THE COASE THEOREM AND THE EMPTY CORE*

VAROUJ A. AIVAZIAN and JEFFREY L. CALLEN
McMaster University

THE Coase theorem, which states that in the absence of transaction costs resource allocation is neutral with respect to liability rules, is usually demonstrated in a two-participant scenario of either two firms or two consumers.[1] This is so in Coase's original verbal discussion, as well as in the more recent literature that treats the Coase argument within the confines of the neoclassical paradigm.[2] As we will show, this is an unfortunate state of affairs, because with more than two participants the Coase theorem cannot always be demonstrated.

The reader's initial response to our claim is undoubtedly skeptical; he may believe that we have underestimated the power and robustness of the zero transaction cost assumption. We wish to emphasize, however, that our result is not predicated on the existence of transaction costs, be they information costs, bargaining costs, or (damage) valuation costs. Our assumptions include those made explicitly or implicitly by Coase.[3] Furthermore, our contentions are not based on novel notions but rather on some well-known results from the theory of the core, a theory which plays an important role in mathematical economics and game theory.[4] The fact that our type of critique of the Coase theorem has not appeared in the literature until now is primarily a function of the slow diffusion of mathematical economics to the remainder of the profession rather than a function of our incisive analysis.

We will show, utilizing a three-participant example, that the Coase

* We wish to acknowledge the very helpful comments of Professor Andrew Schotter on an earlier version of this paper. The usual disclaimers apply.

[1] R. H. Coase, The Problem of Social Cost, 3 J. Law & Econ. 1 (1960).

[2] See, for example, James R. Marchand & Keith P. Russell, Externalities, Liability, Separability, and Resource Allocation, 63 Am. Econ. Rev. 611 (1973); Ken-Ichi Inada & Kiyoshi Kuga, Limitations of the "Coase Theorem" on Liability Rules, 6 J. Econ. Theory 606 (1973); and Peter H. Greenwood & Charles A. Ingene, Uncertain Externalities, Liability Rules, and Resource Allocation, 68 Am. Econ. Rev. 300 (1978).

[3] Coase assumed away wealth effects and bargaining threats.

[4] One of the most comprehensive analyses of the theory of the core and its applications to economics is provided by Lester Telser, Competition, Collusion, and Game Theory (1972).

theorem cannot be proved if the core of the game is empty under one of the liability rules. This is in contradistinction to the two-participant case, for which the core is always nonempty so that the Coase theorem can always be demonstrated. Besides questioning the general validity of the Coase theorem, we will illustrate that it cannot be studied meaningfully (or rigorously) by employing the familiar neoclassical paradigm. The Coase theorem is a statement about nonmarket allocations and the predilection of the extant literature to impose competitive market structure parameters or specific adjustment paths on the bargaining process is limiting at best and, probably, self-defeating.[5] The only tool we are aware of that is capable of providing the appropriate insights into nonmarket allocations is game theory.

In what follows, Section I describes an example for which the Coase theorem cannot be demonstrated. Although the example is somewhat stylized and simplified, it captures the crux of the argument. Section II provides the intuition for the breakdown of the Coase theorem in the context of the example. Section III demonstrates the argument in a more rigorous fashion, while Section IV discusses the implications of our results.

I. Description of the Example[6]

Consider two factories labeled A and B polluting a neighborhood laundry, firm C. The laundry does not emit pollutants. Let the function V denote profits per day. In the absence of negotiations, profits per day for the three firms are thus assumed: $V(A) = \$3{,}000$; $V(B) = \$8{,}000$; and $V(C) = \$24{,}000$.

Since there are no transaction costs, firms will negotiate or, equivalently, merge to mitigate the impact of the pollution. In the spirit of game theory such mergers will be called coalitions and V the characteristic function. Define $V(A,B,C)$ to be the joint profits of the (grand) coalition of all three firms. Similarly, $V(A,B)$ will denote the joint profits of the coalition of firms A and B and so on. We will assume the following joint profits for all potential coalitions: $V(A,B) = \$15{,}000$; $V(A,C) = \$31{,}000$; $V(B,C) = \$36{,}000$; and $V(A,B,C) = \$40{,}000$.

These values should be interpreted to mean that if, for example, A and C negotiate or merge, sufficient pollution can be eliminated so that both firms earn \$31,000. This is in contradistinction to the \$27,000 they would have earned in the absence of negotiations. To simplify matters we will

[5] See the references cited in note 2 *supra*.

[6] The numerics (but not the context) of this example are from Jan Mossin, Merger Agreements: Some Game-Theoretic Considerations, 41 J. Bus. 460 (1968).

assume that the merger or negotiation between A and C, if consummated, will close the polluter A down.[7] Thus, all the potential benefits from mergers between the laundry and either factory comes from the closing down of the polluter. If A and C do merge without B, then B is assumed to keep polluting and earning \$8,000 per day.

Should all three firms negotiate or merge, the resulting grand coalition will earn joint profits of \$40,000 per day rather than the \$35,000 they earned by acting independently of each other. Again we assume that, should the grand coalition form, both A and B stop producing.

Although A's and B's pollution activities do not affect each other's outputs, they are also assumed to have the incentive to merge since $V(A,B) > V(A) + V(B)$. In the context of our example, one can assume that this incentive is due to economies of scale in production.[8] To keep the example tractable, it is assumed that the A,B coalition does not affect the level of pollutants so that the laundry continues to earn $V(C)$.[9]

We have assumed in our example that profits accruing to each of the potential coalitions is greater than what these firms would have earned by acting independently. In the merger literature this phenomenon is known as synergy. In the game-theory literature the characteristic function V is said to be superadditive. Superadditivity implies that $V(A) + V(B) + V(C) \leq V(A,B) + V(C) \leq V(A,B,C)$. Similar relationships would of course hold for the other potential coalitions. Far from being a restrictive assumption, superadditivity is obviously a necessary condition for the bargaining process to be initiated and for the Coase theorem to be valid.

II. Breakdown of the Coase Theorem

We now show that the Coase theorem breaks down for this example. Let us first analyse the simple case in which A and B are liable for the pollution damages to C. It is immediately obvious that the grand coalition solution must obtain and neither A nor B are allowed to produce. Thus, if X_A, X_B, and X_C denote the profit allocations to firms A, B, and C, respectively, the solution in the case where A and B are liable is $X_A = X_B = 0$

[7] Otherwise, we would have to specify each firm's production function or, equivalently, a tableau of outputs for each level of pollution. This assumption does not affect the qualitative nature of our results.

[8] On first blush, one is tempted to argue that, if A and B have the incentive to merge because of economies of scale, they will merge and we are back in a two-firm world. However, this is only one potential merger and there are others. It is because there is more than one merger possibility that the Coase theorem fails to hold in our example.

[9] Again, we could complicate the example by providing a tableau of C's earnings for each level of pollution from the merged firm. This would provide more realism with no more insight.

and X_C = \$40,000. This is the correct solution because the maximum bribe that A and B can offer is \$15,000, but their combined liabilities total \$16,000. Therefore, C obtains the maximum benefits of \$40,000 while A and B shut down.[10]

The alternative case in which A and B are not liable for the pollution damages to the laundry is far more complex. Since A and B are not liable, C has the incentive to bribe either A or B, or perhaps both, to reduce their pollution activities. Suppose C bribes B but not A, so that the distribution of profits is \$3,000 to A and \$36,000 to B and C. Since firm C could earn \$24,000 in spite of the pollution, the maximum bribe C will offer B is \$12,000 and, of course, the minimum bribe B is willing to accept is $V(B)$ or \$8,000.

Let us suppose initially that C captures all the potential gains and B is willing to accept the minimum bribe. This distribution of profits (X_A = \$3,000; X_B = \$8,000; and X_C = \$28,000) is inherently unstable. In a transaction-costless world, firm A has the incentive to offer B an additional \$500 to join in a coalition involving only A and B in which X_A = \$6,500 and X_B = \$8,500. Since A and B are better off in this latter case than before, they can block the original distribution of profits involving the coalition of B and C by forming their own coalition. Of course the story does not end here since C could come back and offer B \$9,000 to join the original B,C coalition. Therefore, B can play A against C until, ultimately, it would appear, B captures all the potential benefits by forming a coalition with either A or C.

Suppose, in fact, that B forms a coalition with C, as initially hypothesized, but now B is offered the maximum bribe so that C earns only \$24,000. This distribution is again unstable since A will now bribe C (say by an additional \$500), blocking the latest coalition. Thus, C can play A against B and, again, try to appropriate the maximum benefit.

Even if a two-firm coalition could be found to be stable, these coalitions are unstable relative to the grand coalition. In our example, every two-member coalition yields total profits of \$39,000 which is less than the benefits from the grand coalition of \$40,000. Thus, given a specific distribution of profits in a two-firm coalition, one can always find an alternative

[10] The discussion throughout the paper implicitly assumes that the characteristic function itself does not depend on the particular liability rule. This assumption can readily be relaxed by assuming that, where A and B are liable for the damages, the characteristic function takes on these values: $V(A) = 0$; $V(B) = 0$; $V(C) = 24{,}000$; $V(A,B) = 0$; $V(A,C) = 31{,}000$; $V(B,C) = 36{,}000$; $V(A,B,C) = 40{,}000$. This means that, if A and B are liable for the damages and if the lost profits to C due to pollution are greater than the realized profits accruing to A and B due to production, then the best A and/or B can do is not to produce. If they do produce, they would be liable to pay C more than they earned by producing. In this example again, the core exists and the unique imputation is $X_A = X_B = 0$ and X_C = \$40,000.

distribution in the grand coalition which makes each firm better off. Therefore, one might be tempted to argue that the grand coalition must result. But in our example even the grand coalition is unstable. To see why this is so consider the grand coalition distribution of profits: X_A = \$4,750; X_B = \$9,750; and X_C = \$25,500. Although each firm earns more than it would in the absence of negotiations, there are many two-member coalitions which could, and presumably would, block this particular distribution. For example, the coalition of B and C where X_B = \$10,000 and X_C = \$26,000 is preferred by B and C to the grand coalition solution given above.

Our discussion thus far has tried to highlight a number of the coalition possibilities and why these coalitions are infeasible. Specifically, a coalition is infeasible if any firm earns less than it would by independent action. In addition, no coalition can succeed if the resulting distribution of profits can be blocked by an alternative coalition. As we prove rigorously below, these two latter conditions cause all the potential coalitions in our example to be unstable. Ironically, in a world of zero transaction costs, the inherent instability of all coalitions could result in endless recontracting among the firms.

III. The Empty Core

In a three-participant scenario, coalitional stability requires that the resulting allocation of profits satisfy the following conditions:

$$X_A \geqslant V(A),\ X_B \geqslant V(B),\ X_C \geqslant V(C) \tag{1}$$

$$X_A + X_B \geqslant V(A,B),\ X_A + X_C \geqslant V(A,C),\ X_B + X_C \geqslant V(B,C) \tag{2}$$

$$X_A + X_B + X_C = V(A,B,C). \tag{3}$$

The set of allocations which satisfies these conditions is called the core.

In our example, the core is empty. To see this, note that if all the two-coalition inequalities in (2) are added together then

$$(X_A + X_B + X_C) \geqslant \frac{1}{2}[V(A,B) + V(A,C) + V(B,C)] \tag{4}$$

so that substituting (3) into (4) yields

$$V(A,B,C) \geqslant \frac{1}{2}[V(A,B) + V(A,C) + V(B,C)]. \tag{5}$$

This last condition is violated by our example since the right-hand side of the inequality turns out to be equal to \$41,000 and the left-hand side to \$40,000.

This situation could never occur in the two-participant case, where the

core is always nonempty. With two participants, say A and B, the core is defined by the constraints

$$X_A \geqslant V(A), X_B \geqslant V(B) \tag{6}$$

$$X_A + X_B = V(A,B). \tag{7}$$

But these constraints must be satisfied by at least one allocation since $V(A,B) \geqslant V(A) + V(B)$.

IV. Discussion and Conclusion

We have provided an example in which liability rules have an asymmetrical impact on resource allocation. It was seen that when firms A and B are liable for damages they will stop producing and firm C earns $40,000 per day. On the other hand, when firms A and B are not liable, negotiations cannot be consummated because the core is empty. In a world of zero transaction costs, negotiations could be endless. Even if some merger is negotiated, it will be the grand coalition only by happenstance.

It could be argued that, if the core is empty and no solution is forthcoming, the participants may agree to accept a set of allocation rules that will yield the desired outcome. The Shapley and Von Neumann-Morgenstern allocations are just two of many game-theory solutions which do not depend on the existence of the core.[11] Moreover, since these solutions are Pareto optimal (by assumption), the grand coalition obtains and the allocative outcome is independent of the liability rule. Thus, it may be argued the Coase theorem is valid once more.

The problem with this approach is that it makes a tautology out of the Coase theorem. If the core does not exist, the participants may accept an alternative solution concept; then again they may simply stop negotiating. It is an empirical question as to what happens when the core is empty. We do not know. Of course, one can impose a Pareto-optimal solution on the participants such as the Shapley allocation. To do so, however, would be tantamount to assuming what Coase was trying to prove, namely, that a Pareto-optimal solution will be forthcoming independent of the liability rule *and* independent of the bargaining structure. We have demonstrated that for one bargaining structure at least, the Coase theorem cannot be proved.

The Coase theorem fundamentally derives its importance from a world with positive transaction costs. It makes it clear that the existence and rationalization of specific contractual arrangements rests on an examination of the transaction costs associated with potential (market or nonmar-

[11] For a brief description of the Shapley and Von Neumann-Morgenstern solutions see Michael D. Intriligator, Mathematical Optimization and Economic Theory (1971).

ket) exchanges which would arise in the absence of such contracts. The nonexistence of the core furnishes yet a further rationale for the existence of particular contractual arrangements in that these contracts may be useful in eliminating potential bargaining situations that defy solution. Contractual arrangements will evolve to force a solution on the bargaining process.

Empirically, however, it is difficult to differentiate between contractual arrangements that arise because of the nonexistence of the core from those that arise from transaction costs when there is a core. The nonexistence of the core will also manifest itself in transaction costs either because of the opportunity cost of (negotiation) time or because of the erosion of the potential exchange opportunity as it is postponed. One should not make the mistake of assuming, however, that once "time is money" this will force a Pareto-optimal solution. No one knows what actually happens when the core is empty. Do the participants agree to a Pareto-optimal solution or will they merely stop negotiating? This is an empirical question.

Finally, we would like to emphasize that the simplicity of our example belies its potential generality. Examples of negotiations where the core does not exist are easy to concoct and may be quite common in practice. External diseconomies, increasing returns to scale, and nonconvex preference functions are all factors which may lead to an empty core.[12]

[12] See Lloyd S. Shapley & Martin Shubik, On the Core of an Economic System with Externalities, 59 Am. Econ. Rev. 678 (1969); and Telser, *supra* note 4, at 81.