The Reasonable and the Rational Capacities in Political Analysis

PAUL CLEMENTS
EMILY HAUPTMANN

The authors employ Rawls's distinction between the reasonable and rational capacities to show why and how rational choice theory cannot provide adequate explanations of human behavior. According to Rawls, the reasonable capacity, associated with the concept of right and the sense of justice, is no less fundamental a moral power than is the rational, associated with the concept of the good and self-interest. Since rational choice analysis presupposes the primacy of rationality, however, those who rely upon it see persons' expressions of conceptions of right as expressions of rationality. The authors argue that in cases ranging from prisoner's dilemma experiments to the analysis of social institutions, rational choice theorists encounter expressions of the reasonable but cannot, because of their theoretical commitments, take systematic account of them. The article concludes by making some tentative suggestions about the form political analysis based on both the reasonable and the rational capacities might take.

I. INTRODUCTION

For the past two centuries there has been no concept more central to Anglo-American political theory than that of rationality. Central themes in this tradition have involved our use of our rational powers to enhance the social good and to maximize our own individual utility. The mantle of this tradition is currently borne by rational choice theorists, some of whom view their theory as providing the proper analytic core for political science overall.[1] There is indeed no theory that can rival rational choice theory in applications across the discipline.[2]

We believe this article has been substantially improved in response to comments and suggestions from David Plotke, Magali Sarfatti Larson, and the other members of the *Politics & Society* editorial board.

POLITICS & SOCIETY, Vol. 30 No. 1, March 2002 85-111

In *A Theory of Justice*, however, John Rawls dissents sharply with the rational choice tradition. Our understanding of justice, he argues, is better grounded in the concept of right than in that of the good.[3] In developing this argument, Rawls builds on a Kantian conception of the person that is deeply at odds with that found in utilitarianism and its analytic cousins, neoclassical economics and rational choice theory. These disciplines see utility maximization, or the efficient promotion of our preferences or interests, at the heart of our social reasoning. According to Kant, however, we regulate the complex capacities that we call sensibility and understanding with maxims or principles that are fundamentally nonalgorithmic.[4] In *Political Liberalism*, Rawls elaborates on this idea of the person in a discussion of two moral powers: the reasonable and the rational.[5] He argues that these are the capacities central to our political and social reasoning; that they are distinct, but they work in tandem, so one cannot be properly understood without the other.

This article presents a critique of rational choice theory and suggests how political analysis based on this Rawlsian conception of the person might be conducted. If the reasonable capacity is indeed as fundamental as the rational, any rational choice theorist who departs from the cool territory of deductive logic to analyze actual social relations must take some account of it. We show how this is generally accomplished on an *ad hoc* basis,[6] by admitting a notion of fairness independent of rationality,[7] and/or by subjecting conceptions of right in the guise of social norms to the calculations of the rational self.[8] In any case, rational choice theory takes the central exercise of reason to occur in choices that maximize across a schedule of preferences. One finds, however, that even in situations artificially constructed to encourage maximizing calculation alone, choice is conditioned and often dominated by conceptions of right.

Acknowledging the reasonable capacity allows for better, more theoretically grounded explanations for certain behaviors, such as in prisoners' dilemma experiments, than rational choice theory has been able to provide. More important, however, are the analytic and investigative strategies that emerge from doing so, and the resulting conception of the social and political world. The reasonable capacity is manifest in principles or rules that condition our responses to external events and that can (in some but not all instances) be understood and altered by the person. Principles are empirical phenomena, structuring relations among persons and between societies and the material world. In this aspect they are subject to analysis in the same general manner and for the same sorts of purposes as any empirical phenomenon. At the same time, principles are constitutive of individual and social identity. In this aspect, the act of analysis takes on a subjective character and significance not possible for analysis oriented, say, to traffic patterns or calorie counts. Rational choice theorists have constructed a grand edifice that occupies many scholars' minds and guides the social reflections of many citizens. We argue that a Rawlsian construction could yield more securely grounded and more penetrating analysis.

The article proceeds as follows. We first discuss the rational and the reasonable as concepts and as cognitive capacities. Next we explore how scholars within and outside the rational choice tradition have come to grips particularly with the reasonable capacity in its various expressions, and the consequences of these strategies for their larger projects. We consider analysts of prisoners' dilemma experiments and rational choice theorists who analyze social institutions. This provides a basis for speculating about the contours of an applied Rawlsian analysis.

II. THE REASONABLE AND THE RATIONAL CAPACITIES

Rational choice theory has come to prominence through parsimony and by permitting a certain deductive rigor. Its rationally self-interested utility maximizing agent is not a full-bodied image such as the lay person observes in the mirror. In its native environment in economics, this agent makes no pretense of describing all economic transactions. Laws of supply and demand require that buyers generally prefer to buy low and sellers to sell high, but they do nothing to stop shopkeepers from giving special deals to their friends. Similarly in political analysis, the assumption of rational self-interest is not expected to be always descriptively accurate. It is intended rather to provide a fair approximation of most political activity and to serve well enough to orient political analysis over its range of topics.

We wish to offer an alternative to the microfoundations for political and social analysis found in rational choice theory's conception of the person. Weber notes that any consistently elaborated intellectual-theoretical or practical-ethical attitude "has and always has had a power over man," and we would argue that it is exceedingly difficult to maintain the assumption of rational self-interest as a mere assumption in the course of extended analysis.[9] It is all too easily projected onto its object. The source of our disagreement lies less in this assumption's strength, however, than in its adequacy for its subject matter. There is something fundamental about politics, another axis on its plane, that utterly escapes rational choice theory. This can be seen, for example, in the limited place of ideas of justice in the theoretical space mapped by rational choice. It is our thesis that the distinct cognitive capacities associated with the concepts of right and of the good ought to be conceived as embedded in one another but operating according to different principles.

We have noted that the central cognitive act for rational choice theory is the choice that maximizes across a schedule of preferences. Here the task of rationality is to assess the range of alternatives that a situation of choice presents in terms of how far each would advance the agent's various interests. (Utilitarianism notoriously recommends the "util" as a unit in which diverse interests can be compared.) This act is a species of consequentialist reasoning involving (1) the identification with more or less clarity of a set of possible futures, based on (2) some notion of the causal relations by which one's actions can influence the state of the

world, and (3) an assessment of how imagined futures satisfy one's interests or preferences. At its most primitive, this model collapses into desire fulfillment. The baby searches for the breast and, feeling hungry, sucks. Full blown rationality, however, clearly involves many complex mental operations. Desires and interests must be ordered and arranged. We must have implicit models as to how a given situation could yield various outcomes. We imagine how our own resources can be deployed. Each of these operations presupposes subsidiary mental abilities, all of which we take quite for granted in daily affairs.

Partly through the constant concatenation of the two concepts in economics, rationality has come to be associated with self-interest. Weber uses the term "rational" to refer to any systematic association of means to ends, such as in "the methods of mortificatory or of magical asceticism," and to the "increasingly theoretical mastery of reality by means of increasingly precise and abstract concepts."[10] Yet he also believes that an economic rationalism has come to dominate civic life in the West, and he takes the rejection of nonutilitarian yardsticks (as by Confucianism) to be a particular mark of a rationalist ethical system. The basis for his categorization of utilitarian yardsticks as "rationalist" appears to be functional; it seems that Weber would consider the idea that one set of ends could be more substantively rational than another to be a kind of category error.[11] Economic rationalism is particularly rationalistic because, more than other ethics, it supports a thoroughly systematic perception and analysis of social relations (e.g., with calculus).

Rawls, like Weber, identifies means-ends reasoning as the primary meaning of rationality. He also sees the rational capacity at work when agents "balance final ends by their significance for their plan of life as a whole, and by how well these ends cohere with and complement one another."[12] For the rational choice framework, this balancing of final ends is prior to and constitutive of the elaboration of a schedule of preferences. Agents reflect rationally on their own purposes based on knowledge of the self, of the constraints inherent in human vulnerability, and of rights and opportunities that the current social milieu affords.

We are all familiar with means-end reasoning. When we say that someone is very good at solving a kind of math problem, fixing a car, or predicting how a recipe will turn out we are acknowledging this particular form of mastery. The idea of progress includes (in part) our increasing mastery over nature (or our learning to cope with nature's limits) through improved means-end reasoning as a kind of collective human project. Yet the centrality of means-end reasoning by itself provides no basis for parsimonious or deductive social theory. Rational choice theory achieves its analytic purchase only by adding the assumption of self-interest—that this reasoning is guided primarily by the desire to enhance one's own wealth or power.[13]

If, for purposes of political analysis, we retain the assumption of rationality but drop that of self-interest, then we have left what we get by generalizing this capac-

ity across a population of interacting humans. Each society consists (in part) of a population of persons with a distribution of ends, or ideas of the good, that they seek to promote, at times, through politics. In promoting their ends, persons employ means-ends reasoning that is itself a social construction. Forms of rationality in any society are constrained to solve basic problems of production and defense—to secure goods such as food and shelter that are needed for whatever ends. We can say with confidence that some populations have more effective reasoning in favor of particular ends. In light of the great variety of ends that contemporary and historical societies present, however, it is clear that natural constraints do not restrict a population's selection of ends very tightly.

To see the distinct cognitive capacity that is the reasonable, it is useful to think of how an action strikes us as inappropriate, improper, or wrong. Watching children at play, we see one strike another. Someone jumps a turn-style without paying. A driver runs a red light. That's not right, it strikes us. Perhaps there are extenuating circumstances, but there is a sense of unease, of a violation. If so, this is a manifestation of what Rawls calls "the reasonable capacity." Suppose I try at something, to get or achieve something I want, and perhaps through bad luck, I fail. There is disappointment. But suppose I expect to succeed and an adversary fiddles the game or misleads the jury. Besides disappointment, there is resentment. As disappointment is evidence of the capacity for a sense of the good, so is resentment evidence of the capacity for a sense of right.

When a particular exercise of political power strikes us as legitimate—or not—this too is evidence of a conception of right established in our minds. As we contemplate some assertive act, we sometimes justify it to ourselves. We see the interest we seek to advance, and we see the terms of our relation with another person or other persons expressed in the act. To affirm these terms and act is to exercise the reasonable capacity, whether it is a teacher giving a failing grade, a slave-owner punishing a slave, or a protester chanting slogans.

It is not just that the reasonable and the rational capacities are both fundamental to our political and social reasoning. According to Rawls (or Kant) they are the fundamental moral capacities. Our political reflections take place in their terms. The political and social world consists substantially (for better or worse) of our reasonable and rational constructions. This is not to say that other cognitive capacities such as the emotions or creativity are not politically important. The emotions can of course be powerfully motivating, and creativity is essential to finding new reasonable and rational solutions to practical problems. As we conceive of our problems and of strategies for solving them, however, as we carry out daily routines that we have found satisfactory, it is the reasonable and the rational capacities that provide the latticework on which we absolutely depend.

The reasonable capacity is constituted by our (closely related) senses of justice, fairness, legitimacy, appropriateness, and propriety. These are words we use to represent various applications of this capacity. To have such a sense is to have a

mental construction that yields those responses. These senses bear a family resemblance to one another but not to our sense of self-interest or to the form of reasoning by which we determine how to promote our ends. In contrast to maximizing across a schedule of preferences, these senses operate in the manner of the application of a principle or rule. When we experience the sense of unease discussed above, the way to articulate its cause is to identify the principle that has been violated in the event we have observed.

When Rawls discusses the reasonable, he emphasizes its positive potential. He describes persons as reasonable when they "are ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so."[14] Thus another manifestation of the reasonable is the disposition to act fairly. When we are moved by a sense of fairness, we are not seeking to promote our own interests, although it may indeed be in our interest to have the disposition to be so moved. Rawls states that reasonable persons "desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept," but this is not to suggest that reasonable persons must have articulated this desire *per se*.[15] This is a better way to understand their motivation than to understand them as seeking to promote their own interests or the general good.

Although Rawls points out a rich and positive manifestation of the reasonable that is essential for his theory, we must emphasize that as a capacity, its range of operation is much broader than this. Persons exercise the reasonable whenever they propose, justify, or assess terms for cooperation, even when these terms are informed by conceptions of right that we find abhorrent. This power is exercised by slave owners invoking the superiority of their race and by inquisitors invoking the will of God as they condemn heretics to the flames.

We understand our relations with others in terms of principles, but these may or may not be articulated as such. This is the basis for the promise and the challenge of a Rawlsian political analysis. Principles (rules) can be identified at various levels: as unarticulated cognitive patterns, as articulated personal principles, and as unstated or stated shared principles. Shared principles may be mere agreements, promises, formal rules that constitute an organization, norms accepted by a community, or laws that a political unit is committed to enforce. Operative principles can be more or less completely acknowledged or recognized, and they can be recognized in different forms by different parties. Every operative principle has a history in the practical problems it has solved or helped to solve.

A challenge that is particularly keen when there is no explicitly articulated principle at hand, therefore, is to identify the salient principles in a given social context. While there may be some use in making lists of principles (e.g., in contexts x and y), it is important to delineate the range and limits of their applications, how they were established and how they are reproduced, and their empirical consequences, for example, compared to other possible principles.

III. MORALITY, JUSTICE, AND POLITICAL ANALYSIS

Neither Rawls nor Kant is engaged in political analysis of the world at large in the first instance. To see how a Rawlsian conception of the person can be applied in political analysis, it is useful first to discuss its employment in its original habitat in the work of these philosophers. Rawls's project is to develop a conception of justice appropriate for a modern, reasonably well-off liberal state. Kant's corpus is wide ranging, but his distinctions between the categorical and the hypothetical imperative and between pure practical reason and empirical practical reason, from which Rawls draws his notions of the reasonable and rational capacities,[16] are found mainly in his work on the nature of moral reasoning. The very possibility of social justice for Rawls rests on the reasonable capacity, as for Kant the possibility of morality depends on categorical imperatives and pure practical reason.

When Kant argues that we employ maxims or principles in our reasoning, he is making an empirical assertion: we reason in terms of rules. (Alternatively, the idea of a rule provides a fair characterization of terms in our reasoning.) Thus the baby's rule, although not understood as such, is to satisfy his or her hunger. A hypothetical imperative, in Kant's lexicon, is just such a rule, one for which an action is a means to something else.[17] Therefore the rational choice schema of maximizing across a schedule of preferences exemplifies the hypothetical imperative. We act to advance our interests, to achieve our goals. A categorical imperative, by contrast, is a rule that conceives of an action as good in itself, by virtue of satisfying a principle. It is a categorical imperative that affirms an action (our own or another's) because it is right, fair, or appropriate.[18]

Kant's distinction between pure and empirical practical reason makes the same point but with reference to the cognitive nature of our reasoning. Practical reason is empirical when it involves some external stimulus, at the present time or in expectation. Kant takes it that the external stimulus is somehow represented in our minds, say as pain or pleasure, such as in anticipation when one puts money in the bank in preparation for a future purchase. By contrast, the term "pure" for Kant refers to an activity of the mind that does not involve impressions from the senses. A categorical imperative is pure in this sense because its assessment of an action does not in the first instance refer to its consequences, but instead establishes its conformance to a principle already established in the mind.[19]

One may observe that all principles that satisfy the criteria for categorical imperatives involve the significance of an action both for ourselves and for other persons. This is clear from Kant's argument that all (morally valid) categorical imperatives can be deduced from one: "Act only on a maxim by which you can will that it, at the same time, should become a general law."[20] It is no surprise that this, (the) categorical imperative, resembles Rawls's description of reasonable persons as ready to propose principles as fair terms and to abide by them willingly, although Rawls adds that one may consider whether others are likely likewise to do so. The important point for us is that Kant takes the categorical imperative to be

the supreme principle of practical morality. Without the capacity to think in terms of and to be moved by categorical imperatives, without pure practical reason, there could be no morality. For Kant a merely rational agent, lacking pure practical reason, can understand the meaning of the moral law but is unmoved by it; "to such an agent it is simply a curious idea."[21]

It is essential for Kant that principles (both hypothetical and categorical) can be objects of decisions. If this were not the case at least for hypothetical imperatives, we would be mere creatures of instinct. Thus it is to persons capable of pure practical reason—persons with a reasonable capacity—that Rawls offers his principles of justice. Given that we each have our own ends and resources are limited, the question arises as to how social cooperation should be organized. Every association is organized on some principles. They guide our actions on an ongoing basis, but they are most often invoked in situations of conflict. So the question can be rephrased, on what principles should conflicts be resolved? Or, to regress a step, if Rawls is right that justice is the first virtue of social institutions, how should principles of justice be identified?[22] To this Rawls answers that we should take into account a general knowledge of social conditions but not our individual circumstances; that's fair. If we do this, he argues that we will conclude that each person should have the most extensive liberty compatible with a similar liberty for others, and that economic inequalities should only be affirmed when they are to the greatest benefit of the least advantaged members of society.[23]

When we view persons as rationally self-interested utility maximizing agents, our political analysis will be oriented to the strategies these agents employ and the external factors that affect their strategic choices. When we take account of both reasonable and rational capacities, however, we envision a network of principles of social organization and a distribution of conceptions of the good. When we analyze political phenomena, we are still interested in agents' strategic choices, but we view these choices as embedded in and possibly altering a network of principles. We are interested in the material results, as some conceptions of the good are realized and others are not. We also want to understand the operative principles a society presents, their empirical consequences, and how these principles change.

IV. PRISONER'S DILEMMA GAMES AND PLAYERS' SENSE OF FAIRNESS

The prisoner's dilemma game appears to lend itself particularly well to making a case for the primacy of the rational power. According to game theoretic analysis, the structure of the game gives each player strong incentives to exercise his or her rational capacity to trump any desire to act cooperatively, especially when the game is played only once or when there is a relatively low probability of its being repeated. The outcome of the game, according to this analysis, is all the more dramatic since if both players respond to the strong incentives the game gives them to act rationally, both will end up worse off than if they had acted cooperatively. The

circumstances in which this is true can be represented by the matrix below, in which the payoffs are ranked $b > a > c > d$.

	Cooperate	Defect
Cooperate	(a, a)	(d, b)
Defect	(b, d)	(c, c)

If the game is played once without any communication allowed between the players after its payoff structure has been revealed, the dominant strategy (for both players) will be to defect, leaving each with the lowly payoff c, an inefficient equilibrium. This result seems all the more striking when one imagines the game played repeatedly, the "winner" being the player with the largest total payoff. Even in this version of the game, game theorists predict an inefficient equilibrium: "the unique equilibrium behavior involves confessing [defecting] at every period, even though confessing is no longer a dominant strategy."[24] This analysis seemed flawed to many social scientists, who insisted that "when players properly understood the game, they would choose to cooperate with one another and not confess [defect]."[25]

Experiments using the basic structure of the prisoner's dilemma began to be conducted in part to address such concerns. Using principally multiple-instance versions of the game, experimenters tried to isolate variables that would significantly raise rates of either cooperation or defection. A recent summary of all public goods provision experiments (of which prisoner's dilemma experiments are a special category) concludes that the only factors with "strong and apparently replicable" positive effects on cooperation were increasing the marginal per capita rate of return (for cooperation) and allowing communication among the players (a change that alters one of the basic premises of the game's design). Repetition of the game, economics training, and experience with similar games had "strong and apparently replicable" negative effects on cooperative behavior.[26]

Because of our concern with showing how the reasonable capacity ought to be understood as independent of, though related to, the rational, we focus here on how experimenters have analyzed the cooperative behavior of their participants. According to some, people cooperate simply because they do not understand the rules of the game or because they have stubborn altruistic preferences. For others, particular types of cooperation (in multiple instance games) can be a rational solution to the utility maximizing problem the game presents. To be sure, what it means to cooperate changes depending on how particular experimenters structure their version of the game. Still, it is generally the case that experimenters do not understand cooperative behavior in prisoner's dilemma experiments as an expression of people's desire to propose and abide by fair terms, but rather as, at best, a rational method for maximizing payoffs or, at worst, a mistake.

Experimental situations are highly artificial by design; they often purposely rely on participants who do not know one another and cannot communicate with

one another and therefore need not reckon with any significant consequences to how they behave during the experiment. We contend, however, that even in these highly artificial situations, analyzing expressions of the reasonable capacity independently of the rational helps make sense of what are otherwise puzzling results. By showing how the reasonable capacity can be used to analyze a situation apparently tailor made for rational choice analysis, we lay the foundation for showing how the reasonable capacity might be fruitfully incorporated into the analysis of social and political life.

We focus here on four distinct types of explanations of cooperative behavior in prisoner's dilemma experiments: one in which cooperation is understood as a sign that the cooperator has made a mistake about the rules of the experiment's game; another in which cooperative behavior is interpreted as arising from altruistic preferences; one in which (a type of) cooperation appears to be the rational solution to the maximization problem posed by a modified prisoner's dilemma; and finally, one that calls in people's desire to be fair to explain higher than expected rates of cooperation.

Cooperation As a Mistake

According to early prisoner's dilemma experimenters, significant instances of cooperative behavior on the part of participants were, above all, a symptom of poor experimental design. Early experimenters criticized one another for allowing too much communication between subjects or for making division of the game's proceeds into fair portions too obvious.[27] On this view, cooperating or being concerned with a fair outcome (rather than a profit-maximizing one) is a cognitive cop-out; presenting people with an experimental situation in which the path to a fair, fifty-fifty division of the profits is readily apparent invites them to choose such a division over the less obvious but more profitable route. If the experimental situation were structured to discourage such thoughtless cooperation, early experimenters expected cooperative behavior to decline considerably, if not disappear.

After some forty years of prisoner's dilemma experiments, however, most commentators would probably concede along with one recent analyst that "[h]ard-nosed game theory cannot explain the data."[28] People cooperate too often in too many experimental settings specifically designed to reduce "mistakes" about how to play the game for cooperation to be dismissed as a mistake about the rules of the game alone.[29]

Cooperation and Altruistic Preferences

Another way to explain cooperative behavior by participants in prisoner's dilemma experiments is to refine the assumption that people are rational util-

ity-maximizers into the claim that for some, their utility depends directly (and positively) on the "payoffs" of others. Once the possibility that some people may have altruistic preferences is introduced into the analysis of prisoner's dilemma experiments, the way even purely self-interested players ought to approach the game changes. In a repeated prisoner's dilemma in which purely self-interested players believe that other players might have altruistic preferences, defecting at every move is no longer the preferred strategy; instead, to maximize his own payoffs, a purely self-interested player should adopt the strategy of cooperating until the final round of the game (and defect only then) or of defecting only after the other player does so.[30]

Rational choice analysts can readily explain how cooperation makes sense for self-interested players maximizing their utility when they suspect that other players might have altruistic preferences. But the altruists themselves, those with an apparent preference for cooperation, remain a puzzle. That some people may "care directly about the payoff of the other player" changes the dynamic of the whole game for all its players;[31] nevertheless, those who have such preferences are still, according to game theoretic analysis, "irrational" or "silly."[32] Altruistic preferences seem so puzzling from the perspective of rational choice theory that a number of experimenters are genuinely stumped by the problem of how to set up a situation in which such preferences could be overridden or rendered irrelevant.

Admitting that for some people, "maximizing utility" means maximizing some combination of their payoffs and those of others seems like a better way to understand cooperative behavior than to see it simply as a failure of rationality. Nevertheless, if one believes that the best way to explain why some people cooperate is by spelling out their altruistic utility functions, then one implies that what distinguishes such people from others are merely their odd tastes rather than their moral perspectives or conceptions of right. These people "get additional utility from mutual cooperation," a satisfying but to others inexplicable "warm glow" that somehow makes up for the lower payoffs their cooperative behavior reaps for them.[33]

The trouble with altruism so conceived is that it becomes the exception that proves the self-interest rule. Even those who criticize economic explanations of altruism often define altruism so narrowly (as Kristen Monroe does when she says that altruism is "action designed to benefit another, even at the risk of significant harm to the actor's own well-being") that rational choice theorists can easily maintain that the rational pursuit of self-interest is nevertheless the norm.[34] And if one conceives of altruism as an approach to maximizing utility, altruists appear unusual because they insist on regarding things, like others' payoffs and others' cooperation, as benefits to the self when, according to the theory, there is no *prima facie* reason to regard these things as such. The concept of altruism performs the function in much of rational choice theory of reinforcing the view that acting on one's self-interest, narrowly construed, is the royal road to maximizing utility.

From the analytic perspective we propose, it is misconceived to understand people's desire for systems of mutual cooperation solely in terms of the benefits they believe they will derive from them. People seek to cooperate with others because they have a desire, related to but independent of their desire to maximize benefits to themselves, to live in a world they believe is fair and just. Of course, establishing systems based on fair cooperation promises benefits far beyond an ephemeral "warm glow"; yet Rawls's distinction suggests that to focus on the benefits people expect to gain from cooperation alone or to conceive of a desire to cooperate as a preference will give us only a partial picture of why and when people seek to cooperate.

Limited Cooperation As a Rational Strategy

One of the most widely known series prisoner's dilemma experiments, reported in Axelrod's *The Evolution of Cooperation*, demonstrated that a limited form of cooperation could be the basis for a successful strategy in a multiple-instance prisoner's dilemma game.[35] The participants in Axelrod's study were not the usual college students, but rather people with expertise both in the prisoner's dilemma game as well as computer programming; all participants designed programs they believed would score well against the entire array of other programs during repeated plays of the game. The winning program (that is, the program that reaped the highest total payoff at the end of the game) in the round-robin tournament used a simple "tit-for-tat" or "copy-cat" strategy; it cooperated in the first instance and then copied the move its opponent had made in the preceding round thereafter. To call the strategy "cooperative," therefore, can be misleading since how often a player using this strategy cooperates depends (after the first instance) on what his or her opponent does. Nevertheless, Axelrod's results are often cited to show that in some instances cooperation is rational, even in a situation so apparently structured to discourage it like the prisoner's dilemma. Axelrod himself, however, is not so sanguine about the conclusions one can draw from his results; indeed, he suggests that tit-for-tat's success was an artifact of the presence of many poorly conceived strategies in the tournament: "Had only the entries which actually ranked in the top half been present, then TIT FOR TAT would have come in fourth after the ones which actually came in 25th, 16th, and 8th."[36] None of the strategies that would have come in ahead of tit-for-tat in such a revised tournament were "nice"; that is, each defected before its opponent did.

While the tit-for-tat strategy's success in Axelrod's tournament is intelligible on the basis of its rationality alone, it is by no means an obviously intelligible rational solution to the problem posed by the game. When presented with the rational problem of developing a strategy that would win the prisoner's dilemma tournament, nearly all participants in the expanded second tournament thought tit-for-tat could still be improved on and exploited, even though it had won a smaller tournament. If the rational, utility-maximizing merits of reciprocity and

cooperation escape most of those skilled in weighing courses of action according to these criteria, it seems unlikely that the best way to understand why most people cooperate when they do is that they believe doing so to be in their rational self-interest. Although we may be able to rely on rational choice analysis to represent cooperative equilibria in some situations, rational choice analysis does little to help us understand what moved people in those situations to act cooperatively. As much as Axelrod believes people could be taught to be more cooperative by being taught how it is often in their self-interest to be so, he does not claim that most people act cooperatively on the basis of this insight now.[37]

Concern for Fairness

The overwhelming majority of prisoner's dilemma experimenters focus on whether cooperative behavior makes sense as a means to the end of maximizing players' utility. The sociologists Gerald Marwell and Ruth Ames, however, called in people's understanding of and concern for fairness to explain experimental results unintelligible under their initial hypotheses.[38] Marwell and Ames set out to determine whether experimental results confirmed the following theoretical claim: in situations in which each individual's interest in the provision of a collective good is less than the cost of the good itself, contribution toward the purchase of the good will be essentially zero (the free-rider hypothesis of the theory of collective action). Although their results confirmed a weak version of the free-rider hypothesis (enough free-riding happens to prevent groups from being able to purchase optimal levels of collective goods), Marwell and Ames showed that in numerous differently structured experiments, participants consistently contributed "between 40 and 60 percent . . . of their resources . . . to the provision of a public good."[39] The experimenters note that they did not expect these results; indeed, they believed that they had so pared down their initial experimental conditions as to "maximize [the] effect [of] the free-rider problem" and, by implication, to occlude "normative factors."[40] But once they saw that their results, even in these conditions, did not confirm a strong version of the free-rider hypothesis, they asked their participants what they considered fair contributions to the public good as well as how concerned they were with being fair. In sum, Marwell and Ames concluded that people's responses to these questions about their conceptions of fairness make much better sense of the levels at which they contributed to the public good than the free-rider hypothesis.[41]

Economists prove to be the strongest exception to the implicitly reasonable rule suggested by Marwell and Ames's results. In a series of experiments under different conditions with different populations, Marwell and Ames replicated their 1979 results—except with a group of first-year graduate students of economics. Although only two of the thirty-two students in the experiment could "specifically identify the theory on which [the] study was based," the mean percentage of private goods contributed to the provision of the public good was markedly lower

among these students than among any other group (20 percent versus 40 percent to 60 percent).[42] Perhaps most telling for the argument we develop here are the responses of a number of these students to the experimenters' questions about fairness; Marwell and Ames reported, "[m]ore than one-third of the economists either refused to answer the question regarding what is fair, or gave very complex, uncodable responses. It seems that the meaning of 'fairness' in this context was somewhat alien for this group."[43] This one specific experimental result might serve as an emblem for our critique as well as for the type of analysis it suggests: even in experimental situations designed according to the main tenets of rational choice theory, many participants (who are not economists) act in ways that the theory can only incompletely or tortuously explain. Only from the perspective of those who have deep intellectual commitments to the primacy of the rational does the reasonable capacity seem like an unnecessary analytic tool. Understanding people's independent desire for fair systems of mutual cooperation as a capacity that works together with their desire for their own good helps us make sense of how people act even in the artificial and hostile environment of the prisoner's dilemma.

V. RATIONAL CHOICE INSTITUTIONALISM AND THE REASONABLE CAPACITY

We have argued that the reasonable capacity contributes no less to our political and social decision making than the rational. If this is true, then any rational choice theorist who engages in social analysis must confront it. We have shown how commitment to the rational choice framework has made it hard for theorists to interpret reasonable behavior in the controlled experimental contexts of prisoners dilemma games. These experiments must, however, be in the service of social analysis in the world at large. In the following section, we consider certain kinds of social analysis that the Rawlsian approach suggests. Here we show how rational choice institutionalists have accommodated evidence of the reasonable capacity: by dealing with each manifestation on an ad hoc basis, by suggesting that principles and norms are determined by interests, or by building norms into the maximization model used to characterize our social reasoning.

It should be remembered that rational choice theory's particular contribution to political science lies in the microfoundations it offers for broader social analysis. Analysts less concerned with microfoundational unity often simply accept principles or ideas of fairness alongside the rational pursuit of interests. For example, March, a sociologist, writes of a logic of consequences associated with rational choice and a logic of appropriateness, "by which actions are matched to situations by means of rules organized into identities. . . . Neither preferences as they are normally conceived nor expectations of future consequences enter directly into the calculus."[44] These logics describe two different approaches to analyzing decisions. The fact that they point to different microfoundations does not trouble

March. Rabin, an economist, suggests that economics and game theory could be extended by incorporating the idea of fairness.[45] This idea could be employed to explain cases in which people voluntarily contribute to public goods or punish "unkind behavior," such as high monopolistic prices, and in so doing fail to pursue their material self-interest. Rabin retains the basic assumption of rational self-interest but calls in the idea of fairness to account for certain behaviors not easily reconciled with this assumption.

To follow in the rational choice tradition, however, is to exclude compromises that keep fairness on hand such as those made by March and Rabin. Thus Bates, in his rational choice analysis of the political economy of agrarian development in Kenya, locates both economic and political sources of historical change firmly in material incentives.[46] He argues persuasively that in the postindependence period, Kenya had higher agricultural growth rates than nearby African states because Kenya's government was dominated by capitalist farmers (an incipient gentry). He predicts correctly, as events have transpired, that agricultural growth rates will decline under the government of Daniel arap Moi (1978 to the present), on the grounds that Moi's political power is based in the poorer western provinces.[47] Bates shows how incentives favor farmers with large holdings in forming institutions to overcome risk and uncertainty and thus to secure the resources needed for investment, yet also how owners of large fixed investments become subject to a politics of predation in the Kenyan context.[48]

Bates provides important empirical analysis of the Kenyan case, but his intentions are also explicitly theoretical. He argues that development economists of the traditional and neoclassical schools need to incorporate a theory of politics, and he offers rational choice institutionalism to serve this purpose.[49] Neither economic nor political interests can be determined outside a particular institutional context, so he employs a rational choice framework to explore the institutions that have been central to Kenya's political and economic development. In developing "the microeconomics of institutions,"[50] in each instance Bates focuses on two primary currencies: money and power. Indeed, whether he is explaining organizations of European settler farmers, Kenyan political parties, or the National Cereals and Produce Board, these incentives are clearly central.

Bates is so accomplished a rational choice analyst that he is able to confine most of his prose to the language of rationality (competition, interests, strategy, negotiation . . .). He does write of land rights and changing principles that validate claims to land, but he explains these changes in material terms.[51] We discuss in the next section how his theoretical commitments influence his choice of topics and the structure of his analysis.

The most striking episode in which the rational choice framework fails him is when he seeks to explain the Mau Mau rebellion that led British colonists eventually to abandon Kenya. Self-interest is adequate to explain why white settlers released large numbers of Kikuyu laborers from their service, and why these

laborers were denied the accommodation they expected from their kinsmen in the reserves. Yet the British were a greatly superior military force to the Kikuyu. Bates points out that in the course of the rebellion more than 14,000 Africans were killed or wounded, but he does not attempt to explain how material interests explain the decision to fight.[52] The key Mau Mau institution, he shows, was that of the oath, and radical politicians sought militants in "the reservoir of those who had lost out in the transformation of property rights in the reserves."[53] It is clear that this is a reservoir of resentment, but to acknowledge resentment as a significant political force exceeds the cognitive boundaries by which rational choice theory defines itself.[54]

Extreme cases like rebellions can illustrate how the reasonable and the rational are mutually embedded. Rationality is sometimes conceived as means-ends reasoning and sometimes as such reasoning oriented to self-interest. To join a rebellion can only be irrational in the second sense, as the first includes no independent criterion for judging ends. (If the aim was Kenyan independence, Bates argues the rebels succeeded.) On one hand the reasonable influences the interests one takes into account. To take an oath for rebellion in good faith is to sever a certain concern for self-interest, and in so doing, to transport the self into a modified emotional habitat and to transform the terms (at least from the rebel's side) of many relationships. There remains the rational question of how to conduct the rebellion, but a threshold has been passed. Since we cannot escape a concern for bodily integrity, this self-denial may have an ongoing psychological impact or cost. Certain emotional states may be more likely now, involving reduced awareness of or attention to interests we normally take for granted. Indeed all of this is to some extent understood and intended in the oathing. We can say without condoning it that resentment may be vindicated or expiated by acts of violence.

While Bates demonstrates the trajectory of a rational choice analysis of a country's political economy, Knight, in *Institutions and Social Conflict*, works on the basic theoretical foundations of rational choice institutionalism. Knight is particularly concerned to show how rational choice theorists should take account of power in analyzing the development of institutions, but in so doing he constructs theoretical foundations for analyzing all institutions.[55] Knight initially argues, like March, that there are two theories of individual action, one based on norms and the other on interests and rational choice.[56] Along the way to his conclusion, however, Knight takes the second approach we have identified by which rational choice theorists deal with norms; that is, he derives them from interests. His main argument is that (given rational choice assumptions) formal and informal institutions must be formed from our attempts to exploit one another. While we agree with the logic of this argument, the inconsistencies in his treatment of norms are symptomatic of limitations in the rational choice framework.

Dividing social theories into two sets, Knight notes that views based on norms stress how institutions allow us to reap benefits from cooperation, while those

based on interests emphasize how certain groups or individuals claim the larger share of these benefits. He does not believe that either of these sets of views is completely satisfactory. "It is reasonable to assume that both norms and rational calculations motivate action in different contexts."[57] Yet he argues that a theory of institutional formation and maintenance based on competing interests is superior to one based on norms.

> The theoretical justification rests on the claim that most social outcomes (at least those social and political outcomes about which we are most concerned in the social sciences) are the product of conflict among actors with competing interests. The rational-choice theory of action is better able to capture the strategic aspects of that social conflict. The practical justification is that the conception of institutional effects derived from the theory of social norms is less successful in explaining these outcomes.[58]

It lies beyond the scope of this article to assess Knight's characterizations of other theorists such as Smith and Marx; yet the notion that a theory of institutional formation must be based either on norms or on interests seems implausible. If one accepts that norms and interests are independently significant, it seems odd then to accept a theory based on a conception of the person as exclusively self-interested. Knight's reasoning can be read as an act of desperation; he seems to view rational choice theory as less inadequate for explaining institutions, say, than March and Olsen's organizational sociology.[59]

A Rawlsian view accepts the reality of deep social conflict, but it rejects the necessity that social theory must be based on norms or on interests. Rather, it takes both into account. It also rejects Knight's suggestion that actions can be divided into separate sets, one guided by norms, the other by interests. There are instances in which we consider it right to pursue our self-regarding interests; market transactions in a modern economy represent a typical case. There are also cases where conceptions of right lead to actions that appear contrary to a common-sense interpretation of rational self-interest, as illustrated by the Mau Mau rebellion. But actions involving great personal risk are unusual. In daily life our notions of fairness and our interests routinely condition one another; prisoner's dilemma experiments provide evidence of "normal" instances (beyond the inevitable abnormality of the experimental context) where people's sense of fairness leads to behavior at odds with what we would expect from rational self-interest alone.[60]

Knight defines institutions as "rules that . . . provide information about how people are expected to act in particular situations [and] can be recognized by those who are members of the relevant group as rules to which others conform in these situations."[61] He argues that institutions are formed out of the resolution of a "standard bargaining problem." It is a situation where

> there are benefits to be gained from social actors working together, sharing resources, or coordinating their activities in some way. These actors need rules to structure their interdependent activities. More than one set of rules can satisfy this requirement, and the rules dif-

fer in their distributional properties. Because of this, people have conflicting preferences regarding the institutional alternatives.[62]

Since he has excluded considerations associated with norms, such as conceptions of right or principles of justice, this problem leaves Knight in the domain of game theory where he immediately confronts the prisoner's dilemma. Yet this same problem is interpreted differently by Rawls. In his view, these conditions define the circumstances of justice:

> There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to try to live solely by his own efforts. There is a conflict of interests since men are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. Thus principles are needed for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These requirements define the role of justice. The background conditions that give rise to these necessities are the circumstances of justice.[63]

In Knight's view, when we encounter a situation that has distributional consequences, we seek the resolution that gives us the largest share. This is what rationality means to Knight. Rational choice theory gains rhetorical mileage by suggesting that what is not rational must be irrational or altruistic. In our view, Knight identifies one possible resolution and Rawls identifies another. People may be reasonable, and this is neither irrational nor altruistic.

Setting out to explain the decentralized emergence of informal institutions, Knight argues that institutions emerge as "a by-product of strategic conflict over substantive social outcomes."[64] Without trying to summarize the steps in his argument, its key feature is "*the fundamental relationship between resource asymmetries, on the one hand, and credibility, risk aversion, and time preferences, on the other*."[65] Yet these will only be the determining features if rational choice assumptions hold. Having initially suggested that some actions are driven by norms and others by interests, he now proposes to build "microfoundations for the informal network of rules, conventions and norms that capture some of the principal ideas of macro-level accounts in the Weberian and Marxian tradition."[66] Indeed it is inevitable that rational choice theory should reach some conclusion like this. Given the ubiquity of norms in social life any credible social theory must explain them somehow. Once the model of the person as a rationally self-interested utility maximizing agent is accepted, norms can only be derived from it.

When the reasonable and the rational are accepted as distinct cognitive capacities, there is no need to derive principles from interests. While principles can be determined by interests, it seems unlikely, on the face of it, that it could be shown that they must be determined in this way. Taking one's principles and one's interests as separate constructions, it is possible to start from either vantage point and reflect on the other. Kant considers principles to be more fundamental than inter-

ests, but for the purposes of political analysis there is no need to accept or reject this view.

A third approach to dealing with norms within the rational choice edifice, although it stretches the theory's boundaries, is to accept that norms are independently significant but then to incorporate them as a term in a maximization framework. "Dual utility" theorists—taking different kinds of utility to arise from satisfying interests and from satisfying principles—represent one variant of this approach. It is also developed with clearly specified theoretical constructions by Ostrom and by Crawford and Ostrom.[67] These theorists are close to our project in many respects. It seems likely that they would be sympathetic to locating the cognitive basis for conceptions of right, propriety, appropriateness, and so on in a single capacity. The basic difference comes from their employing principles and norms merely as terms in a calculus of maximization.

In *Governing the Commons*, Ostrom identifies a scenario in which principles are clearly significant. Common pool resources (CPRs) such as fisheries, pastures, and groundwater are resources owned by no one but used by many. Each individual has an incentive to continue drawing from or using the resource. CPRs typically have a threshold of maximum sustainable use, however, beyond which their regenerative capacity is undermined. If each individual continues to exploit the resource as self-interest dictates, it will eventually be spoiled for all. As for the convicts in the prisoner's dilemma, what is individually rational is collectively irrational. The resulting "tragedy of the commons" can be avoided if rules can be established that effectively govern rates of extraction. Yet rules are public goods, and people may be tempted to "free ride" by covertly taking more than their allotments.

While other theorists had focused on the state or the market as the means for avoiding overuse of CPRs, Ostrom points out that in many instances CPR users have established their own institutions and enforce their own rules for managing CPRs.[68] She explores several cases involving rights for underground water in Southern California to see how people have developed such institutions. The two problems are (1) for rules to be proposed and adopted and (2) for the rules to be followed. To solve these problems, Ostrom suggests that rational action involves an "internal world of individual choice" consisting of four variables: expected benefits, expected costs, discount rates, and internal norms.[69] In some groups, "few individuals share norms about the impropriety of breaking promises, refusing to do one's share, shirking, or taking other opportunistic actions," and in these instances expensive monitoring and sanction mechanisms are needed to protect CPRs.[70] Communities that develop norms involving high levels of trust and reciprocity, however, possess social capital, and these communities are more likely to succeed in building institutions that resolve CPR dilemmas.[71] Norms exclude actions that are considered wrong from the set of strategies that an individual contemplates.[72]

In this instance, although Ostrom literally subsumes norms within the idea of rationality, they have the categorical function of excluding actions from consideration. Later, however, as she discusses water-users' conformance to rules they have established, she returns to the maximization paradigm: "In any repetitive situation, one can assume that individuals come to know, through experience, good approximations of the levels of monitoring and enforcement involved."[73] The implication is that they would break the rules if they thought they could get away with it. Yet if individuals wish to comply because they think it right to do so, they may not reflect on levels of monitoring and enforcement. As studies of compliance with income tax rules suggest, expectations as to levels of monitoring and enforcement can be far from accurate.[74]

In a similar vein, to account for how users change their rules Ostrom asserts that

> individuals compare the net flow of expected benefits and costs to be produced by the set of status quo rules, as compared with an altered set of rules. To explain institutional change, it is therefore necessary to examine how those participating in the arenas in which rule changes are proposed will view and weight the net return of staying with the status quo rules versus some type of change.[75]

Rational choice theorists see such decisions resulting from a calculus of benefits, although the term "view" in addition to "weight" suggests a hint of Kantian flavor. From a Rawlsian perspective, one expects individuals also to employ rules or heuristics—conceptions of right—in altering rules, in modes of thought that are not, in the first instance, oriented to consequences. Questions of consequences are of course not abandoned in thinking about rules. Only it is possible for rules to be conceived, as March notes, in terms of a logic of appropriateness, or as Rawls argues for his "original position," in terms of conditions that express a relevant idea of fairness.[76]

VI. SOCIAL THEORY WITH PRINCIPLES AND INTERESTS

We can give only a brief sketch of some directions social theory that conceives of the person as reasonable and rational might take. Such a theory rejects the idea that principles, norms, and institutions can be derived from interests alone or from any schema of maximization. Taking principles and interests to constitute the basic building materials for our individual mental constructions of the social world, it aims to reveal the structural features of our collective constructions and their empirical consequences in different configurations and contexts. It views each society as presenting a network of principles and a distribution of conceptions of the good.

Rational choice theory takes each person to be a rationally self-interested utility maximizing agent; the same model works for everyone. Although the theory recognizes that some people are particularly adept at pursuing their interests, it

gives little attention to explaining differences. The real action is in explaining the consequences for such agents of different strategic environments—games for game theorists, institutional environments, perhaps with exogenous shocks, for institutionalists.[77] The theory admits changing preferences such as between income and leisure, but the main sources of change with which theorists work are external to the person. Individuals cannot but pursue their interests. When the environment changes, agents mechanically alter their strategies to find their new optimum. Politicians' policy choices, for example,

> depend upon the incentives generated by the institutional context in which they are made. Economic forces thus generate institutions and the structure of these institutions in turn shapes the way in which governments transform their economies. Economy and polity thus interact, generating a process of change. . . . In this way, each society generates its own history.[78]

A Rawlsian view does not imagine each person's construction of principles to be a creative individual act. Institutionalized principles are typically adopted with less than full awareness as we learn to navigate in society. New principles only become institutions when employed to solve practical problems. Nevertheless, a basic feature of a Rawlsian theory is that reflection and even philosophy contribute to social change.[79]

A Rawlsian macro-level analysis aims to identify (portions of) the configurations of principles that a society presents. Each organization consists of both formal and substantive principles, the former determining who has what powers and by what criteria, the latter being the principles manifest in incumbents' programs. While rational choice approaches are described as employing methodological individualism, the principles we are interested in for social analysis are those that are shared at least by politically significant populations. (Principles account for differences among rational agents, explaining why some respond to incentives in one way, some another.) Given that operative principles are often not clearly articulated in their daily employment, it is useful to look for conflicts and transitions when they are likely to be stated explicitly and when their employment can be contrasted with that of different principles addressed to the same problem. For example, central principles for any society determine how government offices are filled and who pays what taxes. One might examine reasons offered when a current pattern was established to replace an earlier practice.[80] To establish the empirical significance of a principle, it is useful to compare the case at hand to cases where essentially the same practical problem is solved with different principles.[81]

The fact that a principle provides the public justification for a certain exercise of power does not guarantee that this is indeed the operative principle. In the first place, multiple principles contribute to the solutions of most practical problems and the emphasis on one or another may vary over time. Second, as Kant emphasizes, we are not reliable reporters of the principles that inform our decisions. Sometimes we are not fully aware of, or prepared to admit, the reasons for our

actions. Third, someone may insist on a principle as a public pretense, aware that other principles guide action in the relevant body of cases. When principles are institutionalized, this kind of ambiguity is all the more significant. To gauge the strength of a principle that is also an institution, the analyst can identify phenomena suggesting alignment or congruence with it or depending on support from it. Since practical problems cannot be solved without recourse to principles (except problems that involve only one person, where principles may be applied only in the breach), to determine which are the most salient, the analyst can set up a sort of competition, considering alternative principles for the problem at hand, spelling out what follows from each, and building a case for and against them as the evidence permits. In many instances no determinate resolution may be possible; this reflects the ambiguity underlying much action in society. The analyst provides a service by delimiting the range of alternatives in play and clarifying the consequences of movement in one direction or another. Once again the comparative method is likely to be fruitful.

At the micro level, a Rawlsian approach has the analyst explore the participant's configuration of principles.[82] To explain the actions of a terrorist or a partisan, for example, one wants to know something about the principles embedded in their earlier way of life and the events in response to which they chose such dangers. One wants to understand the justification they would offer for their actions and changes in conditions that would lead them to different conclusions. In a more ordinary context, one might explore the constellations of principles that lead people to disagree about political matters of the day. By sympathetic identification one reconstructs relationships among the terms in each group's reasoning. The analyst constructs reasoning that is internally consistent with conclusions that map closely to those of individuals who express such principles. Taking two populations similarly situated in terms of material interests but with principles known to differ, it may be instructive to identify questions that are of central concern for one group but of relative indifference to the other.

Given that rational choice theory is poorly equipped to explain social dynamics associated with principles, this is likely to be a fruitful area of investigation. A fully developed Rawlsian analysis, however, should give due weight both to principles and to interests. In regard to analytic strategy, we can conceive of three distinct and mutually embedded structures of causal relations. The first, which we have taken for granted in this article so far, involves the material nature of the problem at hand. As embodied beings, our problems typically have material content, and the shape of the matter for our principles and interests depends on the associated material causality. The second involves the interests at stake, and these can be understood in two ways. Following rational choice theory, one can analyze a situation in terms of generic interests in wealth and power, or one may consider it in terms of the specific interests of the individuals and groups involved as determined by their conceptions of the good. Third is the manner in which the problem

is embedded in the society's network of principles and the specific questions of right that the problem raises. As with interests, we can consider principles as they are understood by participants or we may view a situation with reference to a principle established elsewhere (as with behavior that we may find contrary to human rights).

In this manner, one can analyze ranges of solutions to social problems. Knight's standard bargaining problem and Rawls' circumstances of justice identify a general scenario in which there are conflicting interests that can be variously resolved based on different principles. If we follow Rawls and say that a reasonable solution is one that any affected party (particularly the least advantaged) could freely affirm, then we must examine the range of consequences flowing from the adoption of alternative principles. Often there are institutionalized inequalities, so the interests of affected parties are variously represented in negotiations. Since many unreasonable resolutions may be better for all than no resolution, it is possible for the best to be the enemy of the good. More reasonable (less unreasonable) solutions are likely to (1) employ principles familiar from the society's cultural history, (2) which are identified through a procedure that gives greater than normal voice to disadvantaged groups, and that (3) provide for an adequately thorough working out of the likely consequences for all parties' interests.

General methodological comments aside, the direction of a particular analysis depends on the concerns that drive it. While a rational choice or Rawlsian approach can be applied to any social problem, their different grounds give each affinities with different issues. Given that material goods satisfy many interests, for example, rational choice assumptions are congenial with the goal of economic growth. Analysts often feel no need to justify analyzing a society in terms of the conduciveness of its institutions to economic growth. In analyzing Kenya's political economy, for example, Bates seeks to explain why Kenya had more rapid growth than neighboring countries. Identifying the main cause in the rise to power of farmers with large holdings, it is this that his historical institutional analysis has to explain. A Rawlsian view also takes interests to be fundamental, so it also incorporates this concern for growth. Since it takes principles to be fundamental as well, however, a Rawlsian view would additionally inquire into the justification for the exercise of executive power. It would look to the principles in the constitution, and how far conflicts in society are resolved on the constitution's terms.[83]

While rational choice theory is concerned with wealth and power, a Rawlsian view leads almost as directly to the question of which groups in a society enjoy conditions that support the development and exercise of moral autonomy. Rawls and Kant argue that to the extent that social advantages are distributed randomly they ought not to be given any moral weight. The great importance of randomly distributed advantages is a formidable obstacle to distributive justice. Simply by looking at principles as principles acknowledges their contingency, suggesting

the possibility of criticism. Thus a Rawlsian social theory can be expected to possess a critical edge that rational choice theory lacks. The social analysis that arises from the Rawlsian conception of the person, as well as endorsing a concern for economic growth, is likely to incorporate an abiding commitment to equality.

NOTES

1. Jack Knight, *Institutions and Social Conflict* (New York: Cambridge University Press, 1992); Robert Bates, *Beyond the Miracle of the Market: The Political Economy of Agrarian Development in Kenya* (New York: Cambridge University Press, 1989); Robert Bates, "Area Studies and the Discipline: A Useful Controversy?" *PS: Political Science and Politics* 30:166-9.

2. Jon Elster, ed., *Rational Choice* (New York: New York University Press, 1986); Bates et al., *Analytic Narratives* (Princeton, NJ: Princeton University Press, 1998).

3. While Rawls's use of the term "rationality" and the role of a concept of rationality in his conception of justice have changed over time, (see *Political Liberalism* (New York: Columbia University Press, 1993), 53n), his corpus represents a contractarian alternative to the body of thought including utilitarianism and rational choice theory.

4. Immanuel Kant, *Critique of Pure Reason*, Paul Guyer and Allen W. Wood, trans. (Cambridge, UK: Cambridge University Press, 1998); Onora O'Neill, *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), 19.

5. Rawls, *Political Liberalism*, 48-58.

6. As by Bates, *Beyond the Miracle of the Market*.

7. For example, Gerald Marwell and Ruth E. Ames, "Experiments on the Provision of Public Goods I. Resources, Interest, Group Size, and the Free-Rider Problem," *American Journal of Sociology* 84 (1979): 1335-60; Matthew Rabin, "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83 (1993): 1281-1302.

8. Knight, *Institutions and Social Conflict;* Elinor Ostrom*, Governing the Commons: The Evolution of Institutions for Collective Action* (New York: Cambridge University Press, 1990).

9. Max Weber, *From Max Weber: Essays in Sociology*, H. H. Gerth and C. Wright Mills, trans. and ed. (New York: Oxford University Press, 1946), 324.

10. Weber, *Essays in Sociology*, 293.

11. In "Politics as a Vocation," Weber writes, "Exactly what the cause, in the service of which the politician strives for power and uses power, looks like a matter of faith. The politician may serve national, humanitarian, social, ethical, cultural, worldly, or religious ends. . . . He may claim to stand in the service of an 'idea' or, rejecting this in principle, he may want to serve external ends of everyday life. However, some kind of faith must always exist" (*Essays in Sociology*, 117). Weber is suggesting that politicians who seek power for its own sake are less than honorable, but he does not differentiate among the ends politicians might promote.

12. Rawls, *Political Liberalism*, 51.

13. On the descriptive value of the self-interest assumption, Rawls notes, "Nor are rational agents as such solely self-interested: that is, their interests are not always interests in benefits to themselves. Every interest is an interest of a self (agent), but not every interest is in benefits to the self that has it. Indeed, rational agents may have all kinds of affections for persons and attachments to communities and places, including love of country and of nature; and they may select and order their various ends in various ways" (*Political Liberalism*, 51).

14. Rawls, *Political Liberalism*, 49.

15. Ibid., 50.

16. Ibid., 48n.

17. Immanuel Kant, "Metaphysical Foundations of Morals," in Carl Friedrich, ed., *The Philosophy of Kant: Immanuel Kant's Moral and Political Writings* (New York: Random House, 1949), 163.

18. Although Kant refers to both categorical and hypothetical imperatives as maxims or principles, the meaning usually associated with "principles" is closer to "categorical imperatives," and this is the meaning we generally intend in this article.

19. Similarly Kant, *Critique of Pure Reason*, takes our sense of space and of time to be pure in that these are basic capacities of the mind not drawn from experience but on which experience depends.

20. Kant, "Metaphysical Foundations of Morals," 170.

21. Rawls, *Political Liberalism*, 51n.

22. John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 3.

23. Ibid., 302.

24. Alvin E. Roth, introduction to *The Handbook of Experimental Economics*, John H. Kagel and Alvin E. Roth, eds. (Princeton, NJ: Princeton University Press, 1995), 26.

25. Ibid.

26. John O. Ledyard, "Public Goods: A Survey of Experimental Research," in *The Handbook of Experimental Economics*, John H. Kagel and Alvin E. Roth, eds. (Princeton, NJ: Princeton University Press, 1995), 143.

27. Merill M. Flood, "Some Experimental Games," *Management Science 5* (1958): 16, and Gerhard K. Kalisch, et al., "Some Experimental N-Person Games," in *Decision Processes*, R.M. Thrall, C. H. Coombs, and R. L. Davis, eds. (New York: Wiley Press, 1954).

28. Ledyard, "Public Goods," 172.

29. Michael J. G. Cain, "An Experimental Investigation of Motives and Information in the Prisoners' Dilemma Game" (paper presented at the annual meeting of the Midwest Political Science Association, Chicago, IL, April 1998).

30. James Andreoni and John H. Miller, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," *The Economic Journal* 103 (1993): 572.

31. Ibid.

32. Andreoni and Miller, "Rational Cooperation," 572; Gerald Marwell and Ruth E. Ames, "Economists Free Ride, Does Anyone Else?," *Journal of Public Economics* 15 (1981): 299.

33. Andreoni and Miller, "Rational Cooperation," 582.

34. Kristen Renwick Monroe, *The Heart of Altruism: Perceptions of a Common Humanity* (Princeton, NJ: Princeton University Press, 1996), 4.

35. Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984). Axelrod's results represent an important current not only in prisoner's dilemma experiments but in rational choice theory in general. Advocates of this current in rational choice theory—David Gauthier, *Morals by Agreement* (Oxford: Clarendon, 1986); Russell Hardin, *Morality within the Limits of Reason* (Chicago: Chicago University Press, 1988); Howard Margolis, *Selfishness, Altruism and Rationality: A Theory of Social Choice* (Cambridge: Cambridge University Press, 1982); Robert Frank, *Passions within Reason: The Strategic Role of the Emotions* (New York: Norton, 1988); and Ken Binmore, *Game Theory and Social Contract: Playing Fair* (Cambridge, MA: MIT Press, 1994) and Binmore, *Game Theory and the Social Contract: Just Playing* (Cambridge, MA: MIT Press, 1998)—argue that fair social arrangements and even morality itself are best explained as rational solutions to a host of prisoner's dilemma-like problems. As we shall see in Section V, such

approaches to explaining norms and institutions are eagerly taken up by rational choice new institutionalists.

36. Robert Axelrod, "More Effective Choice in the Prisoner's Dilemma," *Journal of Conflict Resolution* 24 (1980): 402.

37. Axelrod, *Evolution*, 136-9.

38. Marwell and Ames, "Experiments on the Provision of Public Goods I"; Marwell and Ames "Economists Free Ride, Does Anyone Else?"

39. Marwell and Ames, "Economists Free Ride, Does Anyone Else?," 307-308; Table 2 (307) for summary of experimental conditions. Public goods provision experiments resemble prisoner's dilemma experiments because of the way payoffs are structured. If few participants contribute to the public good, those who do receive a "sucker's payoff"; those who choose not to contribute keep the resources allocated to them at the outset and still stand to reap the possible dividends of others' contributions (the "temptation payoff").

40. Marwell and Ames, "Experiments on the Provision of Public Goods I," 1359. The passage cited continues, "Despite isolation, instructions which emphasized the monetary importance of the situation and minimized social factors, and the chance to develop full information regarding the parameters of the situation, normative factors such as fairness seem to have strongly influenced economic decisions" (1359).

41. Ibid., 1357-58.

42. Marwell and Ames, "Economists Free Ride, Does Anyone Else?," 306-307.

43. Ibid., 309. Further on this point, "Those who did respond were much more likely to say that little or no contribution was 'fair'. In addition, the economics graduate students were about half as likely as other subjects to indicate that they were 'concerned with fairness' in making their investment decisions" (309).

44. James G. March, *A Primer of Decision Making: How Decisions Happen* (New York: Free Press, 1994), 57.

45. Rabin, "Incorporating Fairness into Game Theory and Economics."

46. Not all the incentives Bates admits are strictly material. One reason he gives for members of the Kikuyu tribe wanting large families, in the period prior to British colonial rule, is to produce many descendants to care for and commune with the soul after death. See *Beyond the Miracle of the Market*, 15.

47. Ibid., 149.

48. Ibid., 73-84; 86-89.

49. Ibid., 3-6.

50. Ibid., 6.

51. Ibid., 15-16; 28.

52. Ibid., 12.

53. Ibid., 31.

54. One might note that the binding power of oaths is also dependent on the reasonable capacity.

55. Knight, *Institutions and Social Conflict*, 14; 19-20.

56. Knight suggests that Hume, Adam Smith, and March and Olsen fall in the first camp and Marx, Weber, and rational choice theorists fall in the second. Ibid., 4-10.

57. Ibid., 14.

58. Ibid.

59. This interpretation is supported by Knight's concluding remarks, 211.

60. This should not be taken to impugn the behavior of experimental participants who interpret the (trivial experimental) situation as one in which it is appropriate to pursue self-interest.

61. Knight, *Institutions and Social Conflict*, 54.

62. Ibid., 128.

63. Rawls, *Theory of Justice*, 126.

64. Knight, *Institutions and Social Conflict*, 126.

65. Ibid., 129; emphasis in original.

66. Ibid., 125.

67. Ostrom, *Governing the Commons*; Sue E.S. Crawford and Elinor Ostrom, "A Grammar of Institutions," *The American Political Science Review* 89 (1995): 582-600. Crawford and Ostrom write, "A growing body of work considers the mix of normative and material motivations that individuals consider when faced with choices (Coleman 1988; Ellickson 1991; Elster 1989a, 1989b; Etzioni 1988; Hirschman 1985; Knack 1992; Mansbridge 1990, 1994; Margolis 1991; Offe and Wiesenthal 1980; E. Ostrom 1990; V. Ostrom 1986; Udéhn 1993). These works treat the normative aspects of decisions up front as a significant part of the analysis. Margolis argues for the necessity of such an approach: 'If we analyze everything in terms of strict self-interest and then include some social motivation only if we get stuck or if there is something left over, it is not likely to lead to nearly as powerful a social theory as if the two things are built in at the base of the analysis' (1991, 130)." 589-90.

68. Ostrom, *Governing the Commons*, 8-13.

69. Ibid., 37.

70. Ibid., 36.

71. Ibid., 184.

72. Ibid., 35.

73. Ibid., 51.

74. John T. Scholz and Neil Pinney, "Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizen Behavior," *American Journal of Political Science* 39 (1995): 490-512.

75. Ostrom, *Governing the Commons*, 142.

76. Rawls, *Theory of Justice*, 11-12.

77. Bates, *Beyond the Miracle of the Market*, 9-10.

78. Ibid., 154.

79. We show how philosophy can contribute to social change in our Rawlsian analysis of the French Revolution. See Emily Hauptmann and Paul Clements, "The Reasonable and the Rational Capacities and the Analysis of Revolutions" (paper presented at the Annual Conference of the International Society for Political Psychology, Seattle, WA, July 2000).

80. Margaret Levi, *Consent, dissent, and patriotism* (New York: Cambridge University Press, 1997).

81. This is the standard method of comparative politics.

82. And the participant's interests, of course.

83. See Rawls, *Political Liberalism*, 237-40, for a discussion of such issues involving the U.S. Constitution.