

Myerson and Satterthwaite, "Efficient Mechanisms for Bilateral Trading," JET 1983

In many bilateral bargaining situations with asymmetric information, ex post efficiency is inconsistent with incentive compatibility and individual rationality.

One can compute the highest expected surplus consistent with IC and IR. If the equilibrium to a bargaining game yields that surplus, then the game constitutes an optimal mechanism.

Player 1 (seller) owns an indivisible object and has valuation distributed according to the continuous and positive density function, $f_1(v_1)$, over the support $[a_1, b_1]$. Also denote the distribution function as $F_1(v_1)$.

Player 2 (buyer) has valuation distributed according to the continuous and positive density function, $f_2(v_2)$, over the support $[a_2, b_2]$. Also denote the distribution function as $F_2(v_2)$. Types are assumed to be *independent*.

The two (risk neutral) players observe their valuations and enter into a bargaining game. Assume that participation is voluntary (IR). The outcome specifies whether the object is sold and if so, at what price. An implicit assumption is that the players cannot credibly "prove" their valuations to the other player. You can always pretend to have a different valuation (type).

What kinds of mechanisms are most efficient?

A *direct bargaining mechanism* is a game in which each player reports his type to the referee, and the referee chooses whether the object is transferred, and how much the buyer must pay.

As opposed to an indirect mechanism that might be closer to what we think of as bargaining. For example, the buyer and seller name a price; if the bid price is above the ask price, the object is sold for the average of the two prices, and if the ask price is above the bid price, the seller retains the object.

Let $p(v_1, v_2)$ denote the probability that the object is transferred, given the seller reports v_1 and the buyer reports v_2 .

Let $x(v_1, v_2)$ denote the expected payment from the buyer to the seller, given the seller reports v_1 and the buyer reports v_2 .

Because of risk neutrality and separability, it does not matter whether the buyer sometimes pays without purchasing.

A direct mechanism is *Bayesian incentive compatible* (IC) if honest reporting forms a Bayesian equilibrium.

The Revelation Principle: For any Bayesian equilibrium of any bargaining game, there is an equivalent incentive-compatible direct mechanism yielding the same outcome (when the honest equilibrium is played). In this sense, it is without loss of generality to restrict attention to direct mechanisms.

"proof" Start with an equilibrium of the indirect mechanism. Define the direct mechanism as follows. Each player reports his type, and the referee computes what actions would have been chosen in the indirect mechanism, and what the resulting outcome would have been. This outcome is selected in the direct mechanism. Since no player has an incentive to deviate to any allowable action in the indirect mechanism, no player has an incentive to *induce* one of the actions that one of his other types would have chosen. Incentive compatibility is satisfied and the outcome is the same as in the indirect mechanism.

The revelation principle presumes that

1. the referee can commit to abide by the mechanism, even after learning everyone's type,
2. all direct mechanisms are feasible and costless to implement.

Define the following notation:

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

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

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

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

$$U_1(v_1) = \bar{x}_1(v_1) - v_1 \bar{p}_1(v_1)$$

$$U_2(v_2) = v_2 \bar{p}_2(v_2) - \bar{x}_2(v_2)$$

$\bar{x}_1(v_1)$ is the seller's expected revenue, given his valuation

$\bar{p}_1(v_1)$ is the seller's probability of transferring the object, given his valuation

$U_1(v_1)$ is the seller's expected gains from trade, given his valuation

Thus, a mechanism, (p, x) , is **incentive compatible** if and only if for every v_1 and \hat{v}_1 in $[a_1, b_1]$, we have

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

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

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

The mechanism is **individually rational** iff

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

for every v_1 in $[a_1, b_1]$ and v_2 in $[a_2, b_2]$.

Theorem 1. For any incentive compatible mechanism,

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

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

Furthermore, for any "candidate" function $p(v_1, v_2)$, there is a function $x(v_1, v_2)$ such that (p, x) is incentive compatible **and individually rational** iff $\bar{p}_1(v_1)$ is weakly decreasing, $\bar{p}_2(v_2)$ is weakly increasing, and we have

$$0 \leq \int_{a_2}^{b_2} \int_{a_1}^{b_1} \left(\left[v_2 - \frac{1 - F_2(v_2)}{f_2(v_2)} \right] - \left[v_1 + \frac{F_1(v_1)}{f_1(v_1)} \right] \right) \times p(v_1, v_2) f_1(v_1) f_2(v_2) dv_1 dv_2.$$

Proof Sketch: Suppose (p, x) is incentive compatible.
Then

$$U_1(v_1) = \bar{x}_1(v_1) - v_1\bar{p}_1(v_1) \quad (1)$$

$$\geq \bar{x}_1(\hat{v}_1) - v_1\bar{p}_1(\hat{v}_1)$$

$$U_1(\hat{v}_1) = \bar{x}_1(\hat{v}_1) - \hat{v}_1\bar{p}_1(\hat{v}_1) \quad (2)$$

$$\geq \bar{x}_1(v_1) - v_1\bar{p}_1(\hat{v}_1).$$

Then $U_1(v_1) - U_1(\hat{v}_1)$ is at least the last term of (1) minus the middle term of (2), and no more than the middle term of (1) minus the last term of (2):

$$(\hat{v}_1 - v_1)\bar{p}_1(v_1) \geq U_1(v_1) - U_1(\hat{v}_1) \geq (\hat{v}_1 - v_1)\bar{p}_1(\hat{v}_1)$$

This implies that $\bar{p}_1(v_1)$ is weakly decreasing. Dividing by $(\hat{v}_1 - v_1)$ and taking limits, we have

$$U_1'(v_1) = -\bar{p}_1(v_1)$$

and

$$U_1(v_1) = U_1(b_1) + \int_{v_1}^{b_1} \bar{p}_1(t_1) dt_1 \quad (3)$$

Similarly, for the buyer we have

$$\begin{aligned}U_2'(v_2) &= \bar{p}_2(v_2) \\U_2(v_2) &= U_2(a_2) + \int_{a_2}^{v_2} \bar{p}_2(t_2) dt_2\end{aligned}\quad (4)$$

Thus,

$$\begin{aligned}& \int_{a_2}^{b_2} \int_{a_1}^{b_1} (v_2 - v_1) p(v_1, v_2) f_1(v_1) f_2(v_2) dv_1 dv_2 \\&= \int_{a_2}^{b_2} v_2 [\bar{p}_2(v_2)] f_2(v_2) dv_2 \\& \quad - \int_{a_1}^{b_1} v_1 [\bar{p}_1(v_1)] f_1(v_1) dv_1\end{aligned}\quad (5)$$

Now,

$$\begin{aligned} & \int_{a_2}^{b_2} U_2(v_2) f_2(v_2) dv_2 \\ = & \int_{a_2}^{b_2} v_2 [\bar{p}_2(v_2)] f_2(v_2) dv_2 - \int_{a_2}^{b_2} \bar{x}_2(v_2) f_2(v_2) dv_2 \\ = & \int_{a_2}^{b_2} v_2 [\bar{p}_2(v_2)] f_2(v_2) dv_2 \\ & - \int_{a_2}^{b_2} \int_{a_1}^{b_1} [x(v_1, v_2) f_1(v_1) dv_1] f_2(v_2) dv_2 \end{aligned}$$

and similarly,

$$\begin{aligned} & \int_{a_1}^{b_1} U_1(v_1) f_1(v_1) dv_1 \\ = & \int_{a_1}^{b_1} \int_{a_2}^{b_2} [x(v_1, v_2) f_2(v_2) dv_2] f_1(v_1) dv_1 \\ & - \int_{a_1}^{b_1} v_1 [\bar{p}_1(v_1)] f_1(v_1) dv_1. \end{aligned}$$

Thus, we have

$$\begin{aligned} & \int_{a_1}^{b_1} U_1(v_1) f_1(v_1) dv_1 + \int_{a_2}^{b_2} U_2(v_2) f_2(v_2) dv_2 \quad (6) \\ = & \int_{a_2}^{b_2} v_2 [\bar{p}_2(v_2)] f_2(v_2) dv_2 - \int_{a_1}^{b_1} v_1 [\bar{p}_1(v_1)] f_1(v_1) dv_1 \end{aligned}$$

Combining (5) and (6), we have

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

by (3) and (4), this equals

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

(The last step follows from reversing the order of integration.)

From the definitions of \bar{p}_1 and \bar{p}_2 , the last expression equals

$$U_1(b_1) + U_2(a_2) + \int_{a_2}^{b_2} \int_{a_1}^{b_1} [F_1(t_1)f_2(t_2) + (1 - F_2(t_2))f_1(t_1)]p(t_1, t_2)dt_1dt_2,$$

which also equals

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} (v_2 - v_1)p(v_1, v_2)f_1(v_1)f_2(v_2)dv_1dv_2$$

Simple algebra establishes the first part of Theorem 1. It follows from the equation in Theorem 1 that the inequality in Theorem 1 must hold if IR is satisfied. To show that there is a function $x(v_1, v_2)$ such that (p, x) is incentive compatible and individually rational, see the construction in Myerson and Satterthwaite. \square

Ex Post Efficiency

Clearly, a mechanism is ex post efficient iff we have

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

Also, in an ex post efficient mechanism, we have

$$\begin{aligned} \bar{p}_1(v_1) &= 1 - F_2(v_1) \\ \bar{p}_2(v_2) &= F_1(v_2). \end{aligned}$$

That is, the probability of selling is the probability that the buyer's valuation is above v_1 , and the probability of buying is the probability that the seller's valuation is below v_2 .

After a lot of manipulation, it can be shown that the equation in Theorem 1 simplifies to

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

Notice that we are integrating over the range of valuations between the buyer's lowest valuation and the seller's highest valuation.

If there is no such overlap, then the integral is negative, and we can find an IC, IR, and ex post efficient mechanism. For example, fix a price above the seller's highest valuation and below the buyer's lowest valuation, and trade for sure.

If there is an overlap, the integral is positive, so we have

$$U_1(b_1) + U_2(a_2) < 0$$

so any ex post efficient, IC mechanism cannot be IR.

For the overlap case, a third party could create an ex post efficient, IC and IR mechanism by sweetening the pot with a lump sum subsidy. The smallest such subsidy required is

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

Note: proofs rely on a positive density over valuations. Without that assumption, IC, IR, and ex post efficiency is possible. For example, suppose seller valuations are 1 or 4, and buyer valuations are 0 or 3. The mechanism fixes the price at 2 and a sale occurs iff both players agree.

Example: Suppose both valuations are uniformly distributed over the unit interval.

Then IC and IR imply (from the inequality in Theorem 1)

$$0 \leq 2 \int_0^1 \int_0^1 (v_2 - v_1 - \frac{1}{2}) p(v_1, v_2) dv_1 dv_2$$

so the expected difference in valuations, conditional on trade, must be at least one half.

However, the actual expected difference in valuations is only one third if we trade whenever the buyer's valuation is higher.

The minimum subsidy required for efficiency is

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

Maximizing Expected Total Gains From Trade

Expected total gains from trade are

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

To find a mechanism that maximizes these gains from trade, define the following

$$\begin{aligned} c_1(v_1, \alpha) &= v_1 + \alpha \frac{F_1(v_1)}{f_1(v_1)} \\ c_2(v_2, \alpha) &= v_2 - \alpha \frac{1 - F_2(v_2)}{f_2(v_2)} \\ p^\alpha(v_1, v_2) &= 1 \quad \text{if } c_1(v_1, \alpha) \leq c_2(v_2, \alpha) \\ &= 0 \quad \text{if } c_1(v_1, \alpha) > c_2(v_2, \alpha) \end{aligned}$$

Notice that p^0 transfers the object whenever it is ex post efficient to do so, and p^1 is the function that maximizes the integral in Theorem 1.

Theorem 2. If there exists an IC mechanism (p, x) such that $U_1(b_1) + U_2(a_2) = 0$ and $p = p^\alpha$ for some $\alpha \in [0, 1]$, then this mechanism maximizes the expected total gains from trade among all IC and IR mechanisms. Furthermore, if $c_1(v_1, \alpha)$ and $c_2(v_2, \alpha)$ are increasing functions on $[a_1, b_1]$ and $[a_2, b_2]$, and if the interiors of the two intervals have a nonempty intersection, then such a mechanism must exist.

Interpretation: Think of $c_1(v_1, \alpha)$ and $c_2(v_2, \alpha)$ as "virtual valuations" that reflect the informational rents required to maintain incentive compatibility. Trade will take place whenever the seller's virtual valuation is below the buyer's virtual valuation. Sellers adjust their valuations upward and buyers adjust their valuations downward, since they are willing to risk losing some beneficial trades in order to get a more favorable payment. The higher is α , the higher the loss of surplus, but also the higher the integral in Theorem 1. Pick α such that the integral is zero (and no higher) to satisfy the IR constraint.

Back to the uniform example, we have

$$c_1(v_1, \alpha) = (1 + \alpha)v_1 \text{ and } c_2(v_2, \alpha) = (1 + \alpha)v_2 - \alpha$$

Now find the α for which the integral is zero:

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

Since

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

we have

$$\begin{aligned} 0 &= \int_{\alpha/(1+\alpha)}^1 \int_0^{v_2 - \alpha/(1+\alpha)} (2v_2 - 1 - 2v_1) dv_1 dv_2 \\ &= \frac{3\alpha - 1}{6(1 + \alpha)^3} \end{aligned}$$

Thus, $\alpha = \frac{1}{3}$.

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

Chatterjee and Samuelson, "The Simple Economics of Bargaining," Operations Research 1983, study the (indirect) bargaining game mentioned earlier. The buyer and seller name a price; if the bid price is above the ask price, the object is sold for the average of the two prices, and if the ask price is above the bid price, the seller retains the object. It is **not** incentive compatible to set a price equal to your valuation. The equilibrium prices are

$$\frac{2}{3}v_1 + \frac{1}{4} \quad \text{and} \quad \frac{2}{3}v_2 + \frac{1}{12}$$

and the object is sold if and only if $v_1 \leq v_2 - \frac{1}{4}$. Thus, it is impossible to find another bargaining game satisfying IR and yielding higher expected total gains from trade.

The corresponding direct mechanism is

$$\begin{aligned} p(v_1, v_2) &= 1 \quad \text{if } v_1 \leq v_2 - \frac{1}{4} \\ &= 0 \quad \text{otherwise} \\ x(v_1, v_2) &= \frac{v_1 + v_2 + \frac{1}{2}}{3} \quad \text{if } v_1 \leq v_2 - \frac{1}{4} \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Property Rights and Countervailing Incentives

As opposed to the Coase Theorem which says that property rights do not affect efficiency, the initial allocation of property rights matters. We can think of asymmetric information as a "transactions cost" that invalidates the Coase Theorem.

In the uniform example, suppose that each of the players owns half of the good. That is, each is entitled to pay/receive nothing and receive utility $v_i/2$. The following mechanism is IC, IR, and ex post efficient. Players report their valuations. The player, i , reporting the higher valuation pays $v_i/3$ to the other player and buys the other player's share.

$$\begin{aligned} U_i(v_i, \hat{v}_i) &= \int_{\hat{v}_i}^1 \frac{t_i}{3} dt_i + \hat{v}_i \left(v_i - \frac{\hat{v}_i}{3} \right) \\ &= \frac{1}{6} - \frac{(\hat{v}_i)^2}{2} + \hat{v}_i v_i \end{aligned}$$

For any v_i , utility is maximized at $\hat{v}_i = v_i$, so this is IC.

Individual rationality is satisfied, because utility is always greater than $v_i/2$.

The mechanism is ex post efficient, because the object is always consumed by the player with the higher valuation.

Efficiency is restored because of countervailing incentives. The incentive to overstate your valuation when you are a seller is balanced by the incentive to understate your valuation when you are a buyer. Here, players do not know, when they report their valuations, whether they will be a seller or a buyer.