

# PERSPECTIVES ON MECHANISM DESIGN IN ECONOMIC THEORY

*Nobel Prize Lecture, Dec 8, 2007*

by Roger B. Myerson<sup>1</sup>

Department of Economics, University of Chicago, Chicago, IL 60637, USA

<http://home.uchicago.edu/~rmyerson/>

## 1. AN HISTORICAL PERSPECTIVE

Economics began with Xenophon's "Oeconomicus" (c 360 BCE), in which Socrates interviews a model citizen who has two primary concerns. He goes out to his farm in the country to monitor and motivate his workers there. Then he goes back to the city, where his participation in various political institutions is essential for maintaining his rights to own this farm. Such concerns about agents' incentives and political institutions are also central in economic theory today. But they were not always.

Two centuries ago, economics developed as an analytical social science by focusing on production and allocation of material goods, developing methodologies of national-income accounting and price theory. Questions about resource allocation seemed particularly amenable to mathematical analysis, because flows of goods and money are measurable and should satisfy flow-balance equations and no-arbitrage conditions. From this perspective, the classical economic problem was that people's ability to satisfy their desires is constrained by limited resources. The classical economic result was that unrestricted free trade can achieve allocative efficiency, in the sense that reallocating the available resources cannot improve everyone's welfare.

A shift of focus from allocation of resources back to analysis of incentives began from the time of Cournot (1838) when economic theorists began to analyze optimal decisions of rational individuals as a tool for understanding supply and demand in price theory (see Niehans 1990). In

---

<sup>1</sup>I am very grateful to the Nobel Committee for inviting me today. I also want to thank my co-authors and colleagues, at Northwestern University and at the University of Chicago, and my friends and family who have come so far, especially Gina who has come the farthest with me.

the first half of the 20th century, a few mathematicians began to formulate models for analyzing rational competitive decisions in more general frameworks, laying the foundations for game theory (Borel 1921, von Neumann 1928, von Neumann and Morgenstern 1944, Nash 1951; see also Myerson 1999).

Within economics itself, a substantive need for analytical models that go beyond the limits of price theory gradually became evident. In particular, the inconclusiveness of economic theorists' debates about socialism versus capitalism showed the limitations of price theory for evaluating non-price institutions like the socialist command economy (Barone 1935, Lange 1938; Mises 1935, Hayek 1935). Price theory could show (under some conditions) that free markets will achieve allocative efficiency, but such results about free markets did not prove that socialist command economies could not achieve similarly good outcomes. To allow analytical comparison of fundamentally different forms of economic organization, a new and more general theoretical framework was needed. In an widely influential paper, Hayek (1945) argued that a key to this new economic theory should be the recognition that economic institutions of all kinds must serve an essential function of communicating widely dispersed information about the desires and the resources of different individuals in society. That is, different economic institutions should be compared as mechanisms for communication.

Hayek also alleged that the mathematical economists of his day were particularly guilty of overlooking the importance of communication in market systems. But questions about fundamental social reforms require fundamental social theory. In a search for new fundamental theories, the abstract generality of mathematics should be particularly helpful. So the failure that Hayek perceived should not have been attributed to mathematical modeling per se, but it was evidence of a need for fundamentally new mathematical models. Among the mathematical economists who accepted this challenge from Hayek, Leo Hurwicz was the leader.

The pivotal moment occurred when Hurwicz (1972) raised the basic question of incentives to communicate information and introduced the general concept of incentive compatibility. In doing so, he took a long step beyond Hayek in advancing our ability to analyze the fundamental problems of institutions. From that point on, as Makowski and Ostroy (1993) have observed, "the issue of incentives surfaced forcefully, as if a pair of blinders had been removed."

After Hurwicz (1972), many of us jumped into the breach to join the advance. From Harsanyi (1967), we had a general Bayesian model of games where people have different information, and we had Harsanyi's general concept of Bayesian equilibrium to analyze rational behavior in such games. In this framework, we saw Hurwicz's theory of mechanisms as the foundations of a theory about how to design Bayesian games. A coordination mechanism is a plan for how social decisions should depend on people's reported information, and changing the coordination mechanism in a society effectively changes the game that its members will play. Given the information, preferences, and resources that people have in a society, different social coordination mechanisms could yield different games, each of which could have many different equilibria. But remarkably, the set of all possible equilibria of all possible games can be simply characterized by using the revelation principle, which many of us (Dasgupta Hammond Maskin 1979, Harris Townsend 1981, Holmstrom 1977, Myerson 1979, Rosenthal 1978) found independently, building on ideas of Gibbard (1973) and Aumann (1974). With the revelation principle, this feasible set essentially coincides with the set of incentive-compatible mechanisms, which satisfy certain incentive constraints. These incentive constraints express the basic fact that individuals will not share private information or exert hidden efforts without appropriate incentives.

So mechanism theory expanded our general view of the economic problem to include incentive constraints as well resource constraints. Incentive constraints help us to explain many failures of allocative efficiency that we observe in the world. But in this new framework of economic analysis, we also have new concepts of incentive efficiency for evaluating the rules by which resources are allocated (rather than specific resource allocations themselves), taking incentive constraints into account. These conceptual tools now allow us to analyze questions about efficient institutions that were beyond the analytical reach of economic theory in Hayek's day.

## 2. ELEMENTS OF MECHANISM DESIGN THEORY

In society, people have information about their resources and desires, and people choose actions for producing, redistributing, and consuming resources. In markets and other institutions of society, individuals' actions may depend on others' information as it has been communicated in

the market or social institution. This is the perspective that Hayek recommended, that we should view social institutions as mechanisms for communicating people's information and coordinating people's actions. To decide whether we have a good social institution, we want to ask how it performs in this communication and coordination role. If we do not like the performance of our current institutions, then we may want to reform them, to get an institution that implements some desired social plan, where a social plan is a description of how everyone's actions should depend on everyone's information.

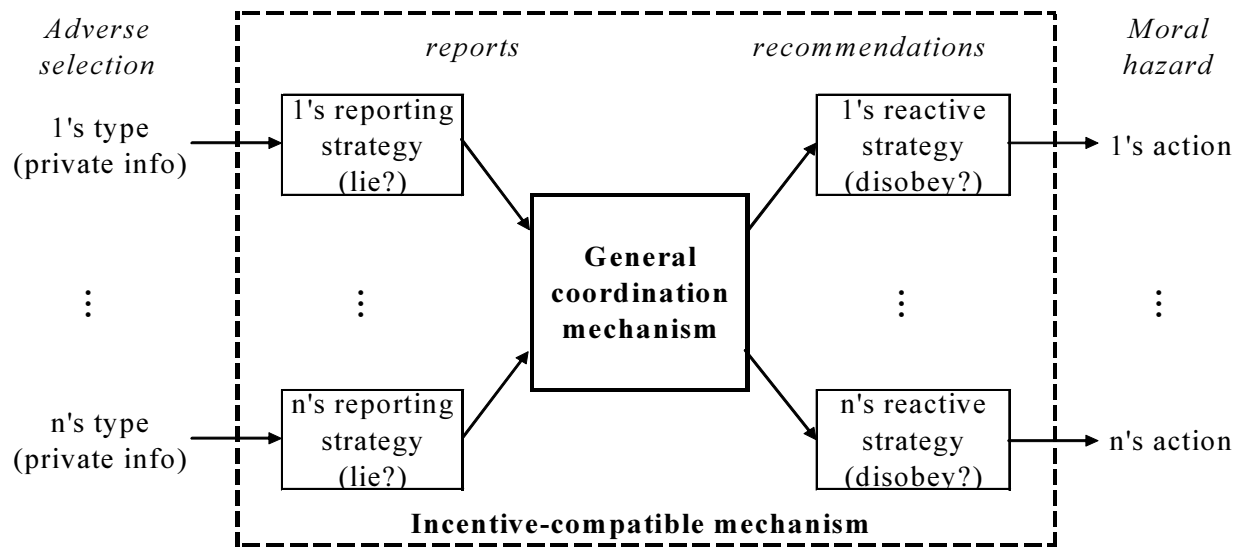
So the crucial question is, what kinds of social coordination plans are actually feasible? A feasible social coordination plan could be implemented by many different social institutions, but it is helpful to begin by considering a very centralized institution where every individual communicates separately and confidentially with a trustworthy central mediator. Suppose first that each individual confidentially reports all his or her private information to the mediator, and then, based on all these reports, the mediator recommends to each individual what actions he or she should take under the plan. But if we allow that individuals can be dishonest or disobedient to the mediator then, as Hurwicz (1972) observed, the social plan must give people incentives to share information and to act appropriately according to the social plan.

First, to the extent that our social plan depends on individuals' private information that is hard for others to observe, we need to give people an incentive to share their information honestly. This problem of getting people to share information honestly is called adverse selection. Second, to the extent that our social plan requires people to choose hidden actions and exert efforts that are hard for others to monitor, we need to give people an incentive to act obediently according to the plan. This problem of getting people to act obediently to a social plan is called moral hazard. If it is a rational equilibrium for everyone to be honest and obedient to the central mediator who is implementing our social coordination plan, then we say that the plan is incentive compatible.

There are two important things to say about such incentive-compatible coordination plans. First, they can be characterized mathematically by a set of inequalities called incentive constraints that are often straightforward to analyze in many interesting examples. Second, although we defined incentive compatibility by thinking about honesty and obedience in communication with a central mediator, in fact these incentive-compatible plans characterize

everything that can be implemented by rational equilibrium behavior in any social institution or mechanism. This assertion of generality is called the revelation principle.

The revelation principle asserts that any rational equilibrium of individual behavior in any social institution must be equivalent to an incentive compatible coordination plan. Given any possible informational reports from the individuals, the equivalent incentive-compatible plan recommends the results of simulated lying and disobedience in the original institution or mechanism, as illustrated in Figure 1. Thus, without loss of generality, a trustworthy mediator can plan to make honesty and obedience the best policy for everyone.



**Figure 1. The revelation principle.**

(To prove the revelation principle, suppose that we are given a general coordination mechanism and an equilibrium that describes rational individual strategies for reporting dishonestly and acting disobediently in this mechanism. We need to describe how a mediator would implement the equivalent incentive-compatible mediation plan where honesty and obedience is an equilibrium. When the mediator has gotten a confidential report of every individual's private information, the equivalent incentive-compatible mediation plan would first compute the dishonest reports that everyone would have sent in the given equilibrium. Then it would compute the behavior that the given mechanism would have indicated for each individual based on these reports. Then it would compute the disobedient action that each individual would have actually chosen in the given equilibrium. Finally it would confidentially recommend to

each individual that he should choose this computed action. If any individual had any incentive to be dishonest or disobedient to the mediator under this plan, then he would have had an incentive also to be dishonest or disobedient to himself under his given equilibrium strategy in the given mechanism. But in a rational equilibrium, nobody can gain by lying to himself or disobeying his own optimal strategy. See Myerson 1982.)

In sections 3 to 5 below, we consider three examples to illustrate the power of mechanism-design theory. First we consider an example of trading in a simple pure-exchange economy, where one seller and one potential buyer are bargaining over the sale of one unique object. This example involves adverse-selection problems, and it illustrates how individuals' incentives to bargain for a better price can prevent allocatively efficient trading. Second, we consider a simple production example, involving moral hazard in management. This example illustrates how incentives for good management may require that managers must have a valuable stake in their enterprise. Third, we consider an example that introduces politics into a productive economy, involving moral hazard in the government. This example shows how unrestrained power of government over the economy can be inefficient, as capital investors require credible political guarantees against the government's temptation to expropriate them. The latter two moral-hazard models here may particularly illustrate the kinds of theoretical frameworks that can be used to exhibit practical disadvantages of socialism, which Hayek sought to show.

### 3. A SIMPLE BILATERAL TRADING EXAMPLE WITH ADVERSE SELECTION

For our first example of mechanism-design theory, let us consider the simplest possible economic transaction: the sale of one single object by one seller who faces one potential buyer. In this example, each individual knows his or her own private value of the object. The object may be worth either \$0 or \$80 to the seller, and it may be worth either \$100 or \$20 to the buyer. For each trader, we may say that the type that is willing to trade at more prices is "weak," and the other type is "strong." So the seller's type is weak when his value of the object is \$0, but he is strong when his value is \$80. The buyer's type is weak when her value is \$100, but she is a strong when her value is \$20. Each trader thinks that that the other trader is equally likely to be weak or strong in this sense; that is, each type of each trader has independent probability 0.5. (This discrete example is from Myerson 1991. Myerson and Satterthwaite 1983 derive stronger

results for models where each trader has a continuum of possible types.)

Let us consider this trading situation from the perspective of a mediator who is assisting the two individuals to negotiate this transaction. Trade would be mutually beneficial unless both individuals are strong, but the range of mutually acceptable prices will depend on how much the object is actually worth to each trader, which each knows privately. So the mediator should ask the traders to reveal this information and should formulate a plan of how the terms of trade may depend on what they report. Based on the reported information, the mediator could either recommend that the object should be traded for some specified price, or the mediator could recommend that they should not trade at all. (For simplicity, let us assume that, whatever the mediator recommends, the buyer and seller will accept and implement the mediator's recommendation, as long as neither one is made worse off by the trade.<sup>2</sup>) Such a mediation plan is, in our theoretical terminology, a mechanism for coordinating the given economic agents.

### 3.1 Failure of incentive compatibility in the simple split-the-difference plan

Figure 2 shows one natural mediation plan, where the mediator recommends trade whenever the buyer's value of the object is more than the seller's value, and the recommended price is always half way between their two values. This mechanism may be called the simple split-the-difference plan.

		Buyer's value	
		[strong]	[weak]
Seller's value	[strong] \$80	0, *	1, \$90
	[weak] \$0	1, \$10	1, \$50

*P(trade), E(price if trade)*

**Figure 2. Split-the-difference mediation plan.**

The four cells in Figure 2 correspond to the four possible combinations of the traders' types. In each cell, the first number listed is the conditional probability of the buyer getting the

---

<sup>2</sup> In a more advanced treatment, we could justify this assumption by arguing that, if either individual announced any other offer after they hear mediator's recommendation, then this offer would be taken as evidence that this individual was weak and would soon make a more generous offer that concedes all his or her gains from trade. See Myerson (1991).

object if the individuals' reported types are as in this cell. The second number in each cell is the expected price that the buyer will pay if they trade when their reported types are as in this cell. (In a cell where the probability of trade is 0, we do not need to specify any price-if-trade because we know that they would not trade if that cell occurred, and so an asterisk is indicated instead.)

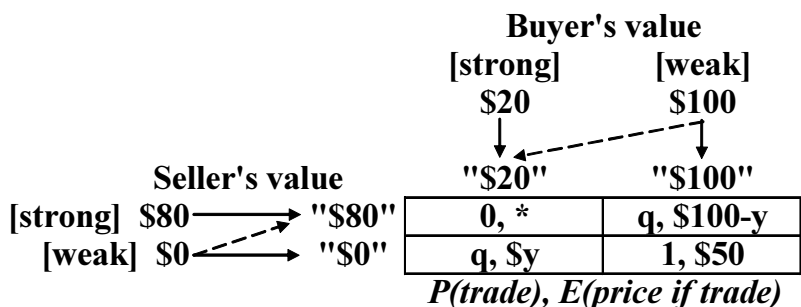
The simple split-the-difference mediation plan might seem a fair way to achieve mutually beneficial trades with probability  $3/4$ . But we are allowing that individuals can misrepresent their types, and unfortunately this plan is not incentive compatible; that is, honesty by both traders is not an equilibrium of this game. For honesty to be an equilibrium in the sense of Nash (1951), it must be that each individual would find honesty to be the best policy when the other is expected to be honest. It is easy to see that a strong type can never gain by claiming to be weak: a seller who thinks that the object is worth \$80 would only be asked to sell at a loss if he pretended that the object was worth \$0 to him. But let us look at the problem from the seller's perspective when he knows he is weak. If the buyer were expected to be honest in this plan, then a weak seller could get a higher expected profit by claiming to be strong. If the weak seller honestly admitted weakness under this plan, his expected profit would be  $0.5(10-0)+0.5(50-0) = 30$ , because the buyer has probability 0.5 of being weak and probability 0.5 of being strong. But if the weak seller claimed to be strong, then he would have a 0.5 probability of getting profit  $\$90-0$ , and so his expected profit from lying would be  $0.5(90-0) = 45$ , which is strictly greater than the 30 that he would expect from honesty. (Throughout we assume here that individuals are risk-neutral, seeking to maximize their expected profits.) Thus, honesty is not an equilibrium of this mediation plan. That is, this simple split-the-difference plan is not incentive compatible.

### *3.2 Incentive constraints for symmetric mediation plans*

Let us now consider other mediation plans. For simplicity, let us consider plans which treat the seller and the buyer similarly or symmetrically. To be specific, let us suppose that the conditional probability of trading when one individual is weak and the other is strong is some number  $q$  that does not depend on which of them is the weak one. Also, let us suppose that the expected profit margin of a weak individual who trades with a strong individual is some number  $y$  that does not depend on which individual is the weak one. For simplicity, let us suppose that two weak individuals, who are maximally eager to trade, would trade with probability 1 at a price \$50, which is half-way between their two values. We can assume that trade does not occur when



both are strong, as the seller would then value the object more than the buyer. So the general symmetric mediation plan with these two parameters  $q$  and  $y$  is as shown in Figure 3.



**Figure 3. General symmetric mediation plan or mechanism.**

For the plan in Figure 3 to be an incentive-compatible mechanism,  $q$  and  $y$  must satisfy three inequalities or constraints. First, the number  $q$  must satisfy the probability constraints

$$0 \leq q \leq 1.$$

(In each cell, the probability of trade could also be interpreted as the conditional expected number of objects that the buyer would get in this case, and so  $q \leq 1$  here can also be interpreted as a resource constraint, expressing the fact that there is only one object that they can trade.)

Under the symmetric plan in Figure 3, a strong buyer with value \$20 would trade only at the price  $y$ , which would be an unacceptable loss for the strong buyer if  $y$  were greater than 20. So for a strong trader (buyer with value \$20 or seller with value \$80) to be willing to participate in this plan,  $y$  must satisfy the participation constraint

$$y \leq 20.$$

For honest reporting to be a Nash equilibrium in Figure 3, we need to verify that each individual would be willing to report his type honestly if he expected the other to be honest. It is easy to see that a strong type of seller or buyer would never want to pretend to be weak under this plan, as it would only result in trading at a loss. But a weak type might be tempted to pretend to be strong, to get a better price. Consider the weak seller, for example. (All calculations for the weak buyer are similar.) Under this plan, against an honest opponent, the weak seller's expected profit if he reports honestly is  $0.5q(y-0)+0.5(1)(50-0)$ , but his expected profit if he falsely claims to be strong is  $0.5(0)+0.5q(100-y-0)$ . Thus, to make honesty an equilibrium,  $q$  and  $y$  must satisfy the informational incentive constraint

$$0.5qy + 0.5(50) \geq 0.5q(100-y).$$

This constraint is algebraically equivalent to the inequality

$$q \leq 25/(50-y).$$

### 3.3 Incentive-efficient trading plans

The incentive constraints  $q \leq 25/(50-y)$  and  $y \leq 20$  together imply that the largest feasible probability of trade is achieved by letting  $y = 20$ ,  $q = 5/6$ . Figure 4 shows this incentive-compatible mechanism. The informational incentive constraint is satisfied as an equality here, as a weak type's expected profit from honesty is  $(0.5)(5/6)20 + (0.5)50 = 33.33$ , which is just equal to the expected profit  $(0.5)(5/6)80$  that a weak type could get from lying.<sup>3</sup>

<b>(y=20, q=5/6)</b>	<b>Buyer's value</b>		
	<b>Seller's value</b>	<b>\$20 [s]</b>	<b>\$100 [w]</b>
<b>EU(str)=\$0</b>	[s] \$80	<b>0, *</b>	<b>5/6, \$80</b>
<b>EU(wk)=\$33.33</b>	[w] \$0	<b>5/6, \$20</b>	<b>1, \$50</b>
		<b>P(trade), E(price if trade)</b>	

**Figure 4. An ex-ante incentive-efficient mechanism.**

One of the basic contributions of mechanism design to economic analysis is an extension of how we think about efficiency in markets and other systems for economic transactions. In Figure 4, the probability of the mediator failing to achieve a mutually beneficial trade is strictly positive (it is  $(2/4)(1/6) = 1/12$ ). Ex post, after the traders have reported their types and one of them has reported weakness, any such failure to realize mutually beneficial gains from trade would be allocatively inefficient, as the object is not being allocated to the person who can make the most valuable use of it. Such was the classical view of efficiency that economists regularly

---

<sup>3</sup>The incentive-compatibility of Figure 4 depends on the two traders communicating separately with the mediator. In general, incentives to lie to the mediator could be exacerbated if one individual learned the other's report before submitting his own. In Figure 4, if the seller learned that the buyer reported her weak type before reporting his own, then the weak seller could get higher expected profit by pretending to be strong, because  $(5/6)80 > 50$ .

applied before mechanism design. But now we understand that, when the participation constraints and informational incentive constraints are taken into account, no incentive-compatible mediation plan can have a lower probability of such ex-post allocative inefficiency here than this mechanism in Figure 4.

When the expected profits are calculated ex ante, before either individual's type is determined, the expected profit for each individual is  $[0+0+(5/6)20+(1)50]/4 = 16.67$  in this mechanism. In fact, there is no incentive-compatible mechanism that could yield a higher ex-ante expected profit for both individuals here. So in this sense (following Holmstrom and Myerson 1983), we may say that this mechanism is ex-ante incentive efficient.

Notice that this concept of incentive efficiency applies to the mechanism for determining how the resource allocation will depend on people's information, not just to a particular resource allocation, and it takes incentive constraints into account in asking whether any other feasible mechanism could be better for these individuals. In ex-ante incentive efficiency, each individual's welfare is evaluated at the ex ante stage, before anyone learns his or her type.

The incentive-efficient mechanism in Figure 4 would not look so good to the seller if he knew he was the strong type, however, because this plan never allows the strong seller to get any positive profit from selling the object for more than his value of 80. So if we care about individuals' welfare as they assess it in the game itself, when each knows his own type, then we need to admit other mechanisms as incentive-efficient.

Figures 5 and 6 show other symmetric mediation plans that have  $y < 20$ , and so allow a strong individual to get a positive profit margin against a weak opponent, but keep the incentive constraint binding with  $q = 25 / (50 - y)$ , so that the trading probability  $q$  is as large as possible given  $y$ . These mechanisms are better for the strong types than Figure 4, and they are interim incentive efficient, in the sense that there is no other incentive-compatible mechanism that would be preferred by each possible type of each individual in the game. (The word interim here refers to the fact that we are evaluating each individual's welfare at a time when he has learned his own type but has not yet learned any others' types; see Holmstrom and Myerson, 1983.) The mechanism in Figure 6 (which is the best among these for the strong types), is the solution identified by a natural generalization of Nash's (1950) bargaining solution, as defined by Myerson (1984).

( $y=10, q=5/8$ )

Seller's value		Buyer's value	
		\$20 [s]	\$100 [w]
$EU(str)=\$3.125$	[s] \$80	0, *	5/8, \$90
$EU(wk)=\$28.125$	[w] \$0	5/8, \$10	1, \$50

*P(trade), E(price if trade)*

Figure 5. An interim incentive-efficient mechanism.

( $y=0, q=1/2$ )

Seller's value		Buyer's value	
		\$20 [s]	\$100 [w]
$EU(str)=\$5$	[s] \$80	0, *	1/2, \$100
$EU(wk)=\$25$	[w] \$0	1/2, \$0	1, \$50

*P(trade), E(price if trade)*

Figure 6. An interim incentive-efficient mechanism (generalized Nash bargaining solution).

3.4 Dishonest equilibria and the revelation principle

The above analysis assumes that the mediator should get information by providing incentives for the traders to be honest. But even for mechanisms that are not incentive-compatible, the traders may convey information by rational strategies that form an equilibrium of their reporting game. For example, consider again the simple split-the-difference plan from Figure 2. We saw that, if the buyer were expected to be honest, then the seller would prefer to always claim to be strong. But if the seller is expected to always claim "strong" in this mediation plan, then the buyer would prefer to be honest, because an honest admission of weakness could at least get  $100-90 = \$10$  profit for the weak buyer. Thus, the simple split-the-difference mediation plan has a reporting equilibrium where the seller always reports strong but the buyer is honest. With this equilibrium of this mediation plan, the outcome of the mediation will depend on the traders' actual types according to table shown in Figure 7.

Seller's value		Buyer's value	
		\$20 [s]	\$100 [w]
[s] \$80		0, *	1, \$90
[w] \$0		0, *	1, \$90

*P(trade), E(price if trade)*

Figure 7. Incentive-compatible mechanism equivalent to an equilibrium of split-the-difference.

The plan in Figure 7 is itself an incentive-compatible mediation plan. In effect, the plan disregards the seller's reported type and has trade occurring at price \$90 if and only if the buyer reports her weak type. It may seem unfair to have trading only at a price that is so favorable to the seller, but honest participation by both individuals would be an equilibrium if the mediator planned to use their reported information according to the plan shown Figure 7.

The simple split-the-difference plan also has two other reporting equilibria. In one equilibrium, it is the buyer who is expected to always report her strong type, and in response the seller is expected to be honest. The outcomes of this equilibrium are the same as the incentive-compatible mechanism shown in Figure 8 (the mirror image of Figure 7). In the other equilibrium, both individuals have a positive probability of lying, each claiming to be strong with probability 0.6 when he or she is weak. For each possible combination of types, the conditional probability of trade and the conditionally expected price if trade occurs under this randomized equilibrium is the same as the incentive-compatible plan shown in Figure 9.

		Buyer's value	
		\$20 [s]	\$100 [w]
Seller's value	\$80	0, *	0, *
	\$0	1, \$10	1, \$10

*P(trade), E(price if trade)*

**Figure 8. Incentive-compatible mechanism equivalent to an equilibrium of split-the-difference.**

		Buyer's value	
		\$20 [s]	\$100 [w]
Seller's value	\$80	0, *	0.4, \$90
	\$0	0.4, \$10	0.64, \$50

*P(trade), E(price if trade)*

**Figure 9. Incentive-compatible mechanism equivalent to an equilibrium of split-the-difference.**

These plans in Figures 7-9 may not seem very fair or efficient, as one favors the seller, one favors the buyer, and the third yields a low probability of trading. But the general point is that, for any equilibrium of the individuals' reporting strategies under any mediation plan, we can find an incentive-compatible mechanism that has the same outcomes for all types. This result is the revelation principle, and it allows us to extend our results from incentive-compatible

mechanisms to all possible equilibria of all possible mechanisms. It allows us to say, for this example, that no equilibrium of any mediation plan, incentive-compatible or not, could have a lower probability of ex-post allocative inefficiency than the honest equilibrium of the incentive-compatible mechanism in Figure 4.

### 3.5 General nonsymmetric mechanisms

Our analysis here has been simplified by focusing only on a class of mediation plans that treat the buyer and seller symmetrically. Without this symmetry, the general class of mediation plans for this example has eight variables, instead of just two. For each  $j$  in  $\{0,80\}$  and each  $k$  in  $\{20,100\}$ , we could let  $q_{j,k}$  denote the conditional probability of trade when the seller's value is  $j$  and the buyer's value is  $k$ , and we could let  $y_{j,k}$  denote the conditionally expected price if trade occurs when their values are  $j$  and  $k$ . These variables must satisfy the probability constraints:

$$0 \leq q_{0,20} \leq 1, \quad 0 \leq q_{0,100} \leq 1, \quad 0 \leq q_{80,20} \leq 1, \quad 0 \leq q_{80,100} \leq 1;$$

the participation constraints:

$$0.5q_{80,20}(y_{80,20} - 80) + 0.5q_{80,100}(y_{80,100} - 80) \geq 0,$$

$$0.5q_{0,20}(y_{0,20} - 0) + 0.5q_{0,100}(y_{0,100} - 0) \geq 0,$$

$$0.5q_{80,20}(20 - y_{80,20}) + 0.5q_{0,20}(20 - y_{0,20}) \geq 0,$$

$$0.5q_{80,100}(100 - y_{80,100}) + 0.5q_{0,100}(100 - y_{0,100}) \geq 0;$$

and the informational incentive constraints:

$$0.5q_{80,20}(y_{80,20} - 80) + 0.5q_{80,100}(y_{80,100} - 80) \geq 0.5q_{0,20}(y_{0,20} - 80) + 0.5q_{0,100}(y_{0,100} - 80),$$

$$0.5q_{0,20}(y_{0,20} - 0) + 0.5q_{0,100}(y_{0,100} - 0) \geq 0.5q_{0,20}(y_{80,20} - 0) + 0.5q_{80,100}(y_{80,100} - 0),$$

$$0.5q_{80,20}(20 - y_{80,20}) + 0.5q_{0,20}(20 - y_{0,20}) \geq 0.5q_{80,100}(20 - y_{80,100}) + 0.5q_{0,100}(20 - y_{0,100}),$$

$$0.5q_{80,100}(100 - y_{80,100}) + 0.5q_{0,100}(100 - y_{0,100}) \geq 0.5q_{80,20}(100 - y_{80,20}) + 0.5q_{0,20}(100 - y_{0,20}).$$

These constraints look very complicated, but they are actually not hard to analyze. They can be linearized by the substitution  $x_{j,k} = q_{j,k} y_{j,k}$ , and then optimal mechanism can be found by linear programming. In particular, the mechanisms that we have studied above in Figures 4-6 remain incentive-efficient in this more general class of nonsymmetric mechanisms.

### 3.6 Incentive constraints as a source of Coase's transactions costs

Coase (1960) argued that, if there were no transactions costs then unrestricted free trade of property rights and resources could achieve allocative efficiency, regardless of the initial

allocation of property rights. Thus, transactions costs are important for understanding the problems of allocative inefficiency, and we need an analytical theory of where transactions costs come from. In situations like this example, the informational incentive constraints can be viewed as a source of Coasean transactions costs. The participation constraints represent the initial allocation of property rights that each can hold without trading (Samuelson 1985). We have shown that ex-post allocative efficiency cannot be guaranteed by any incentive-compatible mechanism when the initial owner is the person whose value may be 0 or 80. But if the initial ownership right was assigned instead to the individual whose value may be 20 or 100, so that the roles of seller and buyer are reversed, then it is easy for a mediator to guarantee that the object ends up with the individual who values it most. An incentive-compatible mechanism that would achieve ex-post efficiency in this case is shown in Figure 10 (where the mediator just proposes the price \$50, and they trade if both accept it). Thus, the theory of mechanism design gives us an analytical framework where the initial allocation of property rights can affect the probability of achieving an outcome that is allocatively efficient.

		Seller's value	
		\$20	\$100
Buyer's value	\$80	1, \$50	0, *
	\$0	0, *	0, *

*P(trade), E(price if trade)*

**Figure 10. An allocatively efficient trading plan with initial ownership reversed.**

#### 4. A RISKY PRODUCTION PROJECT WITH MORAL HAZARD IN MANAGEMENT

Our second example involves production, and here we can introduce problems of moral hazard, because valuable inputs that are required for production may be misused or diverted by the manager of the production process. For simplicity, let us consider a one-time production project that requires an initial capital input worth  $K=100$ , and then returns revenue worth  $R=240$  if the project is a success, or returns no revenue (0) if the project is not a success. The project's probability of success depends on the manager's hidden action. If the manager diligently applies good effort to managing the project, then probability of success is  $p_G = 1/2$ . On the other hand, if the manager behaves badly and abuses his managerial authority in the project then the probability of success is  $p_B = 1/4$ , but the manager can get hidden private benefits worth  $B=30$  from such

abuse of power.<sup>4</sup> With these parameters, the expected returns from the project are greater than the cost of its capital inputs if the manager chooses to be good, as  $p_G R > K$ . But if the manager chose to behave badly, then the capital input cost would be greater than the expected returns plus private benefits, as  $K > p_B R + B$ . So by the expected value criterion, the project can be worth undertaking only if the manager chooses good diligent effort.

Let  $A$  denote the total value of all personal assets that the manager can offer to invest in the project. We may call  $A$  the manager's collateral (although it may also include the value of his time in managing the project). The worst that our social plan can do to the manager, if the project fails, is to take away his collateral, in which case the manager's net payoff would be  $-A$ . We assume that  $A < K$ , so that the manager does not have enough capital to undertake this project on his own, and so capital for this project must be provided by others in society.

Let us consider this problem from the perspective of society at large, that is, of the people other than the manager who must provide the required capital input. Can society at large derive any positive expected benefit from investment in the project? The manager's pay cannot depend on his hidden effort, which is not directly observable, but his pay can depend on whether the project is a success or not. The basic variable in the incentive scheme here is the net payoff  $w$  that society will pay to the manager if the project succeeds. (Here a net payoff of 0 would mean that the manager just keeps his collateral.) If society undertakes the investment in this project, then the social plan should recommend to the manager that he should exert good effort, but to give him an incentive to obey this recommendation, the wage  $w$  from success must satisfy the moral-hazard constraint:

$$p_G w - (1 - p_G)A \geq B + p_B w - (1 - p_B)A.$$

Also, the manager could refuse to participate in the project at all, if he does not get a positive expected net payoff from his participation, and so  $w$  must also satisfy the participation constraint:

$$p_G w - (1 - p_G)A \geq 0.$$

(It can be verified that paying the manager more than  $-A$  in case of failure cannot improve his incentives for participating with good behavior.) The manager's limited assets imply that  $w$  must

---

<sup>4</sup>This is a version of the basic moral-hazard example in section 3.2 of Tirole (2006).



also satisfy the following limited-liability constraint (or resource constraint):

$$w \geq -A.$$

Subject to these constraints, let us maximize the expected net profit for society at large.

When the manager obeys the recommendation to be good, this expected social profit is

$$V = p_G(R-w) + (1-p_G)A - K.$$

Let us consider the case where the manager is not very rich, so that in particular

$$A < Bp_G/(p_G-p_B),$$

that is,  $A < 60$  for our numerical example. Then the optimal incentive mechanism satisfies the moral-hazard constraint as an equality, and has

$$w = B/(p_G-p_B) - A = 120-A.$$

So the manager must be allowed to get a moral-hazard rent that has expected value

$$p_G w - (1-p_G)A = Bp_G/(p_G-p_B) - A = 60-A.$$

Thus, the expected net profit for society at large cannot be more than

$$p_G R - K - (60-A) = A - 40.$$

This amount is negative when  $A < 40$ . So in this example, we cannot get any positive expected profit for society unless the manager himself can contribute assets  $A$  worth at least 40. That is, to deter abuse of power without an expected loss to the rest of society, the manager must have stakes in this project worth at least 40% of the cost of the capital input here. If no one has such a large personal wealth to offer as collateral to this investment, as might be the case in an egalitarian socialist society, then society at large cannot profitably undertake this investment.

Thus, moral-hazard incentive constraints can also provide an analytical framework where the initial allocation of property rights may affect the possibility of productive investments. Indeed, this simple example may provide an analytical perspective on problems of socialism, as Hayek was seeking. Modern industrial production requires integrated managerial control over large scale assets, and whoever exercises that control will have great moral-hazard temptations, which are represented by the parameter  $B$  in this model. When managers have great temptations  $B$ , the moral-hazard incentive constraint cannot be satisfied unless managers have large stakes in success of their projects. If, unlike capitalist entrepreneurs, socialist managers do not have substantial personal assets that they can invest in their projects, then the necessary stakes can only be achieved by allowing socialist managers to take a large share of the benefits from

successful projects. So considerations of moral hazard cast doubt on the egalitarian socialist ideal that profits from industrial means of production should all belong to the general public.

There is one way to extend our model that might help to ward off this specter of a privileged socialist managerial elite which closely resembles the capitalist elite. Suppose that the socialist system allowed the possibility of physically punishing managers in case of failure. In this extended model, the incentive mechanism could have two variables:  $w$  the manager's payment for success; and  $z$  the manager's cost of punishment for failure. Unproductive punishment of a manager would be different from seizing a manager's personal assets, however, as punishment would not yield any benefit to the rest of society. (Any such unproductive punishment would be allocatively inefficient, of course, but we understand that punishment may sometimes be a necessary deterrent in incentive-efficient social rules.) Then the optimal incentive problem for society at large would choose  $w$  and  $z$  to

$$\begin{aligned} &\text{maximize } V = p_G(R-w) + (1-p_G)A - K \\ &\text{subject to } w \geq -A, \quad z \geq 0, \\ &p_G w - (1-p_G)(A+z) \geq B + p_B w - (1-p_B)(A+z), \quad \text{[moral hazard]} \\ &p_G w - (1-p_G)(A+z) \geq 0. \quad \text{[participation]} \end{aligned}$$

For our numerical example when  $A < 60$ , the optimal solution threatens punishment  $z = 60 - A$  for failure and pays  $w = 60$  for success, which wipes out the manager's expected gain but leaves the expected social profit  $V$  equal to  $0.5A - 10$ , which is positive only if  $A > 20$ . Thus, even with punishment, society at large cannot gain from this project when the manager has no valuable collateral.

On the other hand, society at large could benefit from this project with punishment if the participation constraint were dropped. In the pure egalitarian case where  $A = 0$ , society at large could use  $w = 0$  and  $z \geq 120$  to satisfy the moral-hazard constraint and get an expected social profit of 20 in our numerical example. But this seems a terrible solution, as it would correspond to drafting an unwilling manager and motivating him to work hard only by a threat of bodily harm if the project fails.

So our analysis of this simple model of moral hazard in production is enough to show how industrialization under Soviet communism may have necessarily entailed either the creation of a privileged managerial elite or the use of coerced labor from prisoners of the state. Some

insights of this model may be implicit in some of Hayek's informal intuitive arguments, which he could express verbally<sup>5</sup> but which tended to get lost in his analysis as they could not be expressed formally in the economic models of his time.

#### 4.1 An analogous adverse-selection problem

To compare moral hazard and adverse selection, we might consider an analogue of the above problem where the incentive constraint is about the manager's hidden information rather than about the manager's hidden action. In this analogous adverse-selection example, the project's probability of success depends on the manager's hidden type, which may be good or bad. The parameters ( $p_G$ ,  $p_B$ ,  $R$ ,  $K$ ,  $A$ ) all have the same interpretation as above, except that now  $p_G$  [or  $p_B$ ] denotes the conditional probability of success when the manager knows that he is the good talented type [or, respectively, the bad incompetent type], which is now a given fact that only the manager knows (but does not depend on any choices that he will make). Suppose that  $p_G R > K > p_B R$ , so that the project would yield a positive expected net profit for a good manager but not for a bad manager. The outside investors of society at large are uncertain about the manager's true type. Let  $\alpha$  denote the probability that the manager is the good type.

In an incentive-compatible mechanism, we can ask the manager his type, and then plan to undertake the project with some probability  $q_G$  if he reports the good type or  $q_B$  if he reports the bad type. If the project is undertaken, the wage that he receives in case of success could also depend on his report, being  $w_G$  if he reported "good" or  $w_B$  if he reported "bad." We may assume that if the project is undertaken and fails then the manager will forfeit his collateral  $A$ . Under such an investment plan, the expected profit for the investors would be

$$V = \alpha q_G [p_G(R - w_G) + (1 - p_G)A - K] + (1 - \alpha) q_B [p_B(R - w_B) + (1 - p_B)A - K].$$

An incentive compatible plan must satisfy the resource constraints

$$w_G \geq -A, \quad w_B \geq -A, \quad 0 \leq q_G \leq 1, \quad 0 \leq q_B \leq 1;$$

the participation constraints that neither type should expect to lose by participating

$$q_G [p_G w_G - (1 - p_G)A] \geq 0, \quad q_B [p_B w_B - (1 - p_B)A] \geq 0;$$

and the informational incentive constraints that neither type should expect to do better by lying

---

<sup>5</sup>For example, see Hayek (1935) p. 237.

$$q_G[p_G w_G - (1-p_G)A] \geq q_B[p_G w_B - (1-p_G)A],$$

$$q_B[p_B w_B - (1-p_B)A] \geq q_G[p_B w_G - (1-p_B)A].$$

Under socialism, an optimal investment plan should choose  $(q_G, q_B, w_G, w_B)$  to maximize the expected profit  $V$  for society at large, subject to the above constraints. But this problem is easy to solve. Even when the manager has no collateral  $A=0$ , the allocatively-efficient ideal plan with  $q_G=1$  and  $q_B=0$  (financing good managers but not financing bad managers) is feasible with  $w_G = w_B = 0$ . That is, society does not need to pay anything more than the manager's normal cost of time, and the manager has no incentive to lie about his type to get the project funded because he gets no special benefit from managing it.

But under capitalism, competition among investors to finance promising entrepreneurs implies that, in a market equilibrium, uninformed investors cannot make positive expected profits from financing the project. That is, the equation  $V = 0$  is an essential property competitive equilibria under capitalism. Then with  $A=0$ , the ideal  $(q_G = 1, q_B = 0)$  is not feasible in any incentive-compatible equilibrium where  $V=0$  is satisfied.<sup>6</sup> So a potential advantage of socialism here is that a socialist state's monopoly of capital can facilitate honest communication, as bad managers cannot gain from imitating good managers if neither type gets any profits from entrepreneurial management (see Dewatripont and Maskin, 1993).

Thus, in our moral-hazard example, where the problem was to motivate hidden actions, we have found disadvantages of socialism; but in our adverse-selection version of the example, where the problem is to elicit honest reports of hidden information, we have found potential advantages of socialism. In comparison with free-market capitalism, socialism allows individuals to have less private property rights. Giving an individual ownership rights over property can help solve the moral-hazard problem of getting him to exert hidden efforts to manage the property well, but such individual ownership rights also give people different interests which may make it harder for them to communicate honestly with each other, thus exacerbating the informational problems of adverse selection. Conversely, collectivism can often ameliorate adverse-selection problems while exacerbating moral-hazard problems.

---

<sup>6</sup> This is why this adverse-selection example is interesting enough to be analyzed in section 6.2 of Tirole's (2006) corporate-finance book, where competitive financing is assumed.

## 5. A SIMPLE ECONOMY WITH MORAL HAZARD IN THE CENTRAL GOVERNMENT

Proponents of the free-market system do not advocate it merely as an excuse for abandoning egalitarianism. The free market distributes economic power and rights to people throughout the population, and so the free market may be seen as the antithesis of centralized political control of the economy. From this perspective, we may try to derive the rationale for the free market from a model where centralized political control causes economic inefficiency. The costs of unrestrained central power can be understood as problems of moral hazard at the center of government. So our third example involves moral hazard in the government and shows how even an autocratic ruler may prefer to create political guarantees of private property rights, even when such liberalization entails a risk of his losing power. The key is that political liberalization can encourage investments that increase the ruler's tax base.<sup>7</sup>

Let  $Y(K)$  denote the net flow of output (per unit time) that can be produced in a nation when any  $K \geq 0$  is the capital invested in the nation. To produce this output  $Y(K)$ , the capital  $K$  must be used and controlled by many individuals in the general population, whom we may call capitalists, and their control over the capital would enable them to take it abroad at any time. The capitalists' rate of time discounting is  $r$ . So to deter capital flight, the capitalists must enjoy an income flow worth  $rK$  from their capital holdings. We may assume that  $Y(K)$  is net output after input costs and labor, and so the authoritarian rulers of the government can take (in taxes) the remaining flow  $Y(K) - rK$ . Let  $r+b$  denote the rate of time discounting for an authoritarian ruler, which may be different from  $r$  because the authoritarian ruler might face some exogenous risk of losing power, say at a probabilistic rate  $b$  per unit time.

The moral-hazard problem here is that the ruler could, at any time, expropriate the capital and sell it abroad. Of course the capitalists would not invest in this nation if they expected to be expropriated, and so the ruler needs to promise that their capital will not be expropriated. Making this promise credible is the central moral-hazard problem here. If the authoritarian ruler never expropriates, then the expected present-discounted value of his income stream is  $(Y(K) - rK)/(r+b)$ . But without political liberalization, nobody can prevent the absolute authoritarian ruler from breaking his word. If he expropriated the invested capital, the worst that

---

<sup>7</sup>For a broader introduction to such problems of time-inconsistency or political credibility of government policies to protect investors, see also chapter 16 of Tirole (2006).

could happen is that nobody would ever invest in his nation again, and so he would get national income  $Y(0)$  thereafter, but he could still benefit from selling the expropriated capital  $K$  abroad (for use in some other nation where capitalists can still trust the ruler). Thus, it can be credible that the authoritarian ruler will not expropriate the invested capital  $K$  only if

$$(Y(K)-rK)/(r+b) \geq K + Y(0)/(r+b).$$

But now let us allow that the ruler may liberalize his regime to various degrees  $\lambda$ , where a greater liberalization  $\lambda$  means that it is easier for people to organize and overthrow him. To be specific, let us say that a regime has liberalization  $\lambda$  when the probability of the ruler losing power if he tried to expropriate capital (or if he tried to reduce liberalization) would be this number  $\lambda$ , which as a probability must satisfy  $0 \leq \lambda \leq 1$ . Such liberalization  $\lambda$  can help to satisfy the ruler's moral-hazard constraint, because his expected return from trying to expropriate the capitalists would now become  $(1-\lambda)[K + Y(0)/(r+b)]$ .

But liberalization also increases the ruler's political risks. Suppose that there are also false-alarm scandals that occur at some expected probabilistic rate  $a$  per unit time, and people react to such scandals exactly as they would to a genuine attempt to expropriate capital. So for a regime with liberalization  $\lambda$ , when the government is not willfully expropriating anything, still scandals can randomly occur at an expected rate  $a$  per unit time, and so scandals that cause the ruler to lose power will randomly occur at an expected rate of  $a\lambda$  per unit time. So in a regime with liberalization  $\lambda$ , the current ruler should discount future revenue at rate  $r+b+a\lambda$ . Thus, with invested capital  $K$ , the ruler's present discounted value is

$$V(K,\lambda) = (Y(K)-rK)/(r+b+a\lambda).$$

Then the ruler's moral-hazard constraint with liberalization  $\lambda$  is

$$V(K,\lambda) \geq (1-\lambda)[K + Y(0)/(r+b)].$$

In an optimal political regime, the ruler wants to invite investment  $K$  and offer liberalization  $\lambda$  to maximize  $V(K,\lambda)$  subject to this moral hazard constraint, as well as  $K \geq 0$  and  $0 \leq \lambda \leq 1$ .

To be specific in our example, suppose that  $Y(K) = (K+n)^{0.5}$  where the parameter  $n$  denotes the nation's endowment of natural resource, say  $n = 12$ . Suppose that the capitalists' discount rate is  $r = 0.05$ , the basic political-risk rate for authoritarian rulers is also  $b = 0.05$ , and the additional political-risk rate per unit of liberalization is also  $a = 0.05$ . With this  $Y$  and  $r$ ,

if there were no moral-hazard constraint, the ideal capital stock that maximizes the net revenue  $Y(K)-rK$  would have  $K+n$  equal to 100. But with the moral-hazard constraint, this numerical example allows no positive investment if the ruler does not liberalize; that is, if  $\lambda=0$  then only  $K=0$  would be feasible here with  $n=12$ . The ruler's optimal regime for this example has liberalization  $\lambda=0.504$ , so that the ruler can credibly invite capital investment  $K = 52.4$ .

On the other hand, if natural resources were decreased to  $n = 0$ , keeping all other parameters the same, then the ruler's optimal regime would have  $\lambda = 0$  and  $K = 44.4$ . Intuitively, a lack of natural resources makes it more costly for the ruler to lose his reputation for protecting capital, and so he can credibly encourage substantial investments even without liberalizing.

If the endowment of natural resources were increased to  $n = 25$ , however, still keeping all else the same, then the ruler's optimal regime would have  $\lambda = 0$  and  $K = 0$ . Intuitively, a great wealth of natural resources makes the ruler unwilling to accept the additional political risk from liberalization, even though it means that nobody will invest in his nation without liberalization. Thus, an increase in natural resources can reduce the ruling government's incentive to liberalize, and indeed investment may decrease by more than the value of the natural resources themselves. In this sense, a nation may be cursed by too many natural resources.

## 6. CONCLUSIONS

Mechanism design has extended the scope of economic analysis by adding incentive constraints to resource constraints in our definition of the economic problem. Incentive constraints provide an analytical framework for understanding failures of allocative efficiency, showing how such failures may depend on the initial allocation of property rights in a society. But mechanism-design theory changes the basic object of analysis from the resource allocation to the social plan or allocation mechanism that specifies how resource allocations should depend on people's information. Concepts of incentive efficiency can be applied to identify good institutional rules or mechanisms, taking incentive constraints into account.

The cases for collectivism or private ownership may depend on trade-offs between different kinds of incentive problems: moral hazard and adverse selection. Moral-hazard incentive problems particularly are fundamental in any institution, because institutional rules are enforced by actions of leaders and officials, who must be motivated by an expectation of rewards

and privileges as long as they fulfill their institutional responsibilities (see Hurwicz 2007; Myerson 2007, 2008).

Mechanism design and other areas of game theory have contributed to a fundamental change in the scope of economics. Once the scope of economics was defined by the allocation of material goods, but now economists study all kinds of questions about incentives in social institutions. Our theoretical framework now is broad enough to analyze competitive incentive problems in both markets and politics. In the quest for a better analytical understanding of how the wealth of nations may depend on their institutions, economic analysis has returned to the breadth of vision that characterized the ancient Greek social philosophers who first gave economics its name.

#### REFERENCES:

- Robert J. Aumann, "Subjectivity and correlation in randomized strategies." Journal of Mathematical Economics, 1974, 1, pp. 67-96.
- Enrico Barone, "The ministry of production in the collectivist state," in F. A. Hayek, ed., Collectivist Economic Planning, London: Routledge, 1935; translation from Giornale degli Economisti, 1908.
- Emile Borel, "La théorie du jeu et les equation intégrales à noyau symétrique gauche," Comptes Rendus de l'Académie des Sciences, 1921, 173, pp. 1304-1308. English translation by L. J. Savage in Econometrica 21:97-100.
- Ronald Coase, "The problem of social cost," Journal of Law and Economics (1960) 3, pp. 1-44.
- Augustin Cournot, Recherches sur les Principes Mathématiques de la Théorie des Richesses. Paris: Hachette, 1838. English translation by N. T. Bacon, Researches into the Mathematical Principles of the Theory of Wealth, New York: MacMillan, 1927.
- Partha Dasgupta, Peter Hammond and Eric Maskin, "The implementation of social choice rules: some general results on incentive compatibility," Review of Economic Studies 1979, 46, pp. 185-216.
- Mathias Dewatripont and Eric Maskin, "Centralization of credit and long-term investment," in P. K. Bardhan and J. E. Roemer, eds., Market Socialism Oxford, 1993, pp. 169-174.
- Allan Gibbard, "Manipulation of voting schemes: a general result" Econometrica, 1973, 41, pp.



587-601.

Milton Harris and Robert M. Townsend, "Resource allocation under asymmetric information," Econometrica, 1981, 49, pp. 33-64.

John C. Harsanyi, "Games with incomplete information played by 'Bayesian' players." Management Science, 1967, 14, pp. 159-182, 320-334, 486-502.

Friedrich A. Hayek, "The present state of the debate," in F. A. Hayek, ed., Collectivist Economic Planning, Routledge, 1935.

Friedrich A. Hayek, "The use of knowledge in society," American Economic Review, 1945, 35, pp. 519-530.

Bengt Holmstrom, On incentives and control in organizations, Ph.D. thesis, Graduate School of Business, Stanford University, 1977.

Bengt Holmstrom and Roger Myerson, "Efficient and durable decision rules with incomplete information," Econometrica, 1983, 51, pp. 1799-1819.

Leonid Hurwicz, "On informationally decentralized systems," in C. B. McGuire and R. Radner, eds., Decision and Organization: a Volume in Honor of Jacob Marshak, North-Holland, 1972, pp. 297-336.

Leonid Hurwicz, "But who will guard the guardians," to appear in Les Prix Nobel 2007.

Oskar Lange, "On the economic theory of socialism," in B. E. Lippincott, ed., On the Economic Theory of Socialism, U Minnesota Press, 1938.

Louis Makowski and Joseph Ostroy, "General equilibrium and market socialism: clarifying the logic of competitive markets," in K. Bardhan and J. E. Roemer, eds., Market Socialism, Oxford U Press, 1993, pp. 69-88.

Ludwig von Mises, "Economic calculation in the socialist commonwealth" in F. A. Hayek, ed., Collectivist Economic Planning, London: Routledge, 1935; translation of "Die Wirtschaftsrechnung im sozialistischen Gemeinwesen," Archiv fuer Sozialwissenschaften, 1920, 47.

Roger Myerson, "Incentive compatibility and the bargaining problem," Econometrica, 1979, 47, pp. 61-74.

Roger Myerson, "Optimal coordination mechanisms in generalized principal-agent problems," Journal of Mathematical Economics, 1982, 10, pp. 67-81.

- Roger Myerson "Two-person bargaining problems with incomplete information," Econometrica 1984, 52, pp. 461-87.
- Roger Myerson, "Analysis of Incentives in Bargaining and Mediation," in Negotiation Analysis, edited by H. P. Young, University of Michigan Press, 1991, pp. 67-85.
- Roger Myerson, "Nash equilibrium and the history of economic theory," Journal of Economic Literature, 1999, 37, pp. 1067-1082.
- Roger Myerson, "Fundamental theory of institutions: a lecture in honor of Leo Hurwicz," U. Chicago working paper, 2007.
- Roger Myerson, "The autocrat's credibility problem and foundations of the constitutional state," to appear in American Political Science Review, 2008.
- Roger Myerson and Mark A. Satterthwaite, "Efficient mechanisms for bilateral trading," Journal of Economic Theory, 1983, 29, pp. 265-281.
- John F. Nash Jr., "The bargaining problem," Econometrica, 1950, 18, pp.155-162.
- John F. Nash Jr., "Noncooperative games." Annals of Mathematics, 1951, 54, pp. 289-295.
- John von Neumann. 1928. "Zur Theories der Gesellschaftsspiele," Mathematische Annalen 1928, 100, pp.295-320. English translation by S. Bergmann in R. D. Luce and A. W. Tucker, eds., Contributions to the Theory of Games IV, , Princeton University Press, 1959, pp. 13-42.
- John von Neumann and Oskar Morgenstern, Theory of Games and Economic Behavior. Princeton University Press, 1944.
- Jurg Niehans, A History of Economic Theory. Baltimore: Johns Hopkins University Press, 1990.
- Jean Tirole, Theory of Corporate Finance Princeton U Press, 2006.
- Robert W. Rosenthal, "Arbitration of two-party disputes under uncertainty," Review of Economic Studies, 1978, 45, pp. 595-604.
- William Samuelson, "A comment on the Coase Theorem," in Game-Theoretic Models of Bargaining, edited by A. E. Roth, Cambridge University Press, 1985, pp 321-340.
- Xenophon, "Oeconomicus". Translated as "The Estate-Manager" by Robin Waterfield, in Conversations of Socrates, Penguin Classics, 1990.