Crawford, V. (1979). A procedure for generating pareto efficient egalitarian-equivalent allocations. *Econometrica*, *47*, 49–60.

Groves, T. (1976). Information, incentives, and the internalization of production externalities. In Lin, S. (Ed.), *Theory and Measurement of Economic Externalities*. Academic Press, New York.

Groves, T. (1979). Efficient collective choice when compensation is possible. *Review of Economic Studies*, *46*, 227–241.

Groves, T., and Ledyard, J. (1977). Optimal allocations of public goods: A solution to the 'free rider problem'. *Econometrica*, *45*, 783–809.

Groves, T., and Ledyard, J. (1987). Incentive compatibility since 1972. In Groves, T., Radner, R., and Reiter, S. (Eds.), *Information, Incentives and Economic Mechanisms*. University of Minnesota Press, Minneapolis.

Guttman, J. (1978). Understanding collective action: Matching behavior. *American Economic Review*, *68*, 251–255.

Hurwicz, L. (1979). Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points. *Review of Economic Studies*, *46*, 217–225.

Jackson, M., and Moulin, H. (1992). Implementing a public project and distributing its cost. *Journal of Economic Theory*, *xx*, xx.

Maskin, E. (1985). The theory of implementation in Nash equilibrium. In Hurwicz, L., Schmeidler, D., and Sonnenschein, H. (Eds.), *Social Goals and Social Organization*, pp. 173–204. Cambridge Univeristy Press, Cambridge.

Moore, J. (1991). Implementation in environments with complete information. Tech. rep., Suntory-Toyota International Centre for Economics, London School of Economics.

Moore, J., and Repullo, R. (1988). Subgame perfect implementation. *Econometrica*, *56*, 1191–1220.

Moulin, H. (1979). Dominance solvable voting schemes. *Econometrica*, *47*, 1337–1351.

Moulin, H. (1981). Implementing just and efficient decision making. *Journal of Public Economics*, *16*, 193–213.

Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan, London.

Varian, H. (1989). Sequential provision of public goods. Tech. rep., CREST Working Paper.

Walker, M. (1981). A simple incentive compatible scheme for attaining Lindahl allocations. *Econometrica*, *48*, 56–73.

Announcement stage. Each agent i announces $2(n-1)$ real numbers (p^i_{ij}, p^i_{ji}) for $j \neq i$.

Choice stage. Agent i chooses x_i according to the utility maximization problem

$$\max_{x_i, y_i} u_i(x, y_i)$$
$$\text{such that } x_i \sum_{j \neq i} p^j_{ji} + y_i = w_i + \sum_{j \neq i} p^j_{ij} x_j - \| p^i_{ji} - p^j_{ji} \|.$$

As before, we will make the Local Invertibility Assumption that agent i can set a price p^i_{ij} that will induce agent j to make any particular choice x_j. Note that the prices and the choices of agent i may be vectors. The proof that all subgame perfect equilibria are efficient is similar to that of the two-agent case and is omitted.

13. Summary

We have exhibited a general mechanism to achieve efficient allocations in economic environments. Essentially, each agent sets a price for each other agent's choice and for his own choice. Each agent is penalized if the price he sets for his own choice is different from the prices that other agents set for his choice. All subgame perfect equilibria of this game are Pareto efficient allocations. In the case of public goods, they are Lindahl allocations and the equilibrium prices are, essentially, Lindahl prices. In the case of private goods, the prices are competitive prices.

References

Andreoni, J., and Varian, H. (1992). Pre-play contracts in the prisoners' dilemma. Tech. rep., CREST Working Paper.

Arrow, K. (1970). The organization of economic activity: Issues pertinent to the choice of market versus non-market allocation. In Haveman, R. H., and Margolis, J. (Eds.), *Public Expenditures and Policy Analysis*, pp. 59–73. Markham.

Bagnoli, M., and Lipman, B. (1989). Provision of public goods: fully implementing the core through private contributions. *Review of Economic Studies*, *56*, 583–602.

Bergstrom, T., Blume, L., and Varian, H. (1986). On the private provision of public goods. *Journal of Public Economics*, *29*, 25–49.

Clarke, E. (1971). Multipart pricing of public goods. *Public Choice*, *11*, 17–33.

Coase, R. (1960). The problem of social cost. *Journal of Law and Economics*, *3*, 1–44.

Walker (1981) mechanisms implement Lindahl allocations. In the Hurwicz mechanism, each agent proposes an amount of the public good and a Lindahl price; agents pay a quadratic penalty if they announce different levels of the public good. Walker's mechanism avoids such penalty terms. Groves (1979) and Groves and Ledyard (1987) provide a nice survey of these results.

Turning to the more recent literature on simple mechanism, Bagnoli and Lipman (1989) and Jackson and Moulin (1992) examine the special case of a discrete public good with quasilinear utility. The Bagnoli-Lipman mechanism is very simple: each agent offers a voluntary contribution. If the sum of the contributions covers the cost of the public good, it is produced, otherwise the contributions are returned. This mechanism implements the core of the public goods game in undominated perfect equilibria.

The Jackson-Moulin mechanism implements an efficient allocation using undominated Nash equilibria.[3] Their mechanism is reasonably simple and works with a broad family of cost-sharing rules. However, it appears that this mechanism only works for the special case of indivisible public goods and quasilinear utility.

Varian (1989) describes some mechanisms for the public goods problem that are closely related to the compensation mechanism. In the case of two agents with quasilinear utility there is a very simple mechanism that achieves a Lindahl allocation: in the first stage each agent offers to subsidize the contributions of the other agent. In the second stage, each agent makes a voluntary contribution and collects the promised subsidies from the other agent.[4] Varian (1989) also describes some other variations on the compensation mechansim for public goods problems involving many agents and general utility functions.

12. Many agents

Here we describe the generalization of the mechanism to n agents. This is a direct but obvious extension of the previous two-agent case, so we only state the results. Let x_i be the choice of agent i, and let $x = (x_1, \ldots, x_n)$. As usual there are two stages.

[3] They also describe a variation using subgame perfect equilibrium.

[4] This mechanism is related to the mechanism of Guttman (1978), which involves offering to match contributions.

Announcement stage. Agent 1 names p_2^1, how much he is willing to pay agent 2 to cooperate, and agent 2 names p_1^2, how much agent 2 is willing to pay to agent 1 to cooperate.

Choice stage. Each agent choose whether to cooperate or not, receiving payoffs

$$\Pi_1 = u_1(x_1, x_2) - p_2^1 x_2 + p_1^2 x_1$$

$$\Pi_2 = u_2(x_1, x_2) - p_1^2 x_1 + p_2^1 x_2.$$

Note that instead of payments being made by "the regulator" the agents now make the transfers themselves. The prices each agent faces are no longer set by the *other* agent. Despite this, it can be shown that the unique subgame perfect equilibrium involves the efficient outcome. The equilibrium prices are $p_2^1 = 2$ and $p_1^2 = 1$.

It has long been known that the ability to make binding pre-play commitments allows for a solution to the prisoner's dilemma. What is interesting about this example is how simple the first-stage commitments need to be to support efficient outcomes. Andreoni and Varian (1992) have tested this mechanism in the laboratory and find that agents quickly converge to the predicted outcome.

11. Literature review

There is a vast literature on mechanism design that is concerned with how to implement various social decisions. Much of this literature is concerned with whether a particular outcome can be implemented by a decentralized game. Our concern is not so much with the *existence* of a mechanism, but rather finding a suitably simple mechanism. Most of the attempts to find "simple" solutions to externalities problems have been concerned with the case of public goods, so we provide a very brief review of that literature insofar as it relates to the work described here. Moore (1991) provides a thorough review of the recent literature.

The well-known demand revealing mechanism of Clarke (1971) and Groves (1976) implements the efficient amount of a public good via a dominant strategy equilibrium. However, this mechanism only works with quasilinear utility, and is not balanced, even in equilibrium.

The Groves and Ledyard (1977) quadratic mechanism yields efficient Nash equilibria for the public goods problem, but they are not Lindahl allocations. The Hurwicz (1979) and

Prisoners' dilemma

Consider the following asymmetric prisoners' dilemma:

		Column	
		Cooperate	Defect
Row	Cooperate	$5, 5$	$2, 6$
	Defect	$7, 1$	$3, 3$

How can we induce the Pareto efficient outcome? Again, the compensation mechanism does the trick. Let x_i be the strategy choice of agent i; $x_i = 1$ if agent i cooperates, and $x_i = 0$ if agent i defects.

Announcement stage. Agent 1 names p_{12}^1, how much agent 1 should be paid if he cooperates, and p_{21}^1, how much agent 2 should be paid if he cooperates. Similarly agent 2 names p_{21}^2 and p_{12}^2.

Choice stage. Each agent choose whether to cooperate or defect and receive payoffs

$$\Pi_1 = u_1(x_1, x_2) + p_{21}^2 x_1 - p_{12}^2 x_2 - \|p_{21}^1 - p_{21}^2\|$$

$$\Pi_2 = u_2(x_1, x_2) + p_{12}^1 x_2 - p_{21}^1 x_1 - \|p_{12}^2 - p_{12}^1\|$$

Note the sign change: since there is now a positive externality between the two agents, it is natural to subsidize good behavior rather than penalize bad behavior. Using the by now standard argument, it can be shown that the unique subgame perfect equilibrium is for both players to cooperate. The supporting prices satisfy the conditions

$$4 \geq p_{21}^1 = p_{21}^2 \geq 2$$

$$2 \geq p_{12}^1 = p_{21}^2 \geq 1.$$

We could also formulate this problem so as to have each agent announce how the other agent should be fined if he defects. The outcome is essentially the same, but the payoffs are slightly different.

Either form of the compensation mechanism produces an efficient outcome in this game—or in any game, for that matter. However, the prisoners' dilemma has a special structure. It turns out that a related, but simpler, mechanism is available for this special case.

Regulation of duopoly

There are now three agents: the consumer (indexed by 0) and two firms. Firm 1 produces x_1 at cost $c_1(x_1)$, firm 2 produces x_2 at cost $c_2(x_2)$ and the consumer has utility function $u(x_1, x_2) + y_0$. The standard compensation mechanism involves payoffs of the form

$$\Pi_0 = u(x_1, x_2) - p_{01}^1 x_1 - p_{02}^2 x_2$$
$$\Pi_1 = p_{01}^0 x_1 - c_1(x_1) - \|p_{01}^1 - p_{01}^0\|$$
$$\Pi_2 = p_{02}^0 x_2 - c_2(x_2) - \|p_{02}^2 - p_{02}^0\|$$

This is a natural generalization of the mechanism described in the last section. The consumer is setting the price that the firms face, and the firms are setting the price the consumer faces.

However, in the case of duopoly it is natural to think that the firms may know more about each other's technology than the consumer knows. Hence it makes sense for each firm reports the price that the other firm will face. This yields payoffs of the form

$$\Pi_0 = u(x_1, x_2) - p_{01}^2 x_1 - p_{02}^1 x_2$$
$$\Pi_1 = p_{01}^2 x_1 - c_1(x_1)$$
$$\Pi_2 = p_{02}^1 x_2 - c_2(x_2).$$

Note that the consumer chooses both x_1 and x_2; the firms simply choose the price that the consumer faces.

The arguments given earlier show that the competitive allocation is the unique equilibrium of this game. But it is useful to think about this case more directly. Suppose that the two products are substitutes and that each firm has set the price that the other firm faces to be the competitive price. Why wouldn't firm 1, say want, to raise the price that firm 2 faces? If firm 1 raised the price facing firm 2, then the consumer would demand more output from firm 1, which it would be forced to supply. But since the price that firm 1 faces equals its marginal cost this reduces firm 1's profit.

game so that x_1 and x_2 are both positive. Then since x_1 and x_2 are perfect substitutes in consumption, they must have the same price—otherwise we couldn't have an equilibrium.

By inspection of the budget constraint it follows that $(1-p_{21}^2) = p_{12}^2$. Hence the budget constraint facing agent 1 can, in equilibrium, be written as

$$p_{12}^2[x_1 + x_2] + y_1 = w_1.$$

It follows that an equilibrium value of p_{12}^2 is simply the Lindahl price of the public good for agent 1, and the equilibrium allocation is simply a Lindahl allocation. Hence, our mechanism gives a way to decentralize Lindahl allocations: each agent has an incentive to reveal the prices necessary to support the Lindahl allocation.

Regulation of monopoly

Agent 0 is a consumer who consumes and x-good and a y-good and has a quasilinear utility function $u(x)+y_0$. Agent 1 is a firm that can produce the x good at cost $c(x)$; it's objective function is $y_1 - c(x)$. How can we induce the firm to produce the socially optimal output?

If we are only interested in efficiency, this is not terribly difficult: simply have one of the agents dictate a production level and a transfer. Under our full information assumption we will get an efficient amount of x regardless of which agent chooses it; only the transfer will be different.

However, if we want to get a *particular* efficient allocation—the competitive outcome—it is not so obvious how to proceed. However, the compensation mechanism solves the problem quite readily.

Announcement stage. The consumer announces how much he values the good, p_0^0, and the producer announces how much the consumer values the good, p_0^1.

Choice stage. The producer chooses x and the payoffs are

$$\Pi_0 = u(x) - p_0^1 x$$

$$\Pi_1 = p_0^0 x - c(x) - \|p_0^0 - p_0^1\|$$

Applying the standard argument shows that in equilibrium

$$p_0^1 = p_0^0 = u'(x) = c'(x).$$

These are the conditions that characterize the competitive allocation.

in the absence of any transfer mechanism, agent 1's maximization problem takes the form

$$\max_{x_1, y_1} u_1(x_1 + x_2, y_1)$$
$$\text{such that } x_1 + y_1 = w_1$$
$$x_1 \geq 0.$$

The equilibrium of this contribution game has been studied extensively by Bergstrom, Blume, and Varian (1986). In particular, they emphasize the important role of the nonnegativity constraint in examining the comparative statics of the model. The nonnegativity constraint is very natural in a model of voluntary contributions: you may choose to contribute more to a public good, but one is typically not able to make a *negative* contribution.

If we use a compensation mechanism for solving the public goods problem, agent 1 will face the maximization problem

$$\max u_1(x_1 + x_2, y_1)$$
$$\text{such that } -p_{21}^2 x_1 + y_1 = w_1 - x_1 - p_{12}^2 x_2 - \|p_{21}^1 - p_{21}^2\|$$
$$x_1 \geq 0.$$

Note the sign change of the prices. This is more natural in the case of positive externalities since each agent wants to reduce the cost to the other agent of making contributions. Rearranging the budget constraint, we can write

$$(1 - p_{21}^2)x_1 + y_1 = w_1 - p_{12}^2 x_2 - \|p_{21}^1 - p_{21}^2\|.$$

Agent 1's contributions are subsidized at a rate p_{12}^2 which is chosen by agent 2. This subsidy is "recovered" by a tax on agent 2. In the compensation mechanism the taxes and subsidies that each agent are chosen by the other agent. See Varian (1989) for a variation on this mechanism where the agents set some of these rates for themselves.

Since public goods are just a special kind of externality, the proof of efficiency given earlier still applies. Note that the non-calculus proof is the appropriate version here, due to the presence of the nonnegativity condition.

However, given the special form of the public goods externality, we can say a bit more about the equilibrium prices. Suppose that we have an interior solution to the public goods

9. Individual rationality

In equilibrium, each agent does at least as well using the compensation mechanism as he does by simply consuming his endowment. This is true since the mechanism implements a Lindahl equilibrium. However, this is only true in equilibrium; if some agent makes a mistake, other agents may suffer. We may well want to require that agents achieve at least their endowment level of utility *regardless* of the choices made by the other agents.

This is easy to do by adding another stage to the compensation mechanism. After the final allocation has been determined by the mechanism, simply offer each agent his choice between the proposed allocation and his endowment. The proposed allocation is implemented if and only if all agents agree that they prefer it to their endowments.

Since the compensation mechanism implements a Lindahl allocation, each agent will be at least as well off at the equilibrium of the mechanism than at his endowment. Hence, if equilibrium is achieved everyone will (weakly) prefer that to their endowment. If equilibrium fails to be achieved, for some reason, each agent will receive his endowment level of utility.

One could also imagine variations on this unanimity rule: if some agent objects to the allocation, he is excluded, and the rest of the agents play the mechanism again. The Lindahl allocation is still an equilibrium for this mechanism, but there may be others. I have not explored the structure of equilibrium when such unanimity rules are added.

10. Examples of the compensation mechanism

We have described the general form of the compensation mechanism; here we illustrate how it works in some specific cases.

Pure public goods

The special case of a pure public good is of some interest, since it is a well-known and much studied example of a particular type of externality. Think of the public good as "money" to be spent on some public project and let x_1 and x_2 be the two agents' monetary contributions to the public good $x_1 + x_2$. Let y_i be agent i's private consumption. Then, Differentiating agent 1's objective function with respect to his choice variables we have 3 first-order conditions:

$$\frac{\partial u_1(x)}{\partial x_1} - \left(p_{31}^2 + p_{21}^3\right) = 0$$

$$\left[\frac{\partial u_1(x)}{\partial x_1} - (p_{31}^2 + p_{21}^3)\right] \frac{\partial x_1}{\partial p_{32}^1} + \left[\frac{\partial u_1(x)}{\partial x_2} + p_{12}^3\right] \frac{\partial x_2}{\partial p_{32}^1} + \left[\frac{\partial u_1(x)}{\partial x_3} + p_{13}^2\right] \frac{\partial x_3}{\partial p_{32}^1} = 0$$

$$\left[\frac{\partial u_1(x)}{\partial x_1} - (p_{31}^2 + p_{21}^3)\right] \frac{\partial x_1}{\partial p_{23}^1} + \left[\frac{\partial u_1(x)}{\partial x_2} + p_{12}^3\right] \frac{\partial x_2}{\partial p_{23}^1} + \left[\frac{\partial y_1(x)}{\partial x_3} + p_{13}^2\right] \frac{\partial x_3}{\partial p_{23}^1} = 0$$

There are similar sets of first-order conditions for the other agents. It is relatively easy to see that an efficient allocation, with each agent reporting the truth, is an equilibrium: in this case the first-order conditions are all satisfied. Given appropriate invertibility assumptions it can also be shown that this is the unique equilibrium. However, it is much easier to see that the equilibrium is efficient by following a somewhat different argument.

As before we assume that each agent can influence the other agents' choices by setting prices. Since each agent can effectively control the entire vector of choices, optimization implies

$$u_1(x^*) - (p_{31}^2 + p_{21}^3)x_1^* + p_{12}^3 x_2^* + p_{13}^2 x_3^* \geq u_1(x) - (p_{31}^2 + p_{21}^3)x_1 + p_{12}^3 x_2 + p_{13}^2 x_3$$

$$u_2(x^*) - (p_{32}^1 + p_{12}^3)x_2^* + p_{21}^3 x_1^* + p_{23}^1 x_3^* \geq u_2(x) - (p_{32}^1 + p_{12}^3)x_2 + p_{21}^3 x_1 + p_{23}^1 x_3$$

$$u_3(x^*) - (p_{23}^1 + p_{13}^2)x_3^* + p_{31}^2 x_1^* + p_{32}^1 x_2^* \geq u_3(x) - (p_{23}^1 + p_{13}^2)x_3 + p_{31}^2 x_1 + p_{32}^1 x_2,$$

for all x. Adding up the left-hand and right-hand sides of these inequalities gives us

$$u_1(x^*) + u_2(x^*) + u_3(x^*) \geq u_1(x) + u_2(x) + u_3(x).$$

It follows that the equilibrium outcome is efficient.

The proof for nonquasilinear utility functions follows the line of argument used in earlier proof. As before, the argument can be generalized to non-convex environments as long as we use suitably nonlinear prices.

For n agents, the payoff to agent i will have the form

$$u_i(x) + \sum_{j \neq i} \left[p_{ij}^i x_j - p_{ji}^j x_i - \|p_{ji}^i - p_{ji}^j\| \right]$$
$$+ \frac{1}{n-2} \sum_{k \neq i,j} \sum_{j \neq i} \left[(p_{kj}^k - p_{kj}^j) x_j - \|p_{kj}^k - p_{kj}^j\| \right]$$

8. A different information structure

The compensation mechanism described above is appropriate for a "bilateral" information structure: if agent i imposes costs on agent j, both i and j know the magnitude of these costs. Another structure that one might imagine is that there is some third party, k, who knows the magnitude of these costs. In this case, we can use a slightly different type of compensation mechanism to achieve efficient outcomes.

Consider the following example. There are three agents. Agent i chooses x_i, holds "money" y_i, and has a quasilinear utility function $u_i(x_1, x_2, x_3) + y_i$. The prices that support an efficient allocations will have the form $p_{ij} = \partial u_i(x)/\partial x_j$. Let p_{ij}^k denote the report of person k about the appropriate magnitude of the price p_{ij}, and let $x = (x_1, x_2, x_3)$ be the vector of choices.

In this variant of the compensation mechanism, each agent will have payoffs of the form

$$u_1(x) - (p_{31}^2 + p_{21}^3)x_1 + p_{12}^3 x_2 + p_{13}^2 x_3$$
$$u_2(x) - (p_{32}^1 + p_{12}^3)x_2 + p_{21}^3 x_1 + p_{23}^1 x_3$$
$$u_3(x) - (p_{23}^1 + p_{13}^2)x_3 + p_{31}^2 x_1 + p_{32}^1 x_2.$$

Note the payoffs are balanced, even out of equilibrium. No sidepayments or penalties are necessary in this case.

Let us examine agent 1's payoff in detail. Agent 1's choice x_1 imposes costs on agents 2 and 3; p_{13}^2 is agent 2's report about the marginal cost that 1's action imposes on 3. Similarly, p_{21}^3 is agent 3's report about the marginal cost that 1 imposes on 3. In addition to facing these Pigovian taxes, agent 1 is compensated for the choices made by the other two agents. For example, he receives a compensation of $p_{12}^3 x_2$ for agent 2's choice; note that the magnitude of this compensation depends on the report of agent 3.

Suppose that agent 1, say, considers changing the prices he announced. Then he would change x_2^* and, possibly, increase his penalty $\|p_{21}^1 - p_{21}^2\|$. Since (x_1^*, x_2^*, y_1^*) is a utility maximum for agent 1 on his budget set, agent 1 can be no better off by this action. This shows that the choice of prices in the first stage is optimal for agent 1, and a similar argument works for agent 2. ∎

7. Balancing the mechanism

We have seen that in the simple example discussed earlier that the compensation mechanism can be balanced by distributing the budget surplus generated by agent i among the other agents. The same procedure works in general. Here is what the sidepayments look like for 3 agents in the case of quasilinear utility. The line containing the term $u_i(x)$ depicts the basic compensation mechanism. The other two lines indicate the sidepayments necessary to balance the budget.

$$
\begin{aligned}
u_1(x_1, x_2, x_3) - [p_{21}^2 + p_{31}^3]x_1 &+ p_{12}^2 x_2 + p_{13}^3 x_3 - \|p_{21}^1 - p_{21}^2\| - \|p_{31}^1 - p_{31}^3\| \\
&+ [p_{32}^3 - p_{32}^2]x_2 + [p_{23}^2 - p_{23}^3]x_3 \\
&+ \|p_{32}^2 - p_{32}^3\| + \|p_{23}^3 - p_{23}^2\| \\
u_2(x_1, x_2, x_3) - [p_{12}^1 + p_{32}^3]x_2 &+ p_{21}^1 x_1 + p_{23}^3 x_3 - \|p_{12}^2 - p_{12}^1\| - \|p_{32}^2 - p_{32}^3\| \\
&+ [p_{31}^3 - p_{31}^1]x_1 + [p_{13}^1 - p_{13}^3]x_3 \\
&+ \|p_{31}^1 - p_{31}^3\| + \|p_{13}^3 - p_{13}^1\| \\
u_3(x_1, x_2, x_3) - [p_{13}^1 + p_{23}^2]x_3 &+ p_{31}^1 x_1 + p_{32}^2 x_2 - \|p_{13}^3 - p_{13}^1\| - \|p_{23}^3 - p_{23}^2\| \\
&+ [p_{21}^2 - p_{21}^1]x_1 + [p_{12}^1 - p_{12}^2]x_2 \\
&+ \|p_{21}^1 - p_{21}^2\| + \|p_{12}^2 - p_{12}^1\|
\end{aligned}
$$

It is straightforward to verify that the budget is balanced for any set of announcements by the agents and that the equilibrium is efficient. This same construction works even if utility is not quasilinear; the proof given earlier works with essentially no changes.

Note that the logic of this proof is quite general. In particular, the taxation and compensation functions do not need to be linear functions. All that is necessary is that each agent can manipulate the other agent's choice without incurring any cost himself. If economic environment is convex, we only need local manipulability; if the economic environment is non-convex, we need global manipulability.

6. A converse theorem

We have shown that all equilibria of the compensation mechanism are efficient. Here we prove the converse theorem that all efficient outcomes can be equilibria of the compensation mechanism for appropriate distributions of initial endowments. The proof is a simple modification of the Second Theorem of Welfare Economics.

Theorem. *Let (x^*, y^*) be an interior Pareto efficient allocation, and suppose that preferences are convex, monotonic, and continuous. Then there is an allocation of endowments (w_1, w_2) such that the given allocation is a subgame perfect equilibrium of the compensation mechanism.*

Proof. Under the stated assumptions there will exist competitive prices $(p_1^*, p_2^*, 1)$ such that (x^*, y^*) is a competitive equilibrium. Hence, at the equilibrium prices the allocation maximizes the utility of agent 1 on his budget set:

$$\max_{x_1, x_2, y_1} \; u_1(x_1, x_2, y_1)$$
$$\text{such that } p_1^* x_1 + p_2^* x_2 + y_1 = w_1,$$

and likewise for agent 2. This should be compared to the maximization problem for agent 1 in the compensation mechanism, which is:

$$\max_{x_1, y_1} u_1(x_1, x_2, y_1)$$
$$\text{such that } p_{21}^2 x_1 + y_1 = w_1 + p_{12}^2 x_2 - \|p_{21}^1 - p_{21}^2\|,$$

Consider the strategies in which agent 1 announces $p_{21}^1 = p_1^*$, $p_{12}^1 = -p_2^*$, and agent 2 announces $p_{21}^2 = p_2^*$, $p_{12}^2 = -p_1^*$. It is clear from the Second Welfare Theorem that these price choices will lead to (x^*, y^*) being chosen as optimizing choices in the second stage of the compensation mechanism.

Theorem. *Let preferences be convex and continuous. Then every subgame perfect equilibrium of the compensation mechanism is Pareto efficient.*

Proof. Let (x, y, p) be a subgame perfect equilibrium of the compensation mechanism. First we show that in equilibrium $p^1_{21} = p^2_{21}$. To see this, consider the agents' budget constraints:

$$p^2_{21}x_1 + y_1 = w_1 + p^2_{12}x_2 - \|p^1_{21} - p^2_{21}\|$$

$$p^1_{12}x_2 + y_2 = w_2 + p^1_{21}x_1 - \|p^2_{12} - p^1_{12}\|.$$

Note that agent 1 can influence agent 2's choice of x_2 through both the "income term," $p^1_{12}x_1$, and the "price term," p^1_{21}. However, by the Local Invertibility Assumption, any choice of x_2 that can be achieved through the income term can also be achieved by an appropriate choice of the price term, p^1_{12}.

Suppose that there were an equilibrium in which $p^1_{21} \neq p^2_{21}$. Let agent 1 set $p^1_{21} = p^2_{21}$ and adjust p^1_{12} so as to induce the same choice of x_2. This must increase agent 1's utility, contradicting the assumption that we had an equilibrium. The argument that $p^2_{12} = p^1_{12}$ is analogous.

Suppose now that (x', y') is a feasible allocation that Pareto dominates the equilibrium allocation. We will show that the existence of such an allocation leads to a contradiction. By convexity and continuity of preferences, we can assume that (x', y') is arbitrarily close to the equilibrium allocation. According to the Local Invertibility Assumption, agent 1, say, can induce agent 2 to choose x'_2 simply by choosing an appropriate level of p^1_{12}, and of course, agent 1 can directly choose (x'_1, y'_1) himself. If agent 1 decides *not* to choose this preferred allocation, it must be because it lies outside his budget set. The same argument applies to agent 2, and this gives us the inequalities

$$p^2_{21}x'_1 + y'_1 > w_1 + p^2_{12}x'_2$$

$$p^1_{12}x'_2 + y'_2 > w_2 + p^1_{21}x'_1.$$

Summing these inequalities and using the fact that $p^i_{ji} = p^j_{ji}$, we have

$$y'_1 + y'_2 > w_1 + w_2,$$

which shows that the Pareto dominating allocation must be infeasible. ∎

These are precisely the first-order conditions given in (11). Therefore, the subgame perfect equilibrium is efficient.

Note that the equilibrium is a *particular* efficient allocation, namely one that satisfies the budget constraints in the maximization problem. In general, such allocations will be a small subset of all efficient allocations. By analogy with the public goods literature, we may call these allocations *generalized Lindahl allocations.* We show below that when the externalities problem is a public goods problem, the prices in the compensation mechanism are essentially Lindahl prices.

A more general proof

The above proof shows clearly how the individual choice problem results in an efficient allocation. However, being a calculus proof, it doesn't deal very well with corner solutions, additional constraints, nondifferentiabilities, etc. Here is another argument that handles these problems. In this proof we allow for the use of an arbitrary norm $\|p_{ji}^i - p_{ji}^j\|$ rather than restricting ourselves to the quadratic norm.

We need one assumption for our proof, namely that each agent can set a price for the other agent that will induce the other agent to make any desired choice. That is, if agent 1 would like agent 2 to make some choice $\hat{x}_2$, there is some price that agent 1 can set, $\hat{p}_{12}^1$, that will induce agent 2 to make this choice. This is analogous to the assumption that $\partial x_2 / \partial p_{21}^1 \neq 0$ in our previous proof. As in the differentiable case, the demand functions only need to be locally invertible if the environment is suitably convex.

Local Invertibility Assumption: Let $x = (x_1, x_2)$ be the outcome of some set of price announcements. Let $\hat{x}$ be an outcome in a neighborhood of x that agent j prefers to x. Then there is some p_{ji}^j that agent j can announce that will yield this preferred outcome, assuming that the other agent's price announcement does not change.

Local invertibility says that agent 1 can manipulate agent 2's choices through agent 1's price announcements. If the agents' demands are differentiable functions of price, with nonzero derivatives, and preferences are locally nonsatiated, then the Inverse Function Theorem implies local invertibility.

As one might suspect, in equilibrium we must have $p^i_{ji} = p^j_{ji}$. This means that no penalties will be paid and that the payment made by agent i for his action will just be equal to the compensation paid to agent j. Hence, in equilibrium, the aggregate budget constraint will balance.

Let us now show that the equilibrium of this game must be efficient. We will provide two proofs. The first proof simply involves writing down the first-order conditions for the utility maximization problems. There are three choice variables for agent i, x_i, p^i_{ij}, and p^i_{ji}, so we have six first order conditions:

$$\frac{\partial u_1}{\partial x_1} - \frac{\partial u_1}{\partial y_1} p^2_{21} = 0 \tag{12}$$

$$\left(\frac{\partial u_1}{\partial x_1} - \frac{\partial u_1}{\partial y_1} p^2_{21}\right) \frac{\partial x_1}{\partial p^1_{12}} + \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_1}{\partial y_1} p^2_{12}\right) \frac{\partial x_2}{\partial p^1_{12}} = 0 \tag{13}$$

$$\left(\frac{\partial u_1}{\partial x_1} - \frac{\partial u_1}{\partial y_1} p^2_{21}\right) \frac{\partial x_1}{\partial p^1_{21}} + \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_1}{\partial y_1} p^2_{12}\right) \frac{\partial x_2}{\partial p^1_{21}} - 2\frac{\partial u_1}{\partial y}(p^1_{21} - p^2_{21}) = 0 \tag{14}$$

$$\frac{\partial u_2}{\partial x_2} - \frac{\partial u_2}{\partial y_2} p^1_{12} = 0 \tag{15}$$

$$\left(\frac{\partial u_2}{\partial x_2} - \frac{\partial u_2}{\partial y_2} p^1_{12}\right) \frac{\partial x_2}{\partial p^2_{21}} + \left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_2}{\partial y_2} p^1_{21}\right) \frac{\partial x_1}{\partial p^2_{21}} = 0 \tag{16}$$

$$\left(\frac{\partial u_2}{\partial x_2} - \frac{\partial u_2}{\partial y_2} p^1_{12}\right) \frac{\partial x_2}{\partial p^2_{12}} + \left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_2}{\partial y_2} p^1_{21}\right) \frac{\partial x_1}{\partial p^2_{12}} - 2\frac{\partial u_2}{\partial y_2}(p^1_{12} - p^1_{12}) = 0 \tag{17}$$

Here we have assumed that the equilibrium choices in the second stage are differentiable functions of the price announcements made in the first stage. As we shall see in the next section, this is not necessary for the argument, but it does help to see why the method works. Note that when agent 1 chooses p^1_{12}, for example, he recognizes that both his own choice, x_1, and the other agent's choice, x_2, may respond to changes in p^1_{12}.

We must assume that $\partial x_2 / \partial p^1_{12}$ and $\partial x_1 / \partial p^2_{21}$ are not zero. Now simply observe that this assumption and (12), (13) and (14) together imply that $p^1_{21} = p^2_{21}$. Similarly, (15), (16) and (17) imply that $p^1_{12} = p^2_{12}$. Now, combine (12) and (16) to get:

$$\frac{\partial u_1/\partial x_1}{\partial u_1/\partial y_1} + \frac{\partial u_2/\partial x_1}{\partial u_2/\partial y_2} = 0,$$

and combine (13) and (15) to get:

$$\frac{\partial u_2/\partial x_2}{\partial u_2/\partial x_2} + \frac{\partial u_1/\partial x_2}{\partial u_1/\partial y_1} = 0.$$

These conditions simply require that the sum of the marginal-willingnesses-to-pay for activity i should be zero.

Define

$$p_{ij} = \frac{\partial u_i/\partial x_j}{\partial u_i/\partial y_i} \quad \text{for } i \neq j$$

Then we can write the efficiency conditions (11) as

$$\frac{\partial u_1/\partial x_1}{\partial u_1/\partial y_1} + p_{21} = 0$$
$$\frac{\partial u_2/\partial x_2}{\partial u_2/\partial y_2} + p_{12} = 0$$

This form suggests that the efficient allocation can be achieved if each agent faced the correct "price" for his choice. The problem is, how can we determine the correct price?

Here is a description of the general compensation mechanism that solves this problem.

Announcement stage. Agent 1 announces p_{12}^1 and p_{21}^1, and agent 2 announces p_{12}^2 and p_{21}^2.

Choice stage. Each agent chooses x_i and y_i so as to maximize utility subject to a budget constraint:

$$\max_{x_1, y_1} u_1(x_1, x_2, y_1)$$
$$\text{such that } p_{21}^2 x_1 + y_1 = w_1 + p_{12}^2 x_2 - (p_{21}^1 - p_{21}^2)^2,$$

and

$$\max_{x_2, y_2} u_2(x_1, x_2, y_2)$$
$$\text{such that } p_{12}^1 x_2 + y_2 = w_2 + p_{21}^1 x_1 - (p_{12}^2 - p_{12}^1)^2.$$

We show below that the subgame perfect equilibria of this game are precisely the efficient allocations that satisfy the budget constraints. However, before providing that proof, let us make a few observations.

First, each agent i is facing a price, p_{ji}^j, for his own choice x_i. He is also receiving compensation $p_{ij}^j x_j$ for the choice that the other agent makes. Both prices, p_{ji}^j and p_{ij}^j are set by the *other* agent. Each agent i also pays a penalty based on how different his announced price, p_{ji}^i, is from the price that the other agent j announced for i's choice, p_{ji}^j. We have chosen this penalty to be quadratic, but any other norm could be used.

5. A general externalities problem

The above externalities problem is rather special. For example, only one agent makes a choice, and both agents have quasilinear objective functions so there are no income effects. In this section we consider a more general externality problem. For simplicity, we continue to examine a two-agent problem, but it is easy to generalize the argument to an n-agent problem.

In our model there are two choices, x_1 and x_2, and one transferable good, y. Agent i makes choice x_i, and has a quasiconcave utility function $u_i(x_1, x_2, y_i)$. Initially, agent i has w_i units of the transferable good. Think of the transferable good as being money, and the choices as being some action that affects the other agent; e.g., smoking a cigar, playing loud music, or whatever.

Efficient choices

In the absence of any transfers between the agents, each agent i will choose x_i to maximize his own utility. The first-order condition characterizing these choices can be written as

$$\frac{\partial u_1/\partial x_1}{\partial u_1/\partial y_1} = 0$$
$$\frac{\partial y_2/\partial x_2}{\partial u_2/\partial y_2} = 0.$$

These conditions simply say that agent i engages in the action x_i until the marginal-willingness-to-pay for additional x_i just equals the price for x_i, which is zero.

An efficient allocation of choices can be found by maximizing the sum of the utility functions subject to the resource constraint:

$$\begin{aligned} &\max\ u_1(x_1, x_2, y_1) + u_2(x_1, x_2, y_2) \\ &\text{such that } y_1 + y_2 = w_1 + w_2. \end{aligned} \tag{10}$$

The first-order conditions for this problem can be written as

$$\begin{aligned} \frac{\partial u_1/\partial x_1}{\partial u_1/\partial y_1} + \frac{\partial u_2/\partial x_1}{\partial u_2/\partial y_2} &= 0 \\ \frac{\partial u_2/\partial x_2}{\partial u_2/\partial y_2} + \frac{\partial u_1/\partial x_2}{\partial u_1/\partial y_1} &= 0. \end{aligned} \tag{11}$$

which shows that x^* is the socially optimal amount.

This argument shows that all equilibria of the mechanism are efficient. However, in general there will be many equilibria of this game. To see this, observe that if e_1 and e_2 are equilibrium announcements, so are $e_1 + F$ and $e_2 + F$ for arbitrary values of F. In order to get uniqueness of equilibrium, it is necessary to restrict the class of allowable messages.

One way to do this is to parameterize the cost function. Suppose that the set of possible externality costs is $e(x, t)$, where t is an index of type. Suppose that the true type of firm 2 is t_0. In the announcement stage of the game, each firm simply announces the type of the second firm, and firm 1 pays a penalty if its announcement is different from that of firm 2. If t_1 is firm 1's announcement and t_2 is firm 2's announcement, the payoffs will be

$$\Pi_1(x) = rx - c(x) - e(x, t_2) - (t_1 - t_2)^2$$
$$\Pi_2(x) = e(x, t_1) - e(x, t_0)$$

Differentiating with respect to t_1, t_2, and x, we have

$$r - c'(x) - \frac{\partial e(x, t_2)}{\partial x} = 0$$
$$t_1 - t_2 = 0$$
$$\left[\frac{\partial e(x, t_1)}{\partial x} - \frac{\partial e(x, t_0)}{\partial x}\right] x'(t_2) = 0.$$

Assuming that $x'(t_2) \neq 0$, it is easy to see that these equations imply

$$r = c'(x) + \frac{\partial e(x, t_0)}{\partial x},$$

which is the condition for social efficiency. Of course, this argument requires convexity and sufficient regularity so that the derivatives exist.

If the environment is not suitably convex an argument can be constructed along the lines given above in inequalities (7)–(9). However, note that we need to assume that firm 2 can induce firm 1 to choose any desired level of x by choosing an appropriate value of t_2. This is simply a "global" version of the assumption that $x'(t_2) \neq 0$.

messages that the agents send back and forth. In this case it appears that the compensation mechanism may yield efficient outcomes in the case of *ill-informed* agents, as well as in the case of *well-informed* agents.

Nonlinear taxes and compensation functions

The mechanism described above used linear taxes and compensations. This is adequate in the case of a convex environment, but if the environment is not convex, linear prices will not be adequate to achieve efficiency.

However, that does not pose a difficulty for the compensation mechanism. Consider the following generalized version of the mechanism.

Announcement stage. Firm 1 and firm 2 each announce the externality cost function for firm 2. Call these announcements $e_1(x)$ and $e_2(x)$.

Choice stage. Firm 1 chooses x and each firm receives payoffs given by:

$$\Pi_1(x) = rx - c(x) - e_2(x) - \|e_1 - e_2\|$$
$$\Pi_2(x) = e_1(x) - e(x)$$

Here $\|e_1 - e_2\|$ signifies any norm in the appropriate function space. All that is required is that it is positive if the function $e_1(\cdot)$ is not equal to $e_2(\cdot)$.

To see that this works, simply note that in equilibrium, firm 1 will always want to set e_1 equal to e_2. Maximization of profit by firm 1 in the choice stage implies that the chosen level of production, x^*, satisfies the condition

$$rx^* - c(x^*) - e_2(x^*) \geq rx - c(x) - e_2(x) \quad \text{for all } x. \tag{7}$$

However, in the announcement stage, firm 2 can induce any level of x that it wants by appropriate choice of the function e_2. Hence,

$$e_1(x^*) - e(x^*) \geq e_1(x) - e(x) \quad \text{for all } x. \tag{8}$$

Adding these two inequalities together, and using the fact that $e_1(x) \equiv e_2(x)$, we have

$$rx^* - c(x^*) - e(x^*) \geq rx - c(x) - e(x) \quad \text{for all } x, \tag{9}$$

We have assumed that each firm has full information. It is natural to suppose that firm 2 knows the cost of the externality, but firm 1 may not know the magnitude of the costs it imposes, at least with certainty. However, let us suppose that firm 1 may learn the size of the externality by incurring some cost. Will firm 1 have proper incentives to actually make this investment?

Without modeling the information acquisition in detail we cannot give a precise answer to this question. However, roughly speaking, it appears that the answer is yes. Since the efficient output is a strict equilibrium, firm 1 incurs a penalty if it announces $p_1 \neq p_2$. The size of this penalty depends on the magnitude of the α_1. The larger the value of α_1 the more incentive firm 1 has to match firm 2's announcement, even if this requires costly investment in information acquisition.

Adjusting to equilibrium

There is a reasonably natural adjustment process for the compensation mechanism that will lead naive agents to the subgame perfect equilibrium described above. Suppose that two firms play the game over and over again. In period $t+1$ firm 1 sets p_1 to be whatever price firm 2 announced last period, and firm 2 moves p_2 in a direction that increases its profits if firm 1 sets the same price as it did last period. In the choice stage, firm 1 chooses output to maximize profits, given the current prices. This gives us a simple dynamical system:

$$\begin{aligned} p_1(t+1) &= p_2(t) \\ p_2(t+1) &= p_2(t) - \gamma[p_1(t) - e'(x(p_2(t))]. \end{aligned} \tag{5}$$

Here γ is a speed of adjustment parameter. The differential equation analog of this system is

$$\begin{aligned} \dot{p}_1 &= p_2 - p_1 \\ \dot{p}_2 &= -\gamma[p_1 - e'(x(p_2))]. \end{aligned} \tag{6}$$

It is easy to show that that (6) is locally stable; the difference equation version, (5), will be locally stable if γ is small enough to avoid "overshooting."

Note that if the firms use this adjustment procedure neither one needs to know anything about the other firm's technology. All that information is subsumed in the price

Collusion

We have seen that in equilibrium the government's budget balances in the sense that the compensation paid to firm 2 is just equal to the tax paid by firm 1. But out of equilibrium the budget will not necessarily balance unless there are three or more agents and we incorporate the additional sidepayments described earlier. Is it possible that the two firms can collude so as to exploit the regulator?

The sum of the profits of the two firms using the mechanism is:

$$(r + p_1 - p_2)x - c(x) - e(x) - \alpha_1(p_1 - p_2)^2. \tag{4}$$

Ignore the quadratic penalty term in (4) for the moment. Equation (4) implies that the only way the firms can make money is by sending in divergent reports; if they send in the same report, so that $p_1 = p_2$, they simply transfer the same amount of money from one firm to the other.

In general,the firms would like to set p_1 to be as large as possible and p_2 to be as small as possible. That is, firm 1 would want to exaggerate the magnitude of the externality in order to encourage the regulator to pay a large compensation to firm 2. But if firm 2 gets overcompensated for the externality, it will want to report $p_2 = 0$ so as to encourage firm 1 to produce as much as possible. However, this strategy involves making highly divergent reports. Can we use the penalty term to discourage such collusion?

The quadratic penalty is not very good for this purpose since it has a derivative of zero when $p_1 = p_2$. However an absolute value penalty works reasonably well. To see this, let the penalty term be given by $\alpha_1|p_1 - p_2|$. Suppose that we consider increasing p_1 and decreasing p_2. Then the profits of the firm increase by x, and the penalty increases by α_1. Certainly for large enough values of α_1 this will not be a profitable move for the coalition.

Strict equilibria and information costs

Note that the equilibrium described above is a *strict equilibrium* in the sense that each firm is at a strict maximum in equilibrium. This has some interesting consequences for firm behavior.

If we distribute the payments so as to balance the budget out of equilibrium, the payoffs become

$$\Pi_1 = rx - c(x) - [p_{21}^2 + p_{31}^3]x - [p_{21}^1 - p_{21}^2]^2 - [p_{31}^1 - p_{31}^3]^2$$

$$\Pi_2 = p_{21}^1 x - e_2(x) + [p_{31}^3 - p_{31}^1]x + [p_{31}^1 - p_{31}^3]^2$$

$$\Pi_3 = p_{31}^1 x - e_3(x) + [p_{21}^2 - p_{21}^1]x + [p_{21}^1 - p_{21}^2]^2.$$

Using the same sort of arguments as before, it is straightforward to verify that the that the unique equilibrium of this mechanism is the efficient outcome.

Another way to balance the mechanism is to allow each agent k to set the price for the choices affecting agents i and j. This is a bit less natural, due to the informational demands, but yields a very simple mechanism:

$$\Pi_1 = rx - c(x) - (p_{21}^3 + p_{31}^2)x$$

$$\Pi_2 = p_{21}^3 x - e_2(x)$$

$$\Pi_3 = p_{31}^2 x - e_3(x).$$

Differentiating with respect to each of the choice variables shows that the equilibrium of this mechanism is efficient, and it obviously balanced. Both of these constructions work in general as we will see below.

Other forms for the penalty function

The role of the penalty function is simply to ensure that $p_1 = p_2$ in equilibrium. Any increasing function of the difference between p_1 and p_2 will accomplish this goal.

It is also possible to impose a penalty on firm 2 without changing the incentives. For example, we could make firm 2's payoff

$$\Pi_2 = p_1 x - e(x) - \alpha_2(p_2 - p_1)^2.$$

The derivative of this function with respect to p_2 is

$$[p_1 - e'(x)]x'(p_2) - 2\alpha_2(p_2 - p_1).$$

Since $p_1 = p_2$ in equilibrium, the derivative of the penalty term is zero, so it doesn't distort the incentives. Note that firm 1's penalty function can be any increasing function of the difference between the announcements, but firm 2's penalty function must satisfy the additional constraint that its derivative is zero when evaluated at $p_1 = p_2$.

the externality. But this contradicts our original assumption that firm 1 thinks that firm 2 will report a large cost of the externality.

Another way to think about the compensation mechanism is from a dynamic perspective. Firm 1 always wants to announces the same price as the other firm so as to minimize the penalty term. If firm 1 announces a price for firm 2 that is not equal to the marginal externality cost, firm 2 will want to change its announcement of p_1 so as to induce firm 1 to change its production level. Hence the only possible equilibrium is for $p_1 = p_2 = c'(x)$.

4. Remarks and extensions of the basic example

In this section we explore some extensions of the basic example.

Balance

The compensation mechanism, in the form presented above, is balanced in equilibrium, but not out of equilibrium. However, if there are at least three agents it is easy to add sidepayments to balance the mechanism. As Moore and Repullo (1988) point out, we can simply distribute the surplus or deficit generated by each agent's choice among the other agents. Since this lump sum distribution is independent of agent i's choice, there are no resulting incentive effects.[2]

Let's see how this works in the simple example considered earlier. We now suppose that agent 1 imposes an externality on agents 2 and 3. The basic form of the compensation mechanism results in payments of the form

$$\Pi_1 = rx - c(x) - [p_{21}^2 + p_{31}^3]x - [p_{21}^1 - p_{21}^2]^2 - [p_{31}^1 - p_{31}^3]^2$$

$$\Pi_2 = p_{21}^1 x - e_2(x)$$

$$\Pi_3 = p_{31}^1 x - e_3(x).$$

In this mechanism, p_{ij}^k is the price set by agent k; in equilibrium it measures the marginal cost that agent j's choice imposes on agent i.

[2] This idea seems to have been first used by Groves and Ledyard (1977). Since then it has been used by a number of other authors.

This is clear since p_1 has no effect on firm 1's payoff, except through the quadratic penalty.

Consider now firm 2's pricing decision. Although firm 2's announcement has no *direct* effect on firm 2's profits, it does have an *indirect* effect through the influence of p_2 on firm 1's output choice in stage 2. Differentiating the profit function of firm 2 with respect to p_2, and setting it equal to zero we have

$$\Pi_2'(p_2) = [p_1 - e'(x)]x'(p_2) = 0. \tag{3}$$

Since $x'(p_2) < 0$ we must have $p_1 = e'(x)$.[1]

Combining (1), (2) and (3) we have

$$r = c'(x) + e'(x),$$

which is the condition for social optimality. Hence, the unique subgame perfect equilibrium to this game involves firm 1 producing the socially optimal amount of output.

3. Why the compensation mechanism works

The intuition behind the mechanism is not particularly difficult. In the first stage each firm announces the (marginal) cost of the externality: firm 1 announces a cost that will be used to compensate firm 2, and firm 2 announces a cost that will be used to tax firm 1. The essential intuition is that if firm 1 doesn't announce the true marginal cost it imposes on firm 2, then firm 2 will have an incentive to manipulate firm 1's behavior. Hence any misrepresentation of the truth cannot be an equilibrium.

For example, suppose that firm 1 thinks that firm 2 will report a large cost for the externality. Then, since firm 1 is penalized if it announces something different from firm 2, firm 1 will also want to announce a large cost. But if firm 2 thinks that firm 1 will announce a large cost, then it knows that it will be "overcompensated" on the margin for the externality. Hence, it wants firm 1 to produce a large amount of output. But it can give firm 1 an incentive to produce a large amount of output by reporting a *small* cost for

[1] It is easy to check that the second-order condition is satisfied as a strict inequality in equilibrium.

Choice stage. The regulator makes sidepayments to the firms so that the two firms face profit maximization problems:

$$\Pi_1 = rx - c(x) - p_2 x - \alpha_1 (p_1 - p_2)^2$$

$$\Pi_2 = p_1 x - e(x).$$

The parameter $\alpha_1 > 0$ is of arbitrary magnitude.

In this mechanism firm 1 is forced to pay a penalty based on the marginal social cost of the externality as reported by firm 2, and firm 2 receives compensation based on the marginal social cost as reported by firm 1. Firm 1 must also pay a penalty based on the square of the difference between the two reports. This penalty can be any increasing function of difference between the two reports, but we have chosen a quadratic penalty for simplicity. Since α_1 can be arbitrary positive constant, the penalty can be arbitrarily small.

2. Analysis of the compensation mechanism

There are many Nash equilibria of this game; essentially any triple (p_1, p_2, x) such that $p_1 = p_2$ and x maximizes firm 1's objective function is a Nash equilibrium. However, if we use the stronger concept of subgame perfect equilibrium we get a much smaller set of equilibria. In fact, the *unique* subgame perfect equilibrium of this game has each agent reporting $p_1 = p_2 = p^*$ and firm 1 producing the efficient amount of output.

In order to verify this, we must work backwards through the game. We begin with the choice stage. Firm 1 maximizes its profits, given the Pigovian tax announced in stage 1, which implies that firm 1 will choose x to satisfy the first-order condition

$$r = c'(x) + p_2. \qquad (1)$$

This determines the optimal choice, x, as a function of p_2, which we denote by $x(p_2)$. Note that $x'(p_2) < 0$—the higher the tax that firm 2 announces, the less firm 1 will want to produce.

We next examine the price-setting stage of the game. Consider first firm 1. If firm 1 believes that firm 2 will announce p_2, then firm 1 will want to announce

$$p_1 = p_2 \qquad (2)$$

A third class of solutions, associated with Pigou (1920), involves intervention by a regulator who imposes a Pigovian tax. The difficulty with this solution is that it requires the regulator to know the correct level of the Pigovian tax. But if the regulator has enough information to calculate this tax, he may as well simply impose the optimal production level directly—there is no need to use the roundabout method of taxation. Hence, the Pigovian solution is also incomplete. The method proposed below solves this problem, in the sense that it gives the regulator a method to induce the firms to reveal the information necessary to construct the optimal Pigovian tax.

Returning to our example, we note that if the regulator had full information, internalizing the externality would be easy. One solution would be for the regulator to impose the costs of the externality on firm 1 by charging it a "tax" of $e(x)$ if it produces x units of output. Firm 1 would then solve the problem

$$\max_x \; rx - c(x) - e(x).$$

Let x^* be solution to this problem; then x^* satisfies the first-order condition

$$r - c'(x^*) - e'(x^*) = 0.$$

Given our curvature assumptions on $e(x)$, we could just as well set a "Pigovian tax," $p^* = e'(x^*)$ and let firm 1 solve the problem

$$\max_x \; rx - c(x) - p^* x.$$

However, we have assumed that the regulator doesn't know the size of the externality and therefore cannot determine the appropriate value of p^*. The regulator's problem is to design a mechanism that will induce the agents to reveal their information about the magnitude of the externality and achieve an efficient level of production.

Here is a version of the compensation mechanism that solves the regulator's problem.

Announcement stage. Firm 1 and firm 2 simultaneously announce the magnitude of the appropriate Pigovian tax; denote the announcement of firm 1 by p_1 and the announcement of firm 2 by p_2.

We assume that firm 1's choice of output imposes an externality on firm 2. For any choice of x, firm 2's profits are

$$\pi_2 = -e(x),$$

where $e(x)$ is a differentiable, positive, increasing, and convex function of x. All of this information is common knowledge among the agents, but is not known to the regulator.

In general, the level of output chosen by firm 1 will not be efficient, since it ignores the social cost its choice imposes on firm 2. There are three classical solutions to this problem of externalities.

One class of solutions, associated with Coase (1960), involves negotiation between the agents with respect to the externality. Coase argues that if transactions costs are zero and property rights are well-defined, then agents should be able to negotiate their way to an efficient outcome. But this is an incomplete solution since no specific structure for the negotiations is presented. The compensation mechanism described below may be thought of as a structure for such negotiations, and therefore can be viewed as being complementary to the Coase approach.

A second class of solutions, associated with Arrow (1970), involves setting up a market for the externality. If a firm produces pollution that harms another firm then a competitive market for the right to pollute may allow for an efficient outcome. A competitive market may be thought of as a particular institution that allows agents to "negotiate" their way to an efficient outcome. However, as Arrow points out, the market for allocating a particular externality may be very thin—in many cases of interest such markets involve only two participants.

However, a thin market does not necessarily mean a non-competitive market. There are both theoretical and empirical reasons to believe that certain kinds of market interaction can be competitive even though only a small number of agents are involved. A Bertrand model of oligopoly yields a more-or-less competitive outcome with only two firms. The real-life implementation of Bertrand competition—competitive bidding—seems to work reasonably well even if there are only a small number of bidders. This suggests that "markets" for externalities with *price-setting* agents may be a useful model for negotiations among agents. This is the key insight behind the compensation mechanism.

almost any choice rule can be implemented by subgame perfect equilibria. However, as Moore and Repullo point out, "...the mechanisms we construct ... are far from simple; agents move simultaneously at each stage and their strategy sets are unconvincingly rich. We present such mechanisms to show what is possible, not what is realistic." (p. 1198) Moore and Repullo also show that in certain "economic environments" it is possible to use somewhat simpler mechanisms. However, the compensation mechanism appears to be much simpler than the examples Moore and Repullo examined. For a thorough discussion of the recent literature on implementation see Moore (1991).

It should be emphasized that the solution concept of subgame perfection requires the agents to be informed about the technology and tastes of the other agents. This is, of course, more restrictive than one would like. However, there is a broad set of cases for which such mechanisms may be useful. For example, consider a group of agents who must design a mechanism to make group decisions for problems that will arise in the future. At the time the mechanism is chosen, the agents do not know the relevant tastes and technologies, but they will know these things when the mechanism is actually used. In this circumstance, the compensation mechanism may be a reasonable proposal. See Moore and Repullo (1988) and Maskin (1985) for further discussion.

Furthermore, in many cases the subgame perfect equilibrium of the compensation mechanism turns out to be the limit of of a simple dynamic adjustment process. This tatonnement process requires *no* information about the other agents' characteristics.

I first describe a very simple example of the compensation mechanism in a two-firm externalities problem and discuss in an intuitive way why the method works. The following sections show how the method can be extended to work in more general environments.

1. A simple example of the compensation mechanism

Consider the following externality problem involving two firms. Firm 1 produces output x in order to maximize its profit

$$\pi_1 = rx - c(x).$$

Here r is the competitive price of output and the cost function $c(x)$ is a differentiable, positive, increasing, and convex function.

A Solution to the Problem of Externalities when Agents are Well-Informed

Hal R. Varian

Consider an economic environment in which agents take actions which impose benefits or costs on other agents. All agents are aware of the relevant technology and the tastes of all other agents. However the "regulator," who has the responsibility for determining the final allocation, does not have this information. How can the regulator design a mechanism so that the agents will have the proper incentives to reveal their information and achieve an efficient allocation?

In addition to implementing an efficient allocation, one might also want the mechanism to achieve some distributional goals. For example, one might want agents who are injured by an externality to be compensated for that injury. Or, in the case of public goods, one might want the public goods to be paid for by a system of Lindahl taxes.

In this paper I describe a class of simple two-stage games which implement efficient allocations in this sort of environment. In addition to implementing efficient outcomes, the mechanism also achieves the distributional goals just described. In the subgame perfect equilibria of this game, parties injured by the externality are compensated. In the case of public goods, the mechanism implements Lindahl allocations. Because payment of "compensation" is an important feature of the mechanisms I describe, I refer to the general class of mechanisms as *compensation mechanisms.* These mechanisms appear to work in a broad variety of economic environments and do not involve substantial restrictions on tastes or technology. They are also quite simple to describe and analyze.

The fact that sequential games and subgame perfect equilibria may be useful in implementation problems was suggested by Crawford (1979), Moulin (1979, 1981) and extensively analyzed by Moore and Repullo (1988). They show that in economic environments,

This work was supported by National Science Foundation grant SES-8800114 and a Fulbright grant. I wish to thank Mark Bagnoli, Ted Bergstrom, Ken Binmore, Alan Kirman, Diego Moreno, Arthur Robson, Steve Salant, Mark Walker, and Michael Whinston for their comments and suggestions. I also wish to thank the European University Institute and the University of Siena for their hospitality during the period in which I conducted this research.

Preliminary Draft

A Solution to the Problem of Externalities when Agents are Well-Informed

by

Hal R. Varian
University of Michigan

November, 1989
Current version: July 20, 1992

Abstract. I describe a simple two-stage mechanism, the compensation mechanism, that implements efficient allocations in economic environments involving externalities. The compensation mechanism can be used to solve a wide variety of externalities problems, including the standard problem of public goods provision. It requires that that the agents know the magnitudes of the benefits and costs that they impose on other agents, but will also work with naive agents who follow a simple tatonnement.

Keywords. externalities, public goods, sequential games, Lindahl allocations, subgame perfect equilibria

Address. Hal R. Varian, Department of Economics, University of Michigan, Ann Arbor, MI 48109