May 13, 1998 Leonid Hurwicz

# But Who Will Guard the Guardians?

Contrary to what one might have guessed, in posing the famous question,[1] Juvenal was not concerned with affairs of state or politics, but rather trying to convince a friend that marriage is folly, women are not to be trusted, and keeping them locked under guard is not a solution—because the guards could not be trusted either.

But, half a millennium or so earlier, Plato[2] did raise a closely related issue in discussing standards of behavior appropriate for the guardians of the city-state, the best of whom were to be chosen as rulers, thus in the context of ideal structure of governance. Socrates, referring to an earlier statement[3] that "drunkenness is most unbefitting guardians," says: "From intoxication we said that they must abstain. For a guardian is surely the last person in the world to whom it is allowable to get drunk and not know where on earth he is." To which Glaucon, Socrates' interlocutor, replies: "Yes, it would be absurd[4] that a guardian should need a guard." Instead of Juvenal's later pessimism, indeed cynicism, Plato—through Glaucon—expresses the optimistic view that one should be able to trust the city's guardians and rulers to behave properly; that they should require oversight is an absurdity.

Even a casual perusal of daily newspapers should be sufficient to convince us that there is nothing absurd about the present day "guardians"—leaders and officials of political, economic, and social entities—needing, and indeed getting a great deal of oversight.[5] The question is rather as to the extent oversight is, or even can be, effective.

---

[1] "Sed quis custodiet ipsos custodes?" Liber secundus, Satura VI, lines 347–8, p. 325, in *D. Junii Juvenalis Saturarum Libri V*; mit Erklärenden Anmerkungen von Ludwig Friedlaender, Erster Band, Leipzig, Verlag von S. Hirzel, 1895.

[2] Book III, XII, 403E, p. 264 (Greek) and p. 265 (English), in volume I, of Plato, *The Republic* (ΠΟΛΙΤΕΙΑ), with an English translation by Paul Shorey, London, William Heinemann Ltd.; New York: G.P. Putnam's sons, MCMXXX.

[3] Ibid., pp. 246–7, no. 398E.

[4] In B. Jowett's 1908 translation (Ocford, Clarendon Press, 3rd ed., reprinted 1928) "ridiculous" replaces "absurd". [Jowett's rendering seems better because the etymology of the Greek word used by Glaucon appears to be "laughable"—as is that of "ridiculous".] But the recent translation by A.D. Lindsay (Everyman, J.M. Dent, London; Charles E. Tuttle, Vermont, 1992, 1995) again uses "absurd".

[5] Without explicitly mentioning the classical precedents, the Wall Street Journal of May 8, 1998 (pp. B1–B2) carries a story under the headline "Guardians May Need Someone to Watch Over Them." It speaks of court-appointed guardians and conservators who dishonestly dissipate their wards' assets and of the

The publicly expressed attitudes toward government, law enforcement, union and corporation leadership are often more reminiscent of Juvenal than Plato.

The problem is obviously of central importance for political science, but why should it be of interest to economists, and, in particular, to theorists? And if there are good reasons for interest, how should we go about analyzing the problem? It is to these questions that the present essay is devoted.

An answer to these questions lies in the increased importance being attributed to the role of institutions[6] in influencing economic phenomena, and it hinges on the role of implementation (in particular, prevalence of the rule of law and its enforcement) as an essential ingredient in the functioning of institutions.[7]

The economic importance of institutions, stressed at least as far back as the 19th century German Historical School, is not, I believe, controversial. Some of the most basic contemporary policy issues involve choice of institutions: markets versus central planning, the scope and structure of social insurance (unemployment, old age, health), "property rights" as solutions to problems posed by externalities, world free trade, and the degree of economic integration of independent nations are obvious examples. The question is rather whether the role of institutions can be captured by appropriate analytical tools and incorporated into economic models, so as to become an integral part of the theoretical edifice. Metaphorically, whether institutions can be introduced into models as variables, even as unknowns, rather than embedded as fixed parts of the landscape (as is, for instance, perfect competition in so many mainstream models). It is only when such models are available that we can face the issue of incorporating implementation devices, their limitations and potentials.

Much economic analysis is based on the perfectly competitive model. Implicitly, at least, this model (or, more precisely, its applicability) requires strong assumptions

---

difficulties the courts have in exercising their oversight responsibilities. In turn watchdog groups and legislative task forces try to improve the performance of the judicial branch.

[6] In the sense of institutional arrangements("rules of the game") rather than entities such as various types of organizations ("artificial players").

[7] E. Ostrom, Walker and Gardner point out that emphasis on the importance of enforcement is found in Hobbes, although they disagree with his stress on the need for an external enforcing authority (the sovereign). Their stress on intra-group mutual enforcement. Schotter (1981, p. 11) makes self-policing or external policing authority an integral part of the definition of a social institution. He recognizes the possibility of intra-group enforcement through a supergame model (p. 165, note 8).

concerning the information available to "agents" (individuals, firms, etc.) engaging in economic activity as well as the existence of implementation mechanisms such as the enforcement of contracts and absence of collusion. Similarly, conclusions concerning the effects of alternative forms of taxation, subsidies, social insurance depend in an essential way on implementation mechanisms supplying information concerning obligations and entitlements, entities facilitating financial flows, as well as enforcement of payments or disclosure of relevant information. I think it may be fair to say that until recently in economic model building [as distinct from *obiter dicta*] much more attention has been paid to the information requirements (and uncertainty when precise information is not available) than to problems of implementation. Yet if implementation is impossible or prohibitively costly, even the most attractive mechanism remains a utopia.

Progress relevant to the issues being raised here occurred in the 1960's and '70's in connection with the study of informationally decentralized and game-modeled mechanisms.

**Message exchange (non-game-theoretic) processes.** Taking as a point of departure the assumption[8] that information is dispersed among participants in the economic process, a process is defined as informationally decentralized when each economic unit (consumer, firm, etc.) only has information about itself (its preferences, technology, or resources) but not about the characteristics of the other units. The process then requires an exchange of signals (called messages) in order to attain objectives such as efficiency of the system as a whole. Once the amount of information carried by the signals has been quantified,[9] it makes sense to speak of the informational requirements of the process if objectives are to be attained in a specified environment. In rigorously formulated models it has, for instance, been possible to prove that in the presence of increasing returns to scale no finite-dimensional message space is adequate to guarantee the efficiency (in the sense of Pareto optimality) of the equilibrium outcomes of any informationally decentralized

[8] Recognized already in the 1920's in the context of debates about the feasibility of socialism (Lange, Hayek, Mises) but especially stressed by Hayek (1945).

[9] When messages are vectors (*n*-tuples of numbers) the dimensionality of the messages provides such quantification. Specifically, the informational requirements of the process are measured by the dimension of the message space (e.g., the sum of dimensionalities of the messages produced by the various participants), a concept analogous to the capacity of a communication channel. See Hurwicz (1972a), Calsamiglia (1976).

process.[10] A similar result has been obtained for economies with detrimental externalities such as pollution.

These "impossibility" results apply not to a particular mechanism or institutional structure but to all mechanisms qualifying as informationally decentralized. They tell us that even *if one assumes that the participants will be truthful and will abide by whatever the rules of the process prescribe*, no rules can guarantee the attainment of the desired objectives (e.g., efficiency). Such results must be viewed as specifying limitations on implementability due to the dispersion of information even in the presence of perfect enforcement. Thus enforceability of rules is not always an issue.

But, of course, in this essay, we are primarily interested in institutions where one cannot assume that participants will always be truthful. Is it possible to design processes that would give the participants to be truthful? The question was raised by Samuelson in the mid-'50's in his classic articles on public goods, and in particular on the so-called Lindahl solution. He pointed out that for the Lindahl mechanism to work properly one has to rely on the participants' truthful revelation of their preferences which, however, they would find advantageous to misrepresent. Hence the Lindahl solution did not take care of the classic free rider problem. But Samuelson went on to make a stronger claim: he further stated that the same problem would arise with any other decentralized mechanism. Could these two claims be formalized and rigorously justified?

**Game-theoretic framework.** A natural framework for such analysis turned out to be the theory of non-cooperative games formulated by Nash and the equilibrium concept now named after him. The economic process could be viewed as a non-cooperative game[11] in which the strategy of each participant was a statement (not necessarily truthful) about his/her preferences and the participant's utility of the outcome (defined in Samuelson's analysis by the Lindahl formula) as the "payoff". Samuelson's first claim could then be

[10] In the jargon of the field, Pareto optimality cannot be "realized" by a decentralized ("privacy preserving") process with a finite dimensional message space. In addition to informational decentralization, these results presuppose certain mathematical regularity properties of the process rules. These properties involve a strengthening of continuity and are usually referred to as "smoothness". (Lipschitz-continuity is an example.)

[11] In this essay we confine ourselves to non-cooperative games in normal form, although in some contexts extensive form would have been more appropriate. This so, in particular, because rules of a game typically specify which moves (rather than strategies) are prohibited. (However, prohibitions may be aimed at strategies. This seems to be the case in anti-trust law.)

formalized as the proposition that truthful revelation of preferences is not a Nash equilibrium in the Lindahl game.[12] That is, when results (level of public goods and the required individual contributions) are calculated according to the Lindahl formula based on the participants own statements about their preferences, and given that all participants other than the *i*-th participant are being truthful, it would be in general to *i*'s advantage (yield higher utility) to misrepresent his/her preferences. Not surprisingly, thus formalized, Samuelson's first claim turned out to be correct: truth was not a Nash equilibrium.

But Samuelson's second, more general (impossibility) claim, was more difficult to deal with—this for two reasons. To analyze the claim one had to answer two questions: first, how broad a class of mechanism would qualify as decentralized, and, second, what types of the system's performance are viewed as desirable. As to the latter, primary attention was paid to efficiency (Pareto optimality). As to the former, a significant broadening of perspective was provided by Groves and Ledyard (1977) who introduced an element of (previously used in non-game-theoretic models) message exchange processes into the Nash type model and admitted as strategies any signals ("messages"), not necessarily just revelations of one's own preferences. I.e., they moved *beyond* the narrow class of *revelation games* characteristic of earlier work. They showed that this generalization enabled them to construct a (non-revelation) model yielding Pareto-optimal (though not Lindahl) Nash equilibrium outcomes in economies with three or more participants. Later (Hurwicz 1979a) it was shown that even Lindahl outcomes could be obtained as a Nash equilibrium of a suitably designed non-revelation game.[13] If one regards this broader (not necessarily revelation) class of mechanisms as decentralized, Samuelson's second (impossibility) claim cannot be accepted. Hence optimism replaces pessimism. However, is there still a problem of enforcement?

**Are Nash equilibria self-enforcing?** But what about the implementability of mechanisms such as those (mentioned above) designed by Groves and Ledyard or Hurwicz? Is there a problem of enforcement? In fact, one occasionally hears the claim

[12] See Hurwicz (1972b) and Ledyard and Roberts (1974).
[13] This mechanism as well as that in Hurwicz (1979b) had defects involving individual feasibility, but unpublished results Hurwicz (1996b) shows how to repair the defects.

that there can be no enforcement problem because (allegedly) Nash equilibria are self-enforcing. Suppose, for instance, that $\left(s_1^*,\ldots,s_n^*\right)$ is the $n$-tuple of Nash equilibrium strategies in some game (where $n$ is the number of players). By definition, this means that, for any player $i$, a unilateral departure from his/her equilibrium strategy $s_i^*$ to an alternative strategy $s_i$ cannot increase $i$'s payoff. Hence once equilibrium is established, there is no incentive for unilateral departure, and collusions are infeasible in a non-cooperative game, hence no need for an enforcement mechanism.

But this argument has two implicit assumptions. One is that the only strategies that player $i$ would consider as alternatives to $s_i^*$ are members of the admissible strategy domain, say $S^i$, prescribed by the mechanism governing the system, what we might call the "legal" strategies; i.e., that any alternative strategy $s_i$ belongs to $S^i$. But player $i$ may, in fact, have available to him/her some "illegal" strategies, those that are physically feasible but not in the admissible domain, that are more advantageous to player $i$, given that everybody else is behaving legally. Hence the above argument has the implicit assumption that there are no illegal feasible strategies or, at least, none that might be advantageous.

The other assumption implicit in the argument that Nash equilibria are self-implementing is somewhat less obvious. A player's strategic calculations are based on the structure of the payoff function, which in turn depends on the physical outcomes produced by the mechanisms given the strategic choices made by players. These outcomes may involve delivery of goods or payments by various participants. Typically, there must exist some machinery (involving not only enforcement but also information processing and financial procedures) ensuring that the specified outcomes will in fact be produced. An assumption underlying the above claim of implementability of Nash equilibria is that such machinery is in place and in effective operation.

**Formalizing rules of a game as game-forms (mechanisms).** It will be helpful at this point to formalize the concept of a mechanism and its relation to the non-cooperative game (in normal form) defined by it. A game $G$ in normal form for $n$ players is defined by their admissible strategy domains $S^i$ and payoff functions $p^i, i = 1,\ldots,n$. (A payoff function is a numerical-valued function whose domain is the set of admissible *n*-tuples of strategies, i.e., the joint strategy domain, defined as the Cartesian product of the individual domains, i.e., $S = S^i \times \cdots \times S^n$. So $p^i : S \to \Re$, where $\Re$ represents the real numbers.) We write $p = \left(p^1, \ldots, p^n\right)$, and formally define the *game* as $\Gamma(S,p)$. A mechanism shares with the game the strategy domains, but differs in that it involves an outcome function, say $h : S \to Z$ where *Z* is the space of conceivable outcomes (resource allocations in economic models, candidate lists in voting models). Assume further that each player *i* has a utility function $u^i : Z \to \Re$, associating a real outcome with any conceivable outcome.[14] Then the *i*-th player's payoff function is obtained as the composition of the *i*-th utility function with the outcome function. I.e., for any *s* in *S*, and any *i*,

$$p^i(s) = u^i\left(h(s)\right)$$

In special cases $h(s) = \left(h^1(s), \ldots, h^n(s)\right)$, where $h^i(s)$ may represent the monetary payment to player *i* and (1) the *i*-th utility function depends only on the *i*-th payment of which it is (2) a linear function. After a normalization, this can be written as $u^i\left(h^i(s)\right) = h_i(s)$, hence $p^i(s) = h^i(s)$, so that in such special cases the *i*-th payoff function can be identified with the *i*-th component of the outcome function. But in general [when (1) and (2) may not both hold] this identification does not hold.[15] In any case it is important conceptually to distinguish outcome functions from payoff functions, because typically we consider preferences (hence utility functions) as given data (part of the "environment") while outcome functions, as well as strategy domains may be subject to the designer's choice, say as institutional reforms.

Indeed, in a model devoted to the analysis of institutional arrangement it is essential to separate that which belongs to the data ("environment"), thus preferences, endowments, and technologies, from that which is subject to human manipulation, in our model the strategy domains *S* and the outcome functions *h*. It is natural to refer to the

[14] It would be more realistic to define the utility function on a subset of the outcome space and our analysis would be in no way affected, but we make the more extreme assumption to avoid notational complications.
[15] For example, when the mechanism represents a system of voting for the country's president, the outcome function cannot be split into components. But even when individual monetary payments are involved, the non-linearity of utility functions, typical in situations involving risk, would invalidate the identification.

pair $(S,h)$ as the *rules of a game* since the domain $S$ defines the class of legal moves (strategies) and the outcome function $h$ their consequences. On the other hand. In game theory jargon the pair $(S,h)$ is called a *game-form*; in economics, a *mechanism*. The mechanism is not affected by changes in preferences but the payoff functions are. The game-form can be changed directly by legislation or other human actions; the payoff functions only indirectly through changes in the game-form.

To make sense of the need for enforcement, one must admit the possibility of behavior that violates the rules of the game. In a typical game the admissible strategy domains do not cover all behaviors that are physically (or psychologically) feasible. In card games, for instance, certain types of signaling are physically possible, but are prohibited by rules of the game. (E.g., in bridge, the individual strategy domain $S^i$ does not include kicking your partner under the table.) As indicated above, one reason why enforcement may be needed is that actions not included in the admissible domain may be advantageous for some players. After all, that is why it may be desirable, or even necessary, to guard the guardians! But, again as indicated above, in applications the effectiveness of the outcome function may also require either enforcement or other implementation apparatus (informational, financial, etc.).

To formalize, we introduce into the player's domain of choice strategies that are prohibited by the rules of the game. We refer to such strategies as illegal, while those that are permitted by the rules are called legal.

To represent the actual choices facing players we take the radical step of introducing the set of all feasible actions (the *true strategy domain*) to be denoted by $S'$, and correspondingly the *true outcome function* $h': S' \to Z$, representing the consequences of any combination of strategy choices, whether legal or not. That is, we consider the participants to be playing the "*true game*" whose true game-form is the pair $(S',h')$ and the *i*-th payoff functions is $p'^{i} = u^i(h'(s))$. (Here $s$ is $n$-tuple of strategies whose any component may be legal or illegal.) We shall denote the true game (whose rules are $(S',h')$) by $\Gamma'$ while the legal game (governed by the legal strategy domain $S$ and the legal outcome function $h$) is denoted by $\Gamma$. To simplify matters, we shall provisionally[16] assume that all legal strategies of a given mechanism $(S, h)$ are feasible, i.e., that $S$ is a subset of $S'$, and that the two outcome functions coincide when only legal strategies are used. Also, it will be assumed that $S'$ is a Cartesian product of individual feasible domains $S'^i$. Denoting the feasible strategies of player $i$ by $S'^i$, we then have $S' = S'^1 \times \cdots \times S'^n$, with any $n$-tuple of individually feasible strategies (i.e., elements of $S'^i, i = 1, \ldots, n$) being feasible. Furthermore each $S'^i$ is a disjoint union of the legal strategy set $S^i$ with the illegal strategy set denoted by $S\sim^i$.[17]

We are thus in the presence of two games: the true game $\Gamma'$ and the legal game $\Gamma$. Although it is the legal game $\Gamma$ we want the participants to play, they are in fact playing the true game $\Gamma'$.

**Successful enforcement and implementation.** This framework makes it possible to formalize the notions of enforcement and implementation.

*Successful enforcement.* To say that the *legal game rules are being successfully enforced* means that the outcomes of the true game that the use of illegal strategies is less attractive than that of legal strategies. A strong formulation of successful enforcement might require that for every player every illegal strategy is *dominated* by some legal strategy. I.e., that for every $i$, every ($n$-$1$)-tuple $s_{-i}$ of strategies of players other than $i$, and every illegal strategy $s_i$ in $S\sim^i$ available to player $i$, there exists a preferable alternative legal strategy $s\wedge^i$ in $S^i$, so that

$$u^i\left(h'\left(s\wedge^i, s_{-i}\right)\right) > u^i\left(h'\left(s_i, s_{-i}\right)\right).$$

("Weak" domination would permit the replacement of some—but not all—of these strict inequalities by weak inequalities, i.e., by $>=$, for certain choices of $s_{-i}$. I.e., in some situations player $i$ would be no worse but no better by staying within the law.)

[16] As pointed out below, an institutional structure may be necessary to ensure that in fact $h = h'$ when all players use legal strategies.

[17] Certain game models imply that $S = S'$, i.e., that all feasible strategies are included in $S$. Thus in the Prisoner's Dilemma game, the prisoner's two ("legal") choices are sometimes stated as "to denounce" or "not to denounce." Since other aspects of behavior are assumed not to affect the outcome (or at least not to be relevant), in effect the (exhaustive!) class of legal behaviors is coextensive with that of feasible behaviors.

However, this may be asking too much: if everyone else is acting illegally, a player may not find it possible to remain law-abiding. It seems, therefore, more reasonable to adopt a somewhat *weaker concept of successful enforcement* of the rules of a given mechanism $(S,h)$, namely to require that (1) the only Nash equilibria of the true game $\Gamma'$ are *n*-tuples of strategies that are legal for the given mechanism (i.e., using strategies from *S* only); (2) that these Nash equilibria of $\Gamma'$ are also Nash equilibria of the game $\Gamma$ defined by the given mechanism,[18] and (3) that the set of Nash equilibria of true game $\Gamma'$ be non-empty.

*Implementation*.[19] Consider what is involved in making effective an institution such as social security. The desired mechanism $(S,h)$—say as defined by legislation[20]—specifies the class of persons receiving payments, say in relation to previous income and other variables. It may also specify the sources of funding. There is need to formulate a *modus operandi*, verify whether specific applicants are entitled to receive payments and, if so, at what level, and how to collect the required funds. Typically, special agencies are created to accomplish such tasks, both informational and related to enforcement. It is the complex of such activities and arrangements that I think of as the effort to implement the legislation. The implementation activities and arrangements define the true game-form and hence the true game $\Gamma'$. *Implementation is successful* if the equilibrium outcomes correspond to those of the desired game, i.e., those envisaged by the legislation.[21]

[18] Or, at least, that such relations hold for the components of Nash equilibrium outcomes corresponding to behavior subject to enforcement attempt.
[19] The concept of implementation used in the present essay, perhaps closer to common usage, is different from that introduced by Maskin (1977) and used in much of the subsequent literature. [Maskin's implementation concept involves a relationship between Nash equilibrium outcomes of a given mechanism and an optimality criterion ("social choice rule"); it implicitly assumes that no player will use strategies outside of those permitted by the game-form. My concept involves activities designed to make the given game-form effective—in particular to discourage use of prohibited strategies and to make the assumed outcome function a reality. For my concept of implementation, it is the desired game-form that is being implemented, although, of course, when successful, such implementation contributes to the realization of the optimality criterion. Thus the concept of implementation used in this essay involves the relationship of two games—the true game and the desired game, rather than the desired game and the optimality criterion.
[20] In this essay I simplify the discussion by identifying legislated rules with a specific mechanism. But in practice, legislation tends to be rather vague on many specifics, often qualitative rather than quantitative. This can be formalized by considering legislated institutional arrangements as classes of game-forms rather than specific game-forms (see Hurwicz, 1996a).
[21] In the context of most mechanisms that are considered in the literature, the above definition must be further qualified. The reason is that enforcement (or any implementation) requires the use of resources. If

Expressed in this framework, a reason why Nash equilibria are not self-enforcing is that in the absence of enforcement a Nash equilibrium only makes it unprofitable to move to alternative strategies in $S^i$ but not necessarily to those in $S^{\sim i}$. I.e., the fact that a strategy *n*-tuple *s* is a Nash equilibrium for game $\Gamma'$, does not in general imply that every illegal strategy is dominated by a component of *s*.

A reason why Nash equilibria cannot be considered self-implementing is that the assumption of effectiveness of the outcome function $h(\cdot)$ hides the need for institutional arrangements typically required to accomplish this. Thus even if all players behave legally, the assumption that $h = h'$ requires implementing actions. It is insufficient merely to postulate the desired outcome function *h*.

**Back to Juvenal**. Let us now come back to the initial problem, the need for guarding the guardians. Juvenal's cynical question suggest either that there is no way to guard the guardians, hence it is impossible to enforce the wives' desired behavior, or that, in addition to having "guardians of the first order" (those guarding the wives), one must also have "guardians of the second order" to guard the guardians of the first order. But then, if those are also subject to corruption, guardians of the third order are also necessary, and so on. This conjures the image of an infinite regress of guardians, with the guardian of order *k* needed to guard the guardian of order *k*, with $k = 2,3,\ldots$, *ad infinitum*. If an infinity of guardians is not usually available, this seems to preclude the possibility of enforcement!

Why do we care about Juvenal's (or rather the husband's) problem? Mainly because some view it as a parable for the proposition that enforcement is in principle impossible—due to the infinite regress of corrupt guardians.

---

the desired model does not take this into account, it typically asks for efficiency relative to total resource endowment of the economy ("gross efficiency"). But this is infeasible when resources needed for implementation are subtracted. At best one can hope for efficiency relative to resources net of those diverted for implementation ("net efficiency"). Hence even if implementation is successful it can only provide equilibria that are net-efficient and hence, formally, not the gross-efficient equilibria considered in the desired mechanism.

“Casual empiricism” suggests that the pessimism of this proposition is not always justified. We know of many situations where rules are substantially (if not perfectly[22]) implemented and/or enforced. But many factors are in the picture determining whether implementation of rules is possible.

First, and perhaps least important, there may exist methods of implementation that depend on the purely *physical or mechanical* factors and, in fact do not require human guardians. Examples: devices used by some parking lots that result in cutting tires of those who use unauthorized exits or entrances; punishments involving the placing of the culprit on an isolated island without a boat, and too far to swim ashore.

Second, somewhere at a finite end in the chain of guardians there may be guardians (individual or collective) who are in sympathy with the rule (game-form) that makes certain behavior illegal, e.g., whose ethical standards rule out corrupt behavior, and who have the ability (through power, financial assets, personal charisma or status combined with population’s respect for it), as well as the inclination, to act so as to discourage improper behavior of the guardians of lower order. (In some of my writings, I have referred to such individuals or groups and “*intervenors*”.)

In such a situation the rule is likely to be successfully enforced. Well functioning societies try to choose judges and rulers from among such individuals.[23] Juvenal’s pessimistic question suggests “infinite regress” of needed guardians, hence non-existence of intervenors. Glaucon’s comment on the other hand assumes self-control on the part of Plato’s guardians, who in effect qualify as intervenors. Thus, in terms of their views, Juvenal and Plato are at opposite ends of the spectrum.

But we do not have to rely on the presence of intervenors. There are other structures conducive to successful enforcement. Consider, for example, a rule of law that is designed to protect citizens from harmful or dangerous behavior of certain individuals. Suppose that those charged with enforcement of the law (first order guardians) are corrupt or otherwise ineffective, and so are their supervisors (second order guardians). If the latter are holders of elective office, citizens (voters) may be viewed as third order

[22] A more realistic model of performance would allow for imperfections. This could be accomplished by postulating a random choice of enforcement attempts when rule breaking is known or suspected and recognizing that only a fraction of enforcement attempts is likely to be successful.

[23] Some years ago we heard of a city where corruption spread from the police to local judges. It was stopped by a “sting” operation by a higher level uncorrupted judiciary group.

guardians (as well as being guarded by first order guardians). They have both an incentive and the power to intervene by throwing the supervisors out of office. This gives the supervisors an incentive to make sure that first order guardians discharge their duties properly. Effective enforcement is the result. A similar situation, in principle, is if the elective guardians are those of any (finite) order $k$. (In the previous example $k = 2$). Graphically, the situation may be represented as a *closed circle*, with the voters being given two numbers (0 and $k+1$), and so that guardian of order $m$ (with $m = 1, 2, \ldots, k+1$) "guards" the guardian of order $m-1$. (The voters, as guardian $k+1$, guard the top guardian $\#k$. At the same time, as citizens they are guarded by guardian $\#1$.)

Thus every one is a guardian but also guarded. The voters as "guardians" of top guardians is an essential aspect of democracy. This type of structure is also closely related to (but not identical with) the notion of separation of powers.[24]

The circular model can be applied in Juvenal's satire, where, by hypothesis, the husband himself is unable to monitor either the wife's or her immediate (first order) guardian's behavior. But, in principle, it is conceivable that a second order guardian might be found who could effectively supervise the first order guardian's discharge of his duties and be so beholden to the husband that he would in effect make sure the immediate guardian does discharge his duties. This would close the circle.[25]

On the other hand, suppose that such a person of second (or higher order) does not exist. Then the husband's objective is not implementable. In this there is no paradox. It reminds us that, depending on circumstances, implementation may or may not be possible.

When implementation is possible, it can be modeled in terms of the relationship of the true game to the desired mechanism (or game). It is important to note that the equilibria of the true game depend on what is feasible and the actual consequences, as well as on the preferences (sometimes called attitudes or values) of the players. Thus, for instance, an intervener must have preferences that rule out his/her own illegal behavior

[24] E. Ostrom, Walker, and Gardner (1992) stress the possibility of mutual enforcement within a group (analogous to the above circular example), as opposed to the need for enforcing authorities external to the group, with theoretical models as well as empirical examples involving the allocation of common pool resources. It seems clear that such intra-group enforcement is likely to be effective in certain classes of situations but not all.
[25] More generally, it might take a chain of $k$ such guardians.

(especially corruption) and harbor a dislike for others' illegal behavior (at least of the relevant type); these factors would affect the intervener's utility function. On the other hand he/she must have powers, assets, or charisma enabling him/her to influence others in the desired direction; these features come under strategies available to the intervener and the consequences of their use, hence are contained in the true game-form. Depending on the participants in the game (i.e., composition of the society) and their preferences/values, a given desired mechanism may or may not be implementable within that society. The history of experience with prohibition may be an illustration for the negative case, Clearly, implementability depends on the nature of the object or goal of implementation and the attitudes toward the likely outcomes of successful implementation.

What are our conclusions? Juvenal to the contrary, enforcement (or more generally, implementation) is not always impossible. But even when it is possible to construct (a theoretical) mechanism "M-implementing" (i.e., implementing in Maskin's sense[26]) an optimality criterion, it could still be the case that "genuine implementation" (the concept used in this essay) is infeasible (or too costly in required resources if the criterion ignores the costs of implementation).[27]

Furthermore, even when "genuine implementation" is feasible and not prohibitively costly, it may require institutional arrangements not evident from the appearance of the mechanism implementing the optimality criterion in Maskin's sense.

Nevertheless it is my opinion, tinged with a dose of self-interest, that research aimed at discovering mechanisms implementing—in Maskin's sense!—various optimality criteria is important and deserves encouragement. Where such implementation turns out impossible, this should serve as (negative) guidance to reformers, and the problem of "genuine" implementation does not arise. When we do discover M-implementing mechanisms, we must still investigate the problem "genuine"

---

[26] A game-form weakly implements an optimality criterion (social choice rule) for a class of environments in Maskin's sense, if, for every admissible environment, all its Nash equilibrium outcomes are optimal in terms of that criterion, and there exist some equilibria (the set of equilibria is non-empty). Full implementability in Maskin's sense requires that every optimal outcome be an equilibrium outcome.

[27] Since equilibria of the true game depend on preferences, this includes political infeasibility (due, for instance to high costs, value conflicts, or other factors).

implementation (including enforcement where relevant), but at least we have an idea of what might be worth (genuinely) implementing.

## REFERENCES

Calsamiglia, X., "Decentralized resource allocation and increasing returns," *J.E.T*, 14, pp. 263-283.

Groves, T. and J. Ledyard (1977), "Optimal allocation of public goods: A solution to the 'free rider' problem," *Econometrica*, 45: pp. 783-811.

Von Hayek, F.A. (1945), "The use of knowledge in society," *American Economic Review*, 35: pp. 519-530.

Hurwicz, L. (1972a), "On the dimensional requirements of informationally decentralized Pareto-satisfactory processes," in *Studies in Resource Allocation Processes*, Cambridge University Press, pp. 413-424.

Hurwicz, L. (1972b), "On informationally decentralized systems," in *Decision and Organization*, C.B. McGuire and R. Radner, eds., North-Holland, Amsterdam and London, pp. 297-336.

Hurwicz, L. (1979a), "Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points," *Review of Economic Studies*, 46: pp. 217-225.

Hurwicz, L. (1979b), "Balanced outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points for two or more agents," in *General Equilibrium, Growth and Trade*, J.R. Gree and J.A. Sheinkman, eds., New York, Academic Press.

Hurwicz, L. (1993), "Implementation and enforcement," in *Political Economy, Institutions, Competition, and Representation*, W.A. Barnett, M.J. Hinrich, and N.J. Schofield, eds., Cambridge University Press, Cambridge, Chapter 2, pp. 51-59.

Hurwicz, L. (1996a), "Institutions as families of game forms," *The Japanese Economic Review*, 47: pp. 113-132.

Hurwicz, L. (1996b), "Feasible balanced outcome functions yielding constrained Walrasian and Lindahl allocations at Nash equilibrium points in economies with two agents when the designer knows the feasible set," unpublished paper, presented in May 1997 at the Decentralization Conference, Pennsylvania State University.

Juvenal, D. (1895), *Junii Juvenalis Saturarum Libri V*, mit Erklärenden Anmerkungen von Ludwig Friedlaender, Erster Band, Leipzig, Verlag von S. Hirzel.

Maskin, E. (1977), "Nash equilibrium and welfare optimality," working paper, Cambridge, MIT.

[See also, Maskin, E. (1965), "The theory of implementation in Nash equilibrium: A Survey," in *Social Goals and Social Organization, Essays in Memory of Elisha Pazner*, L. Hurwicz, D. Schmeidler and H. Sonnenschein, eds., Cambridge University Press, Cambridge, Chapter 6, pp. 173-204.

Ostrom, E., Walker, J. and R. Gardner (1992), "Covenants with and without a sword: self-governance is possible," *American Political Science Review*, 86: pp. 404-417.

Plato, *The Republic* ( ΠΟΛΙΤΕΙΑ), with an English translation by Paul Shorey. London: William Heinemann Ltd; New York: G.P. Putnam's Sons, MCMXXX.

Plato, *The Republic*, translated by A.D. Lindsay. Everyman, J.M. Dent, London; Charles E. Tuttle, Vermont, 1992, 1995.

Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge University Press, Cambridge.