JOURNAL OF ECONOMIC THEORY 29, 265–281 (1983)

Efficient Mechanisms for Bilateral Trading*

ROGER B. MYERSON AND MARK A. SATTERTHWAITE

J. L. Kellogg Graduate School of Management, Northwestern University, Evanston, Illinois 60201

Received June 8, 1981; revised December 7, 1981

We consider bargaining problems between one buyer and one seller for a single object. The seller's valuation and the buyer's valuation for the object are assumed to be independent random variables, and each individual's valuation is unknown to the other. We characterize the set of allocation mechanisms that are Bayesian incentive compatible and individually rational, and show the general impossibility of ex post efficient mechanisms without outside subsidies. For a wide class of problems we show how to compute mechanisms that maximize expected total gains from trade, and mechanisms that can maximize a broker's expected profit. *Journal of Economic Literature* Classification Number: 026.

1. INTRODUCTION

Vickrey [7] showed the fundamental impossibility of designing a mechanism for negotiating the terms of trade in such a way that (i) honest revelation of supply and demand curves is a dominant strategy for all individuals, (ii) no outside subsidy is needed, and (iii) the final allocation of goods is always Pareto-efficient ex post. D'Aspremont and Gerard-Varet [3] weakened the incentive criterion from dominant-strategy to Bayesian Nash equilibrium, and showed that Bayesian incentive-compatible mechanisms could achieve efficient allocations without outside subsidies. However, the mechanisms of D'Aspremont and Gerard-Varet may give negative expected gains from trade to some individuals. That is, an individual who already knows his true preferences (but still does not know the preferences of the other individuals) may expect to do worse in the d'Aspremont-Gerard-Varet mechanism than if no trade took place. In this paper we prove some results relating to the efficiency properties of Bayesian incentive-compatible mechanisms that are individually rational, in the sense that each individual expects nonnegative gains from trade in any state of his preferences. We restrict our attention here to the simplest trading problems, where two

*Research for this paper was supported by the Kellogg Center for Advanced Study in Managerial Economics and by the National Science Foundation.

0022-0531/83 $3.00

individuals, one of whom has a single indivisible object to sell to the other, attempt to agree on an exchange of the object for money.

Chatterjee and Samuelson [2] have studied this two-person trading problem. For some specific bargaining games, they characterized the Bayesian equilibria. Chatterjee [1] has also studied the impossibility of simple mechanisms satisfying efficiency and individual rationality. In this paper we analyze a more general class of mechanisms, using some techniques similar to those developed in Myerson [6] to analyze optimal auction design. We show, as an application of our results, that if the traders' priors about each other's reservation prices are symmetric and uniform, then one of the games studied by Chatterjee and Samuelson [2] has equilibria that result in maximal expected gains from trade.

In the context of public goods economies, Laffont and Maskin [4] have also studied Bayesian incentive-compatible mechanisms that achieve ex post efficiency. Using a differentiability assumption (that the efficient level of the public goods depends differentiably on the consumers' types), they have shown for a very general class of problems that ex post efficiency and individual rationality may be incompatible. Although their differentiability assumption cannot be used in the trading problems which we study here, we derive an impossibility result (Corollary 1) that is closely related to this result.

The plan of this paper is as follows. In Section 2, we first define the formal structure of the bilateral trading problem. We then present a general characterization of all rules for transferring the object (as a function of the traders' valuations) that can be implemented by incentive-compatible individually-rational mechanisms. In Section 3, we show it is generally impossible to have a mechanism that is incentive-compatible, individually-rational, and ex post efficient, in the sense that it transfers the object to the buyer if and only if his valuation for the object is higher. In proving this result, we also show how to compute the smallest lump-sum subsidy that would be required from an outside party to make such an ex post efficient mechanism possible. In Section 4, we show how to construct mechanisms which maximize the expected total gains from trade, subject to the constraints of individual rationality and incentive compatibility. In Section 5, we consider the case where the traders are intermediated by a broker, who may either subsidize or exploit their desire to trade. We characterize the incentive-compatible individually rational trading mechanisms feasible with a broker, and we show how to construct mechanisms that maximize the broker's expected profit.

2. Incentive Compatibility and Individual Rationality

Let us consider a trading problem where individual #1 owns an object that individual #2 wants to buy. We let $\tilde{V}_1$ and $\tilde{V}_2$ denote the values of the object to the seller (#1) and the buyer (#2), respectively, and assume that these two valuations are independent random variables, with each $\tilde{V}_i$ distributed over a given interval from a_i to b_i. Let $f_1(\cdot)$ and $f_2(\cdot)$ be the probability density functions for $\tilde{V}_1$ and $\tilde{V}_2$, respectively. We assume that each $f_i(\cdot)$ is continuous and positive on its domain $[a_i, b_i]$. We let $F_1(\cdot)$ and $F_2(\cdot)$ be the cumulative distribution functions corresponding to f_1 and f_2 (so $f_i = F_i'$).

We assume that each individual knows his own valuation at the time of bargaining, but he considers the other's valuation as a random variable, distributed as above. Thus, to guarantee that each individual is willing to participate in a bargaining mechanism, the appropriate individual rationality constraint is that the mechanism gives each individual i nonnegative expected gains from trade in the mechanism, regardless of his given valuation $\tilde{V}_i$. Finally, we assume that the individuals are risk neutral and have additively separable utility for money and the object.

These two individuals are going to participate in some bargaining game or mechanism to determine, first, whether the object should be transferred from the seller to the buyer and, second, how much the buyer should pay to the seller. Our general question is: what kinds of bargaining mechanisms can be designed that have good economic efficiency properties?

A *direct bargaining mechanism* is one in which each individual simultaneously reports his valuation to a coordinater or broker who then determines whether the object is transferred, and how much the buyer must pay. A direct mechanism is thus characterized by two *outcome functions*, denoted by $p(\cdot, \cdot)$ and $x(\cdot, \cdot)$, where $p(v_1, v_2)$ is the probability that the object is transferred to the buyer and $x(v_1, v_2)$ is the expected payment from buyer to seller if v_1 and v_2 are the reported valuations of the seller and buyer. A direct mechanism is *(Bayesian) incentive-compatible* if honest reporting forms a Bayesian Nash equilibrium. That is, in an incentive-compatible mechanism, each individual can maximize his expected utility by reporting his true valuation, given that the other is expected to report honestly.

We can, without any loss of generality, restrict our attention to incentive-compatible direct mechanisms. This is because, for any Bayesian equilibrium of any bargaining game, there is an equivalent incentive-compatible direct mechanism that always yields the same outcomes (when the individuals play the honest equilibrium). This result, which is well known and very general, is called the *revelation principle*. The essential idea is that, given any equilibrium of any bargaining game, we can construct an equivalent incentive-compatible direct mechanism by first asking the buyer and seller each to confidentially report his valuation, then computing what each would have done in the given equilibrium strategies with these valuations, and then implementing the outcome (transfer of money and object) as in the given game for this computed behavior. If either individual had any incentive to lie to us in this direct mechanism, then he would have had an incentive to lie to himself in the original game, which is a contradiction of the premise that he was in equilibrium in the original game. (For more on this revelation principle, see Myerson [5] and [6].)

Given a direct mechanism with outcome functions (p, x), we define that following quantities:

$$\bar{x}_1(v_1) = \int_{a_2}^{b_2} x(v_1, t_2) f_2(t_2)\, dt_2, \qquad \bar{x}_2(v_2) = \int_{a_1}^{b_1} x(t_1, v_2) f_1(t_1)\, dt_1,$$

$$\bar{p}_1(v_1) = \int_{a_2}^{b_2} p(v_1, t_2) f_2(t_2)\, dt_2, \qquad \bar{p}_2(v_2) = \int_{a_1}^{b_1} p(t_1, v_2) f_1(t_1)\, dt_1,$$

$$U_1(v_1) = \bar{x}_1(v_1) - v_1\, \bar{p}_1(v_1), \qquad U_2(v_2) = v_2 \bar{p}_2(v_2) - \bar{x}_2(v_2).$$

Thus $U_1(v_1)$ is the expected gains from trade for the seller if his valuation is v_1, since $\bar{x}_1(v_1)$ is his expected revenue and $\bar{p}_1(v_1)$ is his probability of losing the object given $\tilde{V}_1 = v_1$. Similarly, $U_2(v_2)$ is the expected gains from trade for the buyer, $\bar{x}_2(v_2)$ is the buyer's expected payment, and $\bar{p}_2(v_2)$ is the buyer's probability of getting the object, if his valuation is v_2.

In our formal notation, we say that (p, x) is *incentive-compatible* (in the Bayesian sense) iff for every v_1 and $\hat{v}_1$ in $[a_1, b_1]$,

$$U_1(v_1) \geqslant \bar{x}_1(\hat{v}_1) - v_1\, \bar{p}_1(\hat{v}_1),$$

and for every v_2 and $\hat{v}_2$ in $[a_2, b_2]$,

$$U_2(v_2) \geqslant v_2 \bar{p}_2(\hat{v}_2) - \bar{x}_2(\hat{v}_2).$$

These two inequalities assert that neither individual should expect to gain by reporting valuation $\hat{v}_i$ when v_i is true. The mechanism (p, x) is *individually rational* iff

$$U_1(v_1) \geqslant 0 \text{ and } U_2(v_2) \geqslant 0$$

for every v_1 in $[a_1, b_1]$ and for every v_2 in $[a_2, b_2]$. That is, individual rationality requires that each individual have nonnegative expected gains from trade after he knows his own valuation, but before he learns the other's valuation. (We do not require

$$x(v_1, v_2) - v_2 p(v_1, v_2) \geqslant 0 \text{ or } v_2 p(v_1, v_2) - x(v_1, v_2) \geqslant 0 \text{ ex post.})$$

We can now state and prove our main result.

THEOREM 1. *For any incentive-compatible mechanism,*

$$U_1(b_1) + U_2(a_2) = \underset{v_1 \in [a_1, b_1]}{\text{minimum}} (U_1(v_1)) + \underset{v_2 \in [a_2, b_2]}{\text{minimum}} (U_2(v_2))$$

$$= \int_{a_2}^{b_2} \int_{a_1}^{b_1} \left(\left[v_2 - \frac{1 - F_2(v_2)}{f_2(v_2)} \right] - \left[v_1 + \frac{F_1(v_1)}{f_1(v_1)} \right] \right)$$

$$\times p(v_1, v_2) f_1(v_1) f_2(v_2) \, dv_1 \, dv_2. \tag{1}$$

Furthermore, if $p(\cdot, \cdot)$ is any function mapping $[a_1, b_1] \times [a_2, b_2]$ into $[0, 1]$, *then there exists a function $x(\cdot, \cdot)$ such that (p, x) is incentive-compatible and individually-rational if and only if $\bar{p}_1(\cdot)$ is weakly decreasing, $\bar{p}_2(\cdot)$ is weakly increasing, and*

$$0 \leqslant \int_{a_2}^{b_2} \int_{a_1}^{b_1} \left(\left[v_2 - \frac{1 - F_2(v_2)}{f_2(v_2)} \right] - \left[v_1 + \frac{F_1(v_1)}{f_1(v_1)} \right] \right)$$

$$\times p(v_1, v_2) f_1(v_1) f_2(v_2) \, dv_1 \, dv_2. \tag{2}$$

Proof of Theorem 1. Suppose first that we are given an incentive-compatible mechanism (p, x). By incentive compatibility, we know that, for any two possible valuations v_1 and $\hat{v}_1$ for the seller,

$$U_1(v_1) = \bar{x}_1(v_1) - v_1 \, \bar{p}_1(v_1) \geqslant \bar{x}_1(\hat{v}_1) - v_1 \, \bar{p}_1(\hat{v}_1),$$

and

$$U_1(\hat{v}_1) = \bar{x}_1(\hat{v}_1) - \hat{v}_1 \, \bar{p}_1(\hat{v}_1) \geqslant \bar{x}_1(v_1) - \hat{v}_1 \, \bar{p}_1(v_1).$$

These two inequalities imply that

$$(\hat{v}_1 - v_1) \, \bar{p}_1(v_1) \geqslant U_1(v_1) - U_1(\hat{v}_1) \geqslant (\hat{v}_1 - v_1) \, \bar{p}_1(\hat{v}_1). \tag{3}$$

Thus, if $\hat{v}_1 > v_1$, we must have $\bar{p}_1(\hat{v}_1) \leqslant \bar{p}_1(v_1)$, so $\bar{p}_1(\cdot)$ is decreasing. Furthermore, since $\bar{p}_1(\cdot)$ is decreasing, it is Riemann integrable, and so (3) implies that $U_1'(v_1) = -\bar{p}_1(v_1)$ at almost every v_1 and

$$U_1(v_1) = U_1(b_1) + \int_{v_1}^{b_1} \bar{p}_1(t_1) \, dt_1. \tag{4}$$

A similar argument for the buyer shows that

$$(\hat{v}_2 - v_2) \, \bar{p}_2(\hat{v}_2) \geqslant U_2(\hat{v}_2) - U_2(v_2) \geqslant (\hat{v}_2 - v_2) \, \bar{p}_2(v_2).$$

Thus, $\bar{p}_2(\cdot)$ is increasing, $U_2'(v_2)=\bar{p}_2(v_2)$ almost everywhere, and

$$U_2(v_2)=U_2(a_2)+\int_{a_2}^{v_2}\bar{p}_2(t_2)\,dt_2. \tag{5}$$

Equations (4) and (5) imply that $U_1(\cdot)$ is decreasing and $U_2(\cdot)$ is increasing. Furthermore, we get:

$$\begin{aligned}
&\int_{a_2}^{b_2}\int_{a_1}^{b_1}(v_2-v_1)\,p(v_1,v_2)\,f_1(v_1)\,f_2(v_2)\,dv_1\,dv_2\\
&=\int_{a_1}^{b_1}U_1(v_1)\,f_1(v_1)\,dv_1+\int_{a_2}^{b_2}U_2(v_2)\,f_2(v_2)\,dv_2\\
&=U_1(b_1)+\int_{a_1}^{b_1}\int_{v_1}^{b_1}\bar{p}_1(t_1)\,dt_1 f_1(v_1)\,dv_1\\
&\quad+U_2(a_2)+\int_{a_2}^{b_2}\int_{a_2}^{v_2}\bar{p}_2(t_2)\,dt_2\,f_2(v_2)\,dv_2\\
&=U_1(b_1)+U_2(a_2)+\int_{a_1}^{b_1}F_1(t_1)\,\bar{p}_1(t_1)\,dt_1\\
&\quad+\int_{a_2}^{b_2}(1-F_2(t_2))\,\bar{p}_2(t_2)\,dt_2\\
&=U_1(b_1)+U_2(a_2)\\
&\quad+\int_{a_2}^{b_2}\int_{a_1}^{b_1}(F_1(t_1)\,f_2(t_2)\\
&\quad+(1-F_2(t_2))\,f_1(t_1))\,p(t_1,t_2)\,dt_1\,dt_2.
\end{aligned}$$

Equating the first and last of these expressions gives us Eq. (1) of Theorem 1, which in turn implies inequality (2) when the mechanism is individually-rational. Thus, we have proven the first sentence and the "only if" part of the second sentence in Theorem 1.

To complete the proof of Theorem 1, suppose now that $p(\cdot,\cdot)$ satisfies (2), $\bar{p}_1(\cdot)$ is decreasing, and $\bar{p}_2(\cdot)$ is increasing. We must construct the payment function $x(\cdot,\cdot)$ so that (p,x) is an individually rational, incentive-compatible mechanism. There are many such functions which could be used; we shall consider a function defined as follows:

$$\begin{aligned}
x(v_1,v_2)=&\int_{t_2=a_2}^{v_2}t_2\,d[\bar{p}_2(t_2)]-\int_{t_1=a_1}^{v_1}t_1\,d[-\bar{p}_1(t_1)]\\
&+a_2\bar{p}_2(a_2)+\int_{t_1=a_1}^{b_1}t_1(1-F_1(t_1))\,d[-\bar{p}_1(t_1)].
\end{aligned} \tag{6}$$

Notice that, since $\bar{p}_2(\cdot)$ and $-\bar{p}_1(\cdot)$ are monotone increasing, each integral in (6) is a nonnegative quantity (assuming $a_1 \geqslant 0$ and $a_2 \geqslant 0$). The first term in (6) depends only on v_2, the second term depends only on v_1, and the last two terms represent a constant chosen to give

$$U_2(a_2) = a_2 \bar{p}_2(a_2) - \int_{a_1}^{b_1} x(v_1, a_2) f_1(v_1)\, dv_1 = 0.$$

(Notice that $x(v_1, v_2)$ is paid by the buyer even if he does not get the object.) To check incentive-compatibility of (6), observe that

$$\begin{aligned}
&(v_2 \bar{p}_2(v_2) - v_2 \bar{p}_2(\hat{v}_2) - (\bar{x}_2(v_2) - \bar{x}_2(\hat{v}_2)) \\
&= v_2 \int_{t_2=\hat{v}_2}^{v_2} d[\bar{p}_2(t_2)] - \int_{t_2=\hat{v}_2}^{v_2} t_2\, d[\bar{p}_2(t_2)] \\
&= \int_{t_2=\hat{v}_2}^{v_2} (v_2 - t_2)\, d[\bar{p}_2(t_2)] \geqslant 0.
\end{aligned}$$

So the buyer would do better reporting v_2 rather than $\hat{v}_2$ if his true valuation is v_2. (This argument holds even if $\hat{v}_2 > v_2$, in which case the integrand is negative, but the direction of integration is backwards, giving a nonnegative integral overall.) The proof of incentive-compatibility for the seller is analogous.

Thus, since (p, x) is incentive-compatible, (1) applies. Because $U_2(a_2) = 0$, and because we have assumed (2), we must have $U_1(b_1) \geqslant 0$. By the monotonicity properties of $U_1(\cdot)$ and $U_2(\cdot)$, it suffices to check individual-rationality for the buyer's lowest valuation and the seller's highest valuation. Thus, our proof of Theorem 1 is complete. ∎

3. Ex Post Efficiency

A mechanism (p, x) is *ex post efficient* iff

$$\begin{aligned}
p(v_1, v_2) &= 1 \qquad \text{if} \quad v_1 < v_2, \\
&= 0 \qquad \text{if} \quad v_1 > v_2.
\end{aligned}$$

That is, in an ex post efficient mechanism, the buyer gets the object whenever his valuation is higher, and the seller keeps the object whenever his valuation is higher. For such a mechanism, $\bar{p}_1(v_1) = 1 - F_2(v_1)$, which is decreasing, and $\bar{p}_2(v_2) = F_1(v_2)$, which is increasing. Thus, to check whether ex post efficient mechanisms are feasible, it only remains to check whether inequality

(2) in Theorem 1 holds for this $p(\cdot, \cdot)$. In fact, for an ex post efficient mechanism, we get:

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1}\left(\left[v_2-\frac{1-F_2(v_2)}{f_2(v_2)}\right]-\left[v_1+\frac{F_1(v_1)}{f_1(v_1)}\right]\right)p(v_1,v_2)\,f_1(v_1)\,f_2(v_2)\,dv_1\,dv_2$$

$$=\int_{a_2}^{b_2}\int_{a_1}^{\min\{v_2,b_1\}}[v_2f_2(v_2)+F(v_2)-1]\,f_1(v_1)\,dv_1\,dv_2$$

$$-\int_{a_2}^{b_2}\int_{a_1}^{\min\{v_2,b_1\}}[v_1f_1(v_1)+F_1(v_1)]\,dv_1\,f_2(v_2)\,dv_2$$

$$=\int_{a_2}^{b_2}[v_2f_2(v_2)+F_2(v_2)-1]\,F_1(v_2)\,dv_2$$

$$-\int_{a_2}^{b_2}\min\{v_2F_1(v_2),b_1\}\,f_2(v_2)\,dv_2$$

$$=-\int_{a_2}^{b_2}(1-F_2(v_2))\,F_1(v_2)\,dv_2+\int_{b_1}^{b_2}(v_2-b_1)\,f_2(v_2)\,dv_2$$

$$=-\int_{a_2}^{b_2}(1-F_2(v_2))\,F_1(v_2)\,dv_2-\int_{b_1}^{b_2}(F_2(v_2)-1)\,dv_2$$

$$=-\int_{a_2}^{b_1}(1-F_2(t))\,F_1(t)\,dt.$$

Thus, if $a_2 < b_1$ and $a_1 < b_2$, so that the two valuation-intervals properly intersect, then any Bayesian incentive-compatible mechanism which is ex post efficient must give

$$U_1(b_1)+U_2(a_2)=-\int_{a_2}^{b_1}(1-F_2(t))\,F_1(t)\,dt<0, \tag{7}$$

and so it cannot be individually rational. Thus, this quantity

$$\int_{a_2}^{b_1}(1-F_2(t))\,F_1(t)\,dt$$

is the smallest lump-sum subsidy required from an outside party to create a Bayesian incentive-compatible mechanism which is both ex post efficient and individually rational. (Thus far in this paper we have been assuming that no such subsidy is actually available. We will consider the case of trading with a broker more generally in Section 5.)

To summarize, we have shown the following result.

COROLLARY 1. *If the seller's valuation is distributed with positive probability density over the interval* $[a_1, b_1]$, *and the buyer's valuation is distributed with positive probability density over the interval* $[a_2, b_2]$, *and if the interiors of these intervals have a nonempty intersection, then no incentive-compatible individually rational trading mechanism can be ex post efficient.*

It should be noted that our proofs have used the assumption that the valuations have positive density over their respective intervals. Without this assumption the corollary is untrue. For example, suppose that $[a_1, b_1] = [1, 4]$ and $[a_2, b_2] = [0, 3]$, but all the probability mass is concentrated at the endpoints, with

$$\Pr(\tilde{V}_1 = 1) = \Pr(\tilde{V}_1 = 4) = \tfrac{1}{2} = \Pr(\tilde{V}_2 = 0) = \Pr(\tilde{V}_2 = 3).$$

Then the mechanism "sell at price 2 if both are willing (otherwise no trade)" is incentive-compatible, individually-rational, and ex post efficient. But if we admit any small positive probability density over the whole of these intersecting intervals, then neither this mechanism nor any other feasible mechanism can satisfy ex post efficiency.

To apply this theory, consider the case where $\tilde{V}_1$ and $\tilde{V}_2$ are both uniformly distributed on $[0, 1]$. Then, each $F_i(v_i) = v_i$ and $f_i(v_i) = 1$ on this interval, and the constraint (2) becomes

$$0 \leqslant \int_0^1 \int_0^1 ([2v_2 - 1] - [2v_1])\, p(v_1, v_2)\, dv_1\, dv_2$$

$$= 2 \int_0^1 \int_0^1 (v_2 - v_1 - \tfrac{1}{2})\, p(v_1, v_2)\, dv_1\, dv_2.$$

Thus, conditional on the individuals reaching an agreement to trade, the expected difference in their valuations must be at least 1/2. This conclusion holds when these traders use rational equilibrium strategies, no matter what the rules of their bargaining game might be (provided only that either trader could refuse to participate if his expected gains from trade were negative, so that individual rationality is guaranteed). Conditional on the buyer's valuation being higher than the seller's, the expected difference $\tilde{V}_2 - \tilde{V}_1$ would be only

$$\int_0^1 \int_0^{v_2} 2(v_2 - v_1)\, dv_1\, dv_2 = \frac{1}{3}$$

in this problem. Thus, ex post efficiency cannot be achieved by any individually-rational mechanism, unless some outsider is willing to provide a subsidy of at least

$$\int_0^1 (1-t)\, t\, dt = \frac{1}{6}$$

to the traders for participating in the bargaining game.

4. Maximizing Expected Total Gains from Trade

The expected total gains from trade in a mechanism is just

$$\int_{a_1}^{b_1} U_1(v_1) f_1(v_1)\, dv_1 + \int_{a_2}^{b_2} U_2(v_2) f_2(v_2)\, dv_2$$
$$= \int_{a_2}^{b_2} \int_{a_1}^{b_1} (v_2 - v_1)\, p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2.$$

Since ex post efficiency is unattainable, it is natural to seek a mechanism that maximizes expected total gains from trade, subject to the incentive-compatibility and individual-rationality constraints. (Of course, other collective objective functions could also be considered, but here we shall just consider the problem of maximizing expected gains from trade.) With a bit of machinery, we can show how to solve this problem for a wide class of examples.

First, we define some new functions. For any number $\alpha \geqslant 0$, let

$$c_1(v_1, \alpha) = v_1 + \alpha \frac{F_1(v_1)}{f_1(v_1)}, \qquad c_2(v_2, \alpha) = v_2 - \alpha \frac{1 - F_2(v_2)}{f_2(v_2)}.$$

Let $p^\alpha(\cdot, \cdot)$ be defined by

$$\begin{aligned} p^\alpha(v_1, v_2) &= 1 \quad \text{if } c_1(v_1, \alpha) \leqslant c_2(v_2, \alpha), \\ &= 0 \quad \text{if } c_1(v_1, \alpha) > c_2(v_2, \alpha). \end{aligned}$$

Notice that p^0 is the ex post efficient outcome function (transferring the object iff $v_1 \leqslant v_2$), whereas p^1 maximizes the integral in inequality (2).

THEOREM 2. *If there exists an incentive-compatible mechanism* (p, x) *such that* $U_1(b_1) = U_2(a_2) = 0$ and $p = p^\alpha$ *for some* α *in* $[0, 1]$, *then this mechanism maximizes the expected total gains from trade among all incentive-compatible individually-rational mechanisms. Furthermore, if* $c_1(\cdot, 1)$ *and* $c_2(\cdot, 1)$ *are increasing functions on* $[a_1, b_1]$ *and* $[a_2, b_2]$, *respec-*

tively, and if the interiors of these two intervals have a nonempty intersection, then such a mechanism must exist.

Proof of Theorem 2. Consider the problem to choose $p: [a_1, b_1] \times [a_2, b_2] \to [0, 1]$ to maximize

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} (v_2 - v_1) p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2$$

subject to constraint (2), that is,

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} (c_2(v_2, 1) - c_1(v_1, 1)) p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2 \geqslant 0$$

If the solution to this problem happens to give us $\bar{p}_1(\cdot)$ and $\bar{p}_2(\cdot)$, which are monotone decreasing and increasing, respectively, then by Theorem 1, this $p(\cdot, \cdot)$ function will be associated with a mechanism which maximizes the expected gains from trade among all incentive-compatible individually rational mechanisms. If we multiply the integral in the constraint by $\lambda \geqslant 0$ and add it to the objective function, then we get the Lagrangian for this problem; this Lagrangian equals

$$\begin{aligned}
&\iint (v_2 + \lambda c_2(v_2, 1) - v_1 - \lambda c_1(v_1, 1)) p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2 \\
&\quad = (1 + \lambda) \iint \left(c_2 \left(v_2, \frac{\lambda}{1 + \lambda} \right) - c_1 \left(v_1, \frac{\lambda}{1 + \lambda} \right) \right) \\
&\qquad \times p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2 .
\end{aligned}$$

Any $p(\cdot, \cdot)$ function that satisfies the constraint with equality and maximizes the Lagrangian for some $\lambda \geqslant 0$ must be a solution for our problem. But the Lagrangian is maximized by p^α, when $\alpha = \lambda/(1 + \lambda)$; and the constraint (2) will be satisfied with equality if $U_1(b_1) = U_2(a_2) = 0$. This proves the first sentence in Theorem 2.

Now, suppose that $c_1(\cdot, 1)$ and $c_2(\cdot, 1)$ are both increasing. Then for every α between 0 and 1, $c_1(\cdot, \alpha)$ and $c_2(\cdot, \alpha)$ are increasing functions, which, in turn, implies that $p^\alpha(v_1, v_2)$ is increasing in v_2 and decreasing in v_1. So all $\bar{p}_1^\alpha$ and $\bar{p}_2^\alpha$ have the necessary monotonicity properties.

Let

$$G(\alpha) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} (c_2(v_2, 1) - c_1(v_1, 1)) p^\alpha(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2 .$$

Clearly, $G(1) \geqslant 0$, since p^1 is positive only when $c_2(v_2, 1) \geqslant c_1(v_1, 1)$. Furthermore, $G(\alpha)$ is increasing in α. To prove this, observe that

$$c_2(v_2, \alpha) - c_1(v_1, \alpha) = (v_2 - v_1) - \alpha \left(\frac{1 - F_2(v_2)}{f_2(v_2)} + \frac{F_1(v_1)}{f_1(v_1)}\right),$$

which is decreasing in α, so $p^\alpha(v_1, v_2)$ is decreasing in α. Thus, for $\alpha < \beta$, $G(\beta)$ differs from $G(\alpha)$ only because $0 = p^\beta(v_1, v_2) < p^\alpha(v_1, v_2) = 1$ for some (v_1, v_2) where $c_2(v_2, \beta) < c_1(v_1, \beta)$ and so $c_2(v_2, 1) < c_1(v_1, 1)$. Thus, $G(\cdot)$ is increasing.

To prove that $G(\cdot)$ is continuous, observe that, if each $c_i(v_i, 1)$ is increasing in v_i, then each $c_i(v_i, \alpha)$ is strictly increasing in v_i, for any $\alpha < 1$. So given v_2 and α, the equation $c_1(v_1, \alpha) = c_2(v_2, \alpha)$ has at most one solution in v_1, and this solution varies continuously in v_2 and α. Thus, we may write

$$G(\alpha) = \int_{a_2}^{b_2} \int_{a_1}^{g(v_2,\alpha)} (c_2(v_2, 1) - c_1(v_1, 1)) f_1(v_1) f_2(v_2)\, dv_1\, dv_2,$$

where $g(v_2, \alpha)$ is continuous in v_2 and α. So $G(\cdot)$ is continuous, increasing, and $G(1) \geqslant 0$. But $G(0) < 0$, because otherwise p^0 would be an ex post efficient individually-rational mechanism, which is impossible by Corollary 1. Thus there must be some α in $(0, 1]$ such that $G(\alpha) = 0$ and p^α can satisfy the conditions in Theorem 2. This completes the proof of Theorem 2. ∎

For an example, consider again the case where $\tilde{V}_1$ and $\tilde{V}_2$ are both uniform random variables on $[0, 1]$. Then,

$$c_1(v_1, \alpha) = v_1 + \alpha v_1 = (1 + \alpha)\, v_1$$

and

$$c_2(v_2, \alpha) = v_2 - \alpha(1 - v_2) = (1 + \alpha)\, v_2 - \alpha.$$

Both of these functions are monotone increasing when $\alpha = 1$, so we know that the expected gains from trade are maximized by $p = p^\alpha$ for some α between 0 and 1. To get $U_1(b_1) = U_2(a_2) = 0$ for $p = p^\alpha$, we must have (by Eq. (1))

$$0 = \int_0^1 \int_0^1 ([2v_2 - 1] - [2v_2])\, p^\alpha(v_1, v_2)\, dv_1\, dv_2.$$

But in this case

$$\begin{aligned} p^\alpha(v_1, v_2) &= 1 \quad \text{if} \quad (1+\alpha)\, v_1 \leqslant (1+\alpha)\, v_2 - \alpha, \\ &= 0 \quad \text{if} \quad (1+\alpha)\, v_1 > (1+\alpha)\, v_2 - \alpha. \end{aligned}$$

So the above equation becomes

$$0=\int_{\alpha/(1+\alpha)}^{1}\int_{0}^{v_2-\alpha/(1+\alpha)}[2v_2-1-2v_1]\,dv_1\,dv_2=\frac{3\alpha-1}{6(1+\alpha)^3}.$$

So we must have $\alpha=1/3$, and then $p=p^{\alpha}$ implies

$$\begin{aligned}p(v_1,v_2)&=1 \text{ if } v_1\leqslant v_2-\tfrac{1}{4},\\ &=0 \text{ if } v_1>v_2-\tfrac{1}{4}.\end{aligned}$$

So the expected gains from trade are maximized by a mechanism which transfers the object iff the buyer's valuation exceeds the seller's by at least 1/4.

Such a mechanism has indeed been found for this example. Chatterjee and Samuelson [2] have studied a bargaining game with the following rules. The buyer and seller each simultaneously propose a price. If the buyer's price is higher than the seller's, then the object is sold at the average of the two; otherwise the seller keeps the object. This mechanism is *not* incentive-compatible. In fact, Chatterjee and Samuelson have shown that in the equilibrium strategies for this example, the seller proposes price $2/3\,v_1+1/4$ and the buyer proposes price $2/3\,v_2+1/12$. Thus, the object is sold iff

$$\frac{2}{3}v_1+\frac{1}{4}\leqslant\frac{2}{3}v_2+\frac{1}{12}\qquad\text{or}\qquad v_1\leqslant v_2-\frac{1}{4},$$

and the sale price is

$$\frac{1}{2}\left(\frac{2}{3}v_1+\frac{1}{4}\right)+\frac{1}{2}\left(\frac{2}{3}v_2+\frac{1}{12}\right)=\frac{1}{3}\left(v_1+v_2+\frac{1}{2}\right).$$

In this equilibrium, there is no possibility of trade if $\tilde{V}_1=1$ or $\tilde{V}_2=0$, so $U_1(1)=U_2(0)=0$. Thus, this equilibrium of this split-the-difference game gives the highest expected total gains from trade among all equilibria of all bargaining games satisfying individual-rationality for this symmetric-uniform trading problem.

To illustrate the revelation principle, we may point out that the incentive-compatible direct mechanism equivalent to the Chatterjee-Samuelson equilibrium for this example is to have

$$\begin{aligned}p(v_1,v_2)&=1 \qquad \text{if}\quad v_1\leqslant v_2-\tfrac{1}{4},\\ &=0 \qquad \text{otherwise,}\end{aligned}$$

$$\begin{aligned}x(v_1,v_2)&=(v_1+v_2+.5)/3 \qquad \text{if}\quad v_1\leqslant v_2-\tfrac{1}{4},\\ &\quad 0 \qquad\qquad\qquad\quad \text{otherwise.}\end{aligned}$$

That is, the direct mechanism simply implements what the original game would have done after the players used their equilibrium strategies.

5. Trading with a Broker

Thus far, we have assumed that the payments from the buyer must equal the payments to the seller, because there is no outside source (or sink) of funds. Let us now drop this assumption, and allow a third party to either subsidize or exploit the buyer and seller. We shall refer to this third party as the *broker.*

Let us assume that the broker can be a net source or sink of money, but he cannot himself own the object. Thus, a trading mechanism with a broker is characterized by three outcome functions (p, x_1, x_2), where $p(v_1, v_2)$ is the probability that the object is transferred from the seller to the buyer, $x_1(v_1, v_2)$ is the expected payment from broker to seller, and $x_2(v_1, v_2)$ is the expected payment from buyer to broker, if v_1 and v_2 are the reported valuations of seller and buyer. Given such a mechanism, we define $\bar{x}_1(v_1)$, $\bar{x}_2(v_2)$, $\bar{p}_1(v_1)$, $\bar{p}_2(v_2)$, $U_1(v_1)$ and $U_2(v_2)$ exactly as in Section 2, except that we must modify the definitions of $\bar{x}_1(v_1)$ and $\bar{x}_2(v_2)$ to:

$$\bar{x}_1(v_1) = \int_{a_2}^{b_2} x_1(v_1, t_2) f_2(t_2)\, dt_2,$$

$$\bar{x}_2(v_2) = \int_{a_1}^{b_1} x_2(t_1, v_2) f_1(t_1)\, dt_1.$$

In addition, we let U_0 denote the expected net profit for the broker, so that

$$U_0 = \int_{a_2}^{b_2} \int_{a_1}^{b_1} (x_2(t_1, t_2) - x_1(t_1, t_2)) f_1(t_1) f_2(t_2)\, dt_1\, dt_2.$$

The definitions of incentive-compatibility and individual-rationality for a trading mechanism with broker are the same as in Section 2: neither buyer nor seller should ever expect to gain by lying about his valuation, and both traders should get nonnegative expected gains from trade in the mechanism. As before, there is no loss of generality in restricting our attention to incentive-compatible mechanisms, because any Bayesian equilibrium of any trading game with broker can be simulated by an equivalent incentive-compatible direct mechanism.

For any v_1 and v_2, let

$$C_1(v_1) = c_1(v_1, 1) = v_1 + \frac{F_1(v_1)}{f_1(v_1)},$$

$$C_2(v_2) = c_2(v_2, 1) = v_2 - \frac{1 - F_1(v_1)}{f_2(v_2)}.$$

With this formulation we can extend Theorem 1, as follows.

THEOREM 3. *For any incentive-compatible mechanism with a broker, $\bar{p}_1(\cdot)$ is weakly decreasing, $\bar{p}_2(\cdot)$ is weakly increasing, and*

$$U_0 + U_1(b_1) + U_2(a_2) = U_0 + \underset{v_1 \in [a_1, b_1]}{\text{minimum}} (U_1(v_1)) + \underset{v_2 \in [a_2, b_2]}{\text{minimum}} (U_2(v_2))$$

$$= \int_{a_2}^{b_2} \int_{a_1}^{b_1} (C_2(v_2) - C_1(v_1))\, p(v_1, v_2)\, f_1(v_1)\, f_2(v_2)\, dv_1\, dv_2. \tag{8}$$

Proof of Theorem 3. The proof is exactly the same as the proof of Theorem 1 except that the string of equalities after Eq. (5) must begin with

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} (v_1 - v_2)\, p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2 - U_0$$

$$= \int_{a_1}^{b_1} U_1(v_1) f_1(v_1)\, dv_1 + \int_{a_2}^{b_2} U_2(v_2) f_2(v_2)\, dv_2.$$

(That is, the expected gains from trade minus the expected net profit to the broker must equal the expected gains to buyer and seller.) With this one change, the proof goes through exactly as before. ∎

Extending the results in Section 3, Theorem 3 implies that, for any ex post efficient mechanism with broker,

$$U_0 + U_1(b_1) + U_2(a_2) = -\int_{a_2}^{b_1} (1 - F_2(t))\, F_1(t)\, dt.$$

Thus, the minimum expected subsidy required from the broker, to achieve ex post efficiency with individual-rationality, is

$$\int_{a_2}^{b_1} (1 - F_2(t))\, F_1(t)\, dt,$$

even if the subsidy is not lump-sum.

Another interesting question is to ask for the mechanism which maximizes the expected profit to the broker, subject to incentive compatibility and individual rationality for the two traders. That is, if the buyer and seller can only trade through the broker, then what is the optimal mechanism for the broker?

THEOREM 4. *Suppose $C_1(\cdot)$ and $C_2(\cdot)$ are monotone increasing functions on $[a_1, b_1]$ and $[a_2, b_2]$, respectively. Then among all incentive-compatible, individually-rational mechanisms, the broker's expected profit is maximized by a mechanism in which the object is transferred to the buyer if and only if $C_2(\tilde{V}_2) \geqslant C_1(\tilde{V}_1)$.*

Proof of Theorem 4. From Theorem 3 we get

$$U_0 = \iint (C_2(v_2) - C_1(v_1))\, p(v_1, v_2) f_1(v_1) f_2(v_2)\, dv_1\, dv_2 - U_1(b_1) - U_2(a_2),$$

for any incentive-compatible mechanism. To maximize this expression subject to individual rationality, we want

$$\begin{aligned} p(v_1, v_2) &= 1 \qquad \text{if} \quad C_2(v_2) \geqslant C_1(v_1) \\ &= 0 \qquad \text{if} \quad C_2(v_2) < C_1(v_1), \end{aligned}$$

and $U_1(b_1) = U_2(a_2) = 0$. It only remains to construct x_1 and x_2 such that (p, x_1, x_2) satisfies these conditions and incentive compatibility. There are many ways to do this; one is to let

$$\begin{aligned} x_2(v_1, v_2) &= p(v_1, v_2) \cdot \min\{t_2 \mid t_2 \geqslant a_2 \text{ and } C_2(t_2) \geqslant C_1(v_1)\}, \\ x_1(v_1, v_2) &= p(v_1, v_2) \cdot \max\{t_1 \mid t_1 \leqslant b_1 \text{ and } C_1(t_1) \leqslant C_2(v_2)\}. \end{aligned}$$

That is, if there is a trade, then the broker charges the buyer the lowest valuation he could have quoted and still gotten the object (given the seller's valuation), and the broker pays to the seller the highest valuation which he could have quoted and still sold the object (given the buyer's valuation). If there is no trade, then there are no payments.

It is straightforward to show that this mechanism is incentive-compatible if the $C_1(\cdot)$ and $C_2(\cdot)$ functions are monotone, following the argument used by Vickrey [7] to show the incentive-compatibility of the second price auction. For example, if the seller reported a valuation higher than the truth, he would not affect the price he gets when he sells, but he would lose some opportunities to sell when he could have done so profitably in the mechanism. Similarly, if the seller reported a valuation lower than the truth, then he would only add possibilities of selling below his valuation. It is also straightforward to check that $U_1(b_1) = 0$ and $U_2(a_2) = 0$ in this mechanism, since it gives

$$x_1(b_1, v_2) = p(b_1, v_2)\, b_1, \qquad x_2(v_1, a_2) = p(v_1, a_2)\, a_2. \quad \blacksquare$$

For the case where $\tilde{V}_1$ and $\tilde{V}_2$ are both uniform on $[0, 1]$, the broker's optimal mechanism transfers the object if and only if

$$C_2(\tilde{V}_2) = 2\tilde{V}_2 - 1 \geqslant 2\tilde{V}_1 = C_1(\tilde{V}_1), \qquad \text{or} \quad \tilde{V}_2 - \tilde{V}_1 \geqslant \tfrac{1}{2}.$$

So the broker should offer to buy from the seller for $\tilde{V}_2 - 1/2$, and should offer to sell to the buyer for $\tilde{V}_1 + 1/2$, and trade occurs if and only if the traders are both willing to trade at these prices.

Comparing Theorems 2 and 4, we see that the broker's optimal trading mechanism has strictly less trading than the mechanism which maximizes the expected total gains from trade. In this symmetric-uniform example, the maximum expected total gains from trade is achieved by a mechanism in which the object is transferred whenever the buyer's valuation exceeds the seller's by at least 1/4, but the broker's optimal mechanism does not transfer the object unless the buyer's valuations exceeds the seller's by at least 1/2. That is, if a broker is to profitably exploit his control over the trading channel, he must actually restrict trade to some extent.

In both Theorems 2 and 4, monotonicity of the $C_i(\cdot)$ functions is required. This assumption is satisfied for a very wide class of distributions, but it is a restriction. The general case can be analyzed using the methods developed in Myerson [6] to construct optimal auctions in the general (nonmonotone) case.

References

1. K. Chatterjee, Incentive compatibility in bargaining under uncertainty, mimeograph, Pennsylvania State University, 1980. To appear in *Quarterly Journal of Economics.*
2. K. Chatterjee and W. Samuelson, "The Simple Economics of Bargaining," mimeographed, Pennsylvania State University and Boston University, 1979. To appear in *Operations Research.*
3. C. D'Aspremont and L. Gerard-Varet, Incentives and incomplete information, *J. Pub. Econom.* **11** (1979), 25–45.
4. J.-J. Laffont and E. Maskin, A differential approach to expected utility maximizing mechanisms, *in* "Aggregation and Revelation of Preferences," (J.-J. Laffont, Ed.), pp. 289–308, North-Holland, Amsterdam, 1979.
5. R. B. Myerson, Incentive compatibility and the bargaining problem, *Econometrica* **47** (1979), 61–73.
6. R. B. Myerson, Optimal auction design, *Math. Oper. Res.* **6** (1981), 58–73.
7. W. Vickrey, Counterspeculation, auctions, and competitive sealed tenders, *J. Finance* **16** (1961), 8–37.