

Journal of Philosophy, Inc.

Rational Choice and Social Theory

Author(s): Debra Satz and John Ferejohn

Source: *The Journal of Philosophy*, Vol. 91, No. 2 (Feb., 1994), pp. 71-87

Published by: [Journal of Philosophy, Inc.](#)

Stable URL: <http://www.jstor.org/stable/2940928>

Accessed: 23/10/2011 23:37

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Journal of Philosophy, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*.

<http://www.jstor.org>

RATIONAL CHOICE AND SOCIAL THEORY*

THE theory of rational choice plays a central role in the social sciences. The received view among philosophers who study rational-choice theory (both proponents and critics) holds that the theory is psychological and individualistic. Rational-choice theory is taken to be a *psychological* theory in that it explains a person's actions in terms of her mental states. A rational choice or action is one in which the agent takes the best available action given her preferences and beliefs. The theory is also taken to be an *individualistic* theory in that it applies directly only to individuals (that is, only individuals have preferences). The most common philosophical interpretation of rational-choice theory conceives of it as a psychological theory wedded to a reductionist program in the social sciences, where the behavior of a social aggregation is explained in terms of the mental states (that is, the desires and beliefs) of its component individuals and their interactions. On the received view,¹ one cannot be a rational-choice theorist without also being a methodological individualist.²

We believe that the received interpretation is wrong about the nature of rational-choice theory. While rational-choice theory contains statements that can be interpreted in individualist psychological terms, we shall argue that such an interpretation is optional. A theory—a formal syntactic structure—can be interpreted in distinct ways, and how we interpret the theory depends on its larger context. In particular, we believe that rational-choice theory is consistent with nonpsychological interpretations that in some contexts are more plausible. Even where people do act as the theory predicts, the best rational-choice explanation of their actions may not be an individualist, psychological one.

* Thanks to Samuel Bowles, Michael Bratman, David Copp, John Dupré, Geoff Garrett, Peter Godfrey-Smith, Daniel Hausman, Andrew Levine, Daniel Little, Alan Nelson, Alexander Rosenberg, Jonathan Riley, and Elisabeth Wood for comments. Participants at workshops at U.C. Davis, CREA (Paris), and the Public Choice Society Meetings provided valuable responses.

¹ The received view we are criticizing is held primarily by philosophers. Despite some notable exceptions (e.g., G. Becker, *The Economic Approach to Human Behavior* (Chicago: University Press, 1976)) economists have long expressed skepticism about rational-choice theory as a psychological theory. Cf. P. Samuelson, "A Note on the Pure Theory of Consumer's Behavior," *Economica*, v (1938): 61–71. Economists have embraced a purely behaviorist interpretation of the theory, however, and generally seem to endorse what we refer to below as a reductionist program.

² Cf. William Riker, "The Political Psychology of Rational Choice Theory," unpublished manuscript.

In what follows, we undertake a two-part argument. In the first part, we logically detach rational-choice explanations from psychological explanations. The argument of this section is meant to weaken one strategy for connecting rational-choice theory with individualism in explanation. In the second part, we show that some rational-choice explanations gain their explanatory power from features of the agent's environment. The psychology of the agent in such cases is an entirely imputed one: "preferences" are derived on the basis of an agent's location in a social structure.³ We believe that rational-choice explanations are most plausible in settings in which individual action is severely constrained, and thus where the theory gets its explanatory power from structure-generated interests and not from actual individual psychology. In the absence of strong environmental constraints, we believe that rational choice is a weak theory, with limited predictive power. Even a cursory glance at the successful uses of the theory confirms this: the theory of the firm has been a far more productive explanatory theory than the theory of consumer behavior; similarly, rational-choice explanations are more powerful when their object is the behavior of political parties as opposed to voters. The primary reason for the asymmetry of explanatory success is that consumers and voters face less competitive environments than firms and parties. We fully realize the irony of our contention: the theory of rational choice is most powerful in contexts where choice is limited.⁴

The most important consequence of our argument is that structuralism—social theory that offers explanations in terms of irreducible relational or structural properties—can be compatible with rational-choice theory. In this sense, the apparent tension between the reductionism of rational choice and the holism of "structural" theories is superficial.

I

(a) *The received view.* It is common to interpret the theory of rational choice in psychological terms, as instructing an agent maximally to satisfy her overall scheme of preferences. Thus understood, the theory provides only a formal and thin conception of rationality.

³ Marxist political theorists, for example, use rational-choice theory to explain various political outcomes in terms of workers' resources and institutional constraints. Such explanations attribute interests to workers in virtue of their social positions, e.g., their place in the network of economic relations. The explanation does not proceed inductively via *each* worker's individual preferences; it deduces *every* worker's preferences from the environment. Cf. A. Przeworski, *Capitalism and Social Democracy* (New York: Cambridge, 1986), esp. chs. 4 and 5.

⁴ Cf. C. Plott, "Rational Choice in Experimental Markets," *Journal of Business*, LIX (1986): 301–28.

The theory considers only the formal or mathematical properties of the agent's preferences. In particular, the content of those preferences is irrelevant to the theory. It does not matter, for example, what reasons the agent has for her preferences, or indeed why she has come to hold them at all. The theory of formal rationality does not judge the rationality of the agent's preferences: "rationality" pertains only to the relations among her preferences. More specifically, there is a consistency requirement on the preferences of a rational agent. On some accounts, a rational agent's preferences must be transitive: if I prefer *A* to *B* and *B* to *C*, I should prefer *A* to *C*. If I am rational, in the sense required by the theory, I shall order my preferences and act so as to achieve that preference to which I give the highest overall ranking.

There are two points to notice about this interpretation. First, rational-choice theory is seen as describing what is actually going on inside us when we reason. We shall call this the *internalist interpretation* of rational-choice theory. From this perspective, mental entities (for example, preferences and beliefs) are thought to be causally related to choice, in the sense of being reasons for an agent's having made the choice.⁵ The reason that an agent chooses act *x* over act *y* is that the outcome that is believed to follow from *x* is preferred to that which is believed to follow from *y*, and so preferences and beliefs are said to cause the choice.

Second, on the internalist interpretation, rational choice is a normative as well as an explanatory enterprise. It tells us what we ought to do in order to achieve our aims. Insofar as an action expresses some choice the agent has made, that choice can be subjected to moral and prudential scrutiny.⁶ For any choice, we can inquire whether or not that choice stands in the appropriate (maximizing) relation to the beliefs and preferences of the agent.

Most recent criticisms of rational-choice theory have been directed against both the psychological mechanism upon which the theory apparently rests and its adequacy as a normative theory. The psychology of rational-choice theory is so simple that it excludes consideration of many complex cognitive processes of the sort that have occupied the attention of psychologists in recent years. Herbert Simon⁷ has argued against the idea of human beings as friction-

⁵ Cf. Donald Davidson, "Actions, Reasons, and Causes," this JOURNAL, LX, 23 (November 7, 1963): 685–700.

⁶ In David Gauthier's phrase, the transitivity of rational choice holds an individual's choices up to a standard; *Morals by Agreement* (New York: Oxford, 1986), p. 40.

⁷ *Models of Bounded Rationality*, Vols. 1–2 (Cambridge: MIT, 1982).

less and omniscient calculating machines. The computational and information gathering costs of maximizing are simply too high; in many cases people "satisfice" instead. Tversky and Kahnemann have demonstrated that in many contexts people hold intransitive preference orderings. Amartya Sen⁸ has claimed that the theory cannot make sense of the complexity of moral choices: it is unconditionally valid only for agents whom he calls fools. Elizabeth Anderson⁹ has extended Sen's argument, concluding that rational-choice theory's conception of the relationship between choice and preference is too formalist to represent a plausible (or desirable) human psychology.

We find these criticisms of rational-choice theory, taken as a *psychological* theory, quite persuasive. If the theory is taken to specify a psychological mechanism, then these criticisms may be fatal. Human psychology, studied in the laboratory, the psychoanalyst's office, or through introspection, appears to be far more complicated and has much more structure than rational-choice theory allows. The conception of human rational agency in terms of maximizing over a complete and consistent set of preference orderings is not psychologically realistic.

If we are correct about the nature of rational-choice explanations, however, these criticisms will largely miss their mark. Many of the most important uses of the theory need not rely directly on any theory of human psychology. The theory's use in economics and sociology, for example, is largely aimed at illuminating *structures* of social interaction in markets, governments, and other institutions. In many social-scientific explanations, we are not interested in explaining a particular agent's behavior, but in the general regularities that govern the behavior of all agents. In such cases, it is not the agents' psychologies that primarily explain their behavior, but the environmental constraints they face. We shall argue that many of the rational-choice explanations employed in economics, political theory, and game theory do not necessarily depend on psychological foundations.

(b) *The limited autonomy of rational-choice theory from psychology.* Consider two possible responses to the criticisms of the internalist interpretation of rational choice: one that we call *radical externalism* (behaviorism) and the other a *moderate externalism* that we adopt here. The radical externalist denies the existence as well as the

⁸ "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy and Public Affairs*, vi (1977): 317–44.

⁹ "Some Problems in the Normative Theory of Rational Choice with Consequences for Empirical Research," unpublished manuscript.

causal efficacy of mental states. He interprets preferences behavioristically, as choices.¹⁰ Mental entities are simply theoretical constructs inferred from human behavior, or calculating devices, that help us make predictions about behavior; they entail no claims about the agent's psychology at all.¹¹

We agree with the radical externalist that all that formal rationality entails is that an agent's action is explicable *as if* she is maximizing preferences.¹² In this sense, formal rationality is simply the claim that human action is "consistent" with the hypothesis of goal seeking.¹³ It is a coherence requirement on behavior. Saying that people act consistently is not to say why they act; it is not even to say that they have reasons to act. But it is to say that some of their actions can be described as if they had reasons, as if their behavior was goal-directed. The only restrictions regarding the properties of mental states are those required for the consistency of the theory.¹⁴

Despite its attractions, we reject radical externalism. In the first place, we do not deny the existence of mental entities like beliefs

¹⁰ Notice that on the externalist account, rational-choice theory can be applied wherever there is a patterned behavior: for example, it can be applied to the behavior of gas particles when heated, and to the behavior of the temperature of a room when someone enters. In practice, most radical externalists rely on a conception of "choice" which is, strictly speaking, not available to them. (This point emerged in discussion with Bratman.)

¹¹ It should not be forgotten that the original founders of the theory were positivists who wanted to avoid postulating any unobservable psychological entities. They formulated rational-choice theory solely as an external and predictive theory.

¹² Formal rationality may be defined as follows. Let C be a (choice) function that indicates for each set of available alternatives, S contained in X , the set of all alternatives, which subset the organism chooses, $C(S)$. Then (assuming that beliefs are true), we say that C is rational if there is a preference ordering, R , on X such that $C(S)$ is the R -maximal element of S for each S . We say that C is transitive (or T) rational if it can be rationalized by a transitive R . Note that without further restrictions, a rationalization need not be unique.

Furthermore, note that a choice function is a description of how an organism would choose in a variety of situations. As such, it specifies behavior in a number of counterfactual situations. If R rationalizes C , it tells how C would choose in each of these situations, and possibly in other situations as well.

¹³ The following two axioms on choice functions are necessary and sufficient conditions for rationality:

(1) if S is a subset of T then $C(S) \subset S \cap C(T)$; (2) if $x \in C(x, y)$ and $y \in S$ then $x \in C(S)$.

¹⁴ It is easy to show that formal rationality is not tautological. Consider a set of three possible actions, A , B , and C , and consider an organism that chooses A when A and B are available, B when B and C are available and C when A and C are available and A when A , B , and C are all available. Such behavior cannot be rationalized by any preference relation. (Interestingly: consistency conditions may not be necessary components of a behavioral view of rationality. In some contexts (e.g., general equilibrium theory) we can get along with continuity and convexity conditions do not imply consistent acyclic choice.)

and intentions. That is, we hold that there are psychological explanations of phenomena and furthermore that the object of a rational-choice explanation might itself be intentionally characterized. Unless humans are blind automatons, they must actually be motivated to act in the way that the theory predicts. Our moderate externalism is compatible with realism about psychological states. Nonetheless, mental entities need not figure in the best rational-choice explanations of human action. We do not commit ourselves to a particular (and not very plausible) theory of the internal structures of intelligent agency. By bracketing that commitment, moderate externalism opens up the possibility of theoretical uses of rational-choice theory despite the implausibility of it as a psychological theory. It should be emphasized here that there is nothing in the original theory which suggests that its technical apparatus must directly map onto human psychology.

Second, we do not believe that radical externalism can serve the purposes of explanation.¹⁵ We accept the dictum that prediction is not always explanation: some predictions are not explanatory.¹⁶ Moreover, we believe that, if rational-choice theory is to be *explanatory*, it must bear some semantic relation to individual decision making. But we do not think that this makes us internalists about rational-choice theory.

Rational-choice theory can take a perspective external to the agents whose behavior is being explained, and it happens that most of the main results of rational-choice theory, construed as a scientific theory of behavior—indeed, almost all of those of neoclassical economics—require only a view of this sort.¹⁷ The important properties of supply and demand correspondences and the existence and characteristics of competitive equilibria depend only on the hypothesis that behavior can be understood *as if* it is maximizing according to some suitably restricted preference relation.¹⁸

¹⁵ See Daniel Dennett, "Intentional Systems," in *Brainstorms* (New York: Cambridge, 1978), for a defense of an instrumentalist view of explanation.

¹⁶ This is not meant, however, to impugn "mere" prediction, which in many cases can be very useful.

¹⁷ In fact, it is remarkable that much of neoclassical economics remains unaffected by assuming that individuals make decisions by flipping coins rather than by maximizing their preferences. Cf. A. Alchian, "Uncertainty, Evolution and Economic Theory," *Journal of Political Economy*, LVII (1950): 211–21.

¹⁸ The restrictions on preferences include technical requirements such as continuity and some degree of differentiability as well as more substantive restrictions such as convexity and nonsatiation. The substantive restrictions can be stated as coherence requirements on behavior and, as in the case of the maximization hypothesis, they entail no commitment as to the causal power of mental entities.

The externalist rational-choice enterprise, as a theory of behavior, must therefore be seen as having only a remote connection to psychology. While the theory connects behavior to psychological entities, it is not thereby committed to their causal force. In particular, it does not explain behavior in terms of these mental entities; it merely shows that behavior can be interpreted as consistent with them.¹⁹

II

(a) *Moderate externalist rational-choice theory and structuralism.* In what way is our moderate externalism an explanatory theory? How does it help us to understand what an agent is doing to say that her behavior is consistent with the hypothesis of maximization given some “preferences” and “beliefs”? We begin with some general points about explanation in the social sciences, and then turn to examine specific examples.

Many social-science questions—perhaps most—pose questions that cannot be answered at the level of individual psychology. These questions concern the relative stability of certain patterns of behavior. For example, there are two sorts of explanatory statements found in neoclassical economics: statements about the existence of equilibria, and comparative statics statements connecting equilibria to various descriptive parameters of the economy. Existence theory (which establishes conditions for the existence of competitive equilibria) relies only on the hypothesis that individual behavior satisfies externalist coherence tests and shows that it is possible for interacting agents to make mutually consistent choices. Furthermore, existence theory establishes a functional relationship between equilibrium outcomes and parametric descriptions of technology (the productive capacities), and individual “preferences” and endowments. Comparative statics then indicates how equilibrium outcomes change when some of the underlying parameters change. Thus, typical explanatory statements have to do with the effect of increasing some cost to an agent or set of agents (a transportation cost, a tariff, or a tax) on a choice or outcome. In both cases, we explain or predict behavior patterns or aggregations of patterns in terms of observable parameters. Often, as one might imagine, it is impossible to determine the effect of changing a particular parameter since the effect might depend on the “preferences” of individuals or on other parameters in a complicated way. Sometimes,

¹⁹ Studies have shown that pigeons do reasonably well in conforming to the axioms of rational-choice theory. Cf. J. Kagel, “Economics according to the Rat (and Pigeons too): What Have We Learned and What We Hope to Learn,” in *Laboratory Experimentation in Economics: Six Points of View*, A. Roth, ed. (New York: Cambridge, 1987).

however, one can say (unambiguously) what the effect of a parametric shift is no matter what (coherent) preferences exist.²⁰

The above explanations are examples of an equilibrium explanation.²¹ While they do not point to the actual cause of a particular agent's behavior, they do describe a causal structure. In such explanations, we can replace given individuals with others without changing the *structure* in which their actions are embedded. Suppose in one scenario *A* sells a product to *B* for some price *C*. The economy might be said to change slightly if *A* and *B* switched roles. But this change—even if a change in the economy—entails no change in its structure.²²

For some questions, the demonstration of an equilibrium—whatever its cause—can constitute an explanation. There are many examples in which rational-choice theory leads to useful results in economics without necessarily committing itself to the existence of any underlying psychological mechanism.

Consider the following examples that we think illustrate the explanatory power of moderate externalist rational choice. In each of these explanations, “preferences” are imputed to individuals on the basis of their position or location within a surrounding structure.²³

(1) The theory of the firm is a paradigmatic example of an explanatory rational-choice theory that relies on a structural theory of

²⁰ Of course, it is also possible to employ comparative statics methodology to assess the effects of mental shifts on equilibrium outcomes—we can ask what would happen if people came to prefer apples to oranges more than they now do? But, we can interpret such hypotheses either internally (as a shift in preferences or beliefs) or externally (as a shift in “preferences” or “beliefs”). From the external perspective, the explanatory statement simply connects some observable choices to equilibrium outcomes, and entails no statements positing a causal relationship between psychological entities and behavior.

²¹ Cf. E. Sober, *The Nature of Selection* (Cambridge: MIT, 1984), for a discussion of such explanations.

²² Perfectly competitive markets are structures that operate at a different level from individual utility functions. In a perfectly competitive market, strictly speaking, no finite number of agents can influence or deviate from the parametric price structure. Notice that in this case a structural explanation cannot illuminate how this institution or price structure was established in the first place.

²³ Specifying an agent's interests is a complicated matter since we cannot simply define an agent's interests in terms of her existing preferences. Moreover, individuals do not only pursue their own interests, they also look to the interests of others. We shall argue that there are at least some candidate theories of interest that do not seem to rest intrinsically on internalist (psychological) foundations. These theories of interest can not only be completely described in terms of externalist coherence requirements, but also they are causally related to features of the environment of the actor. These theories will include what we have called social theories—theories that construct individual “preferences” out of the social environment.

interests.²⁴ The assumption that the neoclassical firm acts to maximize profits does not apply to the firm in virtue of its internal beliefs and desires. Rather, an environment of competitive capital markets acts as a selector on firms. In a competitive market, only firms that act to maximize profits will survive. The firm's interests may arise from features of its environment and need not arise from any particular psychological facts about its members.

The fact is that profit-maximizing behavior is compatible with various internalist causal stories: profit maximization can be multiply realized. An other-worldly interest in eternal salvation or a crude materialism, or a propensity for innovation and risk taking, are all compatible with such behavior.²⁵ Many sets of individual motivations are compatible with the constraints that competitive market environments place on a firm's behavior. In explaining firm behavior, we often confront causal patterns that hold constant across the diverse realizations of maximizing activity found in Calvinist England and the Savings and Loan community in Texas.

(2) Compare the theory of voting behavior and the theory of party behavior within the theory of electoral competition. In the theory of party behavior, the results are obtained in a purely deductive and theoretical manner. Within plurality-rule systems with two electorally motivated parties, the classical theory of electoral competition makes strong predictions as to what the parties will do in equilibrium. Moreover, within such systems competition among parties for valuable offices encourages them to be relatively electorally motivated, since nonelectoralist parties will tend to remain out of office and be unable to reward their supporters. In this sense, the competitive environment encourages the adoption of electoralist objectives.

For voters, the situation is quite different. How or whether she casts her vote has little effect on the well-being of the ordinary citizen. Electoral outcomes are the results of millions of similar deci-

²⁴ Of course, the neoclassical theory of the firm is now largely regarded as inadequate. More recent work on firms emphasizes the importance of intrafirm organization to a firm's behavior. Rather than view the firm as an "atom," recent work directs attention to the firm as a complex of interacting agents, resources, and information. Yet, in many of these recent approaches, the parts of the atom can still be understood structurally. For example, M. Jensen and W. Meckling argue that corporate management is disciplined by the profit motive (via the need for an efficient capital market); "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, III (1976): 305-60.

²⁵ Cf. Max Weber, *The Protestant Ethic and the Spirit of Capitalism* (New York: Scribner, 1958).

sions, after all, and while (in plurality rule, two-party systems) voting for a dispreferred candidate is a dominated strategy,²⁶ the consequences of so doing are negligible to the ordinary voter. Not surprisingly, therefore, there are no strong reasons to believe that voters will eschew such actions. Moreover, there do not seem to be any competitive forces that would act to shape or constrain the kinds of preferences that ordinary citizens hold. There is no comparable result in the study of voter's behavior. As a result, the theory of voter behavior is a weak predictive theory.

(3) In *The Moral Economy of the Peasant*, James Scott²⁷ derives the peasant's conservatism and aversion to risk from his precarious position—the result of shortages of land, capital, and alternative opportunities. Scott likens the peasant's position to that of an individual up to his neck in the water, where any sudden movement might drown him. Much of the peasant's behavior is then explained by this position. For example, where introducing new farming techniques and crops increases the risk of failure, and therefore threatens a subsistence crisis, Scott argues that peasants will eschew innovation. Rather than maximize wealth or return, Scott argues that the peasant attempts to minimize risk. This risk-minimizing strategy explains many of the peasant's apparent preferences: his preference for crops with low yield which are reliable over those with higher yield which are not; his preference for food crops over cash crops; his preference against specialization. Scott also attempts to explain peasant rebellion and revolution as a "rational" response to the threat of falling below a subsistence minimum.²⁸

In each of these three examples, we could give an internalist account of these phenomena. We could proceed inductively, examining each agent's preferences and beliefs.²⁹ But certain features of the equilibrium explanation are not captured at the microlevel of agent psychology. In particular, the stability of the equilibrium explanation is lost: the inductivist psychological explanation does not illuminate why the same results obtain across putatively different

²⁶ Strategy *X* is said to be dominated by strategy *Y* iff for all possible states of the world, the outcome of *Y* is at least as preferable as the outcome of *X* and at some state of the world the outcome of *Y* is strictly preferred.

²⁷ New Haven: Yale, 1976.

²⁸ There are also problems with Scott's explanation. In particular, he has no theory of aggregation: What ensures that individually rational preferences will yield a collectively rational outcome? For a telling criticism, see S. Popkin, *The Rational Peasant* (Berkeley: California UP, 1979).

²⁹ Furthermore, the Scott example can be viewed as an externalist explanation for the peasants' actual internal states.

microlevel psychologies. In the equilibrium explanation, the stability is explained structurally. Structural conditions constrain and narrow the number of psychological possibilities. These conditions select for compatible microlevel psychologies.

We believe that rational-choice theories are most credible under conditions of scarcity, where human choice is severely constrained. In environments without strong constraints,³⁰ agents will not generally behave as the theory predicts. All of the cases in which rational-choice theory is a strong predictive theory are ones in which the theory is relying on interests that are determined by features of the agent's environment. But in such circumstances, it is best interpreted as an externalist theory, where claims about the actual psychology of its agents are temporarily bracketed.

The crucial paradigm for our discussion is evolutionary biology,³¹ which views nature as a selective structure. The structure of nature selects types with certain properties: those who lack those properties do not reproduce. Although these properties are not necessarily the consequence of the intentional states of the organism, they can often be described in a decision-theoretic way. We can predict the behavior of an organism by assuming that, within constraints, it will behave in ways that will maximize its expected reproductive output.³² We think that there are social analogues to natural selection and that much of rational-choice theory operates in the context of powerful selective mechanisms.

None of this is meant to dismiss internalism. For certain kinds of questions, an internalist perspective is crucial. The normative purposes of rational choice can not be addressed by the externalist theory. Externalism cannot address questions posed at the level of a single individual's behavior or those questions which concern the origin of a certain structure. But such questions are only a part of the domain of social-science inquiry. Many questions in social science require an externalist interpretation of rational-choice theory.

(b) *Objections to structuralism.* Suppose that we are right and that for many explanatory purposes it is most plausible to interpret rational-choice theory as an external theory. It might still be implausi-

³⁰ Bratman's work suggests that plans play a role in practical reasoning such that they provide internal constraints on an agent's available actions; *Intentions, Plans and Practical Reason* (Cambridge: Harvard, 1987), ch. 3, n. 12 and pp. 32–7.

³¹ There are, of course, many disanalogies between biology and economics, as Alan Nelson has forcefully reminded us. Equilibria in economics are idealizations—often with almost no empirical support.

³² Thanks to Peter Godfrey-Smith for discussion about this point.

ble to think of rational choice in terms of structural explanations because, according to some critics, structural explanations are not good explanations. If rational-choice theory is not cast in terms of the semantics of individual decision, then despite bypassing criticisms of its psychological unreality, it stumbles in front of criticisms about its form of explanation. Methodological individualists claim that all valid explanations in the social sciences must ultimately be given in terms of individuals; more particularly, in terms of the desires, beliefs, and opportunities of individuals.³³ The general reasoning behind this claim is as follows:

- (1) All social phenomena are in principle explicable in ways that involve only individuals.
- (2) Individuals are more basic and fundamental to an explanation than structures.

Therefore,

- (3) Where social science is in a position to make a reduction to the level of individuals, it ought to do it.

Many arguments have been given on behalf of premises (1) and (2). While we cannot address all such arguments here, we hope the following remarks will clarify the sources of our skepticism about the above argument's conclusion.

The first premise follows from a commitment to some form of reductionism. In its classical formulation, reductionists assume that there is a hierarchy of sciences where the objects at a given scientific level are composed of the objects at the next lower level. They propose that the higher-level science's laws and objects can be completely derived from the lower-level science, that, for example, chemistry can be derived from physics.³⁴

There are several, well-known difficulties with this position. For example, higher-level laws cannot always be *derived* from lower-level laws. This point has often been illustrated with the property of "fitness," a property which is central to the formulation of laws in biological theory. The property of having a certain level of fitness has many physical realizations; it is multiply realizable.³⁵

The possibility of multiple realizability has led some to adopt a

³³ Some of the most powerful arguments for this position have been given by J. Elster, *Making Sense of Marx* (New York: Cambridge, 1985).

³⁴ Classical reductionists can be contrasted with reductionists who believe that the nonbasic objects do not exist and the nonbasic laws are false. Such reductionists are not interested in reducing by derivation, but rather in reducing by replacement or elimination. For an elaboration of this distinction between types of reduction, see J. Dupre, "The Disunity of Science," *Mind*, xcii (1983): 321-46.

³⁵ This example is adapted from Sober.

principle of “supervenience” in place of a commitment to strong reductionism. There are dozens of kinds of supervenience, but the idea generally is this: a property *X* supervenes on a property *Y* when it is the case that there can be no change in *X* without some change in *Y* (though not vice versa). Supervenience allows for an ordering relation among levels of science while admitting the possibility of multiple realizability. For instance, in the above example there may be multiple ways of realizing *X*.

Is supervenience true? We have no definite view on this matter. But recent work in the philosophy of mind suggests that narrow supervenience—in which one phenomenon narrowly or singly supervenes on another—may be too strong a thesis to hold about the relationship between phenomena in various domains. Tyler Burge,³⁶ for example, has argued that the content of a person’s thought depends on social facts external to the person. Thus, it might be possible for two persons to have identical brain states and think different things. Burge’s work suggests that supervenience between properties cannot be local—it must extend in space and time. Thus understood, supervenience is of little use to the methodological individualist. Even if we view social properties as supervenient on individual psychological properties, Burge’s work shows that characterization of those very psychological properties depends on the surrounding (social) environment. Our account of rational-choice theory is compatible with supervenience, but does not entail methodological individualism.

Many philosophers would probably agree that there are properties that are not locally reducible to the properties of monadic individuals. Nonetheless, they remain committed to make reductions where possible. Is it necessarily good social science to reduce?

This brings us to the second premise in the argument for methodological individualism: individuals are more fundamental than social structures, they are the rock-bottom of any explanation. Therefore, where we have an upper-level explanation that can be eliminated in favor of a microlevel explanation, we should always do so. We should, accordingly, replace external rational-choice explanations with internal explanations whenever we can.

We disagree. While, as Wittgenstein said, all explanations must come to an end somewhere, we see no reason why they all must end up in the same place. Whether individual or structural accounts of social phenomena are appropriate depends, we believe, on the purpose of the explanation. For some purposes, the appropriate focus

³⁶ “Individualism and Psychology,” *Philosophical Review*, xcv (1986): 3–45.

is on individual agency and choice. For many social-science questions, however, the appropriate focus is on how social structures and features of the agent's environment exert constraints on her action. This is not an a priori feature of social-science explanation. Rather, it is a feature of the types of questions in which social scientists are typically interested. In contemplating a reduction, we need always to focus on the question of what purposes our explanation is intended to serve.

Even when a reduction is possible, the microlevel and higher-level explanations are not the *same* explanation.³⁷ The level at which an explanation operates can be crucial to what the explanation is an explanation of. This point can be illustrated by considering Emile Durkheim's explanation of suicide in *Suicide*,³⁸ where he focuses on the question of why the suicide rate in a given country stays relatively constant. This rate, he argued, remained constant even when the motivation of individual suicide victims changed, or when the particular agents who fell "victim" to the "suicido-genetic" currents changed. Durkheim explained the suicide rate in terms of the number of social ties among individuals in society. He held that social positions with too few ties or too many would be associated with higher probabilities of suicide. A microlevel explanation of suicide loses sight of the level on which Durkheim's explanation operates. The appropriate explanation (for Durkheim's questions) is a highly abstract one: only an abstract explanation can isolate and focus on the interrelationship between social structure and individual vulnerabilities. The explanation of why a particular individual commits suicide cannot illuminate the general tendency of certain social-structural positions to produce individual vulnerabilities.

In addition, we do not think that higher-order social facts are always explained by individual behavior. In some contexts, it seems more plausible to claim that social structures are causing the individual behavior, rather than the reverse. Consider, for example, the question: Why do unemployed people tend to be uneducated? (Assume for the moment that the fact alleged in this question is true.) While we could provide an answer to this question in terms of each unemployed individual's "preferences," "assets," and "opportunities," there is another perfectly plausible response to this question. It is: "because there are few jobs available and the jobs that are

³⁷ For arguments along these lines, cf. A. Garfinkle, *Forms of Explanation* (New Haven: Yale, 1981); and F. Dretske, "Contrastive Statements," *Philosophical Review*, LXXXI (1973): 411–37.

³⁸ J. Spaulding and G. Simpson, trans. (New York: Free Press, 1951).

available tend to require a college degree.” In the second answer, we focus on the properties of the environment as a selector. We also say something important about the structural organization of society: if you want uneducated people to have jobs, then make more jobs available with lower entrance requirements.

A somewhat different argument in support of premise (2) concerns the possibility of spurious correlations. Because there are time lags between macrovariables—for example, between a corporation’s inputs at time t and its outputs at time $t + 1$ —we cannot be sure that we have a real as opposed to a spurious correlation. This is a serious problem for any structural theory because it is difficult to isolate structural causes. Comparative research may make us reasonably certain that we have isolated an important structural phenomenon, but our confidence would surely be enhanced if we had an actual microlevel pathway that was sufficient to produce this effect or phenomena.

We believe, however, that our agreement with this point does not commit us to embracing either individualism or reductionism in any significant sense. Rather, we are committed at most to a weak reductionist account: a macrolevel explanation must be compatible with some true microlevel account. Macrolevel mechanisms that cannot be realized on the microlevel are implausible since structures cannot have their effects regardless of individual behavior.

Yet even if we gain more confidence in our explanation where we have a microlevel account that “provides a continuous and contiguous chain of causal or intentional links,”³⁹ we have argued that in the social sciences we cannot simply *replace* a macrolevel explanation with a microlevel one. In the first place, there might be more than one microlevel explanation with which our macrolevel account is compatible. In the second place, the two explanations may not have similar explanatory power. What a given macrolevel explanation is explaining may be obscured or left unanswered when posed in terms of microlevel variables. In particular, for *explananda* where the processes that effect an agent’s desires and beliefs do most of the explanatory work, explanations that focus more or less exclusively on the choices of actions based on those desires and beliefs will have little explanatory force.

All the above could be granted by an internalist who is not a methodological individualist. We do not deny that the explanation of human behavior goes through agent psychologies, in some complicated manner. This in itself is not damaging to our account.

³⁹ Elster, *Explaining Technical Change* (New York: Cambridge, 1983), p. 24.

Rather, we are arguing that for many kinds of questions, the internalist interpretation of rational choice is inappropriate; it is often redundant or misleadingly concrete. Internalism cannot replace externalism for all explanatory purposes, and the philosophical criticisms directed against the internalist interpretation of the theory are not damaging to the theory under its externalist interpretation.

III. CONCLUSION

We have argued that rational-choice theory is compatible with structuralist theory. We have done so by attempting to show that rational-choice theory need not rest on either psychological or individualist foundations, and therefore that it is not, as an explanatory enterprise, intrinsically hostile to social theories that offer explanations in terms of structural properties. In developing our argument, we have offered two perspectives on rational-choice theory—an external one, which explains action merely by appeal to coherence restrictions on choice; and an internal one, which explains action by appeal to real or “real” mental properties.

In our view, rational-choice theorists can proceed in two directions. They can develop an internalist interpretation that identifies an agent’s concrete preferences in particular contexts (and the consistency conditions associated with them); or they can develop an externalist theory that tells us how to infer from structural parameters an agent’s “preferences” and any conditions associated with them. We do not endorse the externalist approach for all purposes. Yet in many contexts, we do think that the externalist approach to understanding an agent’s interests provides the most useful way to proceed for explanatory purposes. Externalism can illuminate structural relations and causes. Moreover, a variety of individual-level mechanisms can be compatible with structural explanations: natural selection, cybernetics, and convention as well as choice.

Our thesis in this paper has been that the link between rational-choice theory and individual psychology is an open one. On our view, rational-choice theory carries no a priori commitment to a particular causal mechanism. We conclude that it would be a mistake to adopt exclusively any single interpretation of rational-choice theory: how the theory is best interpreted depends on the questions it is addressing, and the circumstances in which action is being viewed. Much of what we have written emphasizes the importance of understanding at least some rational-choice explanations as structural explanations. In particular, we have argued that some of the strongest rational-choice explanations we have at present can be best thought of in this way. We have already alluded to the assumption of profit maximization in the neoclassical firm. But profit maxi-

mization is not best understood as an assumption of rational-choice theory; rather, it is the result of supplementing the theory with an externally derived theory of interests. Another example would be the classical theory of electoral competition which rests on the assumption that candidates are motivated to seek office rather than pursue policy goals. In both cases, agents are operating in a competitive environment characterized by extreme scarcity, in which it is plausible to impute interests to the various acting parties. In other cases, where there are fewer environmental constraints, rational-choice theory does poorly at predicting human behavior. The theory of consumer behavior is a notoriously weak theory in this sense; any attempt to predict the choice of a specific marriage partner will also fail.⁴⁰

There is an irony to our account of rational-choice theory. If we are right, then the theory is most illuminating where it can be supplemented with an external theory of interests, where "choice" is seriously constrained. At the same time, we recognize that for many purposes, an internalist, agent-centered perspective on action is crucial. It makes a great deal of difference for normative concerns whether we think people are actually maximizing something (for example, utility) or just behaving as if they are. In fact, we need a background theory to identify in just which contexts a psychological interpretation of rational-choice theory makes sense.

It is important to notice the two possible ways of fleshing out the theory and not to privilege mistakenly the individualist and psychological interpretation of the theory. There is a tendency for rational-choice theory practitioners to read off truths about human psychology from the "successful" uses of the theory. But the behavior of rats, pigeons, hunter-gatherers, and stockbrokers may not tell us much about the nature of our rationality or our moral psychology. It tells us more about the structure of certain environments, environments we hopefully shall not attempt to duplicate in other areas of our social life.

DEBRA SATZ

Stanford University

JOHN FEREJOHN

Stanford University

⁴⁰ For an alternative view, see G. Becker, *A Treatise on the Family* (Cambridge: Harvard, 1991).