Fundamental theory of institutions: a lecture in honor of Leo Hurwicz

Roger B. Myerson

"The economic problem of society is not merely a problem of how to allocate 'given' resources . . . It is rather a problem of how to secure the best use of resources known to any of the members of society, for ends whose relative importance only these individuals know . . . it is a problem of the utilization of knowledge not given to anyone in its totality. This character of the fundamental problem has, I am afraid, been rather obscured than illuminated by many of the recent refinements of economic theory, particularly by many of the uses made of mathematics".

F. A. Hayek, "The Use of Knowledge in Society" (1945).

1 Recognizing the need for a fundamental theory of institutions

In the early twentieth century, economic theorists from left and right (Barone 1908; Lange 1938; Mises 1920; Hayek 1935) argued whether socialist reform of economic institutions was possible without loss of economic efficiency. The inconclusive nature of their debate showed that the existing framework of economic analysis was not adequate to formalize the justifications for the strongly held convictions on each side of this vital argument. To allow analytical comparison of fundamentally different forms of economic organization, a new and more general theoretical framework was needed. In an influential paper, Hayek (1945) argued that a key to this new economic theory should be the recognition that economic

The Hurwicz Lecture, presented at the North American Meetings of the Econometric Society, at the University of Minnesota, on June 22, 2006.
Myerson, R.B. Rev Econ Design (2009) 13: 59. https://doi.org/10.1007/s10058-008-0071-6

R. B. Myerson (✉)
Economics Department, University of Chicago, Chicago, IL, USA
e-mail: myerson@uchicago.edu; http://home.uchicago.edu/~rmyerson/research/hurwicz.pdf

W. Trockel (ed.), *Social Design*, Studies in Economic Design,
https://doi.org/10.1007/978-3-319-93809-7_3

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. From this perspective, 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, and 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 has long been the leader.

Over many years and decades, Leo Hurwicz has worked to show how mathematical economic models can provide a general framework for analyzing different economic institutions, like those of capitalism and socialism, as mechanisms for coordinating the individuals of society. Hurwicz (1973) noted that, in late nineteenth-century economics, the institutionalists were economists who avoided analytical modeling. Today, all this has changed, since Leo Hurwicz set the standard for mathematical economists to study institutions as coordination mechanisms.

The pivotal moment occurred when Hurwicz (1972) introduced the 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." By learning to think more deeply about the nature of incentives in institutions, we have gained better insights into important social problems and policy debates. But as Hurwicz (1998) has observed, there are still basic questions in the theory of institutions that we need to understand better.

As one of many followers in this tradition, I feel privileged to have this opportunity of presenting a Hurwicz lecture. In this lecture, I want to take a broad perspective on the state of these questions and what we have learned about them. First, I will re-examine how modern analysis of incentive constraints can help us to see what was missing in the old socialist debates. Then I will follow Hurwicz (1998) in proposing an abstract general model of how institutions are defined and enforced in a broader social environment. Finally, I will consider more specific models of incentive problems in establishing the fundamental political institutions of a society. Throughout, I will suggest a shift away from Hayek's focus on communication. Although we should recognize the universal significance of informational (adverse-selection) incentive problems in all social systems, I will suggest that strategic (moral-hazard) incentive problems may be even more important for understanding the foundations of social institutions.

2 An old debate and a new theoretical framework

In a polemic against naive dreams of a socialist paradise, Mises (1920) argued that prices from a competitive market equilibrium are necessary for efficient allocation of resources. Countering this argument, Barone (1908) and Lange (1938) saw no reason why socialist managers could not be coordinated equally well by value indexes set by a socialist Ministry of Planning. Mises (1920) and Hayek (1935) expressed great skepticism about the feasibility of such central economic planning without free competitive prices, but their argument on this point remained informal, focusing largely on the intractable complexity of the resource allocation problem. It is hard to be persuasive with such arguments of intractability. After all, if the economy is too complex for our analysis, then how can we be sure that a competitive market will find an efficient solution, or that a socialist planner will not find one? For a convincing argument, they needed a simple economic model in which socialism (suitably defined) could be proven to be less efficient than capitalism.

Of course, the later twentieth century provided much evidence of capitalist economic success and socialist economic failure, but a theorist should not give up a good question simply because there seems to be evidence to answer it empirically. If our theories do not give an adequate answer, then we must continue working to develop theories that can, because one can always propose new institutional structures that do not exactly match those for which we have data. If we have no general theory about why socialism should fail, then we have no way to say that greater success could not be achieved by some new kind of socialism that is different from the socialist systems that have been tried in the past.

Economic theorists today have a strong sense of what was missing from the old debates. The old economists could model resource constraints, but not incentive constraints. Hayek and others made verbal arguments that show a basic awareness of incentive problems, but their arguments remained rhetoric without tight logical support in the absence of any general theoretical framework for analysis of incentives.

In particular, Samuelson (1954) argued that no feasible mechanism could guarantee an efficient allocation of public goods, because asking a person to pay for public goods according to his benefit creates an incentive for him to misrepresent his benefit. This remark seemed consistent with the general view that efficiency is found only in competitive private-good markets. But in trying to formalize this argument, Hurwicz (1972) found that the same incentive problems arise in the allocation of private goods, once we drop the assumptions required for perfect competition. He showed that, with finitely many individuals, no incentive-compatible mechanism can guarantee a Pareto-efficient allocation that is at least as good as autarky for all combinations of individual preferences in a broad class. Thus, the concept of incentive-compatibility was introduced.

The concept of incentive-compatibility developed rapidly after Hurwicz introduced it (Myerson 1982). We have come to understand that there are really two kinds of incentive constraints in the general social coordination problem: *informational incentive constraints* that formalize *adverse-selection* problems of

gathering decentralized information, and *strategic incentive constraints* that formalize *moral-hazard* problems of controlling decentralized activity. As Hayek (1945) emphasized, economic plans must make use of decentralized information that different individuals have about their resources and desires. An individual could not be expected to honestly reveal private information that would be used against his interests, and such adverse-selection problems are formalized in economic models by informational incentive constraints. But economic plans must be implemented by decentralized actions of many different individuals, and there is a problem of getting individuals to accept appropriate guidance and direction when they have conflicting strategic incentives. An individual could not be expected to obediently refrain from opportunistic behavior that would be more rewarding to him, and such moral-hazard problems are formalized in economic models by strategic incentive constraints.

So, although the old socialist debates took place at a time when formal economic models only took account of resource constraints, we have now expanded the scope of economic analysis to take account of informational and strategic incentive constraints. If there was any validity to the intuitive arguments of Hayek and Mises, we should now be much better able to formulate them analytically in our new incentivist framework. Thus, we should ask, what is the simplest model in which we can support Mises's and Hayek's conclusions about socialism's failure?

Mises (1920) saw the essential problem arising in socialist allocation of capital, because state ownership of means of production implies the lack of any capital market. Such questions about mechanisms for allocating capital are a topic of corporate finance. Jean Tirole's *Theory of Corporate Finance* (2006) is full of models applying mechanism design to corporate finance, and we may naturally look to these models for insights into the old debate on socialism. Tirole has many models with many different features, but they are generally based on two simple models: one of moral hazard (Sect. 3.2), one of adverse selection (Sect. 6.2). Each model describes a simple world which we can transform by socialist reforms, and we can see how the efficiency of capital allocation is affected. The result may tell us something about what is truly fundamental in our models.

3 Advantages of socialism in a simple adverse-selection model

In Tirole's (2006, section 6.2) basic adverse-selection model, a manager has private information about the probability of success for a unique investment opportunity. The basic parameters of the model are $(I, A, R, p_H, p_L, \eta)$. Here I denotes the capital investment cost required for new project. The parameter A denotes the value of assets that manager can pledge to forfeit if project fails. The parameter R denotes the returns from the project if it succeeds, but the returns will be 0 if the project fails. The probability of success depends on the manager's type. If the manager's type is high then the project's probability of success in the project is p_H; but if the manager's type is low then the project's probability of success is p_L, where $p_L < p_H$. The manager knows his own type, but it is uncertain to anyone else, and

the manager can lie about his type. Let η denote the probability of the manager being the high type. For simplicity here, let us assume risk neutrality and no discounting of future returns (zero interest rate). We assume that

$$p_H R > I > p_L R \quad \text{and} \quad I > A$$

so that the project is worthwhile only if the manager's type is high, but the manager does not have enough wealth to undertake the project himself.

Under socialism, there is no problem getting the manager to reveal type honestly, because he is willing to report his type honestly when we just pay him a flat wage no matter what he reports. If we want to give him strict incentives to guide social decision-making about the project, the state could pay the manager $\varepsilon(R - I)$ if the project succeeds, but make him pay εI if the project fails. For any $\varepsilon > 0$, this payment plan would give the manager a positive incentive to recommend the project only when its expected social profit is positive. Feasibility requires $\varepsilon I < A$, but for any endowment size $A > 0$, this liquidity constraint can be satisfied when $\varepsilon > 0$ is sufficiently small.

This example is interesting for Tirole (2006) because he is assuming that competition among investors in the financial market always lets the manager borrow at an interest rate such that investors get expected profit equal to zero given their information about the manager. With access to such competitive lenders, low-type managers would want to imitate high-type managers to get their favorable terms of credit. But under socialism, the monopolistic state lender can fully exploit the high-type manager, and then the low type would not want to borrow at all. So we find that socialism may actually have an advantage here, because socialism can flatten the manager's incentives to eliminate his temptation to lie about his chances of success (for other advantages and disadvantages of a monopolistic supply of credit, see Dewatripont and Maskin 1993).

4 Disadvantages of socialism in a simple moral-hazard model

In Tirole's (2006, section 3.2) basic moral-hazard model, the probability of success depends on the manager's actions (instead of the manager's hidden type). Most of the parameters here (I, A, R, p_H, p_L, B) are as in the previous model: the parameter I denotes the capital investment cost required for new project, A denotes the value of assets that manager can pledge to forfeit if project fails, and R denotes the returns from the project if it succeeds, but the returns will be 0 if the project fails. Now p_H is the probability of success if the manager behaves appropriately, but p_L is the probability of success if the manager misbehaves, where $p_L < p_H$, and B denotes the value of private benefits that the manager gets by misbehaving. We assume that

$$p_H R > I > p_L R + B \quad \text{and} \quad I > A,$$

so that the project is worthwhile only if manager behaves appropriately, but the manager cannot undertake the project alone.

As individuals should have only modest wealth under an egalitarian socialist system, let us suppose that the manager's assets are bounded by the inequality

$$A < Bp_H/(p_H - p_L).$$

In a social investment plan, let w denote the wage that will be paid to the manager if the project succeeds. Then a feasible plan must satisfy

$$p_H w - (1 - p_H)A \geq 0$$
$$p_H w - (1 - p_H)A \geq B + p_L w - (1 - p_L)A.$$

Here the first constraint is a participation constraint, that the manager should not expect to lose by participating in the project. (We are assuming that the social investment I includes a payment to the manager for the opportunity cost of his time in managing the project). The second constraint is a strategic incentive constraint, that the manager should not expect better rewards from opportunistic misbehavior. The expected social profit, to be maximized, is

$$Y = p_H(R - w) + (1 - p_H)A - I.$$

The participation constraint implies $w \geq A/p_H - A$, and the moral-hazard constraint implies $w \geq B/(p_H - p_L) - A$. So with our modest-wealth assumption, the lowest feasible wage is

$$w = B/(p_H - p_L) - A,$$

which yields expected social profit

$$Y = p_H R + A - Bp_H/(p_H - p_L) - I.$$

(Because the manager is risk neutral, we could not increase Y by adding payments to the manager when the project fails). Thus, the manager must be allowed to get a moral-hazard rent that has expected value

$$p_H w - (1 - p_H)A = Bp_H/(p_H - p_L) - A.$$

Notice that the expected social profit Y is strictly increasing in the manager's collateral A.

Now let us add the possibility that managers can be punished, and let x denote the punishment cost inflicted on manager if the project fails. Then a feasible mechanism (w, x) must satisfy the participation constraint

$$p_H w - (1 - p_H)(A + x) \geq 0,$$

and the strategic incentive constraint

$$p_H w - (1 - p_H)(A + x) \geq B + p_L w - (1 - p_L)(A + x).$$

The punishment x is not assumed to yield any social value to anyone else. So expected social profit is still

$$Y = p_H(R - w) + (1 - p_H)A - I.$$

The participation and incentive constraints now imply

$$w \geq (A + x)(1/p_H - 1) \text{ and } w \geq B/(p_H - p_L) - (A + x).$$

With modest endowments $A < Bp_H/(p_H - p_L)$, the wage cost is minimized by the punishment

$$x = Bp_H/(p_H - p_L) - A,$$

which allows the wage

$$w = B(1 - p_H)/(p_H - p_L)$$

and so yields the expected social profit

$$Y = p_H R + (1 - p_H)[A - Bp_H/(p_H - p_L)] - I.$$

Thus, punishment of failures can improve social profit. But increasing the manager's private collateral A still helps, even when punishment is allowed.

On the other hand, if there are rich agents who have assets A greater than $Bp_H/(p_H - p_L)$ then we could achieve the ideal social profit $Y = p_H R - I$, by letting the project be managed by such a rich agent for the wage $w = A(1 - p_H)/p_H$ to be paid if the project succeeds, but taking his collateral A if the project fails, with no further punishment ($x = 0$). This wage makes the participation constraint binding ($p_H w - (1 - p_H)A = 0$) and satisfies the moral-hazard constraint with $w + A \geq B/(p_H - p_L)$.

So there are two obvious ways for socialist reformers to achieve full efficiency here. First, they could allow some individuals to hold more wealth, up to $Bp_H/(p_H - p_L)$. Perhaps such favored people could be heroes of the socialist revolution (or of the Norman conquest). Second, they could drop the participation constraint and force people to become managers without compensation for punishment risks. Perhaps such disfavored people might be prisoners or enemies of the state. But either way, socialism looks rather less appealing from the perspective of this moral-hazard model, as it forces us to admit either inequality or coercion or productive inefficiency into our imagined socialist paradise. Indeed, our simple model does not do badly as a source of theoretical insights into the flaws of Soviet communism,

and it formalizes some of Hayek's informal intuitive arguments: "To assume that it is possible to create conditions of full competition without making those who are responsible for the decisions pay for their mistakes seems to be pure illusion" (Hayek 1935, p. 237).

5 Comparing moral hazard and adverse selection

The comparison of these two models suggests that, when we probe the logical foundations of social institutions, moral-hazard problems may be more fundamental than adverse-selection problems. The problems of motivating hidden actions can explain why efficient institutions give individuals property rights, as owners of property are better motivated to maintain it. But property rights give people different vested interests, which can make it more difficult to motivate them to share their private information with each other. Thus, adverse selection might not be so problematic if there were no moral hazard. Socialism differs from capitalism in allowing less property rights for individuals, but moral hazard provides a fundamental economic rationale for some property rights that must apply even under socialism. So adverse-selection problems can be important under socialism, just as under capitalism.

For example, take Tirole's basic moral-hazard model with no punishment ($x = 0$), but now let us add a small probability ε that the manager is a bad type who cannot do better than the p_L probability of success (and cannot get the benefit B). With small collateral $A < p_L B/(p_H - p_L)$, such a bad manager would imitate the good type, to enjoy the positive expected benefits $(p_L B/(p_H - p_L) - A)$ from getting his project financed. So in the presence of moral hazard, the socialist system loses its ability to trivially solve informational adverse-selection problems.

On the other hand, if the uncertainty in the basic adverse-selection model were about the required investment amount I (instead of the success probability p), the socialist planner would have to allow informational rents to low-I type managers. But nobody would even try to take these rents away if the manager were a capitalist entrepreneur.

More generally, even if incentive analysis of other adverse-selection models does not reveal actual disadvantages of socialism, it can help to show that the supposed advantages of socialism may be less than its advocates would have suggested when they failed to recognize the possibility of opportunistic misrepresentations under systems other than capitalism. Analysis of mechanism design with informational incentive constraints has taught us that individuals with unique private information may have to be allowed informational rents in an efficient mechanism. But mechanism design as a conceptual framework can fit capitalist or socialist institutions, and so it can help us to see that the manager of a socialist monopoly who has private information about production costs (and can divert unaudited profits) may extract informational rents that look essentially like the profits of a monopoly in capitalism. Conversely, a capitalist monopoly's profits could be regulated away if its costs were publicly known, and it may be the monopolist's private information about

costs that enables him to fend off such regulation. Thus, mechanism design teaches us that having multiple independent sources of supply may be just as important under socialism as under capitalism, which traditional market models could not show. Soviet planning may have suffered from failing to recognize such benefits of informational decentralization.

6 General theory of institutions enforced in larger games

In recent work, Hurwicz (1998) has focused on questions of how institutions are constructed and how institutional rules are enforced (see also Schotter 1981). Here strategic incentive constraints are at the heart of the problem, so we can focus on games in strategic form where N is the set of players, C_i denotes the set of strategies of player i, and $U_i(c)$ denotes the utility payoff to player i from strategy profile c in $C = \times_{j \in N} C_j$.

To a game theorist, an institutional reform means changing the structure of the game that people play in the institution. So it is common for game theorists to think that institutions are games. But Hurwicz (1998) observes that what we normally mean by institutions (or institutional arrangements) typically does not include the specification of individuals' preferences (nor does it typically include the beliefs that we specify in Bayesian games). So an institution for Hurwicz is more properly to be identified with a *game-form* of Gibbard (1973), specifying only the set of players N, the sets of strategies C_i for each player i, and an outcome function $\Theta : C \to Y$ that defines how outcomes in some set Y would depend on the players' strategies. Such game-forms are mechanisms in Hurwicz's sense. To analyze such a game-form or mechanism, however, we must specify each player i's preferences for outcomes by a utility function $u_i : Y \to \mathbb{R}$ on the outcome set Y. With these outcome-based utility functions, we can then define the strategy-based utility functions $U_i(c) = u_i(\Theta(c))$ that complete the structure of the strategic-form game which corresponds to the institution once preferences are given.

When we ask how an institution is established, we must embed it somehow in a larger game. For example, when two people play a game of chess, typically each of them is physically able to grab the other's king at any time, but is deterred from chess-illegal moves by the damage such behavior could do to one's reputation in the larger game of life. So the chess game seems supported by some kind of reputational equilibrium in a larger more fundamental game. But saying "games are equilibria of larger games" cannot be right, because if chess were embedded as an equilibrium in the game of life, then that equilibrium would specify each player's strategy in the chess game itself.

Hurwicz (1998) explains that, if our *legal game* $G = (N, (C_i)_{i \in N}, (U_i)_{i \in N})$ is embedded in some true game H, the structural relationship must be that $H = (N, (D_i)_{i \in N}, (U_i)_{i \in N})$ has a larger strategy spaces

$$D_i \supset C_i \quad \forall i \in N$$

and has utility functions that extend those of the legal game G to the larger domain $D = \times_{j \in N} D_j$. Hurwicz (1998) then suggests that a strong formulation of successful enforcement could require that, for each player i, each illegal strategy outside C_i should be dominated by some legal strategy in C_i, so that a player's best responses always take him into the legal game, even if others deviate.

Hurwicz (2008) remarks, however, that a normally law-abiding player might not want to remain law-abiding when others are acting illegally, and so a weaker concept of enforcement may be appropriate. Thus, I would suggest that the definition of institutional enforcement should be weakened, to say that G is enforceable in H when

$$\forall i \in N, \quad \forall c_{-i} \in \times_{j \in N-i} C_j, \ \forall d_i \in D_i \backslash C_i, \quad \exists c_i \in C_i \text{ such that } U_i(c_{-i}, c_i) > U_i(c_{-i}, d_i),$$

so that each player's optimal actions are in his legal strategy set when all others' actions are expected to be in their legal sets. That is, G is enforceable when its strategy sets form a *curb set* in H, as defined by Basu and Weibull (1991) (Curb sets are closed under rational behavior).

This weaker definition of enforceability can admit a great multiplicity of enforceable institutions for a given environment, because a big true game H can contain many different minimal curb sets. This multiplicity may seem an annoying indeterminacy, to theorists who believe in economic determinism. But I would argue that the right mathematical model of institutions should admit such a multiplicity of solutions, because real institutions are manifestly determined by cultural norms and traditional concepts of legitimacy, which would have no scope for effect if the economic structure of the true game H admitted only one dominant solution.

For example, legal rules of a political constitution that are written on a piece of parchment in a museum may be enforced in a true game that involves millions of people on a large land-mass. What would prevent anyone from writing another set of rules (on a bigger piece of parchment) and acting according to them instead? Under any political constitution, such an act should be punished as sedition or treason by others who accept the given constitutional rules. But although treason never prospers, the definition of what is treason depends on an arbitrary social consensus. We all understand that a broad failure to agree about constitutional rules and authority can create an anarchy in which everyone suffers. So the social process of identifying what are the constitutional rules of politics and who are the legitimate leaders of our society has the basic structure of a coordination game with multiple equilibria, where the outcome must depend on culture and tradition through Schelling's (1960) focal-point effect.

The essential role of the focal-point effect in the foundations of our basic political institutions has been emphasized by Hardin (1989) and Myerson (2004, 2008). The new theoretical point here is that Schelling's focal-point effect can be extended to questions of selecting among multiple curb sets, just as among multiple equilibria. Once everyone understands that everybody else will be restricting themselves to

strategies in one particular constitutional curb set, it becomes rational for each individual to stay in his or her respective portion of this curb set.

Although people may be symmetric in the true game H, this symmetry can be broken in the curb set G. Indeed, the enforcement of a constitutional curb set may depend crucially on a small group of specially designated individuals (called law-enforcement officials) whose curb-set strategies stipulate that they would punish any deviator who violated constitutional restrictions.

7 Moral hazard and privilege in sovereign political institutions

The preceding model of how institutions are enforced in larger games is very abstract. To move from broad abstractions to practical specifics, we need to think more carefully about the officials who are the guardians of our institutions, as Hurwicz (1998) has emphasized. Let me follow him now in examining the basic question of who guards these guardians, that is, who forces the enforcers to enforce our laws.

Consider again the problem of enforcing the fundamental political institution of a nation, such as the Constitution of the United States. A constitution can be effective only when there are agents who expect to be rewarded for implementing its rules. In particular, it must designate officials who are expected to prosecute sedition and other violations of the constitution, so as to deter the rest of the population from such illegal moves. But what makes these officials do their official function? Of course, a problem of getting people to do what they are supposed to do is what we call a moral-hazard problem. So the basic problem of getting government officials to enforce constitutional rules is a moral-hazard agency problem in the upper levels of government.

Such an agency model has been analyzed by Becker and Stigler (1974). They recognized that powerful officials have regular opportunities to profit from abuse of power, and that such abuse of power can be difficult for others to detect. For abuse of power to be deterred, the official must expect to do better by acting to enforce the rules correctly, and so must expect substantial rewards that would forfeited if evidence of abuse of power were discovered. Assuming risk-neutrality, the magnitude of these rewards must be at least the potential profit that the official could earn from abuse of power divided by the probability that such abuse of power would be discovered. So when temptations are large and probabilities of detection is small, powerful officials may need to be very well rewarded. Thus, we should expect the leaders of fundamental political institutions to be a very well-rewarded elite, highly motivated by the need to preserve their privileges, as Michels (1915) observed even of socialist political parties.

So our concept of a constitution is incomplete if we ignore the essential role of those who expect enjoy the privileges of high office under the constitution and are

therefore well motivated to act to sustain it. From a purely structuralist perspective, it might seem that a political constitution could be fully defined by specifying (1) a set of political offices, (2) the powers, privileges, and responsibilities of these offices, and (3) the procedures for selecting future holders of these offices. But to fully characterize a political constitution as a self-enforcing dynamic system, embedded in a true game where people are symmetric, one must also specify (4) the privileged individuals who actually hold these offices at some initial time (or who expect to be on the short list of serious candidates for these offices in the first elections). In this sense, the specific identity of the small privileged group who are called "Founding Fathers" of the American Republic may be considered an essential component of the American Constitution, as essential as the words written on old parchment in Philadelphia.

If moral-hazard opportunities imply that responsible officials must be well rewarded, then people should be willing to pay for promotion to such offices. In Becker and Stigler (1974) theory, an efficient organization would pass the cost of an official's incentive rewards back to the official ex ante, by charging a fee for promotion to the office. In effect, a candidate for office would be asked to post a bond, which would be returned to the official on retirement if there is no evidence of malfeasance. Such a plan appears to be a simple efficient solution to the fundamental agency problem of government. But it creates a new moral-hazard problem at the highest level, because it implies that the leader who controls appointments to high offices will have an incentive to convict officials of malfeasance and resell their offices. The whole scheme depends on the promise that high officials will be appropriately judged, so that they can expect to be rewarded for correct service and punished for abuse of power, but there may be no neutral party to make such judgments. An official must always be worried that others in the power structure would be tempted to convict him of malfeasance and sell his position to someone else.

Thus, the organizational problem of metering rewards, which Alchian and Demsetz (1972) considered for economic producers, arises even more forcefully for political organizations. Indeed the terms of the problem may be sharpened in the political context, where there can be no question of looking to some higher court for adjudication of contractual relationships.

Hurwicz (1998) recognized that the guardian officials of a sovereign political institution must in some sense be organized into a circle of mutual monitoring and judgment, where the actions of each individual are monitored and judged by others in the circle. But when an individual i is called to monitor the actions of some individual j in such a circle, the monitored actions may include *j*'s monitoring of yet other individuals, which further broadens the scope of activity that individual i must be prepared to observe. So some collective aspect of the fundamental adjudication process seems unavoidable. Within a ruling political faction that admits no higher court of appeal, membership in the faction may require an individual to keep manifestly informed about the general status of other members, perhaps formally by attending regular factional meetings, or informally by staying current in a factional network of gossip.

So the survival of a political institution must depend on its being led by some faction or core group of powerful officials who share a basic trust in each others' judgments. In effect, the members of this group may form a court where they each have a right to be tried before any punishment or loss of privilege. In such a court, evidence of malfeasance against any of them would be commonly heard, so that all members of the group should be able to evaluate whether resulting judgment was reached appropriately. The collective sanction against wrongful judgments in this court could be that the members of this ruling faction would lose trust in each other, so that they would all switch to an equilibrium where each opportunistically abuses his individual power. We may assume that, in a competitive world, a faction would not long hold political power over a large society if members of the faction could not solve free-rider problems in collective actions to defend their power against challenges from other potential factions (Myerson 2008). With this assumption, any one member of a ruling faction could feel protected by the expectation that her colleagues could not mistreat her without risking a general loss of mutual trust within the faction, which would jeopardize all of their privileged positions.

8 Leadership and moral hazard at the center

To be more specific about how such factions are formed, we must recognize the role of leaders as entrepreneurs of institutions. Throughout history, governments have been formed by political leaders whose path to power began by gathering a trusted group of active supporters. When a faction has been organized in this way, privileges of membership in the faction are allocated by the leader. Then the circle of monitoring can be closed by a simple factional rule that the leader should never remove a member's privileges without a process of judgment that is collectively witnessed by other members of the faction. Indeed, rulers throughout history have generally maintained courts or councils, where high officials and others close to the ruler were regularly gathered, and where the ruler's treatment of any courtier could be witnessed and scrutinized by other courtiers. Thus, each individual courtier could feel confident of getting appropriate rewards from the leader, because of the leader's need to maintain a general reputation for appropriately rewarding all courtiers, who are the primary agents of his power.

Popular books on leadership have filled shelves in bookstores, but their descriptions of leadership are often focused on leadership as visionary strategic decision-making (Maxwell 2002). Of course, when people need to coordinate, they may look to a leader for strategic decisions about whether to attack at dawn, or at noon, or not at all. But when we ask what is really the essential function of a leader, I would suggest that the role of strategic planner may be generally less important than the role of honest monitor and reliable paymaster that Alchian and Demsetz (1972) identified. A leader makes a group into an effective team by his reputation for actively monitoring the contributions of individuals in the group and appropriately rewarding their efforts. Such a reputation with a group of supporters, small enough

to be individually monitored but large enough to achieve competitive success by their collective actions, is the essential asset that defines a leader. If a leader loses this reputation for appropriately rewarding the members of his group, then the leader must be replaced or the group will lose its ability to compete with other teams that have better leadership.

This idea dates back at least to Xenophon, whose "Education of Cyrus" (c. 360 BCE) depicts a great leader who establishes a great empire by cultivating a reputation for honestly and generously rewarding captains who serve well in battle. While other leaders think that their power depends on the assets in their treasury, Cyrus understands that his power really depends on his credit with his captains, so that it can be better to pay out generously than to keep anything for himself. In another paper (Myerson 2008), I have analyzed a similar model of the foundations of the state by leaders whose ability to hold power depends on their reputation for reliably rewarding the captains who support them against their rivals in contests for power.

An economic entrepreneur must be able to credibly promise future payments both to the investors who supplied his initial capitalization and to the managers whose moral-hazard opportunities require promises of large future rewards. Similarly, a political leader must be able to credibly promise future rewards both to the supporters or captains whose efforts put him in power and to the high officials or governors through whom his power is exercised.

To further probe the difficulties of maintaining a reputation for appropriately rewarding agents in political institutions, let me describe one more model of moral hazard by high government officials, an extension of the Becker–Stigler model that I have analyzed (Myerson 2015). In this model we consider a high official, whom we may call a governor, in a state that is ruled by a single leader or monarch. At any time, the governor can behave well (govern appropriately), or misbehave (govern corruptly), or openly rebel. The leader cannot directly observe whether a governor is behaving or misbehaving, but he can observe any costly crises that occur in the governor's province. Crises occur as a Poisson process with a low expected rate α when governor behaves, but a high expected rate β when the governor misbehaves, where $\beta > \alpha$. Misbehavior also gives the governor a flow of additional hidden benefits that are worth γ per unit time. The governor observes any crisis in her province shortly before the leader does, but she can be called to court for a brief visit during which rebellion is impossible. Let D denote the expected payoff to the governor when she rebels (which is observable to the leader). Crises and rebellions are very costly for the leader, so he wants his governors to always behave well, that is, to never misbehave or rebel. Each individual is risk neutral and has discount rate δ.

To be deterred from rebellion, a governor must always expect rewards that are worth at least D. Candidates for governor can be asked to pay something for promotion to the office, but any candidate's ability to pay is limited by her wealth, which we denote by A. We assume that a governor's potential gains from rebellion are greater than the private wealth of any candidate for office, so $A < D$. On the other hand, the leader may feel tempted to free himself of his debts to a governor,

by sacking the governor, and such moral hazard at the top is essential to the problem of political leadership. To admit it into our model as simply as possible, we assume an upper bound H on the debt that the leader can be trusted to owe a governor. These parameters $(\alpha, \beta, \gamma, D, \delta, A, H)$ characterize our model.

To minimize the leader's expected cost of paying governors, the optimal incentive plan (derived in Myerson 2015) can be characterized at any time by the expected present discounted value of all future rewards to the incumbent governor, which we may call the governor's credit. To deter hidden misbehavior, any crisis in the province must cause the governor's credit to decrease by a penalty that has expected value

$$\tau = \gamma/(\beta - \alpha).$$

Normally, the sanction for a crisis should be to reduce the governor's credit by this amount τ. But the governor would rebel if her credit ever went below D after a crisis, and so the governor's credit beforehand must never be less than $D + \tau$. So if a crisis occurs when the governor's credit U is less than $D + 2\tau$, then the governor should be called to the leader's court for a trial, where the outcome is either to reinstate the governor at the credit $D + \tau$ with probability $(U - \tau)/(D + \tau)$, or else to dismiss the governor (who thereafter gets 0) and instead appoint a new governor at the minimum feasible credit level $D + \tau$.

Thus, the need to deter both hidden misbehavior and open rebellion requires the leader to make randomized decisions about whether to dismiss or forgive a governor after a crisis. But the leader is not indifferent in such situations, because dismissing the incumbent governor would create an opportunity to resell the office to a new governor for the payment $A > 0$. So the process of deciding a governor's fate in such a situation must be actively monitored by others, because otherwise the leader's ex-post incentive would always be to dismiss the governor. That is, the leader needs to institutionalize a formal trial procedure where others (whose trust he needs to maintain) can observe that he has given the governor an appropriate chance of reinstatement before any dismissal.

The expected discounted value of the leader's cost, at any point in time, is equal to the credit U that he owes to the current governor, plus the expected discounted value of the leader's net cost of promises to other governors who will be promoted into the position after dismissals in the future ($D + \tau - A$ at each promotion). So the optimal plan for the leader should minimize the expected frequency of future dismissals, which can be achieved by keeping governors as far as possible from the low credit range (below $D + 2\tau$) where dismissals occur. Thus, in the optimal incentive plan, a governor should be paid only in credit, not in cash, until the credit bound H is reached. To keep promises to a governor, her credit should increase between crises at the rate $U' = \delta U + \alpha\tau$ until it reaches the bound H on what the leader can be trusted to owe. When the credit owed equals H, the governor should be paid $\delta H + \alpha\tau$ until the next crisis causes her credit to drop to $H - \tau$. In this solution, increasing the trust bound H would strictly decrease the leader's expected discounted cost, as assessed ex ante when a new governor is first appointed. But

with very high H, the leader will ultimately incur large expensive debts to governors who become entrenched in their offices.

That is, even when the leader has the same discount rate as the high officials of his government, the need to deter them from abuse of power creates a motivation for the leader to become a debtor to these officials. Of course this conclusion is just an extension of the results of Becker and Stigler (1974) analysis. Our extended model has been designed only to show how problematic this debt relationship can be, because (to deter corruption) the leader must sometimes actually dismiss officials without paying them their promised rewards, but the circumstances of these dismissals cannot be simply predictable (to avoid rebellions) and so can be verified only by actively monitoring the judgment process, during which the leader's natural incentive is actually to dismiss rather than reinstate (because he can resell the office).

Thus, someone needs to actively monitor the leader's judgments of his high officials and constrain him to act according to an optimal random rule. But who can have such power over the leader of a sovereign political institution? The other high officials on whom his regime depends have such power, because they would rationally misbehave or rebel if they lost trust in the leader's promises of future rewards. (In particular, the leader's problem of deterring misbehavior and rebellion would become infeasible if the amount H that they trust him to owe ever became less than $D + \tau$.) So a sovereign political leader needs a court or council where high officials witness his appropriate treatment of other high officials. Such high councils of government seem universal in political systems. In them, the chief guardian's reputation for rewarding his supporters is collectively guarded by his chief supporters.

Thus, in our fundamental theory of institutions, we should recognize that political institutions are established by political leaders, and political leaders need active supporters. Like a banker, a leader's promises of future credit must be trusted and valued as rewards for current service. The leader's relationship of trust with his inner circle of high officials and supporters requires that they must act collectively to monitor and verify his judgments against any of them. Such a relationship of trust with a group of supporters, small enough for the leader to personally monitor but large enough to effectively control the larger institutions of government, is a political leader's most valuable asset. Furthermore, the members of this group must share a sense of identity, in that each must be confident that the leader's wrongly punishing any one of them could cause all others to lose trust in the leader.

So the establishment of fundamental institutions by political leaders may ultimately rely on a sense of identity among members of a group that is small enough to gather in a court of common judgment to hear a case against any one of them. From this perspective, we can make sense of cases throughout history where powerful political forces have been led by small groups of people who are connected by narrower forms of identity, such as family relationships, or old school ties, or bonds of personal loyalty to their leader, even though these personal connections may seem to have no intrinsic relationship with anyone's position on great questions of national policy. Like the nineteenth century socialists, we may dream of great utopian social reforms, but we should understand that the institutions of any such

brave new world would be built on narrower factional foundations, organized by political leaders whose first imperative is to maintain their reputation for rewarding loyal supporters.

References

Alchian AA, Demsetz H (1972) Production, information costs, and economic organization. Am Econ Rev 62:777–795

Barone E (1908) The ministry of production in the collectivist state. In: Hayek FA (ed) Collectivist economic planning (Routledge, London, 1935); translation from Giornale degli Economisti

Basu K, Weibull JW (1991) Strategy subsets closed under rational behavior. Econ Lett 36:141–146

Becker G, Stigler G (1974) Law enforcement, malfeasance, and compensation of enforcers. J Legal Stud 3:1–18

Dewatripont M, Maskin E (1993) Centralization of credit and long-term investment. In: Bardhan PK, Roemer JE (eds) Market Socialism. Oxford University Press, Oxford, pp 169–174

Gibbard A (1973) Manipulation of voting schemes: a general result. Econometrica 41:587–601

Hardin R (1989) Why a constitution. In: Grofman B, Wittman D (eds) The federalist papers and the new institutionalism. Agathon Press, NY, pp 100–120

Hayek FA (1935) The present state of the debate. In: Hayek FA (ed) Collectivist economic planning. Routledge, London

Hayek FA (1945) The use of knowledge in society. Am Econ Rev 35:519–530

Hurwicz L (1972) On informationally decentralized systems. In: McGuire CB, Radner R (eds) Decision and organization: a volume in honor of Jacob Marshak. North-Holland, Amsterdam, pp 297–336

Hurwicz L (1973) The design of mechanisms for resource allocations. Am Econ Rev 63(2):1–30

Hurwicz L (1998) But who will guard the guardians. University of Minnesota paper, http://www.econ.umn.edu/workingpapers/hurwicz_guardians.pdf, revised for Nobel Lecture in American Economic Review 98(3):577–585 (2008)

Lange O (1938) On the economic theory of socialism. In: Lippincott BE (ed) On the economic theory of socialism. University of Minnesota Press, MN, USA

Maxwell JC (2002) Leadership 101. Thomas Nelson, Inc., Nashville

Makowski L, Ostroy J (1993) General equilibrium and market socialism: clarifying the logic of competitive markets. In: Bardhan K, Roemer JE (eds) Market socialism. Oxford University Press, Oxford, pp 69–88

Michels R (1915) Political parties: a sociological study of oligarchic tendencies in modern democracy. Hearst, NY

Myerson RB (1982) Optimal coordination mechanisms in generalized principal-agent problems. J Math Econ 10:67–81

Myerson RB (2004) Justice, institutions, and multiple equilibria. Chicago J Int Law 5:91–107

Myerson RB (2008) The autocrat's credibility problem and foundations of the constitutional state. Am Polit Sci Rev 102(1):125–139

Myerson RB (2015) Moral hazard in high office and the dynamics of aristocracy. Econometrica 83:2083–2126

Samuelson PA (1954) The pure theory of public expenditure. Rev Econ Stat 36:387–389

Schelling TC (1960) Strategy of conflict. Harvard University Press, Cambridge

Schotter A (1981) Economic theory of social institutions. Cambridge University Press, London

Tirole J (2006) Theory of corporate finance. Princeton Unviersity Press, Princeton

von Mises L (1920) Economic calculation in the socialist commonwealth. In: Hayek FA (ed) Collectivist Economic Planning (Routledge, London, 1935); translation of Die Wirtschaftsrechnung im sozialistischen Gemeinwesen. Archiv fuer Sozialwissenschaften 47

Xenophon (2001) The education of cyrus, translated by Wayne Ambler. Cornell University, Ithaca