March, 1995

Impossible Comparisons and Rational Choice Theory

Richard Warner, *Chicago-Kent College of Law*

ARTICLE

IMPOSSIBLE COMPARISONS AND RATIONAL CHOICE THEORY

RICHARD WARNER*

> It may be that we are witnessing a curious phenomenon in which rational choice theories are fortified in every discipline by reference to their alleged success elsewhere, when a more global view of things would reveal the emperor to be, if not entirely naked, somewhat scantily clad.[1]

Rational choice theory yields empirically confirmed predictions; for example:

> when the wage rate rises, all other things held equal, the supply of labor increases and the demand for labor decreases; when the price of alcohol rises, relative to that of other goods and services, the quantity demanded goes down (if not by much); when the price of a good or service rises, again relative to that of other goods or services, productive effort tends to shift into the supply of that good or service; and when the price of an input rises relative to that of its substitutes, producers tend to use less of that input and relatively more of the substitutes.[2]

Such examples abound, convincing many that "those who question rational-choice theory must simply be ignorant of the theory and its results."[3] But ignorance is far from the sole source of skepticism; for,

---

* Visiting Assistant Professor of Law, University of Southern California; Assistant Professor of Law, Chicago-Kent College of Law. I am indebted to Scott Altman for typically perspicuous comments on an earlier draft, and I also owe thanks to Colin Camerer, Rick Hasen, Richard McAdams, and Matt Spitzer. I gratefully acknowledge the support of the Marshall D. Ewell Research Fund.

1. DONALD GREEN & IAN SHAPIRO, PATHOLOGIES OF RATIONAL CHOICE THEORY: A CRITIQUE OF APPLICATIONS IN POLITICAL SCIENCE 180 (1994).

2. Thomas S. Ulen, *Rational Choice and the Economic Analysis of Law*, 19 LAW & SOC. INQUIRY 487, 489 (1994) (footnote omitted).

3. *Id.*

while rational choice theory has achieved impressive empirical success, it has also encountered disconcerting failure, and—worse yet—the failures occur in those areas most relevant to the law. Much of the disconfirming evidence concerns cooperation, bargaining, and the cognitive limitations that characterize real human agents (as opposed to the idealized agents of rational choice theory).[4] What explains the failures? Different things in different cases, no doubt, and different explanations need not be mutually exclusive; failures may be overdetermined, the result of a confluence of factors each separately sufficient.

An explanation that has so far been ignored forms our focus.[5] Incommensurability provides this explanation. The incommensurability that concerns us is the incommensurability of reasons. Reasons are incommensurable when (and only when) they are *not comparable* as better, worse, or equally good.[6] I make two claims—one conceptual and one empirical—about incommensurability so conceived. The conceptual claim is that, where the incommensurability of reasons plays an essential role in generating actions, rational choice theory makes the wrong predictions. This is not an empirical issue; rather, the point follows from the nature of incommensurability and the structure of rational choice theory. The empirical claim is that incommensurability explains some of the predictive failures of rational choice theory.

---

4. *See id.* at 488.

5. Some may object that this concern with empirical confirmation is really beside the point since we should view rational choice theory not as an empirical theory but as a *normative* model. Indeed, Tversky and Kahneman insist that "the theory was conceived as a normative model of an idealized decision maker, not as a description of the behavior of real people." Amos Tversky & Daniel Kahneman, *Rational Choice and the Framing of Decisions*, *in* DECISION MAKING: NORMATIVE AND PRESCRIPTIVE INTERACTIONS 67, 67 (D. Bell et al. eds., 1988). While Tversky and Kahneman are no doubt correct, we should add that the theory seems to have been originally conceived all along as a *mixed* normative and empirical theory—a theory that predicts the behavior of *rational* agents. Insofar as it defines what counts as rational, it is normative; insofar as it predicts, it is empirical. Moreover, it is standardly used in this mixed way: "Neoclassical economics is based on the premise that models that characterize rational, optimizing behavior also characterize actual human behavior." RICHARD H. THALER, *The Psychology of Choice and the Assumptions of Economics*, *in* QUASI RATIONAL ECONOMICS 137, 137 (1994).

6. In Richard Warner, *Excluding Reasons: Impossible Comparisons and the Law*, 15 OXFORD J. LEGAL STUD. 431 (1995) [hereinafter Warner, *Excluding Reasons*], I use the more accurate term "noncomparative exclusion of reasons" instead of "incommensurability." I use the latter term here since I have not introduced and explained the idea of the nonexclusion of reasons. Many opponents of incommensurability insist that incommensurable reasons cannot be compared *simpliciter*. As I explain incommensurability in the text, reasons are not incommensurable *simpliciter* but only *relative to a commitment*. The advantage of the relativization is that it yields a clear explanation of how and why reasons can be incomparable.

The point of the conceptual and empirical claims is not to reject, but merely to restrict, rational choice theory.[7] Outright rejection is not realistic, for, as Robert Bellah remarks, in commenting on opposition to cost-benefit analysis:

> [O]pponents of the cost-benefit approach to public policy invoke embedded cultural traditions, a sense of moral and religious absolutes. . . . The . . . vision of cost-benefit analysis is countered with a deeply rooted moral individualism that respects the dignity of persons, but has largely lost the social context of the older traditions. *This has little to offer in dealing with the complexity of interrelated choices a modern society must make, and therefore is not an effective alternative to cost-benefit analysis.*[8]

Bellah's point is a powerful one. Economic analysis based on rational choice theory is often an illuminating approach to public policy, revealing possibilities for efficient action and the general improvement of welfare that would otherwise have escaped our notice. But understanding where it fails is essential. We should only use the theory where it gives the right results. Where it does not, we must take another approach, or treat the theory as providing only an inaccurate approximation.

There is a normative issue here as well. Incommensurability plays virtually no acknowledged role in public decisionmaking, in "policy discourse in the heart of government, where crucial questions about our shared future are decided."[9] This is deplorable since in ignoring incommensurability we ignore *ourselves*. To ignore incommensurability is, as we will see, to ignore a feature essential to our identity as persons. Policy discourse that denies incommensurability erases this essential feature from its policymaking calculations. This is wrong. Public policy is policy for *us* and as such it should speak to *our* concerns, not the sanitized psychology of a theoretically-posited decisionmaker shorn of our essential attributes. So, Bellah is mistaken

---

7. Theory revision, not theory restriction, is another alternative. Richard Thaler, for example, envisions a revised rational choice theory that "will retain the idea that individuals try to do the best they can, but these individuals will also have the human strengths of kindness and cooperation, together with the limited human abilities to store and process information." RICHARD H. THALER, THE WINNER'S CURSE: PARADOXES AND ANOMALIES OF ECONOMIC LIFE 5 (1992). In a similar vein, Amos Tversky and Daniel Kahneman propose prospect theory, a mathematical model of decisionmaking that takes explicit account of "the manner in which the choice problem is presented as well as [the] norms, habits, and expectancies of the decision maker." Tversky & Kahneman, *supra* note 5, at 73.

8. ROBERT N. BELLAH ET AL., THE GOOD SOCIETY 120 (1991) (emphasis added).

9. *Id.*

when he contends that moral individualism "has little to offer in dealing with the complexity of interrelated choices a modern society must make." It is no doubt true that "a deeply rooted moral individualism that respects the dignity of persons . . . has largely lost the social context of the older traditions," but this does not mean that such a moral individualism has "little to offer in dealing with the complexity of interrelated choices a modern society must make." An understanding of incommensurability grounds moral individualism, not in the moribund context of defunct traditions, but in our essential nature as persons.[10]

Having noted the normative theme, we will put it aside until the conclusion. Our focus will now turn exclusively empirical. Part I provides a picture of rational choice theory and, in doing so, reveals a fundamental assumption behind the theory. Part II explains and illustrates this assumption. Part III explains the nature of incommensurability, and Part IV combines the results of Parts II and III to show that the conceptual claim is true (where incommensurability of reasons plays an essential role in generating actions, rational choice theory makes the wrong predictions). Part V turns to the empirical claim that incommensurability explains some of the empirical failures of rational choice theory. Part VI contains some concluding normative remarks.

I. WHAT IS RATIONAL CHOICE THEORY?

The obvious first question is, what is rational choice theory?[11] The heart of the theory is the expected utility rule; or, more precisely, *some version* of that rule. Disconfirming evidence has proven fertile ground for the proliferation of revisions.[12] We can, however, put

---

10. I would like to claim Margaret Radin as an ally here. See Margaret Jane Radin, *Compensation and Commensurability*, 1993 DUKE L.J. 56, 56:

> When someone who has lost an arm in an accident receives $100,000 in compensation through the tort system, what does this transaction mean? Does it mean that an arm is "worth" $100,000? . . . I am writing about meaning—the way human interactions are understood by a community sharing a concept in practice.

My ultimate concern is the same, with "meaning—the way human interactions are understood by a community sharing a concept in practice." *Id.* Radin connects this concern with the self in Margaret Jane Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849 (1987).

11. I take rational choice theory to lie at the heart of "positive political theory." There is considerable disagreement over just what positive political theory is. *See* Daniel A. Farber & Philip P. Frickey, *Foreword: Positive Political Theory in the Nineties*, 80 GEO. L.J. 457 (1992).

12. Tversky and Kahneman provide a list of proposed revisions. Tversky & Kahneman, *supra* note 5, at 87. For a detailed review of proposed revisions, see Colin Camerer, *Individual*

these revisions to one side. Our concern is with a fundamental assumption that remains in all revisions, and, in articulating and evaluating this assumption, we can take the simple and convenient course of focusing on the classical form of the expected utility rule.

Stating the rule against the background of certain facts about action throws the rule into sharp relief and reveals the assumption. The facts are basic and uncontroversial. In describing them, I will make free use of the notion of value, and some may object that the notion of value is far too obscure to qualify as "basic and uncontroversial." This is simply wrong. The particular concept of value we will use here has a clear and uncontroversial sense. It has the sense it does when, for example, I say, "I value my daughter's playfulness." To say this is to express a certain *attitude* toward her playfulness, an attitude that typically finds expression in thoughts, feelings, and actions. Talk of value becomes problematic and controversial when we ask whether any mind-independent items—items "out there"—correspond to our valuings and make them true, in something like the way, for example, the cat's being on the mat corresponds to, and makes true, the belief that the cat is on the mat. These issues need not concern us, for our focus is entirely on valuing as an attitude of mind. Some will continue to object, however, on the ground that the notion of value is irrelevant to rational choice theory. Rational choice theory, after all, concerns *preference*, not value. It is a theory about what action a rational person will perform given certain *preferences*. Our exposition of rational choice theory in terms of value is not inconsistent with this fact; quite the contrary, it reveals why we can—apparently—drop talk of value in favor of talk of preference. The "apparently" qualification is crucial, for it is precisely here that we encounter the fundamental assumption.

Let us turn then to the relevant facts about action, described in terms of value. We begin with an example. Suppose you are trying to choose between going to law school to become a lawyer, and retiring to the woods to write your first novel. You choose to go to law school. You do so even though you value being a successful novelist over being a successful lawyer. The explanation for your choice is that the way you value the two payoffs is not the sole determinant of your choice. You also consider the probability of realizing each option.[13]

Decision Making, *in* HANDBOOK OF EXPERIMENTAL ECONOMICS 588 (J. Kagel & A. E. Roth eds., 1995).

13. It does not matter for our purpose whether the probabilities are objective or subjective.

More fully, you *combine* value and probability considerations to obtain a ranking of the actions. You think that you are unlikely to succeed in writing your first novel and highly likely to succeed in graduating from law school. These probability considerations lead you to rank going to law school ahead of retiring to the woods and you act accordingly. Another example: As you stand in front of the ice cream store, the probability that buying an ice cream cone will lead to the pleasure of eating it is high, but you pass up this virtually certain pleasure for the far less certain benefit of entering the bookstore next door. The explanation is that you value reading an interesting book far more than you value the ice-cream-consuming pleasure. Consequently, you rank entering the bookstore higher than buying an ice cream cone.

There is a crucial distinction to draw here. The distinction is between *payoffs* of actions—for example, being a lawyer or being a successful novelist—and the *actions* aimed at realizing the payoffs—going to law school or retiring to the woods. Very roughly, a payoff is what one gets out of performing an action; it is a result the action yields.[14] In one form or another, the payoff/action distinction has a long history. Aristotle, for example, emphasizes the distinction between an action and that for the sake of which all else is performed (what we are calling the payoff).[15] To avoid a misunderstanding, we should note that a payoff (that for the sake of which all else is performed) might itself be an action. What I get out of the action of entering the bookstore is *reading* a book, which certainly qualifies as an action (or, if one disputes this, change the example to make what I get *buying* a book). In general, the payoff/action distinction is relative to the context in which an agent acts. Given the context, and a range of actions open to the agent in that context, we can readily distinguish between action (entering the bookstore) and payoff (reading a book). Given our purposes, there is no need to be more precise than this.

The crucial point is that one ranks payoffs in light of values. For example, you rank being a successful novelist over being a successful lawyer. Talk of "ranking" is talk of *choice*; to say that you rank being

14. It need not be a causal result. Looking at the Cézanne landscapes results in looking at impressionist landscapes since that is what the Cézannes are. I draw some relevant distinctions here in RICHARD WARNER, FREEDOM, ENJOYMENT, AND HAPPINESS 143-45 (1987) [hereinafter WARNER, FREEDOM].

15. ARISTOTLE, NICOMACHEAN ETHICS 14, col. 1097a, ll. 16-24 (Martin Oswald trans., 1962).

a novelist over being a lawyer is to say that, in appropriate circumstances,[16] you would choose the former over the latter. These are not choices under *un*certainty, but choices with certainty; the choice is between the certain payoff of being a lawyer and the certain payoff of being a novelist. In general, given a range of payoffs, we can rank them in terms of which payoff you would choose over which others.[17] Let us call this the *payoff-ranking*. In the lawyer/novelist example, your *values* determine the relevant payoff-ranking, but, of course, values need not be the sole determinant of payoff-rankings. For example, given a choice between vanilla and strawberry ice cream, I opt for vanilla. The explanation is not that I value vanilla ice cream (in the sense in which I value my daughter's playfulness; I hold vanilla ice cream in no such high esteem). The explanation is that I just *feel like* vanilla. Having noted this point, we will continue to focus exclusively on value as the determinant of payoff-rankings, for our concern is with the role of value in rational choice theory.

Now let us turn from payoffs to actions. The payoff-ranking is not a ranking of *actions*. One ranks actions—such as *going* to law school and *entering* the bookstore—by combining probability considerations with the payoff-ranking. So, for example, you rank going to law school over retiring to the woods because of probability considerations. Let us call the result of such combination the *action-ranking* (thus, unlike the payoff-ranking, the action-ranking essentially involves probabilities). Again, talk of "ranking" is talk of choice. To say that going to law school ranks higher than retiring to the woods is to say that one would (in appropriate circumstances) choose to go to law school over retiring to the woods.[18] Insofar as one is rational, one performs the action that ranks highest in one's action-ranking.

16. The "appropriate circumstances" qualification is essential. Suppose you were offered a choice between being a successful novelist and being a successful lawyer in circumstances where you would be killed if you opted for either. You undoubtedly would choose neither. The basic idea is that you are offered a choice where your choice will alter the status quo only to the extent necessary to realize the choice.

17. Such unique ranking may not exist, for we cannot uniquely determine the truth or falsity of the counterfactual statement "You would choose payoff O over payoff O'." Its truth or falsity will vary as we vary the detail and breadth of the background assumptions against which we assess its truth. We will put these issues to one side. Basically, what they mean is that rational choice theorists have to predict action against some fixed set of background assumptions.

18. Again the "appropriate circumstances" qualification is essential (and perhaps more difficult to spell out in this context), and again there may in general be no unique ranking generated by counterfactual claims about what one would choose.

Obviously, then, if we know your action-ranking, we can predict your choice (assuming you are rational). Of course, the action-ranking, as we have just seen, is a construct of two more basic elements: the payoff-ranking and probability considerations. If we could predict what action-ranking a person would construct out of a given payoff-ranking and set of probability considerations, we could predict choices from a knowledge of these two basic elements alone. To do so, of course, we need to know the rule or procedure by which one combines probabilities and payoff-rankings to obtain action-rankings. What rule do we use? Reflection suggests a plausible hypothesis. When we combine value and probability, we give weight to value and weight to probability—the greater the value or probability the greater the effect we give it in determining the action-ranking. There is a simple mathematical way to represent such a procedure—multiplication: The bigger (smaller) the multiplicand, the bigger (smaller) it makes the product.

This is just how expected utility theory represents the rule. The rule represents probabilities by numbers, of course; and it represents a person's payoff-ranking by a utility function. The utility of a payoff is a *number* that indicates the place of that payoff in the person's payoff-ranking. This is all we need to note; the exact mathematical structure of the utility function will not concern us.[19] What we want to focus on is the rule for combining utility and probability. To see how the expected utility rule does this, suppose A is an action with just two possible payoffs, O and *not-O*. Using u for the relevant utility function, their respective utilities are $u(O)$ and $u(\textit{not-}O)$. Using P for the probability that A will yield the payoff O, the expected utility of the action A is

$$(P \times u(O)) + ((1 - P) \times u(\textit{not-}O)).^{20}$$

We can construct an action-ranking by calculating the expected utility of each relevant action.

Rational choice theory asserts that rational agents perform the action at the top of the ranking so constructed, the action with the greatest expected utility. This is the rational choice theory version of the common sense fact that a person performs the action ranking highest in the person's action-ranking. Note that there is no claim that rational agents actually *use* the expected utility rule in constructing

19. For some discussion, see George A. Quattrone & Amos Tversky, *Contrasting Rational and Psychological Analyses of Political Choice*, 82 AM. POL. SCI. REV. 719 (1988).

20. Probabilities must sum to 1; this is why the probability of realizing *not-O* is $1 - P$.

action-rankings. The claim is quite different, and clarity about the claim is essential to our critique of rational choice theory.[21] The key to clarity is to distinguish two claims—a purely predictive claim and an explanatory one.

To see the first claim, return to the book/ice cream cone example. In that example, you pass up the virtually certain pleasure of ice cream for the less certain benefit of entering the bookstore next door. The explanation is that you value reading an interesting book far more than you value the ice-cream-consuming pleasure; consequently, you rank entering the bookstore higher than buying an ice cream cone. The purely predictive claim is this: *The expected utility rule generates the same action-ranking.* Given the same payoff-ranking—represented by the utility function—and the same probabilities, ranking actions by the expected utility accurately captures the rational person's action-ranking. There is no claim here about what rule or procedure rational agents actually use to construct an action-ranking. The claim is that, whatever procedure rational agents use, it is "input-output" equivalent to the expected utility rule. Given the same payoff-ranking and probabilities as input, the rational person and the expected utility rule will yield the same action-ranking as output.

Now let us turn to the explanatory claim. Claims about input-output behavior are typically associated with explanatory claims. To take a simple example, consider two computers both programmed to add numbers but running *different* software programs to do so. There is a clear input-output equivalence here: Given the same addends as input, the computers yield the same sum as output. But we have *different explanations* of their equivalent input-output behavior since one is running one program; the other, another. Another example: Within definable limits, Newtonian mechanics and the theory of relativity are input-output equivalent; given the same initial data, they yield the same predictions. But the theory of relativity gives the (more) correct *explanation* of those predictions. The crucial point for our purposes is that such explanatory claims are essential to scientific practice. Our explanatory understanding of a theory's predictive success guides our testing and revision of the theory and ultimately our transformation of it into a successor theory.

21. It is worth being clear about what the claim is, for this is an area in which considerable confusion still reigns. Jean Hampton clarifies this matter in Jean Hampton, *The Failure of Expected-Utility Theory as a Theory of Reason*, 10 ECON. & PHIL. 195 (1994).

Like any good scientific theory, rational choice theory makes explanatory, not merely purely predictive, claims. Rational choice theory (as we have presented it[22]) posits an underlying psychological reality: People combine payoff-rankings with probability considerations to obtain action-rankings. Increases or decreases in payoff-ranking and/or related probabilities result in a corresponding increase or decrease in action-ranking position. Rational choice theory provides a partial characterization of this combinatory activity in a remarkable mathematical result that lies at the heart of rational choice theory, a result due Von Neumann and Morgenstern.[23] They show that, if a person's payoff-ranking satisfies certain conditions, then the expected utility rule *must* accurately represent the person's action-ranking. The axioms that express the conditions on the payoff-ranking provide a partial mathematical characterization of essential elements of a rational person's combinatory activity.

There is another way in which the axioms are crucial to rational choice theory. The individual axioms that express the conditions on payoff-rankings are intuitive and compelling (at least they can *seem* so; I will argue that the axioms often do *not* hold of rational agents[24]). In this sense, a normative claim underlies rational choice theory—the claim that the axioms capture conditions of rationality, conditions that we *ought* to conform to. This is in no way objectionable; indeed, how could a theory of *rational* choice avoid such a feature? Indeed, revisions of rational choice seem to have forgotten the normative significance of the Von Neumann/Morgenstern axioms. The revisions are designed to correct the empirical failings of the classical expected utility rule, but, in doing so, they pay little or no attention to the normative claim involved in putting forward a theory of *rational* choice.

We consider the Von Neumann/Morgenstern axioms in some detail in the Appendix. At the moment, we should emphasize another, more empirical and less mathematical, reason rational choice theory seems compelling: Namely, the mathematical representation offers considerable precision and explanatory power. By way of illustration, consider a complication we have so far suppressed. We have considered only two payoffs per action. For example, in considering the

22. The presentation is accurate although the explanatory claim is often suppressed in expositions of rational choice theory. The claim is, in any case, essential to the theory's claim to be a scientific theory of human behavior.

23. JOHN VON NEUMANN & OSKAR MORGENSTERN, THEORY OF GAMES AND ECONOMIC BEHAVIOR 1 (1944).

24. *See infra* appendix.

choice to be a lawyer we considered only two possibilities—either you succeed or you fail to become a lawyer. In reality there would be many other possibilities to take into account: You succeed in becoming a lawyer and you are happy; you succeed but are unhappy; you fail but coping with failure ultimately strengthens your self-esteem; you fail and lose your self-esteem; and so on. Going to law school is a complex gamble—a lottery with multiple payoffs associated with varying probabilities of realization. When we construct an action-ranking we are typically ranking the *lotteries* into which our actions enter us. The mathematical framework of rational choice theory provides a perspicuous representation of this aspect of our choices, an aspect that is cumbersome to handle in the nonmathematical language of our ordinary talk and thought about action.

II. THE ASSUMPTION

We began this exposition of rational choice theory to reveal a fundamental assumption behind the theory and the assumption is now clear. It is that only two things matter in constructing the action-ranking: utility (the relative ranking of payoffs) and the relevant probabilities of realizing those payoffs. The crucial point is that value is only relevant to determining the payoff-ranking. Once that is established, *a person's values play no further role in predicting behavior.*[25]

This assumption lies behind the use of the notion of *preference* in rational choice theory, and we can further illustrate the assumption's role by seeing why we can, if the assumption is correct, drop talk of value in favor of talk of preference. To see why, consider that we need to know the payoff-ranking to construct the action-ranking, but all that matters about the payoff-ranking is the order of the payoffs. All we need to know is whether the person would (in appropriate circumstances) choose one payoff over another. The person's values may explain the choice, but we do not care *why* the person would choose the payoff. We only care *that* the person would choose it. So why be concerned with value at all? All it seems we really care about

25. This remains true in all proposed revisions of rational choice theory, with the possible exception of prospect theory. In prospect theory, the role of frame selection and the use of a decision weight instead of probabilities may allow values to play some role in directly determining an action's place in the action-ranking. At this stage in the development of prospect theory, it is difficult to say how this issue will turn out. For an interesting application of prospect theory to the law, see Richard L. Hasen, Comment, *Efficiency Under Informational Asymmetry: The Effect of Framing on Legal Rules*, 38 UCLA L. REV. 391 (1990).

is *preference*, where one prefers *a* to *b* when and only when one would choose *a* over *b* in relevant circumstances.

Of course, there is a point to replacing value with preference only if we have some way to determine a person's preferences *independent of information about a person's values*, but there is an obvious technique here: Ask the person. This is the tactic that contingent valuation takes. We ask how much one would be willing to pay to have one payoff or another. If one is willing to pay more to have *a* than *b*, one would choose *a* over *b*, and thus *a* ranks higher than *b* in the payoff-ranking. This does not mean that value is irrelevant. A person's values do determine a person's preferences, but, as far as rational choice theory is concerned, it seems we can safely relegate value to the undiscussed background.

So it *seems*, but this appearance is—to an extent—an illusion. *We cannot—cannot always—confine value to the construction of the payoff-ranking.* Incommensurability shows that values sometimes play a role in the procedure by which a rational person constructs the action-ranking out of the payoff-ranking and the probabilities. In such cases, rational choice theory makes the wrong predictions, for it constructs the wrong action-ranking. This is the conceptual claim. It follows simply from the nature of incommensurability and the underlying assumption behind rational choice theory. To see that this is true, we first need to explain the nature of incommensurability. Then we can easily see why, in cases involving incommensurability, values play a role in the procedure by which a rational person constructs the action-ranking out of the payoff-ranking and the probabilities.

III. THE NATURE OF INCOMMENSURABILITY

I will simply *assume* that incommensurability exists. I have argued at length elsewhere that it does,[26] and I rely on those arguments here. Those who remain unconvinced, or are otherwise unwilling to concede incommensurability's existence, may regard my arguments as conditional: "Given incommensurability exists, then . . . ." However, even though we will not argue that incommensurability exists, we do need to make its nature clear. Otherwise, we will not be able to see why the conceptual claim is true—the claim that where incommensurability is involved, rational choice theory yields wrong predictions.

26. *See* Warner, *Excluding Reasons*, *supra* note 6.

An example is helpful.[27] Suppose that, as I am out walking with my daughter, a stranger approaches and offers me $1,000,000 if I will turn her over to him and never see her again. I refuse, and the stranger then makes the same offer to Jones. Jones has a reason not to sell his daughter: He would miss her. However, Jones also has a reason to have the money: He would pay off bills and invest the rest; and, since getting the money means selling his daughter, he sees these financial considerations *as a reason to sell his daughter*. He compares his reasons for and against selling his daughter; finds his reason not to sell to be the better one, and accordingly refuses the stranger's offer. My refusal might be thought to rest on similar grounds. I also have a reason not to sell my daughter: I love her. I also have a reason to have the money: I too would pay off bills and invest the rest. However, unlike Jones, my decision results not from comparing, but from *excluding* reasons. I regard the financial considerations as irrelevant to my decision. When I exclude the reason, I proceed exactly as if the money provided *no reason at all* to sell my daughter. So my decision is simple: I have a reason not to sell (my love for my daughter), and no reason—no reason I will consider—to do otherwise. Accordingly, I refuse the stranger's offer.[28]

Some will object that this is not an example of incommensurability as we defined that notion. We defined incommensurability this way: Reasons are incommensurable when they cannot be compared as better, worse, or equally good. The apparent problem is that the most the example shows is that a comparison of reasons cannot *play any role in my decision*; I exclude the reasons. But surely—the objection goes—showing that I *exclude* the reasons is not the same as showing that I cannot *compare* them. Why could I not see the reasons as comparable—although not in a way relevant to deciding whether to sell my daughter? There is no need to settle this issue. What we care about is the comparison *relevant to deciding what to do*. A comparison of reasons *irrelevant to decisionmaking* is hardly worth consideration. Moreover, there is a clear sense in which I really cannot compare the reasons. To see this, we first need to see *why* I exclude the reasons.

I do so because of my commitment to my daughter—my love for her. I would not love her, at least, not in the way I do, if I counted the

27. The material that follows is adapted from Warner, *Excluding Reasons*, *supra* note 6.

28. Of course, I also have moral reasons to refuse: It is morally wrong to sell my daughter. My point is that my commitment *also* leads me to refuse to sell.

money as a reason to sell her. It is *definitive*, *constitutive*, of my commitment that I exclude the reason. As Joseph Raz remarks:

> For many, having children does not have a money price because exchanging them for money, whether buying or selling, is inconsistent with a proper appreciation of the value of parenthood. . . . [B]oth their rejection of the idea that having children has a price and their refusal even to contemplate such exchanges are part of their respect for parenthood, an expression of the very high value which they place on having children.[29]

We care, indeed care a great deal, about whether a father is the sort of person that would see financial gain as a reason to sell his daughter. When differences are sufficiently important to us we often have a concept that marks the difference, and we do here. "Parental love" designates the difference and, of course, it is the difference that matters, not the words that name it. If someone wants to insist that Jones, who would sell his daughter for a sufficiently high price, can be described as "loving" his daughter, we have nothing to argue about as long as we recognize the difference and its significance.

We can now explain the sense in which I cannot compare reasons. I cannot compare in the sense that I cannot do so *consistent with my commitment.* My commitment requires that I exclude the money as a reason and I cannot simultaneously exclude the reason and also compare it to other reasons. To see why, suppose I were to treat the comparison of reasons as relevant to my decision. Suppose, that is, that I see the comparison of reasons as determining, at least in part, whether I will sell my daughter. To so regard the comparison is to treat the financial considerations as a reason to sell my daughter, and that is precisely what my commitment does not allow. To compare a reason is not to exclude it.

A crucial *caveat*: I do not claim that there are no circumstances in which one might sell a child one loves. Suppose Sally has two daughters. One will die if she does not receive medical treatment costing $1,000,000. Sally *might* sell the healthy daughter to raise the money. But she would still not be like Jones. Jones recognizes the financial reasons to have the money as reasons to sell his daughter; Sally recognizes *saving the life of one child* as a reason to sell the other. In the daughter-selling example, the *financial considerations—paying off bills and having money to invest*—are the considerations I reject as reasons to sell my daughter. If my daughter's life would be better

29. JOSEPH RAZ, THE MORALITY OF FREEDOM 348 (1986).

with the stranger, *that*, perhaps, would be a reason to sell her. I am not claiming that nothing can be a reason to sell one's daughter, and, more generally, I do not mean to suggest that what parental love allows and disallows as a reason is well-defined. That is certainly not true; *in general, one discovers case by case what one will and will not count as a reason.* This understanding of the incommensurability-creating commitment of parental love explains why the commitment does not prohibit those household economies in which one trades goods such as health and safety off against a variety of other goals. One's commitment can prohibit transferring one's daughter to the stranger, never to see her again, while allowing one to decide to pay off a variety of bills instead of buying the very safe, but prohibitively expensive, car in which to drive one's daughter. The essential point is that excluding reasons, possibly different reasons for different people, defines in part parental love. Given such a commitment, one cannot, consistently with the commitment, compare the excluded reasons to reasons one recognizes as legitimate bases for action. Of course, one may insist that a relevant comparison of reasons underlies and explains the commitment and its exclusion of reasons; I have answered this objection at some length elsewhere and rely on those arguments here.[30]

30. *See* Warner, *Excluding Reasons*, *supra* note 6. For convenience, I reproduce some of the argument. The objection is that there is an obvious way in which an underlying comparison could explain the exclusion. Consider the commitment itself. I do have reasons for that commitment and, given that I persist in the commitment, I must judge those reasons to be better than the reasons against having the commitment. Why not take this as the comparative judgment that underlies my excluding the money as a reason to sell my daughter? In reply, it is worth noting first that sometimes—often, perhaps—we do not form our commitments for reasons. To speak overtly autobiographically: My commitment to my daughter took hold when I saw her born. The commitment simply happened; no path of reasons led me to it. Of course, we could and sometimes do form commitments as the result of reasoned reflection; and, even in the case of my commitment to my daughter, I can—now—give reasons for it, even if those reasons did not generate it. However, my reasons for having my commitment to my daughter are very general considerations about the pleasure and value of having a daughter; they do not contain or constitute a comparison of reasons to sell my daughter with reasons not to. Indeed, they *could not* constitute such a comparison. Suppose—*arguendo*—that they did. The result would be a contradiction. I would be committed to *not* counting as reasons to sell my daughter the very considerations that, in making the comparison, I *am* counting as reasons. It is essential here that we imagine my commitment to be in full force. We are not imagining a case in which, under the impact of the stranger's offer, my commitment weakens or changes. In the latter cases, I may consider the money as a reason to sell. I am indebted to Scott Altman for emphasizing the importance of these issues about reasons and commitments. My commitment to my daughter creates a noncomparative exclusion; no underlying comparison explains the exclusion.

The point to emphasize here is that such reason-excluding commitments play a central role in the self, as the daughter-selling example illustrates. It would be natural to express the difference between me and the potentially daughter-selling Jones by saying that we are very different people. Commitments—to children, to friends, to ideals—are typically commitments involving incommensurabilities, and they are also typically commitments through which we define who we are.[31] For example, suppose someone suggests that, in an upcoming committee meeting, you should vote as political expediency demands and not as your conscience dictates. You refuse. Your commitment to following your conscience requires excluding reasons of mere political expediency. Indeed, you respond with shock and outrage, "I cannot do that. What would ever make you think I could? What *sort of person* do you think I am?" As the example shows, we define our identities both by the reasons we accept *and* the reasons we exclude. This is not to say that every reason-excluding commitment lies at the center of our self-definition. For example, one might be, as indeed many are, committed to maintaining one's health in a way that excludes financial considerations as reasons to forgo needed treatment. But this commitment need not figure prominently in one's sense of one's identity. Commitments form a continuum, ranging from those at the center of our self-definition to those that lie at the periphery.

There is an objection to consider, one that will turn out to be especially relevant in the context of rational choice theory.[32] The objection is that we have misdescribed my commitment to my daughter.[33] Why not simply say that my commitment is to regard my daughter as worth an *infinite* amount of money? Then we could say that I do compare reasons when I refuse to sell her, but that, as long as

31. I discuss self-defining commitments in detail in WARNER, FREEDOM, *supra* note 14, at 53-118.

32. Donald Regan advances this objection. Regan is "inclined to think all values are commensurable." Donald H. Regan, *Authority and Value: Reflections on Raz's* Morality of Freedom, 62 S. CAL. L. REV. 995, 1056-75 (1989). He thinks claims of the impossibility of comparison arise out of a misunderstanding and misdescription of what is really a comparison. Regan contends that when one claims, for example, that friendship is not comparable in value to money, "such a person is most likely to mean that friendship is more valuable than any amount of money, or in other words, that the value of friendship is incomparably greater." *Id.* at 1058. Of course, by "incomparably greater" Regan does not mean that there is no comparison; he means, *when we do compare the values*, one value so greatly outweighs the other that it creates the *illusion* of a genuine case of the impossibility of comparison.

33. I have recast the Regan objection in terms relevant to our discussion, in terms of commitment and exclusion of reasons.

the amount of money is finite, my reason not to sell my daughter always outweighs those considerations. Only an infinite amount of money could provide a reason that would be better than my reason to keep my daughter, and having an *infinite* amount of money is, of course, impossible.

In reply, first ask what it *means* to say I value my daughter "infinitely"? It means simply I will always refuse to sell her for any finite amount of money; *this is the only meaning given to talk of "infinite valuing" in the objection.*[34] Once this is clear, the objection collapses. We need only ask, *why* will I always refuse? We cannot answer "Because I value her infinitely." That is just to produce as an explanation the very thing to be explained. For, as we just noted, to say that I value my daughter "infinitely" is just to say that I will always refuse to sell her for any finite amount of money. This "objection" does not confront us with an alternate explanation of my refusal to sell my daughter; instead, it is simply a misleading redescription of the phenomenon for which incommensurability provides a genuine explanation. Without an alternate explanation, there is no objection.

IV. THE CONCEPTUAL CLAIM

The conceptual claim is that rational choice theory makes the wrong predictions where incommensurability is involved. The reason is that the way in which the expected utility rule constructs the action-ranking makes it clearly inconsistent with incommensurability. This may seem false. After all, incommensurability simply entails that I prefer keeping my daughter over selling her to the stranger. Isn't this simply to say that keeping my daughter ranks suitably high in my payoff-ranking—suitably high over selling her to the stranger and other similar options? If so, we can capture the effects of incommensurability in a utility function that represents the rankings induced by incommensurability.

This strategy will not work, however. The expected utility rule ranks actions by their degree of expected utility, which is the mathematical product of its utility (its place in the payoff-ranking) and the relevant probabilities. It follows that altering the relevant probabilities moves the associated action up or down in the action-ranking.

34. Some might object that the meaning is that not selling ranks infinitely high in the payoff-ranking. But what does this talk of "infinitely high" mean except that I will not sell her?

The problem is that where the rank of an action is a function of incommensurability, its rank is not responsive to changes in probability. A variant of the daughter-selling example shows why.

Suppose the stranger offers me a lottery. If I enter the lottery, I have a fifty percent chance of getting the million without having to give up my daughter, and a fifty percent chance of getting the million in exchange for my daughter. I refuse. When I refuse, the stranger changes the probabilities. He offers me a seventy percent chance of getting the million and keeping my daughter, and a thirty percent chance of getting the million and giving her up. When I refuse, the stranger offers me .... We can continue the scenario by imagining me refusing and the stranger improving his odds. From the expected utility point of view, what the stranger does is increase the expected value of entering the lottery by offering ever more favorable probabilities. When we construct the action-ranking by the expected utility rule, these increases will move entering the lottery higher and higher in my action-ranking. Assuming (for the moment) that not selling my daughter has some finite expected utility, at some point the stranger will succeed in offering a lottery that has a greater expected utility. The expected utility rule predicts that I will accept the offer at that point.

But this prediction is wrong. I will never accept the stranger's offer since I treat the monetary considerations as simply irrelevant to my decision. Whatever reason they provide to sell my daughter is a reason I *exclude*; instead, I proceed exactly as if the money provided *no reason at all* to sell my daughter. So, insofar as I am rational, I will not sell my daughter, for I have a reason not to sell (my love for my daughter), and no reason—no reason I will consider—to do otherwise. This means that not selling my daughter *will always* rank higher than any action that involves selling her to the stranger *no matter what the relevant probabilities.* Let us express this by saying that the action of not selling my daughter occupies a *rigid* place in my action-ranking. Its position vis-à-vis relevant alternatives cannot be changed by changing the relevant probabilities.

The way in which I *value* my daughter—my commitment to her—explains this action-ranking rigidity. My commitment guarantees that I will never sell her to the stranger. That is, the *action* of not selling her will always rank higher than the *action* of selling her in my *action*-ranking. Incommensurability does not concern the place of a payoff in the payoff-ranking, but the place of an action in the action-ranking.

In such cases, value plays a role in determining directly (and not via the payoff-ranking) the action-ranking: It directly creates a rigidity in that ranking. Value cannot be confined to a role in constructing the payoff-ranking.

The obvious objection is that we can indeed confine value to this role if we are willing to assign an *infinite* utility to the payoff of keeping my daughter. If we assign infinite utility, then no matter what changes one makes in relevant probabilities, the action of not selling will always rank higher than any competing action. In this way, the expected utility rule can yield rigid action-rankings.[35]

To see why this reply is inadequate, recall the two claims we distinguished earlier—the purely predictive claim and the explanatory claim. The purely predictive claim is that the combinatory procedure that rational agents use to construct action-rankings is "input-output" equivalent to the expected utility rule: Given the same payoff-ranking and probabilities as input, the rational person and the expected utility rule will yield the same action-ranking as output. The "infinite utility" reply preserves the purely predictive claim. Using infinite utilities will yield rigid action-rankings and hence generate correct predictions. The problem is that using infinite utilities conflicts with the explanatory claim. The explanatory claim is that the combinatory activity of rational agents explains the input-output equivalence. Rational agents combine payoff-rankings with probability considerations to obtain action-rankings, where increases or decreases in payoff-ranking and/or related probabilities result in corresponding increases or decreases in action-ranking position.

To posit "infinite utility" in order to achieve rigidity in action-rankings is *to concede that, sometimes, probability considerations are simply irrelevant, playing no role in the construction of the action-ranking*. Where incommensurability is involved, the rigidity in the action-ranking is *not* the result of combining payoff-rankings with probability considerations. So, now we have *two* explanations of why the input-output equivalence holds. *Sometimes*, where there is no rigidity in the action-ranking, probability considerations play an essential combinatory role in determining action-rankings, and the expected utility

35. Is assigning infinite utility, strictly speaking, necessary? Isn't all we need to do to ensure that the payoff of not selling my daughter always has a sufficiently high finite utility that the action of not selling my daughter always outranks the action of selling her no matter how the probabilities change? But given any finite utility, take an appropriately small probability and the rule will yield the wrong prediction.

rule—without recourse to infinite utilities—represents this process. *Other times*, where incommensurability is involved, probability considerations *do not matter*. The sort of combinatory activity represented by the ordinary, finite utility instances of the rule is really irrelevant to determining the place in the action-ranking. The rigid place is guaranteed by the way the person values, independent of combination with probabilities.

Of course, we can acknowledge all this and, if we like or insist, still use the infinite utility representation in cases of incommensurability. All I am urging is that our explanatory understanding of the theory has changed dramatically. And, of course, this is to *change the theory to a different one*. A sufficiently large change in explanatory understanding *is* a change to a different theory. It is in this sense that the conceptual claim is true: To incorporate incommensurability into rational choice theory is to change the theory in such a dramatic way that we really have a different theory.

Now let us turn to the empirical claim. The empirical claim is that incommensurability explains some of the empirical failures of rational choice theory. Conclusions here must be tentative as no one has devised or conducted experiments that could reveal incommensurability's role. My point in what follows is that incommensurability is a *possible and plausible* explanation for some of the failures of rational choice theory.

V. INCOMMENSURABILITY AS AN EXPLANATION FOR EMPIRICAL FAILURE

We will focus on two examples: voting behavior and "excess" cooperation. We begin with voting behavior for two reasons. The first is that it clearly illustrates incommensurability's potential explanatory role. Incommensurability provides a simple and satisfying explanation of (some) voting behavior. This contrasts sharply with the dismal failure of rational choice theory to explain voting behavior. The second reason is that voting behavior is an excellent example of the sort of behavior lawyers are concerned with. Voting behavior has both an empirical and a normative dimension. Normatively, democratic theory generally assumes that more voter participation in elections is better than less. The empirical task is, of course, to explain why people do in fact vote. The empirical explanation may reveal why people vote in normatively desirable numbers, or it may reveal why they do not. In the latter case, the empirical explanation may reveal what we

need to do to achieve the normatively desirable result. This marriage of empirical and normative considerations has long attracted legal scholars.[36]

A. Voting Behavior

People vote in the tens of millions in national elections. However, where

> voting is optional and altruism rare, the equilibrium posited [by rational choice theory] for voter turnout in large electorates is one in which very few people, if any, bother to go to the polls. Many scholars, including several working within the rational choice tradition . . ., therefore view voter turnout as a case in which rational choice theory fails empirically.[37]

Rational choice theorists' treatments of voting illustrate

> the characteristic ways that rational choice theorists have reacted to discrepancies between theory and observation. In their resolute determination to declare some variant of rational choice theory victorious over the evidence (or, alternatively, to declare peace with honor through an artful domain restriction), rational choice theorists have trotted out an astonishing variety of conjectures about costs and benefits of voting.[38]

We will consider the "astonishing variety of conjectures" only briefly, but we do need to see why the classical expected utility rule fails to explain voting behavior.

The first step is to calculate the expected utility of voting. Voting is an action that yields payoffs. Three payoffs are relevant. First, one can cast a *decisive* vote (if one had not voted the candidate would not have won); second, one can cast a *nondecisive* vote (the candidate would have won or lost without one's vote); and third, one does one's civic duty by voting and avoids the guilt of not voting, and so on. Let us call this third element the *direct benefits* of voting. Direct benefits are any payoffs of voting other than casting a decisive or nondecisive vote.[39]

To calculate the expected utility of voting, we need a utility function that ranks the payoffs. It is convenient to use a monetary scale where the utility of casting a nondecisive vote is $0. Intuitively, we

36. This marriage attracted many legal realists. See Joseph Singer's excellent review, Joseph W. Singer, *Legal Realism Now*, 76 Cal. L. Rev. 465 (1988) (book review).

37. Green & Shapiro, *supra* note 1, at 47.

38. *Id.*

39. This presentation follows Green & Shapiro, *supra* note 1, at 47-71.

can think of this as representing the answer "$0" to the question, "How much would you be willing pay to ensure that your vote fails to decide the election?" It also simplifies matters to assume that voting yields the direct benefits with absolute reliability, with a probability of 1. Using P for the probability of voting decisively, and 1 for the probability of obtaining the direct benefits, the expected utility of voting is:

$P(u(\text{decisive vote})) + (1 - P)(u(\text{nondecisive vote})) + (1 \times \text{direct benefits})$.

Since $u(\text{nondecisive vote}) = \$0$, we can simplify to:

$$P(u(\text{decisive vote})) + \text{direct benefits}.$$

Now, according to the theory, a rational person will vote only if the expected utility of voting is greater than the expected costs. Using C for the expected costs, the rational person votes when and only when

$$P(u(\text{decisive vote})) + \text{direct benefits} > C.$$

Will a rational person vote? That depends on the values of P, u(decisive vote), the direct benefits, and C. Ignore the direct benefits for the moment. The rational person votes if $P(u(\text{decisive vote})) > C$. Suppose P is .00001, a 1 in a 100,000 chance, which would be quite high for a national election.[40] Now what is u(decisive vote)? What is the utility of voting? One way to capture this is to ask what one would be willing to pay to unilaterally decide the election. Suppose one would pay $10,000. Then the expected value of voting is a dime (ignoring the direct benefits). The cost of voting is certainly greater than that. So in a national election in which everyone assumes that people will vote in the tens of millions, no rational person should vote. But people do vote—in the tens of millions even though the theory predicts they will not.

It predicts this, that is, when we ignore the direct benefits. The obvious way to save the theory is to hold that the direct benefits are sufficiently large to make the expected value of voting outweigh the costs. This suggestion takes a variety of forms depending on just how one conceives of the direct benefits. They could be the pleasure from voting, a sense of doing one's civic duty, the avoidance of guilt from not voting, and so on. These suggestions all claim that we have the payoff-ranking wrong. If we take the direct benefits into account,

40. The 1 in 100,000 chance is the probability chosen by GREEN & SHAPIRO, *supra* note 1, at 49. Game theory treatments allow P to vary, but such treatments prove unsuccessful. *Id.* at 57.

both the payoff of a decisive vote (plus the direct benefits) and the payoff of a nondecisive vote (plus the direct benefits) move up significantly in the payoff-ranking.

The worry here is empirical testability.[41] The suggestion should be accompanied by a way, *independent of observing voting behavior*, to empirically measure the extent of the direct benefit. Then we could empirically test to determine whether the direct benefits were sufficiently large so as to outweigh the cost of voting. Otherwise, how are we to test the theory? Unfortunately, none of those who put forward the "direct benefits suggestion" provide an appropriate way to assess the postulated benefit independently of voting behavior. This does not mean the suggestions are false. On the contrary, we *already know* that some people vote for the pleasure it gives them, some to avoid feeling guilty, some out of sense of civic duty, and so on. This is just a matter of the common sense, background cultural knowledge one acquires as one grows up in our society. The point is that rational choice theory promises more. It promises a mathematically precise, empirically testable theory that can grow and improve in the way scientific theories do. It fails to keep its promise in the area of voting behavior without a way, independent of observing voting behavior, to empirically measure the degree of direct benefits associated with voting. Without this, we are simply translating cultural knowledge we already possess into the language of rational choice theory. Showing how the insights of one way of talking can be expressed with increased clarity and rigor in another way of talking is a traditional philosophical undertaking, and I would be among the last to object to such translations. They are often illuminating. But they are not empirical science.

Now let us turn to the incommensurability explanation. Suppose—*arguendo and for the moment*—that people have a commitment to voting that makes the expected costs of voting irrelevant to their decision to vote. So, just as in the daughter-selling example, the decision is easy: One has a reason to vote (provided by one's commitment) and no cost-provided reason—no such reason one will consider—to do otherwise. Accordingly, one votes. The obvious question, of course, is: Do people really have such a commitment? It should not be too difficult to test the hypothesis that they do. We would need to look for a relevant rigidity in action-rankings. Is the decision to vote appropriately unresponsive to relevant changes in

41. GREEN & SHAPIRO, *supra* note 1, at 47-71, makes this point.

probability? We can in this way "operationalize" claims of incommensurability. It should not be too difficult to devise experiments that would detect action-ranking rigidity, where the design allows us to rule out other possible explanations for the rigidity besides incommensurability.

Testability is not the worry; rather, the problem is that we already know that people *lack* incommensurability-creating commitments to voting. After all, no one disregards *all* the potential costs of voting. Raise the costs high enough—cut off arms and legs, say—and all of us will avoid the polls. There is a clear contrast with the daughter-selling example. In that case, it is not at all implausible to hold that I will not sell my daughter at *any* price. In voting, however, one will "sell" (that is, one will not vote) at some price. So how can we explain voting behavior by an incommensurability-creating commitment?

The answer is that not every incommensurability-creating commitment is as unlimited as my commitment to my daughter. Consider an example of a more limited incommensurability.[42] Suppose Jones and I are avid sailors. We each spend about $400 a month on sailing (yacht club dues, boat storage, and so on). For Jones the choice is between spending the $400 on sailing and saving it for retirement. Jones sails because he takes his reasons to sail to be better than his retirement-saving reasons not to. Like Jones I have reasons to save money toward retirement, but unlike Jones I do not see these reasons as reasons not to sail—even though sailing means not saving. My attitude is similar to my attitude toward my daughter: I am committed to sailing where it is in part constitutive of my commitment that I refuse to recognize saving for retirement as a reason not to spend money on sailing. In this sense, I could literally be said to love sailing. But, in this case, my love is limited. To see how, suppose the cost of sailing suddenly rises to $1000 a month. Jones would stop sailing at this point, for he would take his reasons to save $1000 toward retirement as a reason not to sail, and he would take that reason to be stronger than his reason to sail. I would also stop sailing, but not quite for the same reasons. Confronted with the need to pay $1000, I would abandon—or at least change the nature of—my commitment to sailing. I would—sadly and reluctantly—acknowledge that saving $1000 a month for retirement *was* a reason not to sail. I differ from Jones in

42. The example is adapted from Richard Warner, *Incommensurability as a Jurisprudential Puzzle*, 68 CHI.-KENT L. REV. 147 (1993).

that my decision involves a fundamental change in my attitude toward sailing, a ceasing to love, a change that does not occur in Jones.

Insofar as people have an incommensurability-creating commitment to voting, they surely have a *limited* one of the sailing example sort. Rational choice theorists will be quick to contend that they can easily accommodate such a commitment within the rational choice theory framework. The utility function simply has a certain structure: Voting outranks the relevant payoffs up to a certain point, and fails to do so beyond that point. The problem is the one we noted earlier in discussing the conceptual claim: unresponsiveness to changes in probability. Within its limits, limited incommensurability will be unresponsive to changes in probability. Within the relevant limits, the place of voting in the action-ranking will be rigid. It will not change as relevant probabilities change. So the action-ranking generated by the expected utility rule will not accurately represent the real action-ranking.

Still, one may rightly object that we have offered no positive reason to think that people have even a limited incommensurability-creating commitment to voting. We began by *assuming* such a commitment for the sake of argument. So why think people have such a commitment? To begin with, it is clear that not *everyone* does; widespread cynicism and apathy are sufficient to show that. However, many do recognize a *civic duty* to vote, where this duty should be interpreted as involving a commitment to regard (at least to a considerable extent) the costs of voting as irrelevant to the decision to vote. This is, of course, an empirical claim about prevailing normative attitudes, but it is certainly a prima facie plausible claim. Many find it reprehensible to base the decision to vote on either the likelihood that one's vote will be decisive or on considerations about the time and inconvenience involved in voting. Their attitude is precisely that such expected cost considerations should (within limits) be irrelevant to the decision to vote.

B. "Excess" Cooperation

Now let us turn to "excess" cooperation. Rational choice theory predicts that people will not cooperate in situations in which they in fact do cooperate quite often. Here is a classic example. Imagine two people, call them A and B. We offer them this deal: We will give A $100, or give A and B together $500, *provided that they first agree on how to split the $500 between them.* We also impose the following rule:

A is allowed to propose a split to B, and B may either accept or refuse the offer, but may make no counteroffer. Refusal means that A gets the $100 and B gets nothing. Assuming that the $500 cannot be divided into amounts smaller than $1, rational choice theory predicts that they will agree that A gets $499 and B gets $1.[43] Contrary to this prediction, many people will split the money 50-50. Indeed, more than 2000 experiments since the 1950s show that one fourth to two thirds of the people involved in similar situations cooperate more than rational choice theory predicts they will.[44]

Incommensurability provides an easy explanation—if we assume that many people have a limited incommensurability-creating commitment to fairness that makes the reasons provided by the rational choice theory calculations irrelevant to their decision. (This is not to say that other explanations are wrong; the claim is that incommensurability is *one explanation among others.*) The issues that arise here are essentially the same as those that arise in incommensurability explanations of voting behavior; so, instead of going over the same ground again, let us turn to another well-known situation—the Prisoner's Dilemma—in which rational choice theory predicts less cooperation than in fact occurs. The point of doing so is to illustrate how the standard representations of the Dilemma misrepresent incommensurability-based cooperation.

The Dilemma is very familiar, but, since our concern is precisely with its standard representation, it is worth briefly running through a version of the standard presentation. Suppose you and I have committed two crimes, one serious offense and one relatively minor one. When we are arrested the police interrogate us separately. Unless one

43. The calculation in brief: Note that the relevant payoffs for A and B range from getting $0 to $500, and assume that the dollar value of a payoff represents its utility. Suppose A offers B a split of $499 (for A) and $1 (for B), with the proviso that if B refuses, A will take the $100 and B will get nothing. The expected utility for B of accepting the offered split is $1; the expected utility of refusing is $0. As a rational expected utility maximizer, B will accept if A makes the offer. And A will make the offer as long as A knows or assumes that B is an appropriately informed rational expected utility maximizer. In such a case, offering the $499/$1 split has the highest expected utility for A since any other offer results in A's getting less money.

44. Richard H. McAdams, *Cooperation and Conflict*, 108 HARV. L. REV. 1005, 1011 (1995). It is amusing to note the results of Merrill Flood, a RAND researcher who could arguably be credited with discovering the Prisoner's Dilemma. He informally tested people's willingness to cooperate in situations in which rational choice theory predicts they will not. In 1949, he offered two RAND secretaries the deal described above: He would give one $100, or give them both $500, provided that they first agreed on how to split the $500 between them. They split 50-50. WILLIAM POUNDSTONE, PRISONER'S DILEMMA: JOHN VON NEUMANN, GAME THEORY, AND THE PUZZLE OF THE BOMB 102 (1992).

of us confesses, there is not enough evidence to convict us of the serious crime. However, even if we do not confess, there is enough evidence to convict each of us of the lesser crime, which carries a two-year sentence. The District Attorney offers you a deal. The deal has two parts: (1) if you confess and I do not, you go free, and I get ten years' imprisonment for the serious crime; (2) if we both confess to the serious crime, we both get five years. The District Attorney tells you that, at the same time that he is speaking to you, someone else is offering me the same deal. This matrix represents the situation:

		Me: C	Me: NC
You	C	5 / 5	10 / 0
	NC	0 / 10	2 / 2

The relevant payoffs here are: going to prison for ten years; going for five years; going for two years; not going at all. Assume that the years in prison represent the relative ranks of the payoffs in the payoff-ranking.

It follows that confessing is the action with the greatest expected utility. If I confess, then it is better for you to confess: You get five years instead of ten. If I do not confess, it is better for you to confess: You go free instead of getting five years. I will either confess or not confess, so it is better for you to confess. Since I know that, insofar as you are rational, you will confess, then it is better for me to confess. Thus we both will confess and get five-year sentences.[45]

Now suppose that we both have a limited incommensurability-creating commitment to fairness, and that we both understand this commitment to make the foregoing expected utility calculation irrelevant to our decision. Assuming that we both know and know that we

45. This is the *third* best alternative for each of us. We both prefer the *second* best alternative of mutual nonconfession, but if we each act rationally we cannot realize that alternative. More technically, this shows that the Nash equilibrium is not always Pareto-optimal.

both know that we have and will honor this commitment,[46] neither of us will confess. It is natural—but mistaken—to represent this situation by simply changing the payoffs in the matrix. For example, we might set the value of confessing, and hence not acting fairly, at some suitable high value—say, as equivalent to 100 years in prison:

		Me: C	Me: NC
You	C	100 / 100	10 / 100
	NC	100 / 10	2 / 2

Given these payoff values (and the knowledge assumption made above), rational choice theory correctly predicts that neither of us will confess.

But this is the wrong representation. The years represent relative position in the *payoff*-ranking, not the action-ranking, and, as we argued earlier, incommensurability-creating commitments cannot be captured by alterations in the payoff-ranking. Of course, we could interpret the matrix as representing the *action*-ranking, but this is not the standard interpretation *and* we would still fail to represent the rigidity of the ranking, its unresponsiveness to probability considerations. Rather, the point is not that we need a new representation. The point is that commitments to fairness can take us completely out of Prisoner's Dilemma situations.

If we are to arrive at a fully satisfactory predictive and explanatory theory of human action, we must acknowledge incommensurability-creating commitments. It is absurd not to. If we are to have a science of human action, it had better be *humans* that we study. Of course, rational choice theorists will—correctly—respond that they do not ignore incommensurability-creating commitments. They assign them a role in determining the payoff-ranking. But the issue is

46. More precisely, this must be common knowledge. Something is common knowledge between us when and only when you know it; I know it; you know I know it; I know you know it; and so on ad infinitum (or until some natural limit is reached).

whether we can confine them to that role. We know from the conceptual result that we cannot always do so. These conceptual and empirical observations have normative significance.

VI. NORMATIVE SIGNIFICANCE

To see the normative dimension, recall that commitments involving incommensurability can play an essential role in the self-definition by which we constitute our identities as persons. For this same reason, incommensurability also plays a central role in public decisionmaking. To see why, simply ask, what sort of state do we wish to live in? One that makes public policy by ignoring the self-defining commitments that make us who we are? Or one that takes such commitments appropriately into account? Surely the latter. To live in the former state would be to live in a state that served some anonymous and anodyne abstract "citizen" drained of the self-constituting individuality that defines who we are. When, for example, the state in ordering custody invokes its power to take my daughter from me and to define the terms on which I may see her, on what grounds do I want the decision made? Should the state regard time with my daughter as fungible with money? Should it see visitation as something to be traded off against dollars? In making such decisions, I want the state to give due weight to the fact that, in my eyes, my daughter is not for sale. This commitment is definitive of my relationship with my daughter and to ignore it is to ignore the very relationship that should be a focal point of the custody hearing. The state should serve *us*, not some postulated "citizens" shorn of the commitments that make us who we are.

To bring these observations to bear on rational choice theory, let us first emphasize the fact with which we began: Rational choice theory is a useful tool, often revealing opportunities to enhance overall welfare that would otherwise escape our notice. The normative point is that we should by no means let the tool we have created define us in its limited image—limited because of the excessively confining role to which it consigns value.

APPENDIX

Where the incommensurability of reasons plays an essential role in generating actions, rational choice theory makes the wrong predictions. This means that, in such cases, one or more of the Von Neumann/Morgenstern axioms must be false of rational agents. Otherwise the expected utility rule would accurately predict the rational person's action-ranking. In this Appendix, I examine the axioms and provide relevant incommensurability-based counterexamples to them. I begin by presenting the axioms. The presentation should be accessible even to those unfamiliar with the mathematical details of rational choice theory.

I. THE AXIOMS

The axioms concern preferences.[47] *People* have preferences, and in stating the axioms, we will refer to "the person" meaning any person whatsoever. We begin with two axioms about preferences for payoffs. The first axiom—the completeness axiom—is that, for any two payoffs, either one is indifferent between them, or prefers one to the other. More formally:

Completeness: For any two payoffs a and b, either the person is indifferent between a and b, or the person prefers a to b, or b to a.

The completeness axiom assures us that a person has preferences over a certain range of payoffs. We can represent this range by $a_1, \ldots, a_r$, where r is a number equal to or greater than 1.[48] Here a_1 is the least preferred of all the payoffs, and a_r is the most preferred, and if $i < j$, the person prefers a_j to a_i.

The second axiom—the transitivity axiom—asserts that a person's preferences are transitive. That is, if one prefers a first payoff to a second, and the second to a third, then one prefers the first to the third as well.

Transitivity: For any three payoffs a, b, c, if the person prefers a to b and b to c, then the person prefers a to c.

These two axioms say nothing about *lotteries*. As we noted in the text, our actions enter us into lotteries. We need some axioms explicitly about lotteries. We provide them by considering a special lottery

47. This axiomatization is adapted from Hampton, *supra* note 21.

48. The number of payoffs need not be finite; the theory can be extended to handle a denumerable infinity of payoffs.

involving the least preferred payoff, a_1, and the most preferred payoff, a_r. We will use the bracket notation "[. . .]" to represent lotteries. Thus: $[P(a_1), (1- P)(a_r)]$, where P is the probability of getting the payoff a_1; the lottery with a P chance of a_1 and a 1- P chance of a_r. To see the idea behind the next axiom, let a_1 be paying out one million dollars, and let a_r be receiving one million dollars. Suppose you are offered the following choice: You may have $500 for certain, you may enter a lottery in which there is a probability P that you will have to pay out one million dollars, and a probability $(1 - P)$ that you will receive one million dollars. Assuming you are risk-neutral and have no other relevant objections to gambling, there will be a value for P where you are indifferent between receiving the certain $500 and entering the lottery. The next axiom—the continuity axiom—asserts that given *any* payoff, we can always find such a value for P.

Continuity: For each payoff a, there is a probability P such that the person is indifferent between a and $[P(a_1), (1- P)(a_r)]$.

The next axiom—monotonicity—concerns the behavior of probabilities *in* lotteries. We will not be concerned with this axiom but state it for the sake of completeness.

Monotonicity: With respect to the lotteries $[P(a_1), (1 - P)(a_r)]$ and $[P^*(a_1), (1 - P^*)(a_r)]$, the person prefers the former to the latter only if P is greater than P^*; the person is indifferent between the two only if P is equal to P^*.

To introduce the next axiom, suppose the person is indifferent between a payoff a_i and the special lottery $[P(a_1), (1-P)(a_r)]$. Will this change if a_i and $[P(a_1), (1 - P)(a_r)]$ are themselves part of some more complicated lottery? Rational choice theory answers "no." The following axiom—substitutability—captures this.

Substitutability[49]: Let [. . . $P(a_i)$ . . .] be any lottery involving $P(a_i)$, and suppose that the person is indifferent between a_i and $[P^*(a_1),(1 - P^*)(a_r)]$. Then the person is indifferent between [. . . $P(a_i)$ . . .] and [. . $[P^*(a_1), (1 - P^*)(a_r)]$ . . .], where [. . $[P^*(a_1),(1 - P^*)(a_r)]$ . . .] is the result of substituting $[P^*(a_1),(1 - P^*)(a_r)]$ for $P(a_i)$ in the lottery [. . . $P(a_i)$ . . .].

A final axiom is required because sometimes the payoff of entering a lottery is itself a lottery. For example, I decide to drive to my favorite restaurant despite the snow and ice on the road. The choice

49. This is also called the independence axiom.

enters me into a lottery where the payoffs are as follows: (1) get stuck in the snow and ice and never make it to the restaurant; and (2) arrive at the restaurant with a 50-50 chance that the excellent, as opposed to the merely competent, chef will be cooking. The payoff "arrive at the restaurant" is itself a lottery which we can represent this way: [.5(get there and have the excellent chef), .5(get there and have the merely competent chef)]. I choose to go because I think the probability of getting to the restaurant—call that P—is much greater than the probability of not getting there, $(1 - P)$. So my choice enters me into this lottery: [P([.5(get there and have the excellent chef), .5(get there and have the merely competent chef)]), $(1 - P)$(get stuck)]. We can reduce this compound lottery to the following simple one: [(P x .5)(get there and have the excellent chef), (P × .5)(get there and have the merely competent chef)), $(1 - P)$(get stuck)]. The final axiom asserts that a rational person should be indifferent between the simple and the compound lottery.

To state the final axiom, let $L = [P_1(a_i), \ldots, P_n(a_j)]$ be a lottery. When it occurs as a payoff in another lottery, there will be a certain probability of getting that payoff; we can represent this as follows: [. . . $P^*(L)$ . . .]. We construct the simple lottery by replacing $P^*(L)$ with $Q_1(a_i), \ldots, Q_n(a_j)$ where $Q_i = P^* \times P_i$. The claim is that the person will be indifferent between [. . . $P^*(L)$ . . .] and [. . . $Q_1(a_i), \ldots, Q_n(a_j), \ldots$], where $Q_i = P^* \times P_i$. Now we can state the final axiom—the reduction of compound lotteries.

Reduction of compound lotteries: Let $L = [P_1(a_i), \ldots, P_n(a_j)]$ be a lottery involved in [. . . $P^*(L)$ . . .]. Then the person is indifferent between [. . . $P^*(L)$ . . .] and [. . . $Q_1(a_i), \ldots, Q_n(a_j)$ . . .], where $Q_i = P^* \times P_i$.

When a person satisfies these axioms, then there exists a utility function that represents the person's payoff-ranking, and that, when used in conjunction with the expected utility rule, represents the person's action-ranking.

II. INCOMMENSURABILITY-BASED COUNTEREXAMPLES

We have denied that the expected utility rule does always correctly represent a rational person's action-ranking. So we are committed to the position that the axioms do not always hold of a rational person. Even more strongly, we are committed to the position that

they fail to hold *because of incommensurability*. I offer incommensurability-based counterexamples to four of the axioms. The first two examples are, I contend, compelling; the second two less so, but still interesting enough to discuss. I assume in all cases that the people described in the examples are rational, a claim I have defended elsewhere.[50]

A. Completeness

For a counterexample to the completeness axiom, consider the "Sophie's choice" situation in which a person faces a mutually exclusive choice between two alternatives and does not prefer one to the other. Thus: A sadistic Nazi officer apprehends a mother and her twin five-year-old children, Suzy and Johnny.[51] To satisfy his sadism, the officer says he will kill one of the children and asks the mother to choose which it shall be. If she does not choose, he will kill them both. The mother's dilemma is that it is constitutive of her love for each child that she refuses to recognize saving the life of one as a reason to kill the other. Yet her love for each child gives her a reason to save that child, and both will die if she does not sacrifice one. However, to choose one over the other requires that she cease to love one or at least that she cease to love one in the way she presently does. This is not something she can do at will even if she should want to. This makes rational decision impossible. There is no way she can act for reasons. It is not even possible for her to rationally decide by, for example, flipping a coin. To do so is to count saving the life of one as a reason to choose that the other should die. As long as she loves both children, she does not recognize the existence of such a reason. She does not prefer—would not choose—Suzy's life over Johnny's nor Johnny's over Suzy's.

B. Transitivity

Imagine that you are a political activist trying to rally the people against an exploitative and dictatorial regime. You value your political cause over your life; you would not betray your cause even at the cost of your life. Unfortunately, you are apprehended by government forces that insist that you publicly recant your position; however, instead of threatening you with torture and death, they apprehend an innocent person off the street. Call her Jones. They threaten to kill

50. *See* Warner, *Excluding Reasons*, *supra* note 6.

51. This example forms the central theme of William Styron, Sophie's Choice (1976).

Jones unless you recant. You comply. Your reason is that, while you have pledged your life for your cause, Jones has not pledged hers, and your view is that you cannot decide *for her* that your cause is worth *her* life. You have a commitment to respecting the life of another that excludes achieving political goals as a reason to let another person die. You make this decision even though you would save your own life in preference to Jones' if, for example, you both needed an organ transplant to live and there was just one organ; you would choose the organ for yourself.[52] So, you prefer not betraying your cause to continuing to live; you prefer your continuing to live to Jones' continuing to live; yet you prefer Jones' continuing to live to betraying your cause. This is an intransitive set of preferences.

Of course, we could avoid the intransitivity by describing preferences in a more fine-grained way. Thus, you prefer not-betraying-your-cause to continuing-to-live-when-the-alternative-does-not-involve-anyone-else's-death; you prefer your-continuing-to-live-when-doing-so-does-*not*-mean-betraying-your-cause to Jones'-continuing-to-live. You prefer Jones'-continuing-to-live-to-betraying-your-cause-when-Jones-has-not-pledged-her-life-to-the-cause. There is no intransitivity here. In general, given any particular context of choice, we can describe preferences in a way that makes them transitive. The problem is with predictive power. The goal is to predict choices over a variety of contexts given a determination of the payoff-ranking. The more we contextually define the payoff-ranking (the more we define the preferences with reference to particular contexts), the more we reduce predictive power. As we increase predictive power by rendering our descriptions less context-sensitive, we will encounter relevant intransitivities.

It is worth considering briefly a standard argument that rationality requires transitive preferences. This is the "money pump" argument. Suppose I prefer Faulkner to Hemingway, Hemingway to Fitzgerald, but Fitzgerald to Faulkner. I have a copy of a book by Fitzgerald. You have copies of books by Faulkner and Hemingway. You offer me Hemingway for Fitzgerald plus $1. I agree. Then you offer me Faulkner for Hemingway plus $1. I agree. You then offer me Fitzgerald for Faulkner plus $1. I agree. We are back where we began, and we can start over. You will "pump" dollars out of me in this way. What does the argument show? It is standardly taken to show

52. This violates transitivity—but it must violate some axiom or the expected utility rule must be followed.

that rationality requires transitive preferences, but obviously it does not. What it shows is that intransitivity can be objectionable, not that it *always* is. *When* one encounters a money-pump situation, transitivity is essential, *assuming* one wants to avoid the consequences of being "pumped." Why not maintain intransitive preferences as long as they do not involve us in such situations? As the political activist example illustrates, incommensurability-creating commitments can create such intransitivities, and such commitments often play a central role in our self-definition. This would seem a powerful reason to persist in intransitivity.

C. SUBSTITUTABILITY

It is at least arguable that voting behavior falsifies the substitutability axiom.[53] Assume you have a limited incommensurability-creating commitment to voting that makes the expected cost of voting largely but not entirely irrelevant. You are offered a lottery in which you have a ninety percent probability of voting and getting $10,000 and a ten percent chance of not voting and getting the $10,000. You can simply vote, or enter the lottery. Suppose the $10,000 amount is sufficiently high that your limited incommensurability-creating commitment does *not* result in your excluding the potential gain of $10,000 as a reason not to vote. You consider the reason, and find that you are indifferent between simply voting and entering the lottery.

You enter the lottery and the payoff you get is "$10,000 and not voting." Now you are offered another choice: Vote, or enter a lottery in which you have a fifty percent probability of voting and getting $10,000 and a fifty percent chance of not voting and getting the $10,000. It follows from the substitutability axiom that you should prefer voting to entering this lottery, for you are indifferent between the voting and the first lottery; and, since the probabilities are less favorable in the second lottery, you will prefer the first lottery to the second.[54] So, by the substitutability axiom, you will prefer voting to the second lottery. Nonetheless, you decide to enter the second lottery. The explanation is that you have already decided to gamble with voting. It took the $10,000 lottery to get you to set aside your incommensurability-creating commitment to voting, but *now that you have done so*, it takes less of a payoff for you to be willing to trade voting for an expected gain. This point about the effect of setting aside the

53. This example is adapted from Hampton, *supra* note 21, at 217-18.

54. This follows essentially from the monotonicity axiom.

commitment is, of course, debatable, but the example is worth reflection.

D. Reduction of Compound Lotteries

Many have objected to this axiom. Daniel Ellsberg, for example, contends that the axiom

> implies that the individual is indifferent to the number of steps taken to determine the outcome. On the contrary, a sensible person might easily prefer a lottery which held several intermediate drawings to determine who was still "in" for the final drawing; in other words, he might be willing to pay for the possibility of winning intermediate drawings and "staying in," even though the chances of winning the pot were not improved thereby. A longer time-period of suspense would usually also be involved, but it need not be. The crucial factor is "pleasure of winning," which may be aroused by intermediate wins even if one subsequently fails to receive the prize. Many, perhaps most, slot-machine players know the odds are very unfavorable, and are not really motivated by hopes of winning the jackpot. They feel they have had their money's worth if it takes them a long time to lose a modest sum, meanwhile enjoying a number of intermediate wins—which go back into the machine to pay for the pleasure of the next win.[55]

Ellsberg's example is compelling, but there is, of course, nothing about incommensurability in it. We can easily add this element, however. Assume that some people have an incommensurability-creating commitment to thrill-seeking that makes the odds calculations irrelevant, within limits, to their decisions.

Many find such commitments implausible. It seems all too easy to posit them. My own view, which I will state but not defend, is that incommensurability-creating commitments are pervasive and explain a great deal of behavior.[56]

---

55. Daniel Ellsberg, *Classic and Current Notions of "Measurable Utility,"* 64 Econ. J. 528, 543 (1954).

56. Warner, Freedom, *supra* note 14, would certainly support such a view.