THE SOCIAL CHOICE PERSPECTIVE: A SECOND LOOK

SOCIAL CHOICE THEORY AND THE IMPERFECTABILITY OF A LEGAL ORDER

Allan Gibbard*

The value of a legal order presumably lies in the ways it affects people's lives for better or for worse. Because by this vague standard all legal systems seem to some extent defective, investigating possibilities for improvement is a useful, even necessary task. Axiomatic social choice theory, as it has developed from Professor Arrow's pioneering work, can be interpreted as addressing this issue. As Hobbes argued, a legal order can make the lives of all affected better than they would be in a state of nature—a state with no effective legal constraints on action. That one's life is better than short,

* Professor of Philosophy, University of Michigan. B.A., 1963, Swarthmore College; Ph.D., 1971, Harvard University.


The finall Cause, End, or Designe of men, (who naturally love Liberty, and Dominion over others,) in the introduction of that restraint upon themselves, . . . is the foresight of their own preservation, and of a more contented life thereby; that is to say, of getting themselves out from that miserable condition of Warre, which is necessarily consequent . . . to the natural Passions of men, when there is no visible Power to keep them in awe, and tye them by feare of punishment to the performance of their Covenants, and observation of [the] Lawes of Nature . . . .

Id.
nasty, and brutish is but a poor recommendation for the system that
makes it so; we shall want to know not only that the existence of an
effective legal order can enhance human lives, but by how much.
The conceptual apparatus of axiomatic social choice theory allows us
to pose this question in a highly abstract form, and gives us at least a
negative part of the answer. What the theory says can be put as
follows: Apart from the value of freedom itself, no legal order can
assure as just and efficient an accommodation of the interests of
those it governs as could an omniscient, benevolent ruler who is
obeyed in every particular.\(^3\)

This conclusion will come as no surprise. There is a difference,
though, between suspecting a truth and establishing it, and the
imperfectability in principle of a legal order is well worth establishing.
It is helpful to know that some barriers to the amelioration of legal
systems are barriers in principle, and to know what those barriers
are. Moreover, the history of economic thought lends some initial
plausibility to the opposite conclusion: Only individual irrationality
and defective law stand in the way of achieving efficiency with jus-
tice.\(^4\) From the time of Adam Smith, economic theories have sug-
gested that a legal order can conjure forth an invisible hand—that
under a suitable system of property rights, things would work out for
the best in a world of rational agents.\(^5\) More refined theories have
shown failures of the invisible hand in free markets, but ingenious
incentive-compatible devices have been constructed to remedy some
of those failures.\(^6\) Axiomatic social choice theory, however, reveals

\(^3\) For a full discussion of this theory, see Gibbard, \textit{Social Decision, Strategic Behavior,}
and \textit{Best Outcomes}, in \textit{Decision Theory and Social Ethics} 153 (H. Gottinger & W.
Leinfellner eds. 1978).

\(^4\) Adam Smith argued that national income is maximized in an economy free of mer-
cantile restrictions. \textit{A. Smith, The Wealth of Nations} 397-455 (E. Cannan ed. 1904) (1st
ed. London 1776). A rigorous, somewhat qualified formulation of this claim can be proved for
a perfectly competitive economy. In this form the claim is known as the two fundamental
theorems of welfare economics. See Varian, \textit{Distributive Justice, Welfare Economics and the

\(^5\) Theories such as the two fundamental theorems and the invisible hand fail to account
for economies with imperfect information, with firms large enough to influence the prices they
face, or with public goods or externalities. There is an enormous economic literature on such
market failure. See, \textit{e.g.,} T. Scitovsky, \textit{Welfare and Competition} 241-479 (1971); H.
Varian, \textit{Microeconomic Analysis} 52-79, 197-247 (1978). For a claim that the problem of
externalities can be solved efficiently under idealized conditions by a system of private property
rights and negotiation, see Coase, \textit{The Problem of Social Cost,} 3 \textit{J.L. \& Econ.} 1 (1960). \textit{But see}

\(^6\) Centralized demand-revealing processes or incentive-compatible mechanisms have at-
ttempted to solve the public goods problem but suffer from severe limitations. \textit{See Groves \&
the imperfection not only of the invisible hand of the market, but of
the visible hand of the state and of the cheshire-cat hand at work in
a mixed economy—in any legal order.

ASSUMPTIONS AND THEIR RATIONALE

A great strength of axiomatic social choice theory is that, even
with much left unsaid, significant conclusions may be drawn. Never-
theless, the vague claim of imperfectability that I have broached
here needs elucidation and qualification. What is meant by a “legal
order,” and what are its inevitable limitations? What should we as-
sume about human nature—especially with regard to human motiva-
tion and the sources of intrinsic reward in life? What standard of
value should we use to evaluate a legal order?

First, perfection is not to be understood as a nirvana in which
each person leads the best conceivable life. Rather, it connotes a sys-
tem in which one makes the best use of limited resources. I treat
only those courses of events that are considered technologically feasi-
ble in the economists’ sense.7 A possible course of events is techno-
logically feasible if it would result from everyone’s choosing appro-
priately from among those things available for choice. Put another
way, we could bring about a technologically feasible course of events
if we all so desired and were able to coordinate our actions accord-
ingly. For these purposes, then, perfection equals the best course of
events that is technologically feasible.

The term “best” involves a standard of value. That standard
may include happiness, the satisfaction of preferences, or a pluralis-
tic mixture of good things in life. I shall suppose that we may speak
of a person’s “good” and may order the lives a person would lead
according to the particular value or values he holds most sacred, or
according to the intrinsic reward he would attain from leading such
a life. I shall assume that all the intrinsic value of a course of events
stems from the intrinsic reward of the lives people lead in that course
of events. Furthermore, I shall assume a version of the Pareto princi-
ple: that other things being equal, the more intrinsically rewarding a
person’s life, the better, intrinsically, is the course of events of which

Ledyard, Optimal Allocation of Public Goods: A Solution to the “Free Rider” Problem, 45
ECONOMETRICA 783 (1977); Groves & Ledyard, Some Limitations of Demand Revealing
7. “A feasible allocation is one that is possible; in the pure exchange case, this is just an
allocation that uses up all the goods . . . .” H. VARIAN, supra note 5, at 137 (emphasis in
original).
that life forms a part. I shall suppose that possible courses of events can be ordered by their intrinsic value, or by how well or how justly they accommodate the interests of the people involved. This will be called the postulate of linear ordering. Finally, I shall suppose that anyone's interests may be outweighed; that for each person, there is at least one pair of conceivable courses of events $X$ and $Y$ such that although his own life is more rewarding in $Y$ than in $X$, $X$ is intrinsically better than $Y$ (that is to say, $X$ more justly accommodates his interests with those of everyone else). This will be called the postulate of overridability.

The postulates of linear ordering and overridability may be controversial. Economic theorists often write as though what mattered were not the intrinsic value of a course of events, but its Pareto efficiency, which, in the terminology I am using, may be defined as follows: A course of events $X$ is Pareto-superior to a course of events $Y$ if and only if at least one person leads a more intrinsically rewarding life in $X$ than in $Y$, and no one leads a less intrinsically rewarding life in $X$ than in $Y$. Furthermore, a course of events is Pareto-efficient if and only if it is technologically feasible and no technologically feasible course of events is Pareto-superior to it. Assuming the Pareto principle, if one course of events is Pareto-superior to another, then it is intrinsically better. It follows from the Pareto principle that an intrinsically best, technologically feasible course of events must be Pareto-efficient. It does not follow, however, that Pareto efficiency, if attainable; is worth having—that given any two courses of events, one Pareto-efficient and the other not, the Pareto-efficient one is better. Frequently, economists investigate Pareto efficiency mainly because they despair of coming to any normative conclusions stronger than the Pareto principle. Such moral agnosticism, however, cannot justify the conclusion that the satisfactory status of an institution depends on whether it is Pareto-efficient. If one foreswears all

---

8. To say that the relation "is better than" is a "linear ordering" is to say that its denial is reflexive (a course of events is no better than itself), connected (for any two courses of events $X$ and $Y$, $X$ is no better than $Y$ or $Y$ is no better than $X$, or both), and transitive (if $X$ is no better than $Y$ and $Y$ is no better than $Z$, then $X$ is no better than $Z$). A "linear ordering" here allows ties, but requires that any two alternatives be comparable: Either one is better than the other or they are equally good.

9. Elsewhere I have formulated conditions characterizing the dependence of the relation "is better than" on individual utilities as "ordering," "unanimity," and "no weak dictator." Gibbard, supra note 3, at 162-63.

10. For a review of fallacies involving the Pareto principle, see Sager, Pareto Superiority, Consent, and Justice, 8 HOFSTRA L. REV. 913, 914-30 (1980).
normative convictions except the Pareto principle, then one cannot deny the proposition that rough equality with mild inefficiency is better than gross inequality with perfect efficiency.

To do so would be to express a moral conviction that goes beyond the Pareto principle. The approach here is highly agnostic, in that the only substantive moral principles I am assuming are the Pareto principle and the postulate of overridability. The strong assumption here is one of structure rather than substance; that the relation "is better than" is a linear ordering.

Let us now examine motivational assumptions. Essential to a legal system is that it establish incentives that will keep people from hurting each other in pursuit of their own advantage. Equally important, it must foster beneficial cooperation, assuming a person's lack of concern for the community at large. The point is not that human beings are purely egoistic. People act from a mixture of motives that may include benevolence, moral principle, reciprocity, and spite, as well as self-interest. Nor, of course, are human beings purely rational, nor do all harms stem from one person's rationally pursuing his good in disregard of the good of others. Nevertheless, the rationally egoistic strand in human motivation is strong enough to justify studying the characteristics of a legal system that governs a hypothetical population of rational egoists. By a "rational egoist" I mean a person who rationally pursues his own good, and pursues other goals only inasmuch as their pursuit can promote his own good. Rational egoists can hurt each other, as the parable of the prisoners' dilemma\(^\text{11}\) aptly illustrates, and a central feature of legal systems is that they establish incentives to which rational egoists would be responsive. The question I shall be addressing is whether a legal order can cope with the problems rational egoists pose—whether, within the limits of technological feasibility, a legal order governing a population of rational egoists can ensure a best accommodation of their interests. If it can, then a legal system that fails to do so either must be defective, or must be confronted with problems other than those presented by a population of rational egoists.

Finally, what is being assumed here about the legal order itself? Imperfection is unavoidable if we assume at the outset that some limitations on the legal system are inevitable. For example, when a legal system is created its designers cannot foresee which alternatives will be technologically feasible, or what the system's participants will

\(^{11}\) See R. Luce & H. Raiffa, Games and Decisions 94-102 (1957).
consider intrinsically rewarding. At that point, nothing can be said about how various possible courses of events compare in the intrinsic reward they bring to a given person. The person himself, for all that is being assumed, may know much or little about what is good for him, but the legislator has no prior knowledge of what those subject to his legislation should want for themselves in life. Furthermore, governing officials can respond only to what people do. They cannot tell what is ultimately best for a person, except as he reveals it in the actions he takes. Officials may ask a person what he wants or what is best for him, but if it is to the person's prospective advantage to lie in response, as a rational egoist he will do so. Moreover, the effective commands of the law must always concern outer behavior; whether a person has obeyed the law cannot depend on his internal reaction. Officials may, for all that is being assumed here, have the power to issue commands that will be obeyed invariably, as if by magic, and the law may set up penal sanctions that secure obedience from rational egoists by putting a price on disobedience. The state cannot, however, read minds, or command in terms of the content of minds.

The designer of a legal order is thus confronted with a dilemma. The designer cannot simply order people to act for the best; he does not know what is ultimately better or worse for the various people involved. He can command only external behavior; he cannot effectively command people to give due regard to the good of their fellows. Even if he can get people externally to do whatever he commands, however, he may not know what to command; for what he wants people to do depends on internal values—on what is more and what is less to each person's good. His only hope is somehow to exploit his subjects' self-knowledge and the sources of their intrinsic reward. Nothing has been assumed here about how much the subjects know, but if they know enough about what is to their own good and how the system works, then what they do may well depend on what is ultimately good for them. What is good for a rational egoist affects both what he does and what, from an ethical point of view, constitutes the best accommodation of his interests with those of everyone else. The designer's problem is to find a system of incentives that will make these two effects mesh. He must harness individual interests (whatever they may turn out to be) to the general interest (whatever, in consequence, it turns out to be).

Two extreme strategies for doing this suggest themselves. One strategy is extreme centralization. The designer may have information about what is to the good of each person and establish a plan
whereby the government, on the basis of this information, shall command in detail those actions that would achieve the best accommodation of interests. Such an approach would, of course, raise insuperable problems of complexity; but even ignoring those, and supposing the information can be processed perfectly, the question remains whether it is information or misinformation; for rational egoists may lie about what they want and what they would find most rewarding in life, especially if lying is advantageous to them. The other strategy is decentralization. The designer may set up an incentive system, crafted so that whatever people find to their good they will respond to in a way that will bring about the best course of events. Such a system may be based on the basic incentive of free exchange within a system of property rights, from which might flow more specific incentives such as taxes, subsidies, legal obligations and prohibitions. Many mixtures of these strategies are conceivable, and the propositions in this article apply to all of them.

The assumptions here form a caricature of the problem an actual legislator faces. The caricature can be made mathematically precise, but to learn about the world from it one must exercise an inexact sagacity. It cannot truly describe the world; at most it identifies important tendencies in the world and illuminates their effects.

**THE ARROW IMPOSSIBILITY THEOREM**

One way to centralize decisions is by voting. Suppose individuals are perfectly informed of what life would be like for them in the various courses of events that might come about through alternative governmental policies. Can a system of voting be designed to ensure that so long as each citizen votes his interests the result will be the best that is technologically feasible? Is there a form of democracy that ideally accommodates the interests of the electorate, even though no individual accommodates his interests to those of others in deciding how to vote?

The Arrow impossibility theorem\(^\text{12}\) says that for a wide class of systems of voting, there is no guarantee that the winning alternative will be the best alternative.\(^\text{13}\) The systems are ones in which each person votes an ordering of the technologically feasible alternatives, from the one he most prefers to the one he least prefers. It is assumed that people vote their egoistic preferences sincerely. Addition-

\(^{12}\) Arrow, *supra* note 1, at 227-30.

\(^{13}\) See *id.* at 228.
ally, the assumptions of the Arrow theorem are the ones made here. It is assumed that the relation “better” is an ordering that depends only on how well off each person is under each alternative. The Pareto principle is assumed, as is a principle of non-dictatorship that is slightly weaker than the postulate of overridability that has been assumed here. Finally, there is a postulate of unrestricted domain, which captures the assumption that the legal order must be designed in ignorance both of what will be to anyone’s good and of which alternatives will be technologically feasible. What Arrow demonstrated was that these conditions are not jointly satisfiable. Voting, under any system that restricts information on a person’s ballot to his preference ordering of the technologically feasible alternatives, will not always choose the best technologically feasible alternative.

The Arrow theorem is open to other interpretations. As a mathematical result, it abstractly concerns the properties of a certain class of structures. Disparate subject matter may have abstract structure in common, in which case a theorem that applies to one will apply to the other, and say different things in the two applications. The Arrow theorem applies not only to voting, but, for instance, to systems of scoring in multi-event athletic contests. The interpretation given here, then, by no means exhausts the valid applications of the theorem, and it is not even clear whether I have picked the most illuminating application.

Nevertheless, it can be argued that any application of the Arrow theorem must suffer from an important limitation. The subject matter to which the theorem applies must, in structure, be like the voting of preference orderings. Thus, if the theorem is applied to a social interaction in which an outcome depends on the actions of a number of individuals, those actions must somehow be interpretable as expressions of orderings. With each preference ordering that a person might have, a unique plan of action is taken to express that ordering. Many actions, however, are not purely expressive of preferences, and even when an action does express a preference, it may be that the person could have expressed the same preference in another way, with a different result. Moreover, people may express their preferences inaccurately, and the Arrow theorem takes no direct ac-

14. See note 8 supra and accompanying text.
15. See note 9 supra and accompanying text.
16. See K. Arrow, supra note 1, at 97-100.
count of the possibility of strategic misrepresentation of preferences. Even when it has been determined what plan of action shall count as the expression of each preference ordering that a person might have, a person may discover that better results may be derived by expressing a different preference. On its evident interpretations, the Arrow theorem presupposes that voting is sincere.

GAME FORMS AND STRATEGY

The Arrow theorem is the classical result of axiomatic social choice theory, and it has provided a model for the whole subject. Much could be gained, however, by abstracting from the specific apparatus of the Arrow theorem, and instead using it as the inspiration for constructing a more general apparatus. An applicable device comes from game theory. Instead of regarding the individual actions one is studying as votes of preference orderings, one may speak abstractly of actions, which may or may not be conventional expressions of preferences. Indeed, for the sake of even more generality one may think of each person as having a contingency plan for basing his actions, over a period of time, on what he learns during that time, and in particular on what he learns from other people’s experiences. Such a contingency plan, in maximal detail, is called a “strategy” by game theorists. The outcome of an interaction among agents who have strategies in this sense will then depend on which strategies they have selected. We can let a game form (or outcome function) be the mathematical function giving the outcome of each combination of strategies that the respective agents might select. If we suppose that what happens depends not only on what people decide to do, but also on chance, then we may understand an outcome as a probability distribution over possible courses of events; in that case I shall speak of an indeterminate game form. Conversely, a determinate game form makes an outcome depend on individual strategies involving no element of chance. The widest variety of social situations may be represented by game forms, determinate or indeterminate. Game forms, then, are mathematical devices that allow us to pose general questions about interactions among rational agents.

How will a rational agent choose his strategy? Ordinarily, what

19. Gibbard, Straightforwardness of Game Forms, supra note 18, at 595.
people do depends greatly on what they expect others to do, either because agents wish to coordinate what they do with each other, or because they are trying to outwit each other. We might ask, though, whether the need for such guesswork could be eliminated, and whether a system could then be constructed in which each person, no matter what his preferences were, could choose his strategy without regard to what he expected others to do. Such a game form is called straightforward.20 One example of a straightforward game form would be a strategy-proof voting scheme, in which each person, no matter what his preferences were and no matter how he expected anyone else to vote, would find it to his advantage to state his genuine preferences on his ballot. Straightforwardness, however, is not to be had at a reasonable price, since only an extremely narrow class of game forms is straightforward. Of determinate game forms, the only straightforward ones are either dictatorial, in the sense that there is a person who can secure any outcome he chooses, irrespective of what anyone else does, or duple, in the sense that the outcome is restricted in advance to a fixed pair of alternatives (or a single alternative), irrespective of what anyone does.21 In the case of indeterminate game forms, the possibilities expand only slightly. Any determinate game form that yields Pareto-efficient outcomes must be a random dictatorship—a probability mixture of dictatorial game forms, which amounts to drawing lots for who shall be dictator.22 Straightforwardness, then, is clearly not worth its price.

Thus, gaming cannot be designed out of a legal order that governs rational egoists. Given any system in which agents interact, if the agents are rational egoists and the system ensures a satisfactory accommodation of their interests, there will be possible circumstances in which an agent will base what he does on what he expects others to do. The hope for designing a satisfactory legal order must lie in the exploitation of such gaming. How successfully can that be done? Can the legal order of a society be designed so that the result of self-interested gaming will be a satisfactory accommodation of interests?

To answer this question, one must employ the theory of such gaming. Game theory is an arcane and controversial subject, but two properties of virtually all theories of games can be used to answer the question I have posed. First, on all standard theories, if the

20. Id.
21. Id. at 596.
22. Id. at 596-97.
agents know which alternatives are feasible, the outcome will be independent of their preferences that involve infeasible alternatives. A rational agent's motive for action is to affect what happens, and his preferences regarding impossible events have no bearing on his sensible choices. Second, a rational egoist is taken to act to maximize his expected utility—the expected value of intrinsic reward in his life. In comparisons of expected value, only the relative magnitudes of gains and losses matter. Thus, on all standard theories, only the ratios of differences in intrinsic reward (not absolute magnitudes of intrinsic reward) affect the outcome. These two assumptions will be called the two postulates of game theory.

From the two postulates of game theory and the assumptions laid down earlier, the imperfectability of a legal order follows deductively. No legal order can ensure that the best technically feasible course of events will ensue regardless of each person's individual good.

Let me recapitulate the postulates from which this conclusion follows. First, the two postulates of game theory apply because the individuals in question are rational egoists and no one can read minds. Second, the postulate of unlimited domain—that individuals may have any ordering whatsoever of the alternatives, and any set of alternatives may turn out to be the set of technologically feasible alternatives—holds because, we assumed, the designer of a legal order can foresee neither what will be technologically feasible nor what degree of intrinsic reward a person will find in any possible course of events. Hence, the legal order must cover all possibilities on these matters. Third, it is assumed that the intrinsic value of a course of events, from an impartial point of view, depends on the degree of intrinsic reward individuals find in that course of events, and three postulates govern that dependence: the Pareto principle, the postulate of ordering, and the postulate of overridability. From these postulates, the imperfectability of a legal order follows. This theorem, because distinct from the Arrow theorem but very close to it both in its formulation and in its proof, will be called the modified Arrow theorem.

**DISCUSSION**

The modified Arrow theorem is not, of course, the culmination of a satisfactory normative theory of legal design, but preliminary to

such a theory. Insofar as the theorem's caricature of the problem of legal design illuminates the reality of the problem, the moral will be that legal design must consist of compromises. The need for compromise, the theorem suggests, stems not simply from scarcity of resources but from the limited power of laws, even if they are obeyed invariably, to evoke the best kinds of cooperation from egoists. No degree of legal and administrative refinement could eliminate compromise, even were everyone's ability to cope with complexity unlimited.

How good is the theorem's caricature of the human condition? Four of its elements seem dubious, and it may even be that reasonable relaxations make the impossibility result disappear. First, there is the postulate of unlimited domain. Its rationale was that the designer of a legal order must act in ignorance of what will be the good of those for whom he legislates. In fact, we approach the problem of legal design with a vast lore concerning human nature and human good. We know of the importance of health and length of life, of security, of liberty, and of opportunities. There are, to be sure, important individual differences in the exact shapes these things must take for life to be most rewarding, and in their relative importance for a rewarding life; but we do not proceed in complete ignorance.

A second dubious assumption of the caricature is individual inscrutability; that is, only by observing a person's choices can we identify those things from which he would gain intrinsic reward. We normally suppose that we can tell much about a person's psychic life by our observations of his involuntary emotional responses to life; and while the law is normally limited in the use it can make of such psychological judgments, private individuals frequently base their actions on others' psychological judgments. It is in part the assumption of individual inscrutability that lies behind the second of the two postulates of game theory. If people in their private dealings react to involuntary emotional expressions they observe in others, then the absolute intensities of inner feelings can affect the outcome of egoistic gaming. The modified Arrow theorem, as I have applied it, ignores the possibility that the law may benevolently exploit individuals' powers of psychological observation.

A third dubious assumption is that of rational egoism. The theorem does not determine the degree of success with which a legal order can depend on non-egoistic motivations to get things to work out for the best. The common wisdom is that it can do so considerably in some cases, and with little hope of success in others. The rationale
for the law concerning child custody, for example, relies heavily on a presumption of parental love; commercial law does not depend on a presumption of love among traders.

A final dubious assumption is that possible courses of events range from best to worst in a linear ordering. That may be too much to assume, even if the opposite extreme—complete refusal to go beyond the Pareto principle in one’s normative judgments—is unconscionable. Perhaps we should assume something between the morally agnostic partial ordering by Pareto superiority and a full linear ordering.

On all these points, further theoretical investigation is needed. Axiomatic social choice theory is far from a completed body of knowledge. For the most part, however, the impossibility results of axiomatic social choice theory have been remarkably resilient under reasonable relaxations of assumptions. If that is the case with the modified Arrow theorem, then a main task for the further development of the subject will be an exploration of the compromises that must be made in the design of a legal order.