Arithmetic vs. Morality: Liberalism in Collective Choice

Fred M. Frohock

Arithmetic vs. Morality:

Liberalism in Collective Choice

Fred M. Frohock

Fred M. Frohock has a B.A. and M.A. from the University of Florida and a Ph.D. from the University of North Carolina. He is the author of four books and has published articles in the *American Political Science Review,* the *Journal of Politics, Polity, Political Studies,* and *Social Theory and Practice.* He has twice been a Fellow of the Social Science Research Council. His current research is in bioethics and decision theory.

Anyone who has glanced at the reading material of small children is familiar with riddles like these: "What is white when it's dirty and black when it's clean?" "What gets wetter and wetter the more it dries?" "What is the surest way to keep fish from smelling?" Most children's riddles intrigue (adults as well as children), but usually only for a short time. Once the answer is provided, the puzzle is solved by exposing the anomaly or odd use of words or general trick on which the riddle is formed.

Riddles for some are veridical paradoxes for others. What is puzzling in a riddle may be an apparent paradox resolved by introducing a hidden or more general truth. For example, the true description of a man as being 21 years old yet having had only five birthdays is explained by the fact that he was born in a leap year on February 29th. A more pernicious puzzlement is found in falsidical paradoxes. Here the puzzlement is genuine, but the paradox is not. It is exposed by exhibiting the error on which it is based. For example, logicians now agree that Zeno's paradox of the tortoise and the hare, though a source of puzzlement (and pleasure to hairsplitters) for centuries, is based on the fallacy of supposing that an infinite succession of intervals must add up to an infinite interval.

The tougher, and thus more interesting, sources of puzzlement are the genuine paradoxes. The famous paradox of Epimenides—"All Cretans are liars," uttered by a Cretan (or, more generally, "I am lying now")—and Gödel's theorem are two examples of self-referential paradoxes. They seem to circle back on themselves, creating a contradiction by the maintenance of rules and arguments that, taken singly, are impeccable. The paradox known as Newcomb's problem, though not self-referential, is created by the "pull" of two decision rules. (See Appendix 1.) A genuine paradox seems to force us in two directions at once when we cannot go in both directions or even in one direction

so long as the rules and conditions of the paradox are maintained. In simplest terms, a genuine paradox is a self-contradiction from valid rules and acceptable premises.

Genuine paradoxes still can be managed. They cannot be "solved" like veridical paradoxes, or "exposed" like falsidical ones, but strategies of accommodation are available. Look again at the paradox of Epimenides. One familiar approach to the liar paradox is to avoid it by introducing a hierarchy of truth locutions. True and false can be indicated with numerical subscripts denoting the location of statements in a matrix of true-false types—Quine's solution.[1] We might say, for example, it is true$_1$ that I am telling the truth when I say "I am now not telling the truth$_0$." The paradox is avoided with the recognition of a type of truth function (subscript 1) that can address the truth of other statements (subscript 0)—an arrangement of language familiar to students of the sociology of knowledge, where claims within social practices are judged by truth criteria from outside those practices. The familiar dichotomy in the social sciences between "participant" and "observer" is basically the acceptance of terms for subscripts representing hierarchies of truth locutions.

Scientific anomalies present more complex strategies of resolution for puzzling events. A recent experiment in physics fires subatomic particles through a slit on a screen. The resulting distribution of particles is influenced by whether another slit on the screen, causally independent of the particles, is open or shut.[2] This seeming violation of causality raises questions (as such tests always do) about the retentive power of basic concepts and the validity of critical tests. Some propose replacing classical logic with some new logic accommodating quantum physics.[3] Others maintain that logic is necessary to criticize a theory and subject it to falsification tests.[4] In all cases of scientific anomaly, pressures are strong to explain the events empirically. Test conditions may be discredited. Auxiliary hypotheses are offered to save important theoretical principles (an offering some see as preventing falsification).[5] Failure to accommodate the anomaly inevitably creates additional pressures to change some parts of the theory while maintaining basic principles. Perhaps the basic principles must give way to a new paradigm.[6]

Game theory and collective choice, though no threats to understandings of physical reality, introduce anomalies to recent comprehensions of a political society. Both of these fields are inventions (largely) of the twentieth century. Game theory was devised by a mathematician, John von Neumann. Collective choice theory, though found in inchoate form in the writing of Condorcet in the eighteenth century and Charles Dodgson in the nineteenth, begins its contemporary incarnation with the work of Duncan Black and Kenneth Arrow.[7] Both fields teach us that what counts as a rational choice for any individual is critically affected by how others choose. But it is yet another contribution that I will be concerned with here: the demonstration in both fields that a collection of individual choices can be irrational even when every individual in the collection is choosing rationally. A rational inconsistency between individual and collective is obviously no dilemma comparable to those found today in physics. It is not certain that such rational breakdowns are even genuine paradoxes. Nor is a conflict between individual and collective a dilemma for types of holism.[8] But rational

1. W. V. O. Quine, *The Ways of Paradox* (Cambridge, Mass.: Harvard University Press, 1976), where the distinction between "veridical" and "falsidical" paradoxes is also found. One can also change the reference to avoid self-referential paradoxes, as when a non-Cretan says "All Cretans are liars." But two different people uttering the same sentence may be expressing different propositions. For a discussion of semantic paradoxes, see James Cargile, *Paradoxes: A Study in Form and Predication* (Cambridge: Cambridge University Press, 1979).

2. The speculative discussion of this experiment has even challenged, on the basis of quantum physics, the doctrine of a real world independent of human consciousness. For a nontechnical overview, see Bernard d'Espagnat, "The Quantum Theory and Reality," *Scientific American* 241 (November 1979): 158-81.

3. Hilary Putnam, "Is Logic Empirical?" in Proceedings of the Boston Colloquium for Philosophy of Science, *Boston Studies in the Philosophy of Science*, vol. 5, ed. R. Cohen and M. Wartofsky (Dordrecht: Reidel, 1969).

4. Karl Popper, *Objective Knowledge* (Oxford: Clarendon Press, 1972).

5. I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes," in *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. Musgrave (Cambridge: Cambridge University Press, 1970).

6. Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

7. Black, *The Theory of Committees and Elections* (Cambridge: Cambridge University Press, 1958). Arrow, *Social Choice and Individual Values* (New York: Wiley, 1963).

8. Individuals in corporate wholes who oppose collective outcomes are simply irrational—a pattern of judgment found in the idealistic political philosophies of Plato and Hegel. Collectivist theories in general do not suppose that societies are rational in the same way that individuals are rational.

discontinuity is an anomaly for methodological individualism, the philosophy guiding game theory and collective choice. If we believe (as most Western social scientists believe) that society is no more and no less than a collection of individuals, it is a puzzle of some importance that society can fail tests of rationality passed by every individual constituting the society.

An enormous and expanding body of work has explored the resolutions and implications of these rational conflicts between one and all.[9] But it is still not clear exactly what concepts can be maintained and what must be jettisoned in avoiding these conflicts. My objective here is reasonably modest. I will examine two types of rational breakdown between individual and society—the problems demonstrated in Kenneth Arrow's impossibility theorem and those found in exchanges of goods. My goal is to establish that collective choice theory relies on a number of background concepts that are not recognized in the formal delineations of the theory. These concepts combine with the explicit conditions and axioms of collective choice to produce competing senses of equality and an ambivalent status for the term "individual"—and *these* problems of equality and individualism explain the rational breakdowns in collective choice theory. Indeed, two concepts of community can be found in collective choice. These two communities originate in conflicts between background moral concepts and the more straightforward arithmetical languages of collective choice. In disclosing these two communities, moreover, one gains an understanding both of the liberal assumptions of collective choice theory and of some conflicts within liberalism itself.

[9]. A collection of readings and text on these representations is *Rational Man and Irrational Society?* ed. Brian Barry and Russell Hardin (Beverly Hills, Calif.: Sage Publications, 1982). See also the overview by Dennis Mueller, *Public Choice* (Cambridge: Cambridge University Press, 1979).

Liberal communities

The high standing of methodological individualism in collective choice is celebrated throughout Arrow's theorem. The starting conditions of the original proof require discrete (nonoverlapping) individuals. The collective, in turn, is a combination of separate orderings. Indeed, Arrow extends a tradition of thought which assumes that the definitive question in collective choice is how to aggregate the values of separate individuals to reach a collective outcome. The theorem develops a rational breakdown between individual and whole that questions this tradition and the influence of methodological individualism on collective choice.

The intriguing hold that Arrow's theorem has on us begins with logic. We first ponder orderings like these: voter 1 ranks three alternatives $a > b > c$; voter 2 ranks the same alternatives $b > c > a$; and a third voter orders the alternatives $c > a > b$. If these three rankings are combined, an intransitive ordering is the result: $a > b, b > c$, and $c > a$. The trap is thus set by the famous cyclical majority. We are drawn further into it by Arrow's general theorem. The basic theorem proves that when there are more than three individuals and alternatives, there does not exist any social choice function (or aggregator) that satisfies Pareto, nondictatorship, universal domain, complete and transitive rationality, and the independence of irrelevant alternatives. (See Appendix 2.)

The proofs of the theorem demonstrate how these conditions are incompatible. The most common proof is to suppose that there is a set of individuals decisive over two alternatives, $x > y$. By universal domain,

then, both $x > y > z$ and $y > z > x$ must be admissible orderings. Now $x > y$ on decisiveness, $y > z$ on Pareto, and, on transitivity, $x > z$ follows. The members of the decisive set prefer $x > z$, and so the decisiveness over $x > y$ has spread to $x > z$ by virtue of Pareto and transitivity. This contagion effect of decisiveness can be extended to all pairs of alternatives, demonstrating that local decisiveness (over any pair of alternatives) becomes global decisiveness (over all pairs of alternatives) when Arrow's conditions are maintained. Further, there must always be a decisive set if there is anything short of unanimity in society and society selects (on whatever rule) some ordering over others. For example, if set 1 prefers $x > y > z$, set 2 prefers $y > z > x$, and set 3 prefers $z > x > y$ (the cyclical-majority orderings), majority rule selects $y > z$. Sets 1 and 2 then constitute a set decisive over set 3 and the contagion effect reoccurs. The upshot of the proof is that a society cannot make any collective decisions when global unanimity is absent without either dismissing Arrow's conditions or violating them.

The formal nature of the theorem seems at the outset cause for celebration (assuming that the nonspecialist maintains sanity). One of the attractive features of formal theory is that conditions and axioms are stated explicitly and relationships among basic concepts are demonstrated. The explicit and demonstrable nature of such theorizing provides a clarity and generality of thought that can extend from one level of theory to another, and even among substantive areas sharing the same abstract calculus. But formal theory succeeds only if all of the terms needed to understand a problem are disclosed; and it is by no means certain that the most important conditions, axioms, and relationships are treated in Arrow's proof. If one inspects a deeper layer of assumptions, a different and more general set of components can be uncovered and used to demonstrate conflicts of a different order. These conflicts, moreover, may require languages that do not meet the requirements of formal systems.

One background set of assumptions in Arrow's theorem, for example, defines a liberal model of a political society: individuals are (*a*) moral equals who are (*b*) separate from one another and (*c*) free to pursue their own goals without institutional impediments or interference by others. Yet only the second and third of these three features of liberalism are explicit in the theorem. The first—moral equality—must be inferred from the formal conditions of the proof. If one began and ended an understanding of Arrow's theorem with the surface conditions and axioms, moral equality would never be encountered. Yet the concept of moral equality helps us understand both the logic and the meaning of the theorem as an exercise in social theory.

One indication that there is a deeper layer of concepts in Arrow's theorem is that the demonstration of the problem is not completely describable in logical or rational terms. Three of Arrow's original five conditions are influenced by moral or equity concepts: Pareto, nondictatorship, and universal domain. Each of these three conditions is an effort to fulfill autonomy: Pareto in ensuring that unanimity will be represented at the collective level, nondictatorship in ruling out the dominance of one over all, and universal domain in prohibiting the manipulation of alternatives. If moral concerns are dismissed, no rational problem occurs in the theorem; for transitivity and the independence condition can be satisfied with a violation of any one of

the three equity conditions.

The moral concerns of Arrow's theorem define a common grid on which are found a number of theories that appear disparate and even contradictory on the surface. Both John Rawls's theory of justice and Robert Nozick's libertarian state share the same liberal model underlying Arrow's theorem. Rawls's principles of justice are drawn from the hypothetical choices of individuals in liberal conditions (equal regard, liberty, rationality). Nozick's account of justice begins with a liberal vision of free and (initially) equal individuals each invested with rights that protect autonomy.[10] What is especially intriguing about these three theories is that each is an attempt to reconcile the moral concepts and radical individualism of liberal theory. The two theories that reach collective outcomes by adding individual values fail consistency tests (Arrow's and Nozick's). The one that transforms all into everyone (a holistic noun) remains internally consistent (Rawls's). The more interesting observation, however, is that a general understanding of collective choice is gained by exploring the liberal model as it strains to accommodate arithmetic. Since collective choice produces conflicts between individual and social rationality only on the premise that individuals are moral equals free to set their own goals, an inspection of the liberal model should identify in a more general way what concepts must be modified to avoid contradictions in collective choice.

10. John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971). Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).

Equality in collective choice

The surface, or explicit, axioms and conditions of Arrow's theorem tolerate many forms of inequality. Entries to the aggregation machine (any device for aggregating preferences) can be counted more than once, so that a social state in which one individual has, say, one hundred votes and another only one vote is not ruled out by Arrow's theorem. Also, the individual actors, though required by the conditions to be discrete, do not have to be singletons. They can be sets, collectives, blocs, lumps, whatever. Both United Technologies and an individual citizen of an upstate New York village can be individual actors in the theorem.

The theorem is also silent on any number of other equality measures and criteria. The ratio of participants (those individuals introducing preferences to the aggregation machine) to nonparticipants is not an issue in Arrow's theorem. So, like Aristotle's views on citizenship, exclusionary rules may keep most individuals from participating; and whatever equality exists among individuals in an Arrow society (and in an Aristotelian one) may apply only to a very small subset of individuals. The theorem also says nothing about equality within sets or blocs, so that even if the actors in Arrow's theorem are in some way equal, the members of such units may be unequal to each other and to members of other units. The absence of overlap among actors reinforces whatever inequalities may exist within units, for the possibility of multiple memberships vitiates the more extreme effects of inequality within collectives (allowing individuals to be unequal in one setting while equal in another—as church vicars may find themselves low on the club tennis ladder). Finally, all but one of the theorem's conditions permit inequalities in the distribution of goods (nondictatorship, see below, is the exception). Pareto, for example, is a concept used by

Arrow mainly as a device to guarantee that unanimity will be honored. But the concept itself does not require any distributive equality.[11] If equality is to be found in Arrow's theorem, it must be located in the context of an assemblage of concepts that, on the whole, tolerate a variety of inequalities.

The presence of a concept of equality is suggested by the strong evaluative language in which many of the rational problems of collective choice are typically described. The "free rider," for example, is an individual who benefits from the collective production of public goods without contributing to the collective effort. The phrase "free rider" suggests the stigma that helps form the rational problem. A free rider is a cheat, someone who gets something he doesn't earn. In broader terms, an individual who fails to contribute to a cooperative enterprise in which he is a member does not meet minimal tests of fairness. The free rider is someone who ought to contribute but does not and as a noncontributing member, the free rider is a moral as well as a rational failure. The individual who deserves the public good without contribution—the very ill or the very young, for example—is not a free rider. The rational problem of suboptimal provisions of public goods would look entirely different if noncontributors were justly excused from group participation. In Prisoners' Dilemma, a famous game in which players who choose rationally find that the combination of their choices is subrational, the cell in which one individual secures optimal returns at the expense of the other player is routinely labeled the "exploitive" cell or the "sucker" outcome. Arrow employs evaluative language in an even stronger and more explicit sense: the decisive set consisting of a singleton is a "dictator," one whose orderings are the orderings of all.

Much of the evaluative language in collective choice is of course window dressing, except for Arrow's explicit use of the nondictatorship condition. But the language is still formed by expectations that collective outcomes must fulfill Aristotle's definition of numerical equality as equal shares to and from all relevant individuals; for none of the rational problems is represented by a canon of claims that might rank individual claims on, and obligations to, collective action in some distributive pattern. Instead, unequal outcomes, by virtue of being unequal, are viewed as failures of collective action. The free rider, the nonvoter, the exploitive cell, and Arrow's dictator are regarded as pathologies of collective choice. Now it is an axiom of equality that inequalities in the social unit may be needed to ensure equalities among individuals. If, for example, patient 1 needs 3 units of an antibiotic for a restoration of health and patient 2 needs 5 units to achieve the same result, then, in the table below,

	1	2
a	3	3
b	3	5

social state *b* is a more authentic expression of equal treatment of patients 1 and 2 than is social state *a*.[12] Arrow's theorem, in contrast, looks only to collective outcomes, not to the differences that may obtain among individuals. A theory that views inequalities of outcomes as ipso facto unsatisfactory must assume that the individuals produc-

11. Imagine a graph. Choose any two numbers, say 2,1, to mark a point on the graph. The area Pareto superior to this point is any point in the upper right-hand quadrant of the graph greater than 2,1. The numbers in this space can be radically unequal, e.g., 900,3.

12. Or, individuals sometimes have to be treated differently in order to be treated equally. Douglas Rae, *Equalities* (Cambridge, Mass.: Harvard University Press, 1981). See also the discussion by Felix Oppenheim, "Egalitarian Rules of Distribution," *Ethics* 90 (January 1980): 164-79.

ing and consuming the outcomes are equal. Or, above, if *a* is preferable to *b*, 1 and 2 must be equal in their claims (effect, need, desert, etc.) on the shares distributed.

The assumption of equal individuals, however, proceeds with no information about individuals, except that they are countable units. No interpersonal value comparisons are conducted (except at a late point in the development of Arrow's theorem, and then with "extended sympathy"). No theory of justice is developed to evaluate individual claims. But the acceptance of equal distributions requires the assumption that individuals have equal claims on joint outcomes.

The logic of the nondictatorship condition in Arrow's theorem supports equality of claims and requires equality of effect. The condition rules out a decisive set, which means only that no individual's ordering can be the ordering for all others in the society. The question is, why not? An individual endorsing a more just social distribution can dictate legitimately to others without being a dictator. A judge (in a non-jury trial) can dictate the outcome of court proceedings. Decisiveness as such can be denied only on the assumption that no individual is an authority over all others. Again, equality of claims must be a background concept.

But the point can be taken further. Arrow's theorem proves that local decisiveness is contagious. If a set *S* of individuals is locally decisive for *x* against *y*, then with Arrow's conditions, that set is also globally decisive for *x* against *y*. Or if any individual, *i*, dictates on some pair of alternatives, that individual *i* is decisive for any pair of alternatives.[13] Thus any inequality of effect spreads to dictatorship. Arrow's individuals must all be equal to one another; for if any one is decisive over any other on a single pair of alternatives, that decisiveness extends logically to global decisiveness (or dictatorship as decisiveness over all pairs of alternatives). Nondictatorship is a condition assigned to collective outcomes, and the condition depends on a moral equality (equal claims, equal effects) among individuals.

The two equality assumptions—equal effects and equal claims—are contained in the thought that collective choice theories assume that justice is settled prior to the stage of decision making.[14] Usually the prior settlement involves equity expectations that preferences will be revealed successfully in the rules chosen for decision making. Arrow's theorem represents a breakdown in these expectations. But, also, the two equality assumptions reveal substantive expectations that explain in large measure why the breakdowns occur—because equality assumptions rule out dominance patterns. Note that nothing in the theorem prohibits a distribution of goods or resources in the collective outcome. It is simply that arithmetical composition rules provide no criteria for arriving at distributions. The background moral equality of the theorem can tolerate and even justify distributional inequality if used within some theory of justice. But Arrow's theorem, containing no criteria for fair or just distributions, can only move between absolute numerical equality and absolute inequality (or dictatorship) with no capacity for occupying any intermediate position between these two extremes. And it is precisely this absence of any device to rank claims that compels the theorem to regard individuals as absolutely equal.

13. I will sternly resist walking readers through the exercises proving this point (on the advice of the editors) and instead simply say that the point can be found in Arrow's *Social Choice* and numerous secondary sources. I have found especially helpful on this (and other parts of Arrow's proof) Jerry Kelly's *Arrow Impossibility Theorems* (New York: Academic Press, 1978).

14. See Dennis C. Mueller, *Public Choice* (Cambridge: Cambridge University Press, 1979), chap. 14, for an overview discussion.

Arrow and Rawls

In its reliance on equality as a background concept, Arrow's theorem is similar to Rawls's theory of justice. The theorem and the theory are of course unlike one another in several important ways. Arrow's starting conditions of choice contain discrete and countable individuals who have ordinary knowledge about themselves (their abilities, needs, interests) and their probable locations in a collective outcome. Rawls's individuals, however, choose governing principles in an original position (OP) where they are denied knowledge of their assets and liabilities, their locations in the social practice formed by the governing principles, and a theory of the good. One effect of this veil of ignorance is to suspend that information which ordinarily allows individuals to demarcate themselves from others. Rawls's OP individuals are not discrete and countable, for each is identical to every other. (One is equivalent to everyone.) A second effect of the veil is to set aside aggregation. Arrow's individuals express preferences that are combined by means of arithmetical composition rules. Individuals in the OP have preferences (for distributive principles) that are logically unanimous. The two principles of justice in Rawls's theory—liberty and equal opportunity conjoined with the difference principle—are composed or logically derived from the conditions of the OP. They are not produced from aggregation.[15]

Nor is it clear that the rules governing individual preferences and collective outcomes are congenial in each case. Arrow's five conditions and three axioms (completeness, transitivity, and rationality) fit Rawls's OP only in part. One condition—universal domain—is comfortably joined to Rawls's theory. Individuals in the OP can survey all logically possible distributive principles (including those of utilitarianism) without any restrictions (except those built into the features of rational choice in the OP—which point can be directed against any conditions in rational choice). The domain of social choice may then be seen as consisting of every logically possible combination of individual orderings of the alternatives surveyed (thus satisfying universal domain). Two of the other three conditions, however, do not bear on the OP. Pareto and nondictatorship are useless when applied to the conditions of the OP, for each requires more than one discrete individual for its primary effect. Where, as in the OP, individuals are not rationally distinguishable and, as a consequence, unanimity is logically assured, Pareto and nondictatorship are worthless standards.[16] The independence of irrelevant alternatives, however, is important. It is always worthwhile to ensure that collective choices will not vary on static preferences, even when outcomes are derived rather than aggregated.[17] One axiom of individual choice—binary comparisons—is not used in the OP. (Rawls allows global comparisons of principles.) The other—transitivity—is not mentioned in Rawls's account but can be reasonably expected to apply to the means-ends deliberations he endorses.

But, in spite of these disjunctures and only mild overlaps, a common set of assumptions is shared by Arrow's theorem and Rawls's theory of justice. These assumptions are disclosed in an inspection of the deeper model of a political society in Rawls's method of theorizing. Recall that intuitionism—the establishment of a rank

15. Rawls, *A Theory of Justice*. The principles are produced from a bargaining game in the first model of justice, in Rawls's "Justice as Fairness," in *Philosophy, Politics, and Society*, 2d ser., ed. Peter Lasslett and W. G. Runciman (New York: Barnes & Noble, 1962). But the later amendments to this first model bring out Kantian features of the OP which make bargaining inappropriate and indeed impossible. See Eric Von Magnus, "On Modeling the Original Position," *Reason Papers* 6 (Spring 1980): 25-35.

16. The formal requirements of Pareto and nondictatorship, however, are satisfied with a single individual. Pareto is met when x, $y \in X$ (the set of all alternatives), and the set of all N is decisive for $x > y$. Nondictatorship states that no individual is decisive for x against y for all $x, y \in D$ (D = profile). Now a set, S, is decisive for x against y ($x, y \in X$) if, for every profile D in which (1) $x \geq_i y$ for all $i \in s$, (2) $x >_i y$ for at least one $i \in S$, we have (3) $x \in v \Rightarrow y \notin Cu (v)$. So both Pareto and nondictatorship formally apply to a singleton set. But since Pareto seeks to ensure that unanimity is reflected in collective outcomes, and nondictatorship rules out a single individual dominating all others in the collective outcome, the use of each condition in Rawls's OP would be otiose.

17. The independence condition requires that two distinct profiles whose restriction to an agenda are the same must also have choice functions that act the same, at least on that agenda. Or, put less formally, collective outcomes are to remain the same if individual orderings do not vary. Any theory of collective choice that derives principles from individual choices would be concerned to ensure such noncreativity of composition rules.

ordering of basic principles without benefit of publicly accessible criteria for ranking—is abandoned by Rawls in favor of the social contract. Intuitionism (as in Plato's *Republic*) permits authoritative accounts of political arrangements insulated from challenge by those who have not had the critical intuitive experience or who do not have access to the ranking criteria. Contract theory, on the other hand, presumes that individuals are equal in the formation and ranking of political principles. The deep assumption in Rawls's theory, formed from his reliance on contract theory, is that all individuals, regardless of their status, have a right to be given an equal regard in the establishment of social practices.[18]

The assumption of rights to an equal regard is itself part of a larger set of assumptions. Rights depend on critical separations among individuals and between individuals and the political society. A right, as traditionally understood, is not a constraint on the individual to whom it is assigned; it is a constraint on others not to impede the actions of the one who has the right. Thus a right to vote is a constraint on registrars (and the like) that forbids interference with an individual's effort to vote, but that does not require of the individual with the right to vote that she do anything (even vote). Similarly, rights against the state restrict the state from interfering in the areas protected by rights. It follows that rights presume that individuals are separate and capable of adversary relationships with each other and with the political society. It also follows that freedom is assigned to individuals, for in the absence of freedom there are no rights at all.[19]

When this larger set of assumptions is described, we see more clearly how the OP functions in Rawls's theory. It is an intermediate device that represents and transforms a model of discrete individuals into a hypothetical community of identical rational agents. This hypothetical community is a moral society that fulfills tests of fairness. These tests (primarily impartiality) allow us to view the two derived principles as principles of justice. That the derivation fails to produce a substantive outcome from a formal procedure has been adequately documented.[20] The list of primary goods strongly biases the theory toward liberalism; the ''general facts about society'' condition sets historical limits on the generality of the theory. But this failure only reaffirms Hume's dictum that nothing can be found in the conclusions of a deduction that is not present in the premises. Once the deeper assumptions are produced, the theory is properly seen as a reexpression of a liberal model filtered through the mechanism of an OP. Nowhere is this implicit liberalism clearer than in the derived principles of justice. Liberty is the first principle chosen, and it is shielded through a lexical ordering from economic practices (set by the difference principle). This version of justice is as strong a reexpression of the liberal ideal of a political society as one is likely to find.

Arrow's theorem is also developed on a deep assumption of equality, though the equality is more deeply embedded. The possibility of contagion in decisiveness sets forth the strictest type of equality of effect; and both the general acceptance of equal distributions and the denial of authoritative claims by means of non-dictatorship require equality of claims. Liberal ideals of autonomy are expressed by these conditions. Indeed, all four of Arrow's explicit

18. Ronald Dworkin, *Taking Rights Seriously* (Cambridge: Cambridge University Press, 1978).

19. H. L. A. Hart, ''Are There Any Natural Rights?'' *Philosophical Review* 64 (1955): 175-91.

20. Brian Barry, *The Liberal Theory of Justice* (Oxford: Clarendon Press, 1973); Robert Paul Wolff, *Understanding Rawls* (Princeton: Princeton University Press, 1977).

conditions—universal domain, Pareto, nondictatorship, and the independence of irrelevant alternatives—are features of a liberal society. They suggest a well-known account of autonomous individuals originating social practices without constraints from any external source (natural law, institutions, procedures, or other individuals).

A liberal society is developed on two distinct moral perspectives.[21] One is that the state must be neutral on the values that individuals ascribe to their lives. Another is the view that all members of the political society are to be given an equal regard, without reference to their circumstances. Though these two perspectives can lead to quite different and frequently contrary conclusions on the proper role of the state in regulating the lives of its citizens, both are congenially represented in Arrow's equity conditions. The uses of universal domain and Pareto express the first form of liberalism, neutrality. Universal domain in particular ensures state neutrality in the availability and ordering of alternatives. Pareto grants legitimacy to those alternatives, and only those alternatives, that have unanimous support from individuals (not support from the state). Both conditions depend on a standard justification for state neutrality: a noncognitive theory of value. Any theory of value that ranks moral principles or statements on truth criteria would immediately (*a*) dismiss universal domain by restricting the range of acceptable alternatives for individuals to order, and (*b*) address Pareto from the perspective of a critical morality that would not necessarily accept unanimity as the satisfaction of moral demands.[22] The current labeling of preferences as "tastes"[23] is thus no mere convenience but an expression of the belief that values make no truth claims. In the absence of a noncognitive theory of value Arrow's theorem could not be developed. But the other moral perspective on liberalism is also represented in the theorem. Nondictatorship sets up a procedural condition that expresses, through a logical guarantee, the thought that each individual is to be given an equal regard in, and even have an equal effect on, the collective outcomes of society. Taken together, these two moral perspectives amount to the more complex liberal view that individuals are self-legislating creatures who are morally equal to one another, and that the political society in some way originates in the expressed preferences of these individuals.

The complex liberal view begins to break down in Arrow's theorem with the use of arithmetical methods to reach collective outcomes. The methods are justified by the individualism of liberal theory. "Counting heads" is a way of ensuring that social practices originate with the descriptions of individuals. This individualism is elaborated in the philosophy of methodological individualism—wholes are arithmetical compositions of, and reducible to, their parts. Two features of Arrow's theorem represent this philosophy. The first is the separate and countable status of individuals. Atomism is not too strong a metaphor. Individuals are not, as in Aristotle's polis, conceptually embedded in the political society. Nor is society temporally or conceptually prior to the individual. The opposite is assumed: individuals are the independent variables from which social states are derived. The second feature, complementing the first, is that collective outcomes are no more and no less than the arithmetical sum of individual orderings (a requirement finding one expression in the independence condition). Nothing is to emerge in the collective outcome that is not present in the constituent parts.

21. Ronald Dworkin, "Liberalism," in *Public and Private Morality*, ed. Stuart Hampshire (Cambridge: Cambridge University Press, 1978).

22. I. M. D. Little, "Social Choice and Individual Values," *Journal of Political Economy* 60 (October 1952): 422–32.

23. William H. Riker, *Liberalism against Populism* (San Francisco: W. H. Freeman, 1982).

Rawls's theory, however, is closer to meeting the condition of nonemergence than is Arrow's theorem. Rawls's early work attempted to produce more than the premises allowed—substantive principles from formal procedures. His later, more developed theory of justice introduced substantive conditions to the procedures of choice (e.g., primary goods, the general facts about society) that are necessary if the principles of justice are to be derived, but that nonetheless compromise the effort to maintain the procedures as purely formal devices. The full theory was presented as a general theory of justice. But liberal principles appear in the premises and make their way through the filter of the OP to emerge as (justified) principles of justice. Everything that emerges in Rawls's theory is present in the premises, and the assumptions become principles that are consistent with procedure once the liberal status of the theory is acknowledged. Arrow's assumptions, in contrast, lead to emergents that are self-contradictory: the collective state does not successfully represent individual values. The different methods employed in each approach to collective choice explain the different outcomes. Rawls mediates the arithmetical language of liberalism with a Kantian representation of equal regard. The OP reconciles the atomism of liberalism with its moral needs, ensuring equal regard not through the requirements of methodological individualism, but by means of that collective state formed by the veil of ignorance. Arrow's theorem, in contrast, tries to produce a collective state by arithmetical means— aggregation of separate preferences. The effort fails to meet its own conditions for success.

The contradictions of Arrow's theorem, however, do not occur within the set of arithmetical assumptions. The production of, say, a nontransitive ordering from a collection of transitive orderings is logically intriguing. But such events are natural features of arithmetic and logic and often avoidable with extension or manipulation of the formal systems. In Arrow's theorem, for example, a decisive set satisfies transitivity at the collective level. Rational problems occur only if the theorem is viewed as a representation of social conflict and allocation. The representation must stand for outcomes that are (in Rawls's phrase) ''burdens and benefits'' to members of society.

Such representations are indeed found in collective choice. Look again at the free rider problem. Imagine a group of individuals who cooperate to produce some collective good—musicians, say, who voluntarily provide a concert every Sunday for the neighborhood. If some individuals take up a collection to help the musicians with expenses, contributions would be nice but could not be obligatory even for those who live close enough to the site of the concerts to hear the music without any effort. The concert is produced without benefit of a collective agreement, explicit or tacit. If an individual has not consented to a joint effort, and indeed does not care if the good is provided or not, the evaluative force of the phrase ''free rider'' and the logic of the free rider argument are meaningless. A fairness principle is needed for the free rider argument, and a cooperative community is needed for the principle. If all have consented to a jointly beneficial project, each person has a reason to cooperate. In the absence of a consensual community and the sense of fairness drawn from it, the formal proof of the free rider dilemma cannot be developed.[24]

24. See Robert Nozick's examples and discussion in *Anarchy, State, and Utopia*, pp. 93-94, where he develops the modest point that the benefits of collective action to an individual must be more than his own calculated costs in contributing (in order for fairness to apply), pp. 267-68 for a recognition that some may legitimately refuse to contribute even if all others give to collective action (''they don't care about the ride at all'').

Arrow's conditions require a similar sense of community. Nondictatorship suggests that the community must be a moral community. Unless individuals are taken to be moral agents, the dominance of one over all is not describable as dictatorship. Imagine a faulty roulette wheel that always produces the same number, ensuring the dominance of that number over all the others on the wheel. If the wheel is demonstrated as a curiosity, we would not say that such numerical decisiveness is dictatorship. Dictatorship is the way we would describe the use of the wheel by a flawed dealer to dominate other (human) gamblers by taking their money. Or imagine a decisive star, one whose luminosity is so bright as to render all rivals practically invisible. Or think of a decisive solution to a mathematical problem, or a decisive experiment in science. None of these are dictatorial events. Nondictatorship is a condition that rules out the dominance of one agent over all other agents, a dominance that is without reason or agreement from those dominated.

The two other equity conditions support this moral sense of community. Universal domain ensures that all possible combinations of individual orderings can be considered. One point to this condition is to avoid manipulation of preferences. The attempt fails, of course. Even with universal domain the theorem, and its many successors, demonstrates that manipulation is possible; for outcomes are not independent of the paths to them. But the attempt is senseless if moral agents are not the victims of manipulation. Pareto follows the same logic. The nonperverse expression of unanimity in outcomes is empty as a purely mathematical condition. It does not bear on the social representation of the theorem. Pareto, as a moral concept, guarantees that the unanimous preferences of reasoning agents will be fulfilled at the collective level. Moral agency is in general needed to establish the equity conditions in the absence of which the rational problem of Arrow's theorem does not occur.

A community of moral agents, whatever else it is, consists of individuals who self-legislate, ordering alternatives on reasons. This is a moral condition of liberalism. Such a community is not consistent with an arithmetical community. Prescription is a feature of all reasons. A reason to do *a* rather than some rival alternative is a rational appeal for all to do *a*. A reason for an action prescribes for a class of relevantly similar agents, never just for a particular person or situation. Moral agents are thus never entirely distinct units but always have normative effects on each other by means of the reasons employed for orderings. A moral accord is based on reasoned argument, not aggregation. And reasoned deliberations permit emergent values. For example, a juridical proceeding, one device to accommodate reasoned orderings, is normally seen as defective unless (*a*) individuals are viewed not as discrete units but as members of classes, (*b*) outcomes are produced from rational deliberation rather than arithmetical combinations, and (*c*) decisions can establish new precedents from conventional rules and evidence. The second sense of a community, also present in Arrow's theorem, is holistic rather than numerical, in the sense that individuals are constituent members of social practices established on shared values and that collective outcomes can routinely produce emergent values.

The problem is that the theorem contains no device to express this

second community. Rawls's theory of justice uses the OP to represent a moral community. Rational agents are transformed into moral agents, choosing for everyone, by the veil of ignorance. But unlike Rawls's theory, Arrow's theorem has no filter (like the OP) to mediate between the disjointedness of the liberal model and collective outcomes. Collective outcomes are instead produced directly by aggregating the separate preferences of discrete individuals. This uninterrupted transformation fails to be completed. One or more of the features of liberalism conflict with each other. The equality engraved in Arrow's theorem by logical contagion never leads to a warranted inequality. All dominance is therefore unjustified. Rawls's theory, in contrast, justifies distributions that (in complex ways) favor the worst-off representative person. The OP uses equality as the basis for justifying inequalities. The force and elegance of this hypothetical condition can be appreciated anew. The absence of such a mediating device in Arrow's theorem aggravates the natural tensions in liberal societies between equality and inequality or, more broadly, between arithmetical and moral needs. Aggregation and the moral conditions of liberalism are articulated throughout Arrow's theorem, with no instrument to render them consistent with each other. There is no reason to think that anything resembling an OP would resolve Arrow's problem. But the impossibility result is a failure of consistency between aggregation and morality that Rawls's theory, whatever its liabilities, avoids.

Exchange theory

The liberal community of rational and autonomous moral agents is prominently displayed in exchange theory. Arrow's theorem demonstrates that separate and countable individuals cannot be joined arithmetically to produce collective outcomes meeting simultaneous tests of rationality and equity. The same type of problem occurs in markets, where again efforts to map a collective outcome from discrete individuals fail to fulfill the moral and rational expectations of liberalism.

Let a market be defined as a collection of exchanges.[25] A thought experiment can identify the range of problems both addressed and raised by exchange theory. Think of two rational individuals each with a supply of goods which they freely exchange (for whatever reason, though presumably each benefits). The first thing to notice in the exchange is that transitivity seems to be maintained through the fulfillment of Pareto. Since, in a free exchange, everyone is better off than before the exchange (or some are better off and no one is worse off), a handy altimeter is provided. Each successive social state, if brought about by exchange, must be "higher" (better) than its antecedent. The cycle of cyclical majorities, or the general failure of transitive orderings, is thus avoided. If A (exchange state 1) $> B$ (exchange state 2), and $B > C$ (exchange state 3), then with the altimeter of Pareto, $A > C$.[26]

The two individuals may also believe that they avoid the other failures of collective action prominently displayed in various theorems and proofs. Certainly, their preferences are transformed without interruption into collective outcomes. Each individual gets exactly what he prefers in an ideal exchange. Equity tests also seem to be met by the condition of liberty found in exchanges. If individuals are truly free to exchange goods, nondictatorship is realized. The liberal model of

25. Theories of the market routinely use a number of other defining conditions but rarely agree on them. Neoclassical economics, e.g., develops markers on perfect information, while the tradition identified with Ludwig von Mises, in *Human Action—a Treatise on Economics* (London: Hodge, 1949), abandons conditions of perfect information. I know of no theory of the market, however, that does not contain as a core concept the minimalist definition I offer here.

26. Douglas Rae, "An Altimeter for Mr. Escher's Stairway: A Comment on William H. Riker's 'Implications from the Disequilibrium of Majority Rule for the Study of Institutions,'" *American Political Science Review* 74 (June 1980): 451-55. Rae views the altimeter of neoclassical markets as a device "to underwrite the rights-utility bond" in liberal thought.

a political community is maintained by introducing its defining features directly into an idealized view of markets.

But the requirements of liberalism are not met in market institutions. Theories about institutions can be criticized from two standpoints. One might say that, as applied to reality, they have (logically, empirically) anomalous or contradictory implications. Such a critique of Plato's *Republic,* for example, would concentrate on the general problems of implementation and the effects of partial implementation on one or another of the state's features (what happens, say, to the ideal status of the arrangements if one part fails—no common property for the upper guardians—and all other parts are intact). A second line of criticism might concentrate on the theory as an ideal and trace out problems and inconsistencies in terms of its internal logic. Both lines have been developed in the literature on exchange. An inspection of this literature will suggest (*a*) how markets, like aggregation machines, fail tests of rationality and equity, and (*b*) what changes in the general concepts of collective choice might be needed to avoid these failures.

The failures of markets, however, must also be measured against our expectations. An especially optimistic and persuasive case for the rationality and fairness of exchange is drawn up by Robert Nozick in his *Anarchy, State, and Utopia.* Nozick's account is important in several ways, not least because the moral community fixed at the background level in Rawls's theory of justice and in Arrow's theorem is an explicit and richly described society in Nozick's story of justice. We are asked to imagine a collection of separate and rational individuals in conditions of no authority (a state of nature). Lockean problems occur in these conditions, primarily overestimations of harm that lead to excessive retributions and an endless series of retaliations. Mutual protection associations develop to address these problems. Eventually, a dominant association emerges that provides protection to all who pay for its services.

The state-like entity providing protection to its clients is limited by the moral endowments of individuals in conditions of no authority. Nozick sugggests a hyperplane of moral space around each individual which can be crossed only if the individual consents. (If the state must cross such a moral border in protecting its clients, compensation must be paid to the individual.) Indeed, the unauthorized crossings in the state of nature are precisely what occasion protection associations. The moral status of individuals—separate from each other and with rights to pursue their own goals without interference from others—is unchanged when authority is established. Limitations on state authority are thus set by the premised moral features. The state cannot redistribute resources but only carry out the protective functions for which it was created.

The contrasts between such a limited state and interventionist accounts of authority are well known.[27] Collective distributions are just on Nozick's theory if the individuals are entitled to their shares of the collective product, not if the distribution satisfies some time-slice principle that is indifferent to the way the distribution occurs (e.g., Rawls's difference principle). Put simply, a distribution is just on Nozick's tests if it arises from another just distribution by legitimate means (prior steps that are just). Legitimacy in this case follows the pattern set by the

27. Among the fine anthologies elaborating and criticizing Nozick's theory is the issue of *Arizona Law Review* devoted to a symposium on *Anarchy, State, and Utopia* (vol. 19, no. 1 [1977]).

original moral community. Liberty must be maintained in the acquisition and transfer of holdings; and liberty, according to Nozick, is best secured through local exchanges of goods. The liberal model of discrete and autonomous individuals is maintained consistently throughout the theory of justice that Nozick develops by accepting only those social conditions—in particular exchange—that do not affect its defining features.

But exchange fares no better in maintaining the liberal model than aggregation does. Even the basic distinctions between power and exchange on which market justice is developed cannot always be drawn clearly. Power is an unreasonably complex term. It can be expressed as an actual occurrence, an ability, and the successful achievement of intended results, among other things.[28] When distinguished from exchange transactions, the asymmetry of power is seized by analogy: "*A* has power over *B*" is like "*A* causes *B*" to do something, where the flow is in one direction, *A* to *B*. (If *A* has power over/causes *B* to do *a*, then *B* does not have power over/cause *A* to do *a*).[29] Power is also unlike exchange in its capacity to be assigned to environmental or ecological control, whereby *A* can effectively get *B* to do *a* by affecting some set of conditions, *c*, without any communication or direct contact with *B*, or *A*—*C*—*B* (*a*).[30]

An exchange transaction is, in contrast, a relation between *A* and *B* characterized by a transfer of items (goods, behaviors, etc.). The customer buying a dozen eggs with ready cash has engaged in a social exchange characterized by a medium of general value (money). When set apart from "power," the symmetrical nature of exchange is stressed: equity in outcome, volition, and effect is characteristic of social exchange. The flow of action is reciprocal, from *A* to *B* and *B* to *A*. Economics is frequently said to be concerned with social exchanges, politics with power relations—a division of labor assumed on distinctions between the two types of events.[31]

Critics, however, have pointed out a rich area of overlap between power and exchange. In general, each concept seems robust enough to include almost all members of the other: (*a*) Coercion (or negative sanctions), long a defining component of power, can successfully be viewed as a feature of social exchange in which *B* does *a* in order to avoid sanctions (in other words, he exchanges his behavior for the non-occurrence of the sanction). (*b*) The nonvolitional nature of power, in which the respondent acts against his will (preferences, interests), is uninterpreted in the absence of opportunity costs. Thus the provision of sufficient rewards can at once get *B* to do *a* against his will and also complete an exchange transaction. (*c*) An imbalance in outcome, volition, or effect can also be found in social exchanges; for an exchange unfavorable to one or some of the parties (non-Pareto in outcome) is still nonetheless an exchange.[32] Or, power can be viewed successfully as an exchange, and exchange looks remarkably like power. One intriguing effort to transform exchange to power fragments exchange into a series of power relations: (*a*) *A* gets *B* to do *j* (hand over a dozen eggs), and (*b*) *B* gets *A* to do *k* (hand over the ready cash), with *A* and *B* each occupying (temporarily) the role of power authority vs. respondent.[33]

The only clear distinction between the two concepts might be drawn

28. See, respectively, the entry on "power" by Robert Dahl in *International Encyclopedia of the Social Sciencies* (New York: Macmillan, 1968); by Quentin Gibson in *Philosophy of Social Science* 1 (1971): 101–12; and by Stanley Benn in *Encyclopedia of Philosophy*, ed. Paul Edwards (New York: Macmillan, 1967).

29. E.g., James March, "An Introduction to the Theory and Measurement of Influence," *American Political Science Review* 49 (June 1955): 431–51; Robert Dahl, "power"; Jack Nagel, *The Descriptive Analysis of Power* (New Haven, Conn.: Yale University Press, 1975).

30. The first recognition of "ecological" power was, so far as I know, by Dorwin Cartwright, "Influence, Leadership, Control," in *Handbook of Organizations*, ed. James March (Chicago: Rand McNally, 1965), pp. 1–41.
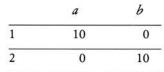
31. See Harry Eckstein, "Authority Patterns: A Structural Basis for Political Inquiry," *American Political Science Review* 67 (December 1973): 1142–61, for a slightly different rendition.

32. For a review of recent literature (and some helpful contributions to it), see David Baldwin, "Power and Social Exchange," *American Political Science Review* 72 (December 1978): 1229–42.

33. Baldwin, "Power and Social Exchange"; Felix Oppenheim, "'Power' Revisited," *Journal of Politics* 40 (1978): 589–608.

Syracuse Scholar (1979-1991), Vol. 6, Iss. 1 [1985], Art. 2

up on differences between actions that are direct (individuals act on each other) and those that are oblique (individuals act on attendant conditions or collateral agents). Authorities who control through positive or negative sanctions are exercising power in a direct fashion. When control is consummated by restricting the agenda of choices or by failing to make decisions for alternative social arrangements, power is oblique—and thus distinct from exchange.[34] Note that the condition of universal domain—one guarantee that agendas will not be controlled by authorities—is embedded in a satisfactory definition of liberty.[35] So to the degree that exchanges fulfill liberty, oblique power will be excluded from exchange. But direct forms of power are less easily demarcated from exchange.

Suppose, however, that the critics can be satisfied with criteria demarcating exchange from power. Even then the concept of exchange falters in decisive ways on a settlement of the rational problems surveyed here. Imagine now in the thought experiment that a third party, though not involved directly in the exchange, is part of the social practice of exchange between the first two individuals. Suppose also that the exchange results in a loss to this third individual, that some of the costs of the transaction are passed off to him. Then Pareto is an illusion, maintained only by ignoring the external effects of an exchange. And transitivity is lost when the wider effects of exchange are calculated. Shifting costs to those outside a transaction is a violation of equity however conceived (since the ''outsiders'' are neither responsible for, nor benefit from, the exchange). But the rational point is more important: a comprehensive perspective on exchange can deny one of the principles—Pareto—that justifies exchange as superior to power or authority. The altimeter that cyclical majorities deny in majority rule is also, when externalities occur, missing in exchange.[36]

Nor is coordination among individuals guaranteed in exchanges. Experimental efforts at institution building in game theory frequently use bargaining and side payments—the introduction of exchange—as a solution to coordination problems. Let a simple case be part of the thought experiment on exchange. Two individuals consider alternatives a and b. Let each individual assign a utility to the alternatives, represented by the integers below.

	a	b
1	10	0
2	0	10

A Prisoners' Dilemma game occurs with advantage going to the player whose preferred alternative is considered first. Since exchange depends on the realization of a and b, the conditions for successful exchange may be exactly those avoiding coordination and assurance dilemmas: either (1) a cohesive social unit established by stable players; repetitive alternatives; a continuing framework of rewards and penalties; small, face-to-face social relations, and so on;[37] or (2) an external guarantee of compliance with agreements (e.g., the coercive state, long accepted as the guarantor of property rights and contracts). Exchange theory, in

34. Peter Bachrach and Morton Baratz, ''Two Faces of Power,'' *American Political Science Review* 56 (December 1962): 947-52.

35. Universal domain ensures only negative liberty, or the absence of obstacles to free choice. No condition in the theorem guarantees or even refers to positive liberty, which requires the provision of those conditions in the absence of which effective liberty is impossible. I refer here to Isaiah Berlin's famous distinction in *Two Concepts of Liberty* (Oxford: Clarendon Press, 1958) and *Four Essays on Liberty* (New York: Oxford University Press, 1969). See also the synthesis of negative and positive liberty by means of the concept of ''constraint,'' in Gerald MacCullum, ''Negative and Positive Liberty,'' in *Contemporary Political Theory*, ed. Alan de Crespigny and Alan Wertheimer (New York: Atherton, 1970). The theorem's concentration on negative liberty indicates yet again the liberal view that individual autonomy depends on noninterference. The possibility that a different, and perhaps more robust, sense of autonomy occurs in communal arrangements, especially those guaranteed by the state, is simply never considered.

36. William H. Riker and Steven J. Brams, ''The Paradox of Vote-Trading,'' *American Political Science Review* 67 (December 1973): 1235-47. The technical literature on this issue, as with the others treated here, is vast. See, e.g., Thomas Schwartz, ''Vote Trading and Pareto Efficiency,'' *Public Choice* 24 (1975): 101-10, for one among several qualifications to the claim that externalities can make everyone worse off in exchange (in this case, vote trading).

37. Peter Bernholz, ''Prisoners' Dilemma, Logrolling and Cyclical Group Preferences,'' *Public Choice* 29 (Spring 1977): 73-84.

short, does not so much solve problems of coordination as restate the conditions needed for a solution.

If we maintain in the thought experiment the image of numerous individuals exchanging goods to be aggregated into a collective outcome, another version of the fallacy of composition can occur with exchanges. A decision to exchange items in individual transactions is not equivalent to, or the condition for the derivation of, a preference for the distribution resulting from the aggregated transactions. Professors can consistently (*a*) choose to pay $10 to see their favorite soccer team play, while (*b*) not preferring the inequitable distributions of moneys to star players that result (e.g., $1 million per year salaries). Champions of the market (like Nozick) inevitably point out that a modification of the aggregate distribution will restrict the liberty of individuals to dispose of, and accumulate, their resources as they wish. But this observation does not touch the point on rationality: that what is rational for the individual in single transactions can be irrational for the individual in aggregate form.[38]

What does touch on liberty, however, is a widely-held point on the empirical operation of markets. Exchange transactions are notorious for producing inequitable distributions of resources. Even the most cursory glance at the operations of the free market will reveal enormously unequal distributions. Again, however, the logic of exchange suggests why unequal patterns are possible. Repeating an exchange over time while maintaining liberty provides no check on the pattern of resource distribution. Outcomes can, and empirically do, result in unequal accumulations. And unequal outcomes affect the premises of exchange. Full freedom to exchange goods depends in the most obvious ways on equality of starting resources. If individual 1 is wealthy, individual 2 not, it is a matter of little dispute that coercive results are both possible and likely. And if dictators, those who dominate others on the disproportionate accumulation of wealth, are the products of free markets as well as of political institutions, exchange cannot be a solution to the equity problems of collective choice.[39]

These brief points on markets are of course compressed critiques of a complex set of theories. To be reasonable, they have to be joined to empirical studies of the market, and, of course, expanded. Even in compressed form, however, they state what only blind advocates of the market can deny: that markets do not fulfill the rational and moral criteria drawn from the liberal model on which they are developed. Markets may (1) fail to transform preferences into collective outcomes (assurance and coordination failures can occur) and (2) fail both consistency tests (the fallacy of composition holds for markets as well as for aggregation devices) and equity criteria (Pareto and nondictatorship). Markets, in short, seem as vulnerable to rational and moral breakdowns as aggregation machines and so are members of that species of rational problem represented by Prisoners' Dilemma and Arrow's theorem. It is important to note that market failures occur because there is no feature of exchange that will guarantee the conditions set out by the rational and moral criteria of collective choice theory. The logic of exchange, though altering several of the conditions found in collective choice (successfully substituting, for example, various cardinal scales for Arrow's orderings), still permits rational

38. This is a well-known critique of Nozick's main argument in *Anarchy, State, and Utopia.* The recent discussion of vote trading in Riker, *Liberalism against Populism,* pp. 157–67, establishes in more general ways that exchange will not avoid composition fallacies. See also the discussion of market disequilibria by Norman Schofield, "Instability and Development in the Political Economy," in *Political Equilibrium,* ed. Peter C. Ordeshook and Kenneth A. Shepsle (Boston: Kluwer-Nijhoff, 1982).

39. This last point is an irresistible observation on the empirical operation of markets that is much disputed in the literature. The conceptual point, however, is indisputable: liberty in exchange is strongly conditioned by the prior distribution of property rights, and no continuing system of exchange can guarantee equity in the distribution of property. Indeed, markets provide no criteria for the normative task of assigning property rights, which nonetheless is required prior to the operations of the market. An especially helpful discussion of these points is Charles Lindblom's *Politics and Markets* (New York: Basic Books, 1977). See also the nice summary in Charles Schultze, *The Public Use of Private Interest* (Washington, D.C.: Brookings Institution, 1977).

breakdowns between liberal premises and collective outcomes.

The failure of markets, moreover, is due to the same conflict between arithmetical and moral languages found in Arrow's theorem. Nozick's unexamined assumption is that individuals are self-legislating creatures. Yet markets provide no institutional arrangements to address individual claims or justify aggregate outcomes. Like aggregation machines, markets accept only reasonless entries, and market outcomes are ungoverned collections of such entries. The absence of reasons has a double edge in exchange theory. On the one hand, there are advantages. Exchange is effective in reaching joint outcomes in large measure because shared reasons are not required to reach an agreement. (Even the definition of a gain or loss can vary with each individual in an exchange.) Outcomes satisfactory to all parties that leave intact the variety of reasons contributing to dispute are obviously less demanding than those requiring an agreement on reasons. Indeed, to demand that each individual in an exchange agree on reasons may be pathological, rupturing the selective compromise distinctive of exchange. A Pareto optimal outcome can be reached with each individual having a different reason to support the outcome.

On the other hand, the absence of reasons makes rational agreement difficult, perhaps impossible, on a number of issues. If reasons are entirely excluded from settlements among individuals, those issues requiring a reasoned resolution—such as moral issues—cannot be included in such settlements.[40] This limitation is generally recognized in the use of exchange. In ordinary language, items are sometimes ''priceless.'' One import of such thoughts is that certain goods are regularly excluded from the marketplace. Children, it is frequently noted, are not bought and sold as exchange commodities. They are allocated, when allocation is needed, on reasoned grounds by adoption agencies. Although the items uncomfortable with exchange settlements vary with conventions, those items with moral status resist market transactions (moral language requiring reasoned orderings). Thus life-maintenance items and basic rights are often outside the pale of exchange in Western societies. Technical resolutions also, since they require reasoned outcomes, cannot be settled by exchange. The limited usefulness of exchange as an instrument to resolve moral and technical issues is the natural consequence of reasonless orderings. And these limitations restrict the capacity of markets to represent that liberal community of reason-giving individuals that provides rational and moral criteria in collective choice for evaluating institutions.

The limitations of exchange are generally recognized in the development of regulating devices that meet liberal tests. Dworkin, for example, introduces the liberal model in order to elaborate theories of equality.[41] A hypothetical collection of immigrants (rational, autonomous, and discrete individuals) is faced with the problem of distributing bundles of resources among themselves. A distribution is equal if no one prefers anyone else's bundle of resources to his own. An auction is the device that Dworkin suggests to effect an equal distribution of resources: prices are set so that all lots (resources) clear the market (there is only one purchaser). The auction addresses the problem of dissatisfaction with resource bundles due to different tastes or needs by distributing each lot in terms of how important the

40. An idea expressed first (and best) by Marx. See, e.g., the lovely phrases in the opening pages of ''Needs, Production, and Division of Labor,'' in *Karl Marx: Early Writings*, ed. T. B. Bottomore (New York: McGraw-Hill, 1964). See also the classic study by Charles Titmuss, *The Gift Relationship* (London: Allen & Unwin, 1971). The tradition of excluding items from markets on moral grounds is long and varied. Even libertarians join in. Mill, in *On Liberty*, rules out slavery as a possible outcome from exchange.

41. Ronald Dworkin, ''What Is Equality?'' pts. 1 and 2, *Philosophy and Public Affairs* 10 (Summer 1981): 185-246.

resource bundle is to all individuals in the community. The auction thus assigns importance to each individual preference by comparing it to all other preferences.

Unequal talents require additional devices, however. Nozick's noninterventionist state must permit any inequality that might follow from the initial conditions of equality. Dworkin also allows inequality in the distribution of resources. But, unlike minimal-state theories, Dworkin's account requires that inequalities be ambition sensitive (in reflecting industry, effort) but not endowment sensitive (luck, genetics). The state must permit the first type of inequality and compensate for the second. The device used to reestablish a distributive balance is the progressive income tax at rates set by a hypothetical insurance market. Suppose each individual knows his own talents and the income distribution but not his location on the distributive matrix. He can then choose an income level and pay the premium set for that level. The insurance would then pay the difference between the actual and chosen income levels. Devices such as these are designed to compensate for unequal distributions of talent without penalizing unequal expenditures of effort. Nozick's star athlete (Wilt Chamberlin) will in this way be taxed progressively at rates deemed fair by a theory that marks off warranted and unwarranted inequalities.

An inspection of exchange tells us that, as with aggregation, additional devices must be introduced if the moral features of liberalism are to be successfully represented in institutions. If the liberal model is interpreted simply in terms of numerical units, integers to be arithmetically combined, nothing compels us to compensate for inequalities. But the liberal community of moral agents in a hypothetical setting of autonomy both guides and justifies the combination rules designed to extend liberalism to institutional forms. The transformation by markets is incomplete without regulating devices to ensure that the moral conditions of the hypothetical community are realized in the actual conditions of society.

Models of a political society

A general view of aggregation and exchange submits this to our understanding: that the surface and background assumptions of collective choice fail to generate an arrangement of individual parts that is comprehensible in terms of the assumptions.

The failure of the assumptions is traceable to a basic incompatibility between moral and arithmetical languages. A liberal model of society is a background assumption in Rawls's theory of justice and Arrow's theorem, a foreground description in Nozick's market version of justice. This liberal model sets out certain moral conditions that are to be fulfilled in collective outcomes, but these conditions cannot be fulfilled with the use of arithmetical composition rules. The moral language of Arrow's problem surfaces in conflicts between rationality and equity. Arrow's dictator, for example, satisfies the requirements of rationality as set out in the conditions of the theorem. (Indeed, decisive or dominant individuals violate no rule of rationality known to collective choice theory.) Equity, however, is offended by dictatorship or exploitation. As we have seen, the rejection of decisive individuals requires that no individual override any other on any preference ordering; and this

avoidance of inequality in all areas of collective choice requires that individuals be absolutely equal. No collective outcome can then be generated without violating at least one of the theorem's other conditions. The starting point for this eventual paralysis is the same liberal model of morally autonomous individuals that leads to the failures of exchange.

The moral conditions of liberalism seem to be incompatible with any procedural or technical effort to forge social institutions. Imagine two individuals. Now (*a*) combine by counting some set of values discovered or ascribed to the two individuals, or (*b*) witness the two individuals transferring a unit of value. Suppose that either event starts and stops with the physicalist fact (a description of aggregation or transfer). Nothing can be inferred from the physicalist fact that will establish moral and rational criteria. The individuals might be machines, the action a mechanical process. Now imagine that the two individuals have been invested with moral and rational qualities (they are autonomous; they feel, suffer, think, calculate; they are moral equals). Neither aggregation nor exchange has the capacity to address the claims (reasoned preferences) these individuals might make on the outcomes of their joint actions or even on the practices within which their actions occur. Both aggregation and exchange fail moral and rational tests because each is a numerical form of interaction and the tests are drawn from moral conceptions of human life. The conflicts in each case occur between counting rules, emergent outcomes, and the moral conditions of a liberal community.

It is a truism that all theories depend on assumptions. It is another truism, though a more profound one, that there are two models of a political society that order and interpret even the most basic of assumptions: one is holistic, the other arithmetical. Aristotle, developing his political theory on holistic terms, fuses items that arithmetical models maintain as separate: individual, state, society. One consequence of this fusion is that a number of concepts and theories characterizing liberal and libertarian political theory cannot be developed on the Aristotelian version of the polis. Among these are individual rights, anarchism, and civil disobedience—each of which requires an adversary relationship between the individual and society that is impossible to conceive in Aristotle's political society. Hobbes, in contrast, separates individual, state, and society on a more nearly numerical model of association. From this separation follow theories of individual rights, the intelligibility (though not rationality) of anarchism, and the possibility of civil disobedience.

Concepts change their sense from one model to the other. Liberty, for example, is a communal fulfillment in Aristotle, a negative freedom from state regulation in Hobbes. But the stronger effect of the two models on political thought is seen in the realization that some concepts exist only in terms of one or the other model. Anarchism, for example, is not rejected by Aristotle (as it is by Hobbes). The question asked by Hobbes, Ought there to be any authority? is simply excluded on the conditions of Aristotle's political theory. Anarchism is literally unintelligible on Aristotle's organic model of civil society.

The contradictions of aggregation and exchange are more fully understood if framed in terms of a conflict within liberalism between

the holistic needs of morality and the discrete logic of arithmetical languages. The liberal model accepts individuals as separate and countable units, yet endows them with moral features that are more comfortably elaborated in holistic political theories. The failures of collective choice represented by Arrow's theorem and exchange theory are the result of contradictions within liberal theory. These contradictions are produced because there is no mediating device to transform the disjointed features of liberalism into the type of moral community liberalism seeks. The analysis of background concepts tells us in general that the formal terms and conditions of collective choice do not cut deeply enough, and that there is another layer of conflict where theories with competing needs (in this case, liberalism) must be amended to produce noncontradictory collective outcomes.

Perhaps the original flaw in liberalism is that it is best elaborated as a series of shields that insulate individuals from collective regulation. Certainly the failures of aggregation and exchange testify to the difficulty of extending the hypothetical community of moral agents into the area of social practice. Yet the very minimalism of these two combining forms, which at first thought seems congenial with liberal communities, is the source of failure. Neither form is robust enough to express the moral and rational criteria that liberalism seems to require. More substantial structures are needed to move the liberal assumptions in collective choice to the explicit level of a rational political society.

It is important to see the problems of aggregation and exchange as a conflict between two different kinds of languages in order to recognize those institutions that do meet liberal needs. MacKay, for example, views Arrow's theorem as an infinite regress paradox requiring a familiar resolution: introduce a first cause, in this case a dismissal of unlimited scope through a restriction on the pattern of preferences.[42] Since it is well known that single-peaked preferences avoid cyclical majorities,[43] transitivity can be achieved and the paradox dissolved. But if the breakdowns originate in background moral expectations, unlimited scope cannot be dismissed without first in some way preserving autonomy, one of the features of moral agency that creates the problem in the first place. The identification of background assumptions tells us that some reconstruction of basic concepts is needed for a satisfactory solution to the problems of collective choice. The prime candidates seem to be the concept of rationality, the primitive term "individual" (currently both a count noun and a holistic unit in collective choice), and the composition rules producing collective outcomes. Or, put in more unsettling language, the rational breakdowns between individual and society elaborated in collective choice may force us to paradigm-level changes in political theory.

42. Alfred F. MacKay, *Arrow's Theorem: The Paradox of Social Choice* (New Haven, Conn.: Yale University Press, 1980).

43. Black, *The Theory of Committees and Elections*.

Appendix 1

The original version of Newcomb's problem was first described by Robert Nozick in his "Newcomb's Problem and Two Principles of Choice," in *Essays in Honor of Carl G. Hempel*, ed. Nicholas Rescher (Dordrecht: Reidel, 1969). See also Martin Gardner's column, "Mathematical Games," in *Scientific American* 230 (March 1974): 102-9. In the problem, an individual faces two opaque boxes. In one box there is always $1,000. In the second box there is either $1 million or nothing. A superior being places the $1 million in the second box if he correctly

predicts that the actor will choose only the second box (choice A). The being does not place the $1 million if he predicts that the individual will choose both boxes (choice B). Should the individual choose A or B? Newcomb's problem is a conflict between expected utility (EU) and dominance (D) in making the choice of A or B.

(a) On EU: Assuming a probability of 1 for the being's correct predictions, then $A = [1 \times \$1,000,000] + [0 \times 0] = \$1,000,000$; and $B = [0 \times \$1,000,000] + [1 \times \$1,000] = \$1,000$. On certainty of prediction, EU $= A > B$ (choose only the second box). Note that even with a successful prediction rate as low as .5005, EU still produces $A > B$.

(b) On D: Taking only one box has outcomes of $1 or $0. Taking both boxes has outcomes of $1,001,000 or $1,000. The money is either in the second box or not. In either case, the choice of both boxes leads to results superior to those of taking only one box.

Embellishment for dominance principle: Imagine a friendly observer behind the boxes, able to see into them. Whether the second box is empty or stuffed with $1 million, the friend would urge the actor to take both boxes; for $1,001,000 > \$1,000,000$ and $\$1,000 > \$0$. Dominance requires $B > A$.

Appendix 2

Let the following definitions be accepted as Arrow's meanings:

Decisiveness: A set of individuals is decisive for x against y ($x, y \in X$, where $X = $ a set of mutually exclusive alternatives) if, for every profile u in which (1) $x \geq_i y$ for all $i \in s$; (2) $x \geq_i y$ for at least one $i \in s$; we have $x \in v$ (where $v = $ a nonempty subset of X) $= > y \notin Cu(v)$ (or, y is excluded from the choice function, Cu, over the agendas, v).

Put in looser language, if x is in the agenda, y will not be chosen. Decisiveness is thus an exclusionary power.

Pareto: For all distinct $x, y \in X$, the set of all n (where $n = $ the number of individuals) is decisive for x against y.

Nondictatorship: There is no one individual who alone is decisive for x against y for all distinct $x, y \in X$.

Universal domain: (1) The domain of f consists of all logically possible profiles, u (roughly, orderings of alternatives). (2) At every u, the domain of $Cu = f(u)$ includes all finite nonempty subsets of X. If $u = $ a logically possible profile and $v \in X$, v is finite and $v \neq \emptyset$, $Cu(v) \neq \emptyset$.

Or, the domain is unrestricted in the sense that every logically possible combination of individual orderings of alternatives in X must be the domain of the social choice function.

Independence of irrelevant alternatives: Let u and u' be two distinct profiles whose restriction to v are the same. If f assigns choice function Cu to profile u and Cu' to u', then $Cu(v) = Cu'(v)$.

Completeness: $x\ Ru\ y$ or $y\ Ru\ x$ for all x, y.

Transitivity: $x\ Ru\ y$ and $y\ Ru\ z => x\ Ru\ z$.

Rationality: For every profile u, Cu has a binary relation Ru such that Cu can be explained by Ru as $Cu(v) = [x \in v\ (X\ Ru\ \text{for all } y \in v)]$.