-i}, \theta_i))$. Respectively, at the moment when i decides on joining a coalition, i *believes* to get at least $E_{\tilde{\mu}_i} E_{\mu_i}(u_i(\theta_i, s^*((\hat{\theta}_{-i}, \theta_i)) + x_i((\hat{\theta}_{-i}, \theta_i))$ if he refuses; which puts a lower bound on i's interim utility in equilibrium. Taking expectation over θ_i, one gets (1) for $k = i$.

Since (1) holds for any agent k, the total ex ante utility of all the agents satisfies $E_\theta \sum_{k=1}^n (u_k(\theta_k, s^*(\hat{\theta}(\theta))) + m_k(\theta) \geq \sum_{k=1}^n E_\theta u_k(\theta)$. The total monetary transfer to all agents $\sum_{k=1}^n m_k(\theta)$ is non-positive due to Lemma 1 and $\sum_{i \in C} y_i(\theta'_C) \leq 0$. Therefore, the ex ante total *payoff* of all the agents $E_\theta \sum_{k=1}^n u_k(\theta_k, s^*(\hat{\theta}(\theta)))$ weakly exceeds its ex ante efficient value of $\sum_k E_\theta u_k(\theta)$, and the efficient social outcome is achieved with probability 1 in equilibrium. *Q.E.D.*

Remark 2 *Note that Theorem 3 guarantees the standard property of full efficient implementation, if the third party is absent. In addition, the property that all equilibria are efficient, will be obviously satisfied if one considers more stringent notions of equilibria - for example, with restrictions on off-path belief formations.*

5 Direct symmetric mechanism

The ordering of the agents' sequential 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 4 *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:

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

Taking expectation over truthful reporting of types $\theta_1, ..., \theta_{i-1}$, 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.*

If agents submit their reports simultaneously, it is possible to make the transfer scheme symmetric. The direct mechanism will still satisfy the properties of the sequential mechanism:

Corollary 1 *The statements of Theorems 1, 2, 3 hold for each of the two direct symmetric efficient mechanisms: Take transfers given by Definition 4 (or Definition 5) for each ordering of the agents, and take the average across all orderings.*

With the first scheme from Corollary 1, agent i pays the externality that other agents' reports impose on his payoff; and gets paid the Shapley value of marginal externalities that his report imposes on the total payoff. In other words, each player gets paid the marginal externality that his report imposes on the total payoff, given that the report of a random subcoalition has been revealed before. This payment scheme has similarities with the coalitional games and Shapley value (Shapley (1953)), in the latter each agent gets his marginal contribution to a total surplus generated by a random subcoalition.

REMARK 3 *The social planner may be interested in alternative transfer schemes, for example, agents may have budget constraints. One can get other budget-balanced transfer schemes that satisfy the statements of Theorems 1, 2, 3, as follows. Take any measurable payoff functions $v_i(\theta)$, estimate the two transfer schemes from Corollary 1 for v instead of u, take the difference between the two obtained schemes, and add this difference to the first transfer scheme from Corollary 1 estimated for the original utilities $u_i(\theta)$.*

6 Individual rationality

The current mechanism is constructed under the assumption that no agent can quit the mechanism. 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. However, due to Myerson–Satterwaite impossibility theorem, if the agents could decide on participation after learning private types, [15] they might choose to quit.

Despite the general impossibility of achieving interim individual rationality, a different question may be asked: if there is an efficient mechanism M which satisfies interim incentive compatibility and individual rationality, and ex post budget balance, then, under certain conditions, there exists an alternative mechanism that satisfies all the above properties, and coalition-proofness and full implementation. More precisely, I consider the setting by Krishna and Perry (2000). The set of possible social choices S is finite and has k elements. The agents $i \in \{1, ..., n\}$ have independently distributed private types, with each type θ_i having a continuous density function, with full support on a compact and convex subset Θ_i of Euclidean space R^k.

PROPOSITION 5 *There exists a budget-balanced mechanism which satisfies the state-*

[15] I assume all agents decide on participation in the mechanism before getting an offer to join a coalition (if any), and before the first report in the case of sequential reporting.

ments of Theorems 1, 2, 3, and interim individual rationality, in each of the following two cases:

a) there exists an efficient mechanism M in the setting of Krishna and Perry (2000);

b) The AGV mechanism satisfies interim individual rationality.

7 Conclusion

This paper provides an efficient mechanism with agents having independent private types 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 mechanism requires the private type distribution to be common knowledge, which may seem to be too strong an assumption, as related to the Wilson doctrine. However, it may be plausible for each agent to willingly report his type distribution to the social planner. With this information the social planner can remove the ex ante externalities imposed on the agent and guarantee that the agent will get his efficient payoff. This guarantee would be attractive to any risk- or ambiguity-averse agent, or any agent who is struggling to form beliefs about others' strategies, and would incentivize him to truthfully report his type distribution to the social planner. The possibility of agents willingly reporting their type distribution could be an interesting direction for future work.

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..

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.

D'Aspremont, C., and L. Gerard-Varet. (1979): "Incentives and Incomplete Information," *Journal of Public Economics*, 11, 25–45.

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 (2015): "Full Implementation and Belief Restrictions," under revision in *American Economic Review*.

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.

Postlewaite, A. and D. Schmeidler (1986): "Implementation in Differential Information Economies," *Journal of Economic Theory*, 39, 14–33.

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.