Journal of Institutional and Theoretical Economics (JITE) 144 (1988), 245–266
Zeitschrift für die gesamte Staatswissenschaft

Externalities and the Coase Theorem: Hypothesis or Result?

by

URS SCHWEIZER *

0. Introduction

A substantial part of the economic analysis of law and the theory of property rights is based on a famous paper by COASE [1960]. The conclusion of that paper is usually referred to as the Coase Theorem, though, as is well-known, Coase left it to others to express his idea as a theorem. The Coase Theorem deals with decentralized-decision making by parties whose activities impose external economies or diseconomies on each other. For such situations, PIGOU [1932] claimed that the discrepancy between private and social costs would lead to a waste of resources unless remedied by appropriate governmental intervention. The Coase Theorem forcefully challenged this view by asserting that the final outcome will be efficient, provided that the initial assignment of property rights is well-defined and that transactions involving the exchange of rights are costless. The debate then appears to centre upon the efficiency versus inefficiency of the outcome if externalities are involved.

The present paper treats the issue of efficiency by relying on concepts and notions drawn from game theory. Even in the formal literature, some studies exist which establish the inefficiency (market failure literature), others the efficiency of outcomes, resulting from decentralized decision taking. Since most of these studies appear to be correct in the mathematical sense, the question arises as to the precise reason for such disparate predictions. In this paper, this question will be addressed in the following way.

In Section 1, a simple allocation problem including external diseconomies between two parties is introduced which serves, throughout the paper, as a unifying framework for investigating various approaches to the efficiency issue. The model captures, in essence, most of the examples discussed by COASE [1960].

Section 2 deals with Coase's analysis [c.f. COASE [1960, Section VIII]] of the uncompensated damage done to surrounding woods by sparks from railway

* The author wishes to thank Eric van Damme, Martin Hellwig and an anonymous referee for useful comments on an earlier version of this paper. Financial support by the Deutsche Forschungsgemeinschaft through Sonderforschungsbreich 303 is gratefully ac-

engines. Coase revisited the railway example first studied by Pigou in order to show that Pigou was mistaken in his economic investigation of the case. Since Coase approaches the example as if it were one of very high transaction costs, his findings are not really of significance so far as the Coase Theorem is concerned. But it provides some hints as to how situations of high transaction costs should appropriately be modeled according to Coase's own approach. I take these hints to indicate that, in the language of game theory, the noncooperative Nash equilibrium of the game for which parties' activity levels are the only strategies should be the predicted outcome. In any case, Coase does not deny the outcome to be inefficient. Due to the reciprocal nature of externalities, however, making the railway compensate for damages caused by fires – as Pigou apparently recommended – would not restore efficiency. Rather, for suitable parameter configurations, the regime of uncompensated damages could turn out to be less wasteful. Therefore, shifting liability must carefully be distinguished from a Pigouvian scheme of taxation because, under the above solution concept, such a scheme would indeed correct market failure if tax rates were determined properly.

Section 3 explores the idea that, by introducing a sufficient number of competitive markets, efficiency could also be restored. ARROW [1969] has shown how the argument leading to the welfare theorems can formally be extended to cover cases of external economies. But the assumption of competitiveness seems rather implausible because, even for more general cases than the one considered in the present paper, Arrow's method has only one agent on each side of the market. The approach is interesting but, apparently, it does not really attempt to formalize the ideas lying behing the Coase Theorem.

In Section 4, cooperative solution concepts of game theory are considered as a way of illustrating the efficiency claim of the Coase Theorem. The cooperative approach has received much attention in spite of the fact that, as cooperative solutions are efficient by definition, it changes the Coase Theorem from a theorem into a mere hypothesis on the solution concept. Even worse, if situations of high transaction costs have to be captured by noncooperative solutions and those of low transaction costs by cooperative solutions, then an undue dichotomy with respect to the solution concept arises: for it then has no reasonable alternative left to it to handle cases of intermediate transaction costs.

In Section 5, we suggest the way in which such a dichotomy may be avoided by relying exclusively on the noncooperative Nash equilibrium as the proper solution concept, regardless of the level of transaction costs. It is shown that, by introducing steps of the bargaining procedure among the strategies, the Nash equilibrium has a fair chance of leading to an efficient outcome. For any particular bargaining procedure, some formal proof will be needed to establish efficiency. Since a proof is thus involved, the Coase Theorem may become truly a theorem under the noncooperative approach.

Traditionally, the notion of transaction costs refers both to strategic aspects and to tangible forms of such costs. Under the noncooperative approach, however, all strategic aspects are relegated to the solution concept whereas any tangible form of transaction costs will be written directly into the game's extensive form. Section 6 discusses an example of tangible costs. Since the same solution concept is used for all levels of costs, comparative statics analysis can be carried out with respect to that level.

The last two sections deal with settings of incomplete information. It definitely appears as one of the main virtues of the noncooperative approach that it permits such situations to be handled. In this context, again, private information does not have to be considered as some abstract source of transaction costs. Rather, strategic interaction between parties is captured by the solution concept and it remains a matter of formal analysis whether the outcome is efficient or not. In Section 7, the case is considered in which the party suffering the damage is fully informed whereas the party causing it does not know its true magnitude. It is shown that, in this setting, basically the same bargaining procedure as under complete information will lead to an (ex post) efficient outcome.

In Section 8, finally, information is assumed incomplete with respect to the utility function of the party who causes the damage: this party is fully informed whereas the other party is not. In this setting, efficiency cannot generally be achieved, regardless of which bargaining procedure is implemented. With reference to a numerical example, it is established that there exists no direct mechanism which is incentive-compatible, individually rational and ex post efficient. The revelation principle then implies that there is no bargaining procedure, no matter how sophisticated, that can ever restore efficiency. The main conclusion of the paper is that it is possible for the Coase Theorem to enter into the class of formal theorems only if the noncooperative approach is pursued. Moreover, its chances of doing so are, even under that approach, greater if information is complete!

1. *A Simple Externality Model*

Consider two parties A and B engaging in some activities at levels x and y, respectively. A's action is assumed to have a harmful effect on B. The net utility of the two parties, which is assumed measurable in monetary units, amounts to

$$\text{(N)} \qquad U = A(x) \quad \text{and} \quad V = B(y) - S(x,y), \qquad \text{respectively}$$

The functions $A(x)$ and $B(y)$ denote utility if external effects are absent and are supposed to be unimodal. At activity levels x and y, however, external diseconomies amounting to $S(x, y)$ arise which, if uncompensated (N = no liability), have to be borne by B as the victim. We assume throughout that

$$S(x,0) = S(0,y) = 0 \quad \text{and} \quad S_x(x,y) > 0,\ S_y(x,y) > 0.$$

This assumption emphasizes the reciprocal nature of the externality, as damages arise only if both activities simultaneously operate at positive levels.

The above model allows for various interpretations. In particular, it covers the two main examples discussed by COASE [1960]. As regards the first, party A would be the cattle raiser whose straying cattle destroy crops on neighboring land and party B would correspond to the farmer who cultivates that neighboring land. The damage amounts to the value of the (annual) crop loss as caused by the straying cattle. The second example concerns a railroad (party A) whose activity causes damage, by sparks from engines, to party B's property adjoining the railway line. Here, the damage consists of the crops destroyed by fires.

The model also captures situations related to the law of accidents as first studied by CALABRESI [1965]. For instance, party A could be thought of as a car driver operating at average speed x whereas B would be a pedestrian crossing streets at frequency y. The expected damage from accidents would again depend on the activity levels at which the two parties chose to operate. If no particular rule of liability were imposed, damages would have to be borne by party B. Ex ante, pedestrians do not know whether they will be involved in an accident and, if they are, who will be the driver to cause damage to them. Therefore, one might prefer to think of parties A and B as mean representatives of each class. But the situation would still be captured by the formal model.

For later reference, we point out that the efficient levels of activity are determined by the first order conditions

$$\text{(E)} \qquad A_x(x^E) - S_x(x^E, y^E) = 0 \quad \text{and} \quad B_y(y^E) - S_y(x^E, y^E) = 0.$$

Utility being measurable in monetary terms, transfer payments would allow damages to be divided among parties in many ways. Note, however, that the efficient levels would be independent of any particular scheme of transfer payments.

2. *Rules of Liability at High Transaction Costs*

In this section, the formal model is used to revisit Coase's analysis of the railroad example (c.f. COASE [1960, Section VIII]) which led to the conclusion that forcing the railways to compensate those who suffer from sparks would not necessarily improve efficiency. Unfortunately, Coase does not argue whether this is a situation of high or low transaction costs. He only mentions that, if it were one of low (vanishing) transaction costs, bargaining would lead to an efficient outcome no matter what particular rule of liability were imposed. Therefore, in terms of efficiency, liability only matters in conditions in which it is too expensive for such bargains to be made. As for Section VIII, he simply assumes that such conditions prevail. While he does not explicitly use the language of game theory, his argument seems well captured by the following line of reasoning.

Under a system of no liability, parties find themselves in the situation of a game whose normal form is given by equation (N). Transaction costs, by assumption, are high enough to prevent bargaining between the parties from taking place. Therefore party A will play his dominant strategy whereas party B is simply left to react in an optimal way to that move by A. The corresponding activity levels are to be determined from the equations

$$A_x(x^N) = 0 \quad \text{and} \quad B_y(x^N) - S_y(x^N, y^N) = 0$$

which form the first order conditions of the (noncooperative) Nash equilibrium of game (N). Obviously, at this equilibrium, conditions (E) necessary for efficiency can never be met and, hence, the equilibrium must be inefficient. From the railway's point of view, a divergence exists between private and social net utility which causes the inefficiency. Coase, however, is right to point out that a change from a regime of no liability to one of strict liability would not restore efficiency. Strict liability would mean that the situation corresponds to the game with normal form

$$\text{(L)} \qquad U = A(x) - S(x, y) \quad \text{and} \quad V = B(y).$$

In this game, it is party B who has a dominant strategy. The Nash equilibrium leads to activity levels (x^L, y^L) which can be calculated from the first order conditions

$$A_x(x^L) = S_x(x^L, y^L) \quad \text{and} \quad B_y(y^L) = 0.$$

This Nash equilibrium cannot be efficient either, as the first order conditions (E) of efficiency will once again be violated. The divergence between private and social net utility now lies simply with party B. It is not generally possible to compare, on efficiency grounds, the solutions induced by the two rules of liability considered above. In any case, parameter configurations can easily be shown to exist such that the equilibrium under no liability is more efficient than the one under strict liability. Therefore, if Pigou really suggested – as Coase claims he did – that making the railways liable for damage caused by fire would necessarily be desirable, Pigou was indeed mistaken in his economic analysis. Comparing the two rules of liability amounts to comparing two inefficient allocations. DIAMOND and MIRRLEES [1975] have investigated this problem in detail.

Shifting liability, however, is not the only idea associated with the name of Pigou. Our simple model also allows us to explore Pigouvian taxation (c.f. BAUMOL and OATES [1975] e.g.), which is a scheme to restore efficiency. As the one who causes the damage, party A is required to pay a tax proportional to its activity level. Compensation of the victim, however, need not take place. Let t denote the tax rate of the scheme. Then the situation is captured by the game with normal form

$$\text{(P)} \qquad U = A(x) - tx \quad \text{and} \quad V = B(y) - S(x, y).$$

The Nash equilibrium of this game is to be found as solution of the first order conditions

$$A_x(y^P) - t = 0 \quad \text{and} \quad B_y(y^P) - S_y(x^P, y^P) = 0.$$

It follows that a governmental agency could, by imposing tax rate $t = S_x(x^E, y^E)$, restore efficiency. As a matter of fact, the public interest theory of regulation (see POSNER [1974]), which explains governmental interventions as attempts to correct market failure, would predict that the agency will search for this rate. Needless to say, Coase and the new political economy strongly object to such a view. As far as Pigouvian taxation is concerned, government might neither be able nor have any incentives to find the proper tax rate.

This section deals with the case of high transaction costs. Various games in normal form have been discussed whose common feature is that, except for the Pigouvian scheme of taxation, inefficiency cannot be resolved if the noncooperative Nash equilibrium is taken as the predicted outcome. Coase has not used the strict language of game theory. But given that one of the players has a dominant strategy, most economists would probably agree that the Nash equilibria considered above are very plausible. Rather, disputes arise with the case of low or vanishing transaction costs as the condition under which the Coase Theorem is supposed to hold. For our formal model, the Coase Theorem claims that, regardless of the liability rule, bargaining between parties will always lead to an efficient arrangement. Since all of the above Nash equilibria not relying on governmental intervention turn out to be inefficient, either the Coase Theorem is wrong or, else, these equilibria do not aptly describe situations of low transaction costs.

To rescue the theorem and to achieve efficiency, at least three options are available. First, markets for externalities could be established which, if competitive, would ensure efficiency. Second, it could be argued that situations of low transaction costs are best described by cooperative games and solution concepts. Many authors seem to follow this route at least implicitly (POLINSKY [1983], COOTER [1982] and, quite likely, COASE [1960]) if not explicitly (ORDESHOOK [1986] who has parties playing solutions from the core). Since cooperative solution concepts such as the core or Nash's arbitration point lead, by definition, to efficient outcomes, efficiency would be restored though not as a result of but as a hypothesis on the solution concept. Third, efficiency could also be achieved in a noncooperative framework. NASH [1953] was the first to point out such a possibility. For a noncooperative Nash equilibrium to be efficient, however, the game's normal form must be different from that of the games studied above. Nash's idea consists of making the players' steps of negotiation in the cooperative game become moves in the noncooperative model. The next three sections deal with these options in detail.

3. *Competitive Market Solution*

ARROW [1969] has shown that, under the hypothesis of universality of markets, any competitive equilibrium must be efficient. He has also established the precise meaning of universality if externalities are involved. His idea can easily be adopted to our formal model introduced in Section 1. The quantity of the externality would have to be measured by the activity level of party A. Party A "supplies" the externality whereas party B "demands" the externality as an "input". Moreover, let r denote the "price" of the externality. In other words, the externality is treated as a commodity in the usual sense, except that prices might be negative if external diseconomies exist as in our model. If the market of externalities is competitive, both parties behave as price takers. In this case, supply and demand functions can be defined as the solution of each party's optimization problem. Therefore, at price r, supply amounts to

$$x^s(r) = \arg\max A(x) + rx$$

whereas demand and activity level of party B are given as

$$[y(r), x^d(r)] = \arg\max B(y) - S(x, y) - rx.$$

Suppose that demand equals supply in this market. Then the corresponding activity levels x and y are to be calculated as

$$x = x^d(r) = x^s(r) \quad \text{and} \quad y = y(r).$$

The solution is easily shown to be efficient because, for any other activity levels x' and y', it holds by definition that

$$A(x) + rx \geqq A(x') + rx'$$

and

$$B(y) - S(x, y) - rx \geqq B(y') - S(x', y') - rx'.$$

Hence, total utility at activity levels x and y must be at least as high as under any other pair of activity levels x' and y'. The proof is adopted from that of the first welfare theorem in a straightforward way.

Thus, in a formal sense, efficiency can be achieved by such a competitive market solution. But that the solution does not appear to be plausible has been pointed out by, among others, Arrow himself. One problem arises form the fact that, for many cases of externalities, commodities which represent external effects in the above way might not be appropriable. The other problem has to do with thinness of markets. As Arrow has shown for externalities in general, a single buyer faces a single seller on each market. Such a situation seems to be closer to a bilateral monopoly than to a competitive market. Therefore, the assumption of price-taking behavior can hardly be defended. STARRETT [1972], finally, argues that externalities typically lead to nonconvex production sets. Even in a formal sense, consequently, equilibrium might not exist in such markets.

To conclude, any attempt to establish efficiency by referring to the competitive market solution leads to serious difficulties and, hence, cannot be considered as a valid approach to a "proof" of the Coase Theorem. Actually, there is apparently no passage in COASE [1960] which suggests reliance on competitive markets. It seems better, therefore, to explore the other options.

4. *Cooperative Bargaining Solution*

A formal bargaining situation between two parties is described by a set R of feasible utility allocations and a threat point r of R which denotes the utility allocation resulting from no agreement being reached. A cooperative solution consists of a function f which associates with every formal bargaining situation (R, r) some outcome $f(R, r)$ belonging to the set R of feasible utility allocations. The function f is assumed to satisfy certain axioms, one of which is efficiency. Therefore, by definition, a cooperative solution of the bargaining problem will always be efficient. The best known cooperative solution concept is Nash's arbitration point (see NASH [1950] or OWEN [1982], e.g.).

Let us explore these notions in the framework of our externality model. We first have to associate a formal bargaining situation with the game whose normal form is given by equation (N). One might propose the set R to consist of all utility allocations (U, V) for which (N) holds for at least one pair (x, y) of activity levels. In Figure 1, the upper bound of this set is depicted as NN. If, however, this really were the proper utility frontier then the problem of inefficient noncooperative solutions would not arise because the Nash equilibrium (x^N, y^N) turns out to be efficient with respect to the frontier NN! Pareto improvements can only be achieved if transfer payments between parties are allowed for. In this case, the set R rather consists of all utility allocations (U, V) for which

$$U + V \leqq A(x^E) + B(y^E) - S(x^E, y^E)$$

holds. In any case, it is only with respect to this frontier that the Nash equilibrium fails to be efficient.

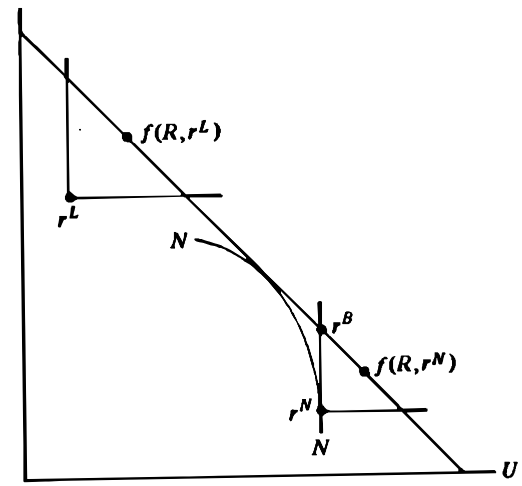

Figure 1
Nash equilibrium, the bargaining solution and the efficiency frontier

Next, a threat point must be specified. To begin with, suppose that party A is not liable for any damage caused to party B (c.f. the rule (N) of no liability). If no agreement can be reached, party A then has a dominant strategy left to play in the noncooperative game to which B has to react in an optimal way. The outcome would be the Nash equilibrium (x^N, y^N) as described in Section 2. It thus seems reasonable to use the utility allocation

$$r^N = [A(x^N), B(y^N) - S(x^N, y^N)]$$

resulting from this Nash equilibrium as the threat point of the cooperative game. For similar reasons, if the rule were such that party A is strictly liable for any damage caused to B, the utility allocation

$$r^L = [A(x^L) - S(x^L, y^L), B(y^L)]$$

arising with the corresponding noncooperative Nash equilibrium of the game with normal form (L) would naturally serve as the threat point of the cooperative game. It follows that the rule of liability determines the threat point (see Figure 1) but has no effect on the set R of feasible utility allocations. If vanishing transaction costs are taken to imply that a cooperative solution such as Nash's arbitration point is played, the predicted outcome in terms of utility allocations will be either $f(R, r^N)$ or $f(R, r^L)$, according to which liability rule applies. For both rules, however, the outcome has to be efficient and hence the rule has significance only with respect to the distribution of income (see Figure 1). This conclusion is well in line with the Coase Theorem and the example of the straying cattle which COASE [1960] discusses in great detail. Cooperative solution concepts are a way of illustrating the theorem but one which will never lead to a proof because efficiency remains a matter of definition under this approach. Put differently, the Coase Theorem would only survive as a hypothesis on the solution concept if cooperative bargaining theory is taken as the proper framework to model situations of vanishing transaction costs.

To be sure, if such a hypothesis were valid it would not remain without consequences for policy conclusions. Suppose that a governmental agency attempts to correct alleged market failure due to external diseconomies by imposing a Pigouvian scheme of taxation with tax rate t. Moreover, following traditional views, it does not compensate the victim. Then the formal bargaining situation faced by the two parties depends on the tax rate t and consists of the set $R = R(t)$ of utility allocations (U, V) for which

$$U + V \leqq A(x) + B(y) - S(x, y) - tx$$

holds for at least one pair (x, y) of activity levels. The threat point would reasonably be assumed to be the utility allocation

$$r^P = r^P(t) = [A(x^P) - tx^P, B(y^P) - S(x^P, y^P)]$$

arising with the corresponding noncooperative Nash equilibrium (c.f. Section 2). If parties A and B can bargain at zero transaction costs, and if the above hypothesis is correct, the predicted outcome would be the utility allocation $f[R(t), r^P(t)]$. This outcome is, of course, on the frontier of the set $R(t)$ and hence satisfies the first order conditions

$$A_x(x^P) - S_x(x^P, y^P) = t.$$

But there is still no positive tax rate t at which the overall efficiency condition (E) will be met. In other words, if parties play – instead of the noncooperative Nash equilibrium – a cooperative solution, implementing any Pigouvian scheme would distort the outcome which, otherwise, would be efficient. This point has been made previously by BUCHANAN and STUBBLEBINE [1962].

In Section 3, the noncooperative Nash equilibrium of game (N) has been predicted as the inefficient outcome if high transaction costs are involved. In the present section, the efficiency claim of the Coase Theorem is restored by having resort to solution concepts from cooperative game theory. The approach is not without merit because it neatly illustrates the basic conclusion of the Coase Theorem. Nevertheless, the approach suffers from a serious defect because it leads to an undue dichotomy with respect to the solution concept. If situations of high transaction costs should be approached by noncooperative and of low transaction costs by cooperative concepts, it remains unclear how all those realistic cases of intermediate transaction costs are to be handled. Therefore, to avoid such a dichotomy, the next section will deal with an attempt to restore efficiency in a noncooperative framework.

5. *Noncooperative Bargaining*

NASH [1953] was also the first to deal with bargaining problems in a noncooperative way. He proposed that any given formal bargaining situation with set R of feasible utility allocations and threat point r could be reduced to a noncooperative game in the following way. As strategies, each of parties A and B must "demand" their utility levels U' and V', respectively. If the pair of utility levels chosen is feasible with respect to the set R, both receive the demanded level. Otherwise, payoffs of the game are given by the threat point r. It is easy to show that any efficient utility allocation above the threat point corresponds to a noncooperative Nash equilibrium of the associated demand game. Hence, in principle, efficient outcomes can indeed be implemented as noncooperative equilibria.

ARROW [1979] has proposed a slightly different noncooperative approach for the achievement of efficient points. WINDISCH [1975] reports further attempts which have appeared in some of the earlier literature. The basic idea essentially amounts to introducing moves of the bargaining procedure as strategy variables of a noncooperative game. To illustrate the idea, consider the following game in extensive form. At stage 1, party B is proposing a contract to A in which he offers payments Z to A provided that A does not extend his activity level x beyond the proposed limit X. At stage 2, party A decides whether to accept such a contract or, else, to extend his activity level beyond the proposed limit. At stage 3, finally, party B has to adjust its activity level y to the one chosen by A at stage 2 of the game.

Without stage 1, the Nash equilibrium would still be the inefficient pair (x^N, y^N) of activity levels associated with game (N) (c.f. Section 2). But introducing stage 1 leads to a very different (subgame perfect) equilibrium which is to be calculated as follows (for the notion of subgame perfect equilibrium the reader is referred to SELTEN [1975]): at stage 3, party B can observe the activity level x chosen by A at an earlier stage. No matter whether A keeps to the limit or not, B's best reaction at this stage will always be

$$y = y(x) = \arg\max\ B(y) - S(x, y). \tag{1}$$

At stage 2, party A decides whether to accept the contract $[X, Z]$ which B has proposed at stage 1. If A accepts, he must keep the limit X in order to receive payments Z. Otherwise A will extend his activity level up to x^N which yields the highest payoff if the contract is to be refused. Therefore, Party A's reaction if offered contract $[X, Z]$ must be

$$x = x(X,Z) = X \quad \text{if } A(x^N) \leqq A(X) + Z \tag{2}$$

and

$$x = x(X,Z) = x^N \quad \text{if } A(x^N) > A(X) + Z. \tag{3}$$

Strictly speaking, if $A(x^N) = A(X) + Z$, party A would be indifferent between accepting and refusing the contract. For convenience, however, it is assumed that A always accepts in such cases of indifference.

At stage 1, it is B's turn to offer a contract. Obviously, it never pays for B to propose payments Z which strictly exceed $A(x^N) - A(X)$ because A would already keep to the limit at lower payments. If, however, B proposes a contract which is not accepted by A, then B's payoff amounts to

$$V = V^N = B(y^N) - S(x^N, y^N) \tag{4}$$

which would be the same as if game (N) were played (c.f. Section 2). On the other hand, if B proposes a contract $[X, Z]$ which is accepted because it satisfies

$$A(x^N) = A(X) + Z$$

then B's payoff amounts to

$$V = B(X) - S[X, y(X)] - Z.$$

The highest payoff of B to be achieved in this way arises, of course, with the contract

$$X = x^E \quad \text{and} \quad Z = A(x^N) - A(x^E) \tag{5}$$

which, due to the inefficiency of the pair (x^N, y^N), exceeds the payoff (4) resulting from any contract which A will not accept. Therefore, in subgame perfect equilibrium, party B necessarily has to propose contract (5) which A accepts and which leads to an efficient allocation. The corresponding utility allocation

$$r^B = [A(x^N), B(y^E) - S(x^E, y^E) - A(x^N) + A(x^E)]$$

is depicted in Figure 1.

To summarize, the noncooperative Nash equilibrium is widely accepted as a solution concept. It has been argued in Section 2 above that Coase implicitly used it in discussing the railway example. Moreover, there can be no doubt that people opposing the Coase Theorem also rely on it to establish all kinds of market failure (c.f. JOHANSEN [1982] for further justification). The present section has established that even the efficiency claim of the Coase Theorem can be reconciled with the noncooperative approach provided that steps of the bargaining procedure are included among the moves of the noncooperative game. Yet to establish the efficiency of a Nash equilibrium, a proof such as the one given above will be needed and, a formal proof being involved, it now seems fair indeed to call Coase's claim a theorem.

The noncooperative approach to the Coase Theorem has the advantage that the notoriously difficult notion of transaction costs can be dispensed with as a tool to select the proper solution concept. The crucial question no longer hinges on the level of transaction costs but rather on what would be accepted as a reasonable noncooperative game to capture the features of a given situation. The game must be different if parties sitting in separate cells were facing a prisoners' dilemma as compared to the case where parties can freely communicate and possibly make binding commitments. The example of car drivers and pedestrians mentioned in Section 1, where, ex ante, parties do not even know each other might well resemble the prisoners' dilemma as far as the proper game's extensive form is concerned. Obviously, it is not only the allocation problem which matters but also the possibility for players to communicate. Under the noncooperative approach, however, strategic interactions will always be described by Nash equilibrium regardless of the level of transaction costs.

6. *Intermediate Transaction Costs*

While it has been proposed that strategic aspects of transaction costs be relegated to the solution concept, the noncooperative approach may nevertheless directly take into account any tangible form of transaction costs such as fees paid to an attorney in order to make a commitment credibly binding. To illustrate this point, let us modify the three-stage game introduced in Section 5 in the following way. If party B wants to propose a contract it has to pay a positive fee F which can only be avoided by not offering a contract. Except for this fee, the game is the same as before.

The (subgame perfect) equilibrium is to be determined as follows. At stage 3, B has observed the other party's activity level x and reacts according to (1). At stage 2, if A is offered no contract, it chooses activity level x^N which yields the highest payoff under these circumstances. If, however, A is offered contract $[X, Z]$ then it responds according to (2) and (3), respectively. At stage 1, finally, B either proposes contract (5) which would lead to payoff

$$V = B(y^E) - S(x^E, y^E) - F \tag{7}$$

or, else, it avoids paying the fee by proposing no contract to A, in which case its payoff would be as in (4). The option it chooses in equilibrium depends on the magnitude of the fee. If fees are low such that payoff (7) exceeds that of (4), then B will propose contract (5). Note that, in this case, the efficient activity levels will be chosen but the outcome still fails to be efficient, since the fee is lost. If, however, fees are high enough such that payoff (4) exceeds payoff (7) then B prefers to propose no contract. Here, no fee will be lost but the outcome is inefficient because activities are operated at inefficient levels. But, in this example, inefficiency would be of no concern to the Coase Theorem because strictly positive transaction costs are involved.

Under the noncooperative approach which avoids any dichotomy as far as the solution concept is concerned, comparative statics analysis with respect to the magnitude of the fee is conceptually feasible. As the fee approaches zero, the outcome is predicted to move towards the efficiency frontier whereas, as the fee approaches infinity, the outcome moves to the Nash equilibrium of the game with normal form (N) (c.f. Section 2). These limiting cases would also be predicted by the dichotomous approach discussed in Sections 2 and 4. It is, however, the merit of the noncooperative approach that it allows intermediate cases as well to be dealt with, as has been illustrated by the above example.

7. *Bargaining Under Incomplete Information*

The remainder of the paper will be devoted to investigating the externality model in a setting of incomplete information. To begin with, it is shown that the three-stage game of Section 5 can easily be extended to include incomplete information with respect to the damage caused by the negative externality. The extended version of the game still has a Nash equilibrium which is (ex post) efficient. One might be tempted to conclude from this result that the Coase Theorem may possibly hold even under incomplete information (c.f. ARROW [1979]). In general, however, efficiency cannot be achieved, as follows from the impossibility result of MYERSON and SATTERTHWAITE [1983] as well as from related findings due to SAMUELSON [1985]. These authors have investigated the exchange of an indivisible commodity between parties each of which does not know the other's valuation. Ex post efficiency cannot be achieved in their framework provided that information is incomplete on both sides. In the next section, we shall argue that, even under one-sided incomplete information, inefficient outcomes cannot necessarily be avoided in the situation described by the externality model if it is party A's utility which is unknown to B.

For the present section, however, information is assumed incomplete with respect to the damage caused by the externality. The magnitude of this damage is known only to party B who is also the one to suffer from it because, by assumption, the rule still applies under which party A will not be liable. Parties are fully informed about all other aspects of the model. Incomplete information is to be expressed in the following way. The damage $S = S(x, y, t)$ depends not only on activity levels x and y but also on the type t of player B. At stage zero, nature is assumed to select B's type randomly. While B does learn its type before it has any moves to make, its true type remains unknown to player A. At stage 1, B proposes a contract $[X(t), Z(t)]$ which may now depend on its type t. At stage 2, A only sees the contract which is offered but A does not know B's type. But B's type does not really matter for A's payoff. Therefore A's equilibrium strategy is again given by (2) and (3). At stage 3, finally, player B, still knowing its type t, reacts to the activity level x previously chosen by A by adjusting its own activity level y according to

$$y = y(x, t) = \arg\max B(y) - S(x, y, t).$$

It follows that, for the equilibrium to be sequential in the sense of KREPS and WILSON [1982], player B proposes at stage 1 to pay

$$Z(t) = A(x^N) - A[x^E(t)]$$

to party A if A does not extend its activity beyond the level $X(t) = x^E(t)$ where $x^E(t)$ and $y^E(t)$ denote the ex post efficient activity levels given that B is of type t. These levels can easily be calculated from first order conditions

$$A_x[x^E(t)] = S_x[x^E(t), y^E(t), t] \quad \text{and} \quad B_y[y^E(t)] = S_y[x^E(t), y^E(t), t].$$

The sidepayments offered by B are such that A ends up with the same payoff as if it had refused to bargain and, instead, were playing the dominant strategy x^N of the simple game (N). If information is incomplete only with respect to the damage suffered by party B, then A's utility from playing its threat strategy will not depend on the type of player B. It turns out to be this particular feature which allows the attainment of an efficient outcome under incomplete information.

8. *Inefficiency Due to Incomplete Information*

In this section, information is assumed incomplete with respect to party A's utility function which is known to A but not B. All other aspects of the model are common knowledge. Therefore, the setup is still one of one-sided incomplete information but what at first glance appears to be a minor modification turns out to impede efficiency in a substantial way, as we now wish to show.

The investigation of the (sequential) equilibrium of the three-stage game introduced in Section 5, but adapted to the new setting of incomplete information, raises no difficulties. At stage 3, party B has observed A's activity level x from which B may or may not learn A's type. In any case, it is sequentially calculation would basically be the same as in the complete information version of the game (see (2) and (3)) except, of course, that the decision depends on the type, which is known to A. The major difference appears at stage 1, where party B has to propose a contract without knowing A's true type. At stage zero, nature selects the type t of player A which means that A's utility function is $A(x,t)$. A sequential equilibrium of this game, however, cannot be ex post efficient, for the following reason: regardless of nature's move, party B will propose a single contract. Whenever this contract is refused by some types of player A, efficiency obviously cannot prevail. But efficiency also fails to hold if the single contract is accepted by all types because, in this case, the same limit of A's activity level will be kept irrespective of its type. Such an outcome would not be efficient either.

It might be argued that inefficiency is caused by the game's peculiar structure, which does not allow party B to learn anything about A's type from its strategic behavior before B has to propose a final contract. Or else, one might prefer to model the setting as a game where A as the informed player has to make proposals for contracts. Whatever the game of the bargaining procedure, however, the revelation principle (see MYERSON [1979]) will now be shown to lead to the conclusion that efficiency cannot in general be restored. Given an equilibrium of any bargaining game, there exists some direct and incentive-compatible mechanism whose truth-telling outcome corresponds to the given equilibrium. To establish that efficiency is not to be achieved, it will thus be sufficient to focus on the class of direct and incentive-compatible mechanisms. This is the immediate consequence of the revelation principle. Certainly, efficient and incentive-compatible direct mechanisms exist for our problem of incomplete information, and in fact all of them will be described below. Difficulties only arise from the additional requirement that the mechanism should be freely accepted by both parties because, for any reasonable bargaining procedure, parties will always have the opportunity not to communicate at all. In other words, for a direct mechanism to be meaningful in the context of voluntary bargaining, it must be individually rational.

In our setting, a direct mechanism has the following appearance. The informed party A is asked its type, which it may or may not report truthfully. Depending on A's reply, a pair of activity levels will be implemented. The mechanism might also specify transfer payments between parties. The scheme is incentive-compatible if it never pays not to tell the truth. In the following, only efficient mechanisms will be considered which, as a consequence of the truth-telling property, means that the mechanism must implement the pair of activity levels

$$[x^E(t), y^E(t)] = \arg\max\ A(x,t) + B(y) - S(x,y)$$

if A reports type t. Efficient mechanisms, however, may differ with respect to the

$$U(t',t) = A[x^E(t'),t] + Z(t')$$

denote A's utility if its true type is t while reporting type t'. For the mechanism $Z(t)$ to be incentive compatible, the condition $U(t,t) \geqq U(t',t)$ must hold for any pair of types t and t'. In particular, it follows that

$$A_x[x^E(t),t]\,\mathrm{d}x^E(t)/\mathrm{d}t + \mathrm{d}Z(t)/\mathrm{d}t = 0$$

must hold if the efficient mechanism $Z(t)$ has to be incentive compatible. If A reports type t', then B's utility amounts to

$$V(t') = B[y^E(t')] - S[x^E(t'), y^E(t')] - Z(t').$$

Incentive compatibility and the first order conditions associated with the efficient pair of activity levels immediately imply that $\mathrm{d}V(t)/\mathrm{d}t = O$ for all t. In other words, as a necessary condition for incentive-compatibility, party B's utility level $V(t) = v$ does not depend on nature's move. The condition can easily be shown to be sufficient as well. Therefore the class of efficient, incentive-compatible direct mechanisms is fully characterized by transfer payments of the form

(IC) $$Z(t) = B[y^E(t)] - S[x^E(t), y^E(t)] - v.$$

Such mechanisms differ only with respect to party B's constant utility level v.

Next, the requirement of individual rationality must be discussed. To begin with, consider the informed party A. If A refuses to bargain it always has the option left to it to choose, after having learned its type t, the activity level $x^N(t) = \arg\max\ A(x,t)$ which would guarantee the utility level $U^N(t) = A[x^N(t),t]$ to A. Therefore, from A's view, the mechanism will only be individually rational if, for all types t,

(RA) $$U^N(t) \leqq U(t,t).$$

As for party B, a condition of individual rationality could be arrived at in the following way. Suppose A is given no chance to bargain. As above, A would then plan to play strategy $x^N(t)$. Party B, which does not know A's type, would then have to choose some activity level y^N which would lead to expected utility

$$EV^N = B(y^N) - ES[x^N(t), y^N]$$

where the expected value E is taken with respect to A's type. Therefore, from B's

(RB) $$\text{Max } EV^N \leqq v$$

where the maximum is taken with respect to B's activity level y^N.

The question now arises whether there exists any mechanism which simultaneously satisfies conditions (IC), (RA) and (RB). In general, the answer turns out to be that there is not, which means that, for any incentive-compatible direct mechanism which is individually rational in the above sense, the possibility must exist that it is inefficient. To illustrate this point, consider the following numerical example:

$$A(x, t) = x(t - x/4);$$
$$B(y) = y(2 - y) \quad \text{and} \quad S(x, y) = xy/2.$$

The efficient pair of activity levels for types $1/2 \leq t \leqq 2$ can easily be calculated as

$$x^E(t) = 4(2t - 1)/3 \quad \text{and} \quad y^E(t) = 2(2 - t)/3.$$

Therefore, truthful reporting by A under mechanism (IC) leads to the utility level

$$U(t, t) = 4(t^2 - t + 1)/3 - v$$

provided A's true type is t. If A were not given a chance to bargain, his strategy and resulting utility would amount to

$$x^N(t) = 2t \quad \text{and} \quad U^N(t) = t^2.$$

It then follows from condition (RA) of individual rationality that B's utility level cannot be positive, i.e. $v \leqq 0$. Since, however, B could refuse to bargain and instead choose an activity level such as $y^N = (2 - Et)/2$ which would lead to expected utility

$$EV^N = (2 - Et)^2/4 > 0$$

the above mechanism, at $v \leqq 0$, cannot be individually rational from player B's view! Consequently, this numerical example does not allow for an incentive-compatible direct mechanism which is (ex post) efficient and satisfies the condition of individual rationality for both players. The impossibility result hinges on the fact that, for the numerical example, uncertainty introduces enough dispersion. Moreover, the uninformed player's utility level must be constant in order to ensure the mechanism's incentive-compatibility. As a consequence, the condition of individual rationality with respect to the informed player allows only low values for the uninformed player's constant utility level which, however, would not be individually rational from this player's view point. The revelation principle then implies that no bargaining game will exist which has an ex post efficient equilibrium outcome. In such cases, incomplete information must be considered as a very strong impediment to efficiency.

Concluding Remarks

Based on Pigou's analysis, external economies and diseconomies are widely believed to cause market failure in the sense that resources will be wasted under the laissez-faire outcome. Moreover, governmental intervention is recommended to correct such failure. Coase strongly objected to this view for well-known reasons. To be sure, in cases involving high transaction costs, he would not challenge the view that efficiency might not be reached under a regime of laissez-faire. But, of course, he would also question the government's ability (and willingness) to correct such market failure. The main thrust of the Coase Theorem, however, is that no intervention will be needed in cases of low transaction costs.

Observing reality does not directly reveal whether a given outcome is efficient or not. Welfare analysis can only be carried out within models of reality. Such models must specify which actions are, in principle, available to the involved parties and which outcome within the model will result from strategic interaction. In the language of game theory, first, a formal game must be specified and, second, a hypothesis as to the proper solution concept will be needed. Only then does it become possible to investigate the efficiency properties of the predicted outcome.

At transaction costs sufficiently high to prevent any reasonable communication from taking place, the literature in both the Pigou as well the Coase tradition seems to basically accept the noncooperative Nash equilibrium as the proper solution concept. In particular, I am not aware of any studies of market failure which would not at least implicitly if not explicitly rely upon it. It seems rather to be the case of low transaction costs which has led to controversies. For such cases, the Coase Theorem claims an efficient outcome to prevail under laissez-faire and hence that governmental intervention would not be required and, worse, could even be harmful. The theorem may or may not be correct. But since it makes a welfare statement, it can only be approached in the context of both a model and a solution concept. It has been argued in this paper that a substantial part of the literature on the Coase Theorem ensures efficiency of the outcome by having resort to solution concepts from the theory of cooperative games such as the core or Nash's arbitration point. Common to these concepts is that efficiency is part of their definition and, hence, the efficiency claim of the Coase Theorem would not be a theorem in the formal sense but rather a hypothesis as to the solution concept. The approach has another serious deficiency. Suppose we agree indeed that situations of very high transaction costs should be captured by noncooperative Nash equilibria whereas those of low or vanishing transaction costs might best be approached by cooperative solution concepts. Then the difficulty arises as to which solution concept should be relied on if the situation is one of intermediate transaction costs.

To avoid such a dichotomy with respect to the solution concept, the present paper has advocated the noncooperative approach to the Coase Theorem. The idea that efficient outcomes can, in principle, be obtained as noncooperative Nash equilibria is not new. NASH [1953] established a corresponding result before any debate about the Coase Theorem could conceivably have taken place. But, in my view, his idea has not received enough attention. Instead, there developed a large body of literature on market failure which takes the noncooperative view and which has led many to believe that noncooperative Nash equilibria have some intrinsic tendency to be inefficient. NASH [1953] has established that this belief is not true and we have tried to reinforce his view. The conclusion then seems rather to be that, in the future, as much effort should be spent in exploring the set of games which have efficient noncooperative equilibria as has been devoted in the past to games whose noncooperative equilibria happen to be inefficient. To give Nash equilibria a chance of being efficient, steps of the bargaining procedure must explicitly be included among the moves of the game. This approach has the further advantage that tangible forms of transaction costs can readily be taken into account by structuring the game properly, whereas all strategic aspects of transaction costs are relegated to the noncooperative solution solution concept.

Some of the existing literature recognizes incomplete information as a source of transaction costs for which strategic aspects are of particular importance. The noncooperative approach allows settings of incomplete information to be handled in a logically consistent way, as has been explored in the last two sections of the paper, where the following conclusion was reached. In a setting of one-sided incomplete information, if the uninformed party has an outside option which ensures, irrespective of the opponent's true type, a constant utility level, then incomplete information need not impede efficiency. Efficiency, however, cannot generally be restored if the utility level associated with the uninformed party's outside option depends on the informed party's private information. Such results would not be obvious if one tried to cope with incomplete information by referring to some abstract notion of strategic transaction costs. The noncooperative approach, fortunately, does not require such a notion because all strategic aspects are handled by the noncooperative solution concept.

Summary

External effects are widely claimed to cause market failure. Formal studies establish such failure by proving the inefficiency of the noncooperative Nash equilibrium of corresponding games. The Coase Theorem takes a more optimistic view as far as the involved parties' ability to restore efficiency is concerned. Some authors have attempted to illustrate the theorem by referring to cooperative solution concepts. But to avoid any undue dichotomy with respect to the solution concept, and to give the Coase Theorem a chance to become a theorem in the formal sense, the present paper advocates the noncooperative approach. For a Nash equilibrium to be efficient, steps of the bargaining procedure must be included among the moves of the game. The noncooperative approach is also shown to allow for investigating incomplete information as a potential impediment to efficiency.

Zusammenfassung

Externalitäten und das Coase Theorem: Hypothese oder Ergebnis?

Externe Effekte verursachen einer weit verbreiteten Meinung entsprechend Marktversagen. In formalen Studien wird dieses Versagen mit dem Hinweis auf die Ineffizienz gewisser nichtkooperativer Nash-Gleichgewichte begründet. Das Coase Theorem schätzt die Möglichkeit der Individuen höher ein, das Effizienzziel aus eigener Initiative zu erreichen. Einige Autoren veranschaulichen die Aussage des Theorems dadurch, daß kooperative Lösungskonzepte unterstellt werden. Um jedoch eine ungebührliche Dichotomisierung bezüglich der Lösungskonzepte zu vermeiden und um dem Coase Theorem eine Chance zum formalen Resultat aufzusteigen, stellt diese Arbeit den einzuräumen, zum formalen Resultat aufzusteigen, stellt diese Arbeit den nichtkooperativen Zugang in den Vordergrund. Nash-Gleichgewichte können durchaus effizient sein, wenn Verhandlungsschritte in die Strategienräume einbezogen werden. Die nichtkooperative Betrachtungsweise eignet sich auch für Situationen unvollständiger Information, welche sich als potentielle Behinderung der Effizienz herausstellt.

References

ARROW, K. J. [1969], "The Organization of Economic Activities: Issues Pertinent to the Choice of Market Versus Non-market Allocation", pp. 47–64, in: *The Analysis and Evaluation of Public Expenditures: The PPB System*, Joint Economic Committee of the Congress of the United States, Washington, D.C.

-- [1979], "The Property Rights Doctrine and Demand Revelation Under Incomplete Information", pp. 23–39, in: M. Boskin (ed.), *Economics and Human Welfare*, New York.

BAUMOL, W. J. and OATES W. E. [1975], *The Theory of Environmental Policy*, Englewood Cliffs.

BUCHANAN, J. M. and STUBBLEBINE, W. C. [1962], "Externality", *Economica*, 29, 374–384.

CALABRESI, G. [1965], "The Decision for Accidents", *Harvard Law Review*, 178, 713–732.

COASE, R. [1960], "The Problem of Social Cost", *Journal of Law and Economics*, 3, 1–44.

COOTER, R. [1982], "The Cost of Coase", *Journal of Legal Studies*, 11, 1–33.

DIAMOND, P. A. and MIRRLEES, J. A. [1975], "On the Assignment of Liability – the Uniform Case", *Bell Journal of Economics*, 6, 487–516.

JOHANSEN, L. [1982], "On the Status of the Nash Type of Noncooperative Equilibrium in Economic Theory", *Scandinavian Journal of Economics*, 84, 421–441.

KREPS, D. M. and WILSON, R. [1982], "Sequential Equilibria", *Econometrica*, 50, 863–894.

ORDESHOOK, P. C. [1986], *Game Theory and Political Theory*, Cambridge.

OWEN, G. [1982], *Game Theory*, Second Edition, Orlando.

NASH, J. F. [1950], "The Bargaining Problem", *Econometrica*, 18, 155–162.

-- [1953], "Two-person Cooperative Games", *Econometrica*, 21, 128–140.

MYERSON, R. B. [1979], "Incentive Compatibility and the Bargaining Problem", *Econometrica*, 47, 61–73.

-- and SATTERTHWAITE, M. A. [1983], "Efficient Mechanisms for Bilateral Trading", *Journal of Economic Theory*, 29, 265–281.

PIGOU, A. C. [1932], *The Economics of Welfare*, 4th edition, New York.

POLINSKY, A. M. [1983], *An Introduction to Law and Economics*, Boston.

POSNER, R. A. [1974], "Theories of Economic Regulation", *Bell Journal of Economics*, 5, 335–358.

SAMUELSON. W, [1985], "A Comment on the Coase Theorem, pp. 321–339, in: A. E. Roth (ed.), *Game-Theoretic Models of Bargaining*, Cambridge.

SELTEN, R. [1975], "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games", *International Journal of Game Theory*, 4, 25–55.

STARRETT, D. A. [1972], "Fundamental Nonconvexities in the Theory of Externalities", *Journal of Economic Theory*, 4, 180–199.

WINDISCH, R. [1975], "Coase-Paradigma Versus Pigou-Paradigma: Über Information und Motivation als Grundlagen dezentralisierter Umweltkontrolle", *Zeitschrift für Nationalökonomie*, 35, 345–390.

Professor Dr. Urs Schweizer
Institut für Gesellschafts-
und Wirtschaftswissenschaften
Universiät Bonn
D-5300 Bonn 1
Bundesrepublik Deutschland