Exploitation and Rational Choice

Loren King Wilfrid Laurier University

Introduction

Whatever we might say about the usefulness of rational choice approaches in the human sciences (and many people do not say kind things), what are we to make of the intuitions that life typically goes well when, at the very least, our desires do not lead us in circles nor respond in self-defeating ways to seemingly irrelevant features of the world around us?

These, very roughly, are the intuitions captured by venerable consistency postulates that lurk behind rational choice approaches in the social sciences. These postulates ask that our preferences be ordered in certain ways in light of our aims and expectations. Yet these postulates have not fared well under decades of philosophical scrutiny and experimental challenge.

Ridiculed by some, discarded by others, I will nonetheless argue that these postulates are defensible as integral to an attractive account of *rational moral character*, and that this account resonates strongly with a wildly popular position in political thought: the wrong of exploitation. If you think that exploitation is typically a bad thing, then you ought to find these rational consistency postulates attractive as part of an account of self-understanding that mitigates against certain forms of exploitation.

If this argument persuades, then I think we can understand the normative aspirations of rational choice in a more constructive and ambitious way than has been the norm. The normative—that is, evaluative

Acknowledgments: I have offered variations of this argument at the 2004 annual meeting of the Canadian Political Science Association and to subsequent workshops at Wilfrid Laurier and McMaster Universities. I am grateful to participants at these sessions for much thoughtful discussion, and to Rhoda Howard-Hassman and Jan Narveson for helpful comments on an early draft. I also thank (without implicating) Peter Stone and Alex Voorhoeve for helpful conversations about rationality, and three anonymous reviewers from this journal for their constructive criticisms and suggestions.

Loren King, Department of Political Science, Wilfrid Laurier University, 75 University Avenue West, Waterloo, Ontario, Canada N2L 3C5, lking@wlu.ca

Canadian Journal of Political Science / Revue canadienne de science politique
44:3 (September/septembre 2011) 635–661 doi:10.1017/S0008423911000527

and prescriptive—character of rational choice has long been recognized (see, for instance, Hardin 2001) but has typically been understood in nonmoral terms (for example, McClennen, 1990), often as a kind of pragmatic, "common-sense" instrumentalism ("if you want this, then you ought to do that") which strives to be agnostic on substantive moral questions ("I never said that was the right thing to do!"). Some have tried to find a richer conception of morality within rationality (Gauthier, 1986; Narveson, 2010), but the result has been controversial (Buchanan, 1990; Christiano, 2004; Skyrms, 1996: 38–42). In contrast, my aim is not to derive the moral from the rational or to make moral judgments answer at the altar of rationality. I simply show that core elements of the rational choice framework are attractive by virtue of moral reasons rather than mere empirical accuracy, explanatory power or morally anemic pragmatism, and that these moral reasons can expect to find assent across disparate camps in political thought.

Reading much of the critical literature on rational choice in political science, we might be tempted to dismiss the approach as either a muddle of flawed explanations desperately seeking empirical relevance or a logically rigorous but morally vapid elaboration on Hume's adage that reason serves the passions. I think the first indictment misses the mark but for reasons that seem to buttress the second complaint. That complaint, however, ignores a fruitful way of understanding the rational choice framework as a rigorous part of political philosophy, helping to clarify and confront some of our moral intuitions and arguments about moral character and the wrong of exploitation.

The Trials of Rational Choice

"Rational choice theory" is an informal designation encompassing several related assumptions about, and explanations of, human behaviour and social processes. In political science, rational choice explanations are most often associated with "public choice" and "positive political economy" approaches, although they have been offered for a wide range of phenomena, including arms races (Brams et al., 1979; Kydd, 1997), voting and legislative behaviour (Austen-Smith and Banks, 1988; Baron and Ferejohn, 1989; Black, 1958; Downs, 1957), and peasant rebellions (Popkin, 1979). Problems with the approach have long been recognized and debated (see, for example, Elster, 1979; Jervis, 1988; Ostrom, 1998; Sen, 1985; Simon, 1955), and several critics impugn the approach as a feeble explanatory framework, failing most dramatically to offer plausible and interesting explanations of, and accurate predictions about, voting behaviour and other varieties of collective action (Green and Shapiro, 1994; Shapiro, 2005).

Abstract. Critics fault rational choice theory for dubious assumptions and limited explanatory power. The aims of rational choice are, however, as much normative as explanatory, and I argue that an abiding concern of political thought—the wrong of exploitation—gives moral weight to some of the more substantive assumptions underlying many rational choice prescriptions.

Résumé. Les critiques reprochent à la théorie du choix rationnel d'avancer des hypothèses douteuses et d'offrir des explications restreintes. Les objectifs du choix rationnel sont, cependant, aussi bien normatifs qu'explicatifs. J'affirme qu'une préoccupation centrale de la pensée politique – le mal de l'exploitation – donne une signification morale à certains postulats du choix rationnel.

Another critic, James Johnson (2010), worries about those who embrace game theory and its decision-theoretic foundations as the theoretical backdrop of a genuine science of politics. These advocates chant positivist mantras of "rigour," "clarity" and "verification" as they claim to derive testable models from clear assumptions; yet decades of philosophy have challenged these aspirations, painting instead a rather different picture of how science and scientists actually work (for example, Hacking, 1983; Keller, 2000; Johnson, 2006; Lane, 1996). Johnson suggests that, with their emphasis on clarity and predictive power, these advocates still cling to the idea—now deprecated in philosophy of science—that we can make largely uninterpreted "observation reports" about the world, and then use these to test hypotheses derived from axiomatic foundations. For one influential and supposedly rigorous application of game theory in political science, however, Johnson shows that key concepts and claims depend on an implicit set of interpretations and sociocultural contexts that do not fall obviously out of model assumptions and formal derivations. For Johnson—and indeed, as he stresses, for several prominent game theorists—game theory, to be at all useful in clarifying intuitions about what rationality is and what it explains, needs an interpretation that elaborates the context in which specific claims of rationality make sense.

My aims here are distinct from Johnson's, but complementary. I am interested in a moral interpretation that lets core decision-theoretic postulates do useful work in clarifying (moral) intuitions and evaluating (moral) arguments, consistent with the modest descriptive and explanatory aims that Johnson endorses for rational choice theory.

Rationality, interests and choice

At the most general level, rational choice explanations share a sense that we can best explain social practices and institutions by looking to the motivations and expectations of individual agents who desire to satisfy their interests. My interpretation grounds rational choice explanations in individual psychological states, often taken to be revealed by the choices agents make.[1] These agents are self-interested in the weak sense that their motivations typically reflect their interests, which I take to be their settled needs and desires. These agents are rational insofar as their actions tend to be roughly consistent with their interests, often consciously tailored to meet their needs and satisfy their desires, given what they know about themselves and the world around them.

Very roughly, then, an appeal to rational choice involves explaining social outcomes in terms of individual choices, which in turn reflect orderings of interests. These orderings yield preferences for some outcomes over others.[2]

Interests needn't be narrowly conceived, and self-interest, as I've defined it here, is not synonymous with selfishness or pure egoism. We can and do have interests in the welfare of others, as suggested by the burgeoning literature on social preferences, which finds mounting evidence of genetic, neurological and cultural roots of prosocial, other-directed preferences. In a variety of experimental settings, subjects often seem to favour fairer over selfish choices and are often willing to incur considerable costs to punish greedy players if they can identify them (see Frohlich and Oppenheimer, 2006; Henrich et al., 2006, 2010; Ostrom, 2000, 2006; Tricomi et al., 2010). The limits of these experimental findings, and especially the degree to which they can be generalized into satisfying explanations of real-world behaviour, is a matter of some dispute (see Bardsley, 2008; Binmore and Shaked, 2010; Levitt and List, 2007; Woodward, 2008) but the extent of evidence across a variety of cultural and socioeconomic contexts, and the persistence of the result even after taking into account selection and framing effects (for example, Fessler, 2009) is impressive (Eckel and Gintis, 2010).

Yet, while we may often care deeply about what others think of us and assess our needs and desires in light of our commitments to family and friends, community and nation, these are nonetheless our preferences, our desires, our aspirations. The individual agent is, then, on this view, an explanatory primitive—which is entirely consistent with our being irreducibly social creatures.

Origins

The conceptual foundations of what is sometimes called *canonical* rational choice theory were laid centuries ago, although the idea that social outcomes reflect individual actions is arguably ancient.

There is a reasonable case for Thomas Hobbes as the first modern systematic rational choice theorist of politics—a case most often associated with David Gauthier (1969). Hobbes offered a reductionist, axiomatic account of social outcomes in terms of individual incentives and actions, but he also recognized the critical importance of the passions in motivating choices, and he saw clearly the conflict between an axiomatic normative account of rationality, on the one hand, and subjective assessments of interests, coloured by those passions, on the other.

Daniel Bernoulli's 1738 solution to the "St. Petersburg paradox" is often given as a point of origin. The paradox exposes a problem with simply using expected value—that is, the value of an outcome weighted by the likelihood of occurence, $E(e) = p(e) \cdot v(e)$ for event e—as grounds for choice. Suppose I invite you to join the following game, for a price. I flip a fair coin. If heads, you win \$2. If tails, I flip again and double the value of the prize. We continue until you win. How much should you pay to join this game? The expected value of the game, $E(g) = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8$ ..., is infinite, but should you really pay a vast sum to play a game with even odds that you'll "win" \$2 in the first round? Bernoulli's solution introduced the concepts of a utility function and diminishing marginal utility that have become ubiquitous in decision theory and economics ever since.

While these antecedents are important, the rational choice theory attacked by contemporary critics is the direct legacy of twentieth-century work in the foundations of probability theory and economics by the likes of Frank Ramsey, Bruno de Finetti, John von Neumann and Oskar Morganstern, Leonard Savage, and Richard Jeffery. The cumulative result has been an influential approach to choice under uncertainty: subjective expected utility (SEU).[3] In its various mathematical formulations, SEU formalizes the intuition that rational consistency involves ordering our desires and beliefs in ways that are not ultimately self-defeating. The idea is to define a rational choice in terms of constraints on how we order our desires in light of expected gains and beliefs about the world.

Foundations

The first constraint is an ordering principle for preference relations: among other things, our preference relation must satisfy *completeness* (for all options we can say of any pair that we either prefer one over the other or we are indifferent between them) and *transitivity* (if we prefer option a over b and b over c, then we also prefer a over c).

The second constraint, *independence*, admits of several related but distinct formulations. For instance, independence of irrelevant alternatives is a common postulate in axiomatic decision theory and, famously, in the social choice literature after Kenneth Arrow (1951): suppose I prefer a over b and am indifferent between c and d (but prefer a or b to either). I violate this form of independence if I select a from $\{a, b, c\}$ but b from $\{a, b, d\}$.[4] In what are sometimes called "substitutability" (Myerson, 1991) or "sure thing" principles (Savage, 1954),[5] independence requires that our preferences be ordered consistently in the face of risk and uncertainty:[6] expected utility must be linear in probability.

Very roughly, we could say that the first principle disciplines our choices to represent them in terms of a function to be maximized. The second principle then tells us what must be true of our preference ordering if our choices are to maximize *expected* utility, which incorporates degrees of belief about states of the world.

Critics and Caveats

Critics charge that this individualistic stance, emphasizing utility maximization by sufficiently informed bearers of coherent preferences, is too often unrealistic and unhelpful in explaining human behaviour. After all, we often face limited information of uncertain quality, and such limits must surely hamper our ability to make informed judgments about how best to achieve our goals. Indeed, we may sometimes (often?) be uncertain about what, precisely, our goals are, let alone how best to achieve them; we are rarely transparent, even to ourselves. I am not sure how often we are even *translucent* to ourselves, in the sense that David Gauthier (1986) means when he argues for constrained maximization as a rational stance in the Prisoners' Dilemma. True, we often have a reasonably high degree of confidence that others will honour their commitments, but such interpersonal "translucence" is likely attributable to repeated interactions within various social structures, a mechanism not available to Gauthier, who wants to show that morally satisfying solutions to problems of co-ordination and conflict are rational choices in those games, whether they are iterated and culturally embedded or not.[7]Our choices are made within tapestries of meanings and associations that give sense and coherence to our needs and preferences (Geertz, 1973; Taylor, 1971), and our positions in these rich tapestries shape our judgments about what ends are worth pursuing. These contextual beliefs may not be satisfactorily described by the axioms and derivations of Bayesian probability theory, central to the decision-theoretic roots of rational choice approaches. Indeed, as Binmore (2009: ch. 7) and Joyce (1999: 70-77) have noted, one of the seminal contributors to Bayesian decision theory, Leonard Savage, imagined it applying in so-called "small worlds."

For their part, rational choice advocates readily admit the implausibility of assuming full information, formidable cognitive skills and even full understanding of our own needs and desires. As I'll explain, several scholars have laboured mightily (and with some measure of success) to incorporate imperfect information, cognitive constraints, adaptive heuristics and emotional responses into their models.

Having made such concessions and refinements, however, these advocates quite reasonably argue that if, at the end of the day, our choices do not reflect some underlying coherence to our preferences[8] and if our beliefs about the world do not roughly conform to the dictates of probability theory, then we are unlikely to be successful in many facets of our lives. We will be prone to self-defeating choices and vulnerable to manipulation by those who discover our incoherent beliefs and systematic errors in judgment. From an evolutionary perspective, our desires, beliefs and strategies for dealing with the world will be of little use to us or anyone else and so would be unlikely to persist over time.

Critics can rejoin that the set of feasible choices we face, and the expected gains we associate with our actions, are not merely functions of the preferences and capabilities of self-interested utility maximizers (or satisficers, or constrained maximizers) but are importantly shaped by the norms and traditions that we have been raised into.[9] Our preferences and expectations about how others will behave are, these critics suspect, irreducibly social and historical, best understood by interpreting the (culturally embedded) narratives of our lives, rather than by deriving utility functions and calculating core points in games. To imagine that our choices can be understood chiefly in terms of utility functions over complete and transitive preference orderings, given degrees of belief about states of the world, is to misunderstand what is most interesting about human behaviour and social phenomena, namely their richly historical and meaningful characteristics.

On this view, our social context largely determines what strategies are desirable, and so the effort to frame explanations in terms of individual desires, degrees of belief about the world and the adaptive value of strategies inevitably depends upon a rich understanding of the social context within which those attributes, expectations and strategies were formed and against which they have meaning for their bearers. It is this rich contextual understanding that does the heavy explanatory lifting in our scholarship, not our formal model of rationality (see Simon, 1985: 298–300; and again Johnson, 2010).

To see the force of this sort of criticism, consider a hypothetical situation in which you have studied my behaviour for several years. I have repeatedly stated my desire to achieve some goal (home ownership, say), yet I routinely act in ways that seem obviously contrary to that goal (squandering my earnings on wine and gambling, for instance, or repeatedly making speculative and absurdly risky short-term investments). As a social scientist working within a rational choice framework, what would you do?

Most scholars would, I suspect, look for factors that give sense to my apparently self-defeating behaviour. Perhaps I have misunderstood certain key facts about the world. I may be insufficiently nimble of mind; in spite of my best efforts, my reasoning is mired in error. Alternatively, my powers of reasoning may be sound, but I have been tricked by other (rational) agents, who have deliberately misinformed me for their own purposes. Or perhaps I am sufficiently informed, able to reason correctly and free from manipulation by others, but my will is extraordinarily weak. If so, then I am something of a tragic figure, able to affirm my deepest aspirations and to formulate plausible plans to achieve them but doomed never to succeed. Or I may be informed, competent and strong-willed, but misunderstood: I may have spoken metaphorically, and you lack familiarity with the narratives and rituals required to make sense of my world view. Or I may hold moral beliefs that oblige me to forgo certain desires and thus, although I have stated a particular desire, I also possess an unstated and higher order desire to do only what I believe is morally right, even at some cost to myself.[10] Or, finally, it is possible that I was being sarcastic or deceptive, concealing my true desires and intentions, thus making me either obscure or deceitful but not necessarily irrational.

What work is being done by a formal concept of rationality here? Certainly an intuitively familiar conception of rationality is, in some respect, central to our explanatory efforts in this exercise, because what we are explaining is my apparent failure to live up to this very standard. The final hypothetical (where I am deliberately deceiving you about my true intentions) seems ideally suited to study in terms of formal rationality and strategic behaviour. In contrast, the penultimate cases (where we are confused about metaphors or morals) seem ideally suited to more historical and interpretive methods, indeed these are the critic's strongest cases. Think, for instance, of Clifford Geertz's famous analysis (1972) of the Balinese cockfight, showing how seemingly irrational betting behaviour is intelligible when we understand the broader social meanings of those gambles.

In the other hypothetical cases, our explanations would variously rely on understanding psychological factors and social structures. These explanations would, if developed with sufficient care, presumably involve careful exploration of neurological, developmental and social-historical processes that shape our beliefs, desires, abilities and expectations. In such cases we may well be less interested in the formal logic of subjective expected utility and rational choice as a tool for building explanatory models, and far more interested in the empirical facts of how we actually frame our desires and expectations, and how we think through our choices.

Chastened choice theory: better explanations, less rationality

This has been a longstanding dispute in the study of choice behaviour: early on in the development of SEU as an approach to decision making under uncertainty, Maurice Allais (1953) and Daniel Ellsberg (1961) noticed that people could violate the independence postulate in ways that did not seem intuitively unreasonable (Weber, 1998). In the Allais scenario, players became more risk averse in their choices between two essentially identical pairs of lotteries if one of the pairs had an option with guaranteed returns.

For an example of what I am calling an Allais scenario, consider two pairs of lotteries, $\{S, G\}$ and $\{S', G'\}$, over three possible states of the world with probabilities $\{p_1 = 0{:}01;\ p_2 = 0{:}10;\ p_3 = 0{:}89\}$ and expected payoffs $E(S) = \$500 \cdot p_1 + \$500 \cdot p_2 + \$500 \cdot p_3$ and $E(G) = \$0 \cdot p_1 + \$1000 \cdot p_2 + \$500 \cdot p_3$, and $E(S') = \$500 \cdot p_1 + \$500 \cdot p_2 + \$0 \cdot p_3$ and $E(G') = \$0 \cdot p_1 + \$1000 \cdot p_2 + \$0 \cdot p_3$. $E(S) < E(G)$ and $E(S') < E(G')$, but we violate independence if we claim to prefer S over G and G′ over S′. Allais, however, found that many respondents would do just this.[11]

In the Ellsberg scenario, similar behaviour emerged with the introduction of uncertainty (unknown odds) to one of two options with comparable expected returns, suggesting (independence-violating) preferences for lotteries with known odds. Further work by psychologists, perhaps most famously Amos Tversky and Daniel Kahneman (1979), has produced a rich body of experimental evidence that key tenets of SEU are routinely violated. Respondents tend to evaluate prospects against a reference point, such as the status quo, toward which they are often biased. Depending on how choices are framed, many respondents seem to be more sensitive to losses than gains, and they sometimes reverse their stated preferences (Loomes et al., 1991; Tversky and Kahneman, 1981). Such findings suggest that any satisfying explanation of real-world choices will be grounded not in elegant formalism and mathematical analysis but in empirical particulars of our brains, their evolution, and how they tend to work in particular choice settings.

For their part, choice theorists have laboured mightily to accommodate these and related experimental results. Several theorists have profitably turned to evolutionary models, weakening rationality requirements and instead attending to mechanisms of transmission, adaptation and the differential success of strategies given environmental factors and population characteristics (for example, Axelrod, 1986; Binmore, 1998; Skyrms, 1996, 2004). Others have sought to remove suspect rules from axiomatic choice theory, for example, by dispensing with independence (Machina, 1982), weakening transitivity and doing away with completeness (Fishburn, 1983) or by discarding even very weak consistency requirements (Sugden, 1985). Some theorists have offered alternative postulates which capture features we intuitively think of as essential to the idea of rationality, such as our attention to *reasons* for and against a course of action, our *intentional* stance with respect to what we desire, and *dominance*: we ought to choose from the set of options that we definitely prefer over others, however we rank (or fail to rank) the contents of that set (Anand, 1993: ch. 8).

Some critics complain that, in spite of the wealth of anecdotal and experimental evidence against the realism of SEU postulates as behavioural assumptions and regardless of all this subsequent tinkering with axiomatic foundations, formal models of political and economic phenomena continue to proliferate, grounded in the same suspect SEU foundations! Worse yet, these models see endless theoretical refinement instead of careful empirical testing (Green and Shapiro, 1994).

I think this line of criticism—however popular it remains in some quarters of political science—is now unfair. Many economists and political scientists pay far more attention than was once the norm to how, precisely, choice- and game-theoretic models are to be tested. For instance, the seminal work of McKelvey and Palfrey (1995 and 1998) on statistical solution concepts has provided a foundation for empirical tests of game-theoretic models (for example, Signorino, 1999), and fruitful debate has emerged over so-called "analytic narratives," involving historical and qualitative tests of formal models (Bates et al., 1998). Furthermore, many scholars are increasingly sensitive to how our reasoning reflects, and is importantly constrained by, historically durable norms (Chong, 2000) and deep cognitive regularities (McDermott, 2004). These norms and regularities may reflect adaptive learning, and some may be rooted in evolved neurophysiological traits; in either case they amount to adaptively valuable heuristics for coping with informationally complex problems (Bendor et al., 2003; Gintis, 2000, 2009; Gintis et al., 2003; Jones, 2001; McDermott, 2004; Ostrom, 1998).

Beyond Explanatory Choice Theory: Rationality and Normative Questions

When we weaken rationality requirements and attend to mechanisms of learning, adaptation, transmission and evolutionary dynamics, we gain explanatory purchase in modelling how agents actually make decisions. My aim, however, is to show how key elements of the *rational* choice approach can be interpreted as morally attractive *prescriptions*. Alas, neither recent innovations in explaining choice behaviour, nor earlier efforts of rational choice theorists, seem at first blush to stand on their own as morally attractive normative frameworks.

Take, for example, Brian Skyrms's demonstration that weakly dominated "fair" strategies can persist as viable solutions to classes of social interactions (1996: 28–33). Imagine a strategic interaction with players facing a set **S** of feasible strategies. A strategy $s_d \in$ S is weakly dominated if there is at least one other strategy in **S** that would do as well or better than s_d. A strategy $s_{ess} \in$ S is *evolutionarily stable* if a population playing s_{ess} cannot be successfully invaded by a competing strategy. In the ultimatum game, a prize is offered to one player on condition that it be shared with another: if the other player does not accept the share offered, then both players receive nothing. The dominant strategy is to offer the smallest possible increment when in the first role and to accept whatever is offered when in the second role (after all, accepting any positive offer is better than nothing). Yet this is rarely observed in practice; in several experiments involving a variety of players and budgets, reasonably fair offers are the norm, and unfair offers are often rejected.[12] Fair strategies that punish greed (that is, offer somewhere near 50 per cent and reject offers deviating too much from equal division) are weakly dominated by a strategy that makes fair offers but accepts any positive offer. In Skyrms's simulations, strategies propagate as a function of relative success and, in spite of being weakly dominated, "punishing" fair strategies can do rather well, persisting in a population depending on their number and spatial distribution with respect to other strategies at play.

A morally satisfying prescription of fairness, however, demands more than simply explaining how we can be fair and not face repeated losses. More generally, moral arguments for fairness must do more than explain why conformance to, or violations of, canonical rational choice assumptions occur; when and how conformance advances interests; or how violations may be advantageous (given, say, certain selection and transmission mechanisms, such as punishment norms, formative cultural institutions and evolved emotional responses to risk or insult). Certainly if we think that "ought implies can" then we may be pleased to discover that innovations in evolutionary game theory allow us to do what we ought to do, in spite of violating seemingly plausible postulates of Bayesian decision theory, but we still need to argue in favour of that "ought" claim. We are still faced, that is, with *the normative questions* of what we ought to do and why we ought to do it (Korsgaard, 1996: esp. 14–16).

Why take up such irreducibly normative questions? An entire generation (at least) of social scientists was raised on the mantra that their job is to explain choices, not to judge them, and that we can safely take preferences as largely exogenous to our models of choice behaviour. Why not simply allow that the normative content of decision and game theory is no more and no less than rigorous variations on Hume's adage that reason is a slave to the passions? So long as players can rank their ends in a coherent fashion and make informed estimates about the world around them, then decision and game theory can serve the modest prescriptive role of clarifying means to desired ends ("if you want A then you ought to do x, y and z"), and that is the proper limit of the approach's normative aspirations.

Some game theorists are content with this Humean concession (for example, Binmore, 2009: 16–18); indeed, to them it is no concession at all: why should game theorists be in the business of moral and political philosophy? Even supposing that there are irreducibly normative questions (and some philosophers dispute this), we all benefit from an intellectual division of labour: social scientists can leave to philosophers the moral evaluations and prescriptions of particular means and ends. We can then limit the normative content of explanatory choice theory to only those means and ends thus scrutinized and endorsed.

In contrast, I think we can be less modest about the moral appeal of some rational choice prescriptions. If rational choice theory is to provide morally appealing prescriptions, then their appeal must ultimately rest upon some value judgment about what ends our choices ought to be oriented toward and what means are morally acceptable in pursuit of those ends. These needn't be arguments about particular means and ends; they can be more generic: "avoid ends that leave you regretting your choices" or "avoid means that leave you vulnerable to being exploited by others in particular ways." The arguments that sustain such value judgments will, then, determine both the plausibility and attractiveness of the prescriptions offered, and I think an argument is available that makes the foundations of rational choice theory fertile ground for attractive prescriptions about our desires and choices.

How to make this case? One approach would be to enumerate a series of actual applications of rational choice models, interpreting them not as explanations of choice behaviour, but as endorsements of particular strategies. My burden would then be to show why these endorsements are morally attractive. Here, too, I might contrive specific examples of how violations of rational choice postulates, on the one hand, and derived prescriptions with respect to strategies, on the other, leave us vulnerable to exploitation, or persistent regret (or, likely, both).

I will offer some passing examples of how improperly structured preferences might be exploited by clever salesmen and campaign strategists, whose interests can be served by provoking consumer and voter preference reversals through framing effects or by complicating some issue space in ways that provoke citizens to use independence-violating heuristics to select a favoured policy or associated candidate. My method, however, is ultimately deductive, not inductive, and so while these examples may clarify matters, the argument does not depend on them. My thesis is that some rational choice postulates are implicated in an attractive account of how we ought to shape our own character, in light of self-knowledge and our encounters with the world around us. The point is not that there is something intrinsically virtuous about suitably structured preferences, or that all of our decisions ought to be fully informed and strictly utility-maximizing. Rather, what makes my account attractive is (inter alia) that it helps us avoid certain forms of exploitation. Furthermore, this robustness against exploitation ought to be attractive even if we think exploitation is unlikely in practice.

It follows that my argument doesn't endorse particular strategies in game-theoretic models. I need not take sides, for instance, on the question of whether or not there is something generically (im)moral about the dominant strategy of defection in the Prisoners' Dilemma (PD), or the risk-dominant equilibrium in the Stag Hunt (SH). The answer to whether or not we should avoid vulnerability to exploitation by defecting in the Prisoners' Dilemma (or whether instead we ought to be loyal to our fellow players and co-operate) will depend on the particulars of the strategic interactions we are modelling (explaining) and a range of moral considerations in which particular PD or SH type scenarios are embedded. Thus, for instance, counselling defection to avoid exploitation will likely be a strange recommendation if the other player is our beloved and devoted husband or wife! This is no surprise, however, nor any great difficulty for my argument. I simply show that some of the foundational postulates that ground many such explanatory efforts may have something morally appealing to offer in many such settings, even to those who are inclined to criticize the explanations and entirely consistent with the reality that other moral considerations (loyalty to cherished friends and lovers, for example) will often come into play when we ponder decisions.

Finally, I am not claiming that the rational moral character I discuss is the only, or even the best practical way to avoid exploitation by crafty politicians and other sorts of tricksters. As I will demonstrate, there are several promising paths to take if we wish to avoid these traps. We should not favour my account of rational moral character because it is the best practical response to actual cases of exploitation in the marketplace or election campaigns, but because my favoured response is morally attractive in a way that these other solutions are not.

There is a rough analogy here with concerns about intransitivity in the social choice literature on voting and legislatures, most notably associated with the intellectual legacy of William Riker. Kenneth Arrow (1951) famously demonstrated that there is no way to aggregate (even complete and transitive) individual preferences into a social preference ordering that simultaneously satisfies several apparently reasonable liberal and democratic constraints (including voter sovereignty, equal responsiveness and transitivity). Barring certain favourable cases (for instance, where preferences are "single peaked"), either some voter preference orderings must be restricted or majorities may end up cycling through options, voting for A over B, B over C, but then C over A. For Riker (1980), the chief worry here is agenda manipulation: clever and well-informed legislators may anticipate cycles and influence the order of voting and the introduction of amendments in favour of their desired outcome. Gerry Mackie (2003: chs. 8–15) has, however, painstakingly demonstrated that no one has yet found a significant example of an actual majority cycle, anywhere, in *any* democratic political system!

Mackie has arguably demonstrated that, in the real world of democratic politics, Arrow's postulates are regularly violated in ways that avoid majority cycles. Put that thought aside, however, and ask yourself, if we had to select between two political systems that are roughly equal in every respect save that one of them allows majority cycles, whether we would be justified in favouring the more secure system—again, all else roughly equal—even if we knew that such cycles would never emerge in practice. My argument here answers an analogous question: if there are several ways to avoid exploitation of intransitive and probability-violating preferences, all of which are similarly attractive in key respects (what I mean here will become clearer in the course of argument), but one of those solutions involves making ourselves invulnerable to exploitation in the first place, then why not favour that solution, even if we have good reason to believe the the other solutions would also work reasonably well in practice?

I will try, then, to demonstrate that some of the more important postulates underlying many rational choice models in the social sciences may themselves often be poor descriptions of actual behaviour, yet are nonetheless morally attractive maxims for guiding choice behaviour as a general rule, given the broad possible sweep of human needs and aspirations. These may not be morally compelling maxims, but they are appealing enough to demand our consideration when making evaluative and prescriptive judgments about means and ends, just because they secure us against possible exploitation in ways that other solutions do not—and that argument should persuade even if we suspect that, in practice, we are unlikely to be exploited in these ways. Or so I will argue.

Rational Moral Character

I want to argue that transitivity and independence—so fundamental to rational choice, yet so contested—together are morally attractive maxims for guiding our evaluations of what we need and desire in the world. These postulates are bound up with an ideal of rational moral character. Rather than understanding weak ordering (completeness, transitivity) merely as a set of consistency requirements imposed on preference orderings, we ought also to understand these ordering conditions as consonant with, and encouraging of, an ideal of reflection upon the scope and substance of our tastes, interests and aspirations. Similarly for independence: when taken as a maxim, rather than merely a postulate, independence is part of an ideal whereby we reflect carefully upon what we know about the world and ourselves and weigh carefully the costs of violating reasonable expectations informed by plausible evidence. Together, these principles encourage modest virtues implicated in (but not strict requirements for) a sufficiently examined life, involving self-understanding and awareness of our place in the world around us. And that sort of self-understanding is, I will show, attractive for reasons a great many people—including some of the most vociferous critics of rational choice—ought to accept.

Transitivity

Philosophers delight in constructing perverse counterexamples to the transitivity of much-cherished and seemingly unambiguous relations, such as "worse than," claiming to demonstrate that apparently well-ordered preferences can lead us to surprising and unpleasant results. Yet much turns on how we interpret these thought experiments. Apparent violations of transitivity may instead reflect confusing or extraordinarily difficult tradeoffs—for instance, between unbearable pain from sudden injury, on the one hand, against gradually worsening pain over the course of one's life, on the other. In spite of several recent philosophical efforts to find intransitivity in such balancing problems—involving pains and rewards of strikingly different intensities, durations, and qualitative features (see Andreou, 2006; Quinn, 1990; Rachels, 1998 Temkin, 1996), or complex conflicts between norm-related preferences and constraints (see Philips, 1989) —I am inclined to agree with Binmore and Voorhoeve (2003) that these are instead difficult constrained maximization problems along two or more important dimensions and are not best interpreted as violations of transitivity.[13] Furthermore, there may be good reasons to question preferences over such difficult tradeoffs: as Voorhoeve (2008) argues, common heuristics employed in such settings may lead us to error but that hardly justifies embracing intransitivity.

Suppose, however, that we really do hold binary preferences P_{ab}, P_{bc}, and P_{ca} for outcomes a, b, c, where these are commodities or states of the world for which more really is, in some unambiguous sense, better for us. Advocates of the transitivity condition often cite the danger of exploitation and regret in so-called "Dutch book" or "money pump" scenarios, where clever exploiters discern our preference structure ($u(a) > u(b) > u(c) > u(a)$) and engage us in circular exchanges with modest transaction costs. Thus the clever exploiter could trade their b for our c, then their a for our b, but then trade back c for a, leaving us exactly where we started, but somewhat poorer.

Admittedly the money pump scenario seems rather contrived: how often could a salesman or marketer or campaign strategist be able to infer reliably the intransitive structure of our preferences and then find ways to engage us in repeated face-to-face exchanges (of dollars, or votes) which exploit that instability? Consider, however, if such a strategy were implemented through some impersonal mechanism, such as the market. Some recent analysis suggests that non-rational preferences could be persistently exploited in such settings (Rubinstein and Spiegler, 2008;

although consider Laibson and Yariv, 2007). More generally, the persistence of psychological research into framing effects for commercial and political applications ought to give us pause. We know that a variety of widespread biases are routinely exploited in product marketing and in political campaigns and can invoke preference reversals by framing a product, candidate or issue in contrasting ways. I am not concerned, for the purposes of this example, with the very interesting empirical question of when preference reversals reflect errors in reasoning or violations of an independence principle or genuinely intransitive preferences. My point here will be that attention to transitivity as a maxim invites the kind of introspection that will also identify errors in reasoning and exploitable violations of independence.

If our aim is simply to avoid this sort of exploitation, then transitivity is not the only plausible prescription. We might instead counsel choosers to be decisive and resolute (for example, McClennen, 1990; see Cubitt and Sugden, 2001: 138–41; Gauthier, 1997), refusing to revisit their choice once they have made a decision, even if a tempting offer is made ("No thanks. I've settled on this candidate"). Failing this—aware, perhaps, of difficulties with resolve, such as those raised by Uzan-Milofsky (2009)—we could, following Elster (1979: ch. 2), treat intransitive preferences with the same self-binding treatment deployed when we anticipate weakness of will. Like Ulysses desiring both to hear the sirens' song but also to avoid self-destruction, we find ways of binding ourselves to a chosen course.[14]

So, to the bearer of intransitive preferences, the advocate of commitment counsels that we be resolute and, if we think necessary, find ways to bind ourselves at some point in the choice sequence before the critical move that lets the exploiter take advantage of our cyclical preferences. We do this because we have some coherent plan (see Bratman, 1987) which gives force to exclusionary reasons (Raz, 1975: ch. 1) that we invoke when tempted by the exploiter.[15] Yet we recognize informational complexities, the likelihood of conflicting first- and second-order desires (Frankfurt, 1971; Taylor, 1977), and the very real possibility of weakness of resolve in the face of temptation; thus we find ways to commit to our longer-term plans and more comprehensive understanding of our interests.

While this may appear to solve the problem of exploitation in the money pump or strategic framing scenarios, the appeal to commitment as part of rationality brings with it the potential for serious difficulties in other choice settings, where commitment binds us to incredibly self-defeating actions (for example, Skyrms, 1996: ch. 2, esp. 38–42). More critically, the very notion that some broader coherent plan justifies an exclusionary reason (that is, not to engage in repeated exchanges given intransitive preferences) invites reflection not only on our plan and the second-order desires it reflects, but also upon our preferences themselves—not least why they are intransitive in the way they are. It is this sort of self-understanding that a commitment to transitivity encourages.

To be sure, we might endorse exclusionary reasons in the case of exploitable intransitive preferences without appealing to coherent plans (Bratman), second-order desires (Frankfurt) and strong evaluations (Taylor). We may instead simply exhort the potential victim of a Dutch book to look at the comprehensive costs of engaging in repeated exchanges according to their preferences and recognize that the burden of cycling among the same options is prohibitively high. Yet if our preferences really are intransitive, then pointing to the high costs of satisfying those preferences simply asks us to frustrate some of our desires to satisfy our presumably (and very plausibly) higher order desire to avoid exploitation and poverty. While this is certainly a plausible answer to the worry about exploitation in money pump and framing scenarios, it is unsatisfying in the following respect: a normative account of preferences and choice ought surely to do more than simply counsel us to weigh some preferences against others in ways consistent with our interests, broadly understood. We might reasonably ask of our normative account that it encourage reflection not only on the broader field of interests (especially our higher order interest in not being impoverished by exploitative transactions), but also on why we hold our specific preferences in the first place.

The underlying point here is that appeals to plans, comprehensive outcomes, precommitment and resolute choice all seem to direct the bearer of intransitive preferences to attend to some consequence of their choices, taking their intransitive preferences as more or less given. Yet by constructing and reflecting upon coherent plans, second-order desires and strong evaluations over those plans and desires, we are pulled toward better self-understanding. The advocate of transitivity seems more than willing to direct our attention to the preference ranking itself and, through such attention, to the substance of our preferences and the underlying personal character that gives them sense. Transitivity, understood as a maxim for evaluating our needs and desires, invites examination of why we hold a certain ranking, and why that ranking might change if we removed a particular option from consideration, or understood our choice situation and range of options in some other plausible way.

Surely this sort of critical introspection is truer than alternative recommendations to the richness of rationality as a sort of meaningful consistency to our lives, a sense that, as Charles Taylor (following Frankfurt and consistent with Bratman) has aptly put it, there is some "perspicuous order" —over time and across our varied moods and inclinations—to our thoughts, desires, aspirations and actions, and that we can, with enough reflection, make this order clear to ourselves and others (Taylor, 1982). Indeed, this sort of introspective urge seems truer to the account of rationality as richly contextual and bound up with plans, second-order desires and strong evaluations of those plans and desires, evaluations that reflect emerging articulacy about our preferences (Taylor, 1977: 25).[16] By attending to transitivity we secure ourselves against exploitation in ways consistent with rational self-reflection on what, upon due reflection, we want our character to be.

Independence

A similar case follows for independence. Do violations of independence constitute irrational behaviour? As with transitivity, our judgments here depend on how we understand the scenarios that purportedly encourage reasonable violations. Thus, for instance, Patrick Maher argues that, when we include purported preferences for certain gains (in the Allais case) and known odds (in the Ellsberg case) in the specification of the choice scenarios, the corresponding preferences do not violate independence (Maher, 1993: ch. 3). And this comports with the intuitive appeal of independence as a postulate; we ought to avoid choices that go against what seem to be uncontentious facts about the world, and so we ought to figure out those facts and their varied implications for our actions.

To be sure, I have in mind here relatively unambiguous cases of clearly specified lotteries such as Allais-type scenarios. Matters can be made more complex, such as preferences contingent on combinations of outcomes or information specific to particular lotteries (for instance, the counterexamples to strong independence offered by Machina, 1982, and Sen, 1985). In the case of actual politics, we could easily imagine campaign strategists deliberately complicating public coverage of some policy issue, for instance by aggressively presenting some set of complications associated with the expected costs and likely benefits of the policy at stake. But the very fact that such counterexamples depend on complex conditionals seems to undercut their normative force as challenges to independence. If we can become aware of the subtle ways in which our preferences are conditioned on particular choice contexts, then we will likely seek to satisfy independence by informing ourselves and becoming aware of which issues are most important to us, which outcomes are most likely to obtain and what costs and benefits are associated with those anticipated outcomes.

It is vital to my endorsement of independence that, if a game were played over time involving repeated choices in an Allais-like scenario, then the player who respects independence will almost certainly receive higher monetary returns over time than the player who violates the rule. The case is less clear for more complex situations, but I think an analogous line of reasoning holds.

Consider, for example, a voter who, whenever faced with clashing partisan voices seeking to complicate a policy debate, always responded to complex and uncertain evidence in ways that could, upon more careful reflection, be shown to violate independence. Shouldn't we be concerned about that? In the case of campaign strategists complicating some issue space, the relevant calculations may well be ridiculously complex and the evidence of dubious quality; in the face of this, citizens may routinely turn to some heuristic to simplify their choices. Yet we should worry about uncritically accepting heuristics that reliably violate independence. Our efforts as citizens ought to be guided by our informed judgments, both about the facts and about what heuristics are most helpful in particular complex choice settings. Our informed judgments ought roughly to track evidence about likelihoods. These sorts of efforts push us toward respecting independence, even if we may fail to live up to that postulate in practice. We should be reluctant to describe the rejection of plausible probability estimates as rational, or even reasonable conduct, however understandable that rejection may be in certain settings and however durable such violations might be as behavioural norms. This is especially so if our aims are as much prescriptive as explanatory.

True, we may be uncomfortable with a prescriptive theory of choice that counsels us to ignore entirely our settled (and perhaps sometimes evolutionarily advantageous) habits and heuristics—for certainty and social acceptance, say—or to forsake pleasurable and relatively harmless acts of whimsy, based on culturally acquired tastes ("the casino is fun!").[17] Yet we surely do not want a prescriptive theory of choice that, while descriptively accurate and thus sensitive to the complexities of our (socially constituted and richly historical) identities, nonetheless offers little more by way of counsel than "do what seems appropriate, given who you are, where you are from and how you feel at the time." Rather, we want a prescriptive theory of choice that encourages us to reflect on our needs and desires in ways that respond to our changing understandings of the world and especially to new evidence we acquire and integrate into our understandings.

Attention to independence as a maxim encourages just this sort of awareness, by disciplining us to think carefully about our degrees of belief vis-a-vis the world around us, scrutinizing and incorporating new evidence, trying to understand the causal relationships that affect our needs and desires and that condition the feasibility of various courses of action.

Conclusion: Exploitation and Rational Moral Character

So what? Why should critics of rational choice *explanations* be impressed that we can tell a moral story in (modest) praise of two of the more substantive SEU postulates?

I think they should care because the moral story I've told resonates deeply with a value that most all of them share: the wrong of exploitation. Whatever these critics think of claimed innovations in choice-theoretic explanations, they ought to find rational consistency postulates morally attractive insofar as they help mitigate against exploitation.

Exploitation is an abiding concern of political theorists. Marxists are unrelentingly critical of exploitation understood as structural inequalities in ownership of the means of production, allowing a class of property owners to exploit a class of workers who must sell their labour power to survive but do not receive in wages the full value of their efforts. Egalitarian liberals worry that the lucky might exploit their good fortune, securing favourable prospects while being indifferent to, and perhaps also gaining at the expense of, those who are less fortunate through no fault of their own (for example, Rawls, 1971, at least as some egalitarians read him). When civic republicans such as Pettit (1997) talk about freedom as the absence of domination, it is not much of a stretch to conclude that "freedom from the arbitrary will of another" includes "avoiding situations where someone else can exploit us, in the sense of using us simply to achieve their goals, irrespective of our needs and interests." Many feminists are arguably worried about this sort of exploitation: systematic biases—in attitudes and institutions—that reflect and reproduce male advantage are objectionable at the very least because they permit men to use women as means to their own ends, indifferent to the distinctive needs and interests of women and unresponsive to their reasonable arguments and aspirations.

The language I've used here is, of course, roughly Kantian; to use others merely as means to one's own ends violates the categorical imperative. Furthermore, something very much like this sort of exploitation has been a prominent concern among some political scientists. Consider, for example, James Scott's discussion of exploitation and legitimacy in peasant societies (1976: ch. 6) and his intervention in a longstanding debate about oppressive regimes and strategies of resistance (1985).

To be sure, many of these theorists and researchers would argue over particulars and nuance. But whatever, precisely, exploitation is, whoever is generally involved, and why, specifically, it matters (morally and practically speaking), exploitation surely involves some people being used by others and for purposes that are chiefly or solely those of the exploiter, rather than the exploited. And this odious means–ends relationship is often secured by deception and subtle forms of coercion. Nuance and particulars aside, then, it is difficult to see how the political theorists I've just described could consistently resist an overlapping consensus that the roughly Kantian form of exploitation I've summarized here is generally undesirable and indefensible, especially when it is grounded in deception and coercion.

This is not to suggest that other, more subtle or less clearly defined forms of exploitation are less important, morally speaking; it is, however, to assert that most any form of exploitation that we find troubling on moral grounds is likely to be such that we can interpret it in this (roughly Kantian) way, as a relationship in which one party is being used by another for reasons that do not track the former's interests and would not be freely accepted upon informed reflection by the exploited party.

Certain core features of rational choice models are, then, given moral weight by one of the distinctive concerns of normative political theory: the avoidance of exploitation. This may appear a strange concession to ask of rational choice advocates, because it seems to suggest that rationality involves not simply preferences, utility maximization and degrees of belief, but as importantly our moral character—specifically, our being the sort of persons who are not easily exploited by others through deception and manipulation, because we know ourselves and can articulate that self-knowledge with sufficient depth and confidence. Endorsing rational choice prescriptions can be tantamount, I am suggesting, to asserting that we should not allow ourselves to be used merely for someone else's ends, for reasons we would not find plausible upon informed reflection about who we are and want to be. Taking a stand on desirable character traits is not familiar territory for most decision and game theorists!

Some, perhaps many, students of politics must also make a concession here, however, even if their inclination is to criticize rational choice approaches. If you think that this sort of exploitation is generally a bad thing, then it is hard to resist the conclusion that you should also find some of the prescriptive aims of rational choice theory to be attractive with respect to governing our desires and choices. Choosing rationally is morally important just because in settings where decisions are complex and uncertain, intransitivity and systematic violations of independence make us vulnerable to being used by others in ways that, upon informed and careful reflection, we would not freely accept.

This endorsement of independence and transitivity as maxims obviously does not exhaust our moral concerns with exploitation: we can find good reasons to condemn an exploiter's motives and actions, not simply chide the exploited for their vulnerability. Yet there is no inconsistency in condemning both the act of exploiting and intransitive preference orderings that make us vulnerable to such acts. Insofar as most attractive theories of justice do say something substantive about citizen's responsibilities and their legitimate expectations (an excellent model here is Rawls), transitivity and independence seem consonant with a vision of citizens who treat each other as free, equal, and responsible moral agents.

Even if my argument persuades, we might nonetheless reject rational choice as a source of morally weighty prescriptions if we find moral pathologies or chronically ambiguous and morally murky conclusions about how we ought to act. The moral appeal of transitivity and independence should give us pause; however; if these critical components of canonical rational choice theory can be shown to resonate with a powerful line of ethical argument in political thought, then perhaps we should evaluate the rational choice approach not only (or even primarily) as an explanatory framework, but also as a morally rich normative framework that clarifies some (but not all) of our intuitions about interests and agency, responsibility and legitimacy, and that helps us examine, with considerable rigour, various arguments about how we ought to choose and what institutions ought to regulate our interests and activities. Normative rational choice is not a full-fledged theory of virtues or justice, obviously; but it can be interpreted as an interesting and important part of any such theory that is concerned not only with condemning those who exploit others, but as importantly with showing citizens how to secure themselves against such exploitation.

Notes

[1] In contrast, Satz and Ferejohn (1994) define rational choice explanations in terms of structural constraints that *select* for behavioural regularities, but Hausman (1995) suspects that implicit assumptions about mental states and actions ultimately make this an agent-based, rather than a structuralist account.

[2] Some think of preferences in behavioural terms as stable dispositions to choose in particular ways, other things being equal (see Lee, 1984; also Maher, 1993); others tend to emphasize *preference* as an ordering relation, the elements of which are actions and their anticipated consequences. These are not mutually exclusive understandings—again, the point is one of emphasis.

[3] For overviews see McLennen (1990), Anand (1993), Maher (1993), Joyce (1999), and Binmore (2009).

[4] In an elegant formulation by Herbert Gintis (2009: 4–6), a preference relation is consistent if it satisfies completeness, transitivity and this form of independence.

[5] The relationships between various versions of *these* independence postulates are interestingly complex; see Fishburn and Wakker (1995).

[6] The distinction is most often attributed to Frank Knight (1921): risk involves known odds, whereas uncertainty involves unknown objective probabilities.

[7] On difficulties with Gauthier's account, see Franssen (1994) and Uzan-Milofsky (2009).

[8] A relevant distinction here is between more volatile, context-sensitive, *instrumental* preferences, on the one hand; and deeper, more settled, *intrinsic* preferences, on the other; see Binmore (2009: 18ff).

[9] Regardless of background norms and traditions, however, we would not be rational in concluding that, because we cannot achieve some goal, then that goal must be undesirable, or that, because we can feasibly do one thing rather than another, the available option must be better *because* it is feasible; see, for example, Binmore (2009: 17–18) and also Elster (1979: ch. 3).

[10] This hardly applies to the present example. Is there any conceivable moral framework that would, *as a matter of moral principle*, consistently have me thwart my desire for homeownership, yet embrace drunken gambling? Then again, people believe curious things.

[11] Apparently (according to Binmore, 2009: 34), including Leonard Savage himself.

[12] For an intriguing exception see Henrich (2000), and more generally, see Henrich and colleagues (2010) on intercultural variation in Ultimatum game responses.

[13] On the importance of being able to distinguish our preferences along such dimensions, see Binmore (2009: 55ff).

[14] Could this really work for voting? And if it does, shouldn't we be concerned? Can we credibly commit to voting in a particular way at some future point and then go listen to the sirens of opposing views? Ulysses at least had a solid mast and secure binds. That said, perhaps people do find ways to commit, in spite of the siren songs: Simon Jackman and Paul Sniderman (2006) find evidence that, in everyday political argument, exposure to opposing arguments and evidence doesn't seem to affect subsequent beliefs. More likely, people who have committed to a particular ideological position and associated views on issues and candidates simply avoid engaging with bearers of different viewpoints (see Mutz, 2006).

[15] "A *second-order reason* is any reason to act for a reason or to refrain from acting for a reason. An *exclusionary reason* is a second-order reason to refrain from acting for some reason" (Raz, 1975: 39). In the case at hand, we might believe that we have a reason not to act on our intransitive preferences (where those are understood to be reasons for acting toward their satisfaction).

[16] This is what I had in mind earlier when I posed the question: What if competing recommendations were similarly attractive in all but one respect? All of these alternative solutions to exploitation arguably share, or are at least not unfriendly to, an intuitive and broadly attractive sense of rationality as intelligibility paired with a tendency to bend toward a reasonably examined life. If, however, you are already committed to such an account of rational character, then why not bring reflection to bear on why our preferences can be exploited?

[17] Or as Paul Samuelson (1952: 671) put the point, "When I go to a casino, I go not alone for the dollar prizes but also for the pleasures of gaming—for the soft lights and sweet music"! More seriously, our theory ought to recognize, as David Houghton (1995) suggests, that there are some goods for which full information diminishes their desirability, and that we often pursue happiness not by strictly maximizing utility along a single dimension. But, as we've seen, most any sophisticated choice theorist would readily agree.

References

Allais, Maurice. 1953. "Rational Man's Behavior in the Presence of Risk: Critique of the Postulates and Axioms of the American School." *Econometrica* 21: 503–46.

Anand, Paul. 1993. *Foundations of Rational Choice under Risk.* Oxford: Oxford University Press.

Andreou, Chrisoula. 2006. "Environmental Damage and the Puzzle of the Self-Torturer." *Philosophy and Public Affairs* 34: 95–108.

Arrow, Kenneth J. 1951. *Social Choice and Individual Valuess*. 2nd ed. New Haven CT: Yale University Press, 1963.

Austen-Smith, David and Jeffrey Banks. 1988. "Elections, Coalitions, and Legislative Outcomes." *American Political Science Review* 82: 405–22.

Axelrod, Robert. 1986. "An Evolutionary Approach to Norms." *American Political Science Review* 80: 1095–1111.

Bardsley, Nicholas. 2008. "Dictator Game Giving: Altruism or Artifact?" *Experimental Economics* 11: 122–33.

Baron, David P. and John A. Ferejohn. 1989. "Bargaining in Legislatures." *American Political Science Review* 83: 1181–1206.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast. 1998. *Analytic Narratives*. Princeton NJ: Princeton University Press.

Bendor, Jonathan, Daniel Diermeier and Michael Ting. 2003. “A Behavioral Model of Turnout.” *American Political Science Review* 97: 261–80.

Bernoulli, Daniel. 1954 [1738]. “Exposition of a New Theory on the Measurement of Risk,” trans. Louise Summer. *Econometrica* 22: 23–36.

Binmore, Ken. 1998. *Just Playing: Game Theory and the Social Contract*, vol. II. Cambridge MA: MIT Press.

Binmore, Ken. 2009. *Rational Decisions*. Princeton NJ: Princeton University Press.

Binmore, Ken and Alex Voorhoeve. 2003. “Defending Transitivity against Zeno’s Paradox.” *Philosophy and Public Affairs* 31: 272–79.

Binmore, Ken and Avner Shaked. 2010. “Experimental Economics: Where Next?” *Journal of Economic Behavior and Organization* 73: 87–100.

Black, Duncan. 1958. *Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Brams, Steven J., Morton D. Davis and Philip D. Straffin, Jr. 1979. “The Geometry of the Arms Race.” *International Studies Quarterly* 23: 567–88.

Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge MA: Harvard University Press.

Buchanan, Allen. 1990. “Justice as Reciprocity versus Subject-Centered Justice.” *Philosophy & Public Affairs* 19: 227–52.

Chong, Dennis. 2000. *Rational Lives: Norms and Values in Politics and Society*. Chicago: University of Chicago Press.

Christiano, Thomas. 2004. “Is Normative Rational Choice Theory Self-Defeating?” *Ethics* 115: 122–41.

Cubitt, Robin P. and Robert Sugden. 2001. “On Money Pumps.” *Games and Economic Behavior* 37: 121–60.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.

Eckel, Catherine and Herbert Gintis. 2010. “Blaming the Messenger: Notes on the Current State of Experimental Economics.” *Journal of Economic Behavior & Organization* 73: 109–19.

Ellsberg, Daniel. 1961. “Risk, Ambiguity, and the Savage Axioms.” *Quarterly Journal of Economics* 75: 643–69.

Elster, Jon. 1979. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.

Fessler, Daniel M.T. 2009. “Return of the Lost Letter: Experimental Framing Does Not Enhance Altruism in an Everyday Context.” *Journal of Economic Behavior & Organization* 71: 575–78.

Fishburn, Peter C. 1983. “Ellsberg Revisited: A New Look at Comparative Probability.” *Annals of Statistics* 11: 1047–59.

Fishburn, Peter C. and Peter Wakker. 1995. “The Invention of the Independence Condition for Preferences.” *Management Science* 41: 1130–44.

Frankfurt, Harry G. 1971. “Freedom of the Will and the Concept of a Person.” *Journal of Philosophy* 68: 5–20

Franssen, Maarten. 1994. “Constrained Maximization Reconsidered: An Elaboration and Critique of Gauthier’s Modelling of Rational Choice in a Single Prisoners’ Dilemma.” *Synthese* 101: 249–72.

Frohlich, Norman and Joe Oppenheimer. 2006. “Skating on Thin Ice: Cracks in the Public Choice Foundation.” *Journal of Theoretical Politics* 18 (3): 235–66.

Gauthier, David. 1969. *The Logic of Leviathan: the Moral and Political Theory of Thomas Hobbes*. Oxford: Oxford University Press.

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Clarendon.

Gauthier, David. 1997. “Resolute Choice and Rational Deliberation: a Critique and Defense.” *Nous* 31: 1–35.

Geertz, Clifford. 1972. “Deep Play: Notes on the Balinese Cockfight.” *Daedalus* 101: 1–37.

Geertz, Clifford. 1973. "Thick Description: Toward and Interpretive Theory of Culture." In *The Interpretation of Cultures*. New York: Basic Books.

Gintis, Herbert. 2000. "Strong Reciprocity and Human Sociality." *Journal of Theoretical Biology* 206: 169–79.

Gintis, Herbert. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton NJ: Princeton University Press.

Gintis, Herbert and Samuel Bowles, Robert Boyd and Ernst Fehr. 2003. "Explaining Altruistic Behavior in Humans." *Evolution and Human Behavior* 24: 153–72.

Green, Donald and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory*. New Haven: Yale University Press.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. New York: Cambridge University Press.

Hardin, Russell. 2001. "The Normative Core of Rational Choice Theory." In *The Economic World View: Studies in the Ontology of Economics*, ed. Uskali Maki. Cambridge: Cambridge University Press.

Hausman, Daniel M. 1995. "Rational Choice and Social Theory: A Comment." *Journal of Philosophy* 92: 96–102.

Henrich, Joseph. 2000. "Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon." *American Economic Review* 90: 973–79.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer and John Ziker. 2006. "Costly Punishment across Human Societies." *Science* 312: 1767–70.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer and John Ziker. 2010. "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment." *Science* 327: 1480–84.

Houghton, David. 1995. "Reasonable Doubts about Rational Choice." *Philosophy* 70: 53–68.

Jackman, Simon and Paul M. Sniderman. 2006. "The Limits of Deliberative Discussion: a Model of Everyday Political Arguments." *Journal of Politics* 68: 272–83.

Jervis, Robert J. 1988. "Realism, Game Theory, and Cooperation." *World Politics* 40: 317–49.

Johnson, James. 2006. "Consequences of Positivism: a Pragmatist Assessment." *Comparative Political Studies* 39: 224–52.

Johnson, James. 2010. "What Rationality Assumption? Or, How 'Positive Political Theory' Rests on a Mistake." *Political Studies* 58: 282–99.

Jones, Bryan D. 2001. *Politics and the Architecture of Choice: Bounded Rationality and Governance*. Chicago: University of Chicago Press.

Joyce, James. 1999. *Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Keller, Evelyn Fox. 2000. "Models of and Models for: Theory and Practice in Contemporary Biology." *Philosophy of Science* 67: S72–S86.

Knight, Frank. 1921. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin.

Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Kydd, Andrew. 1997. "Game Theory and the Spiral Model." *World Politics* 49: 371–400.

Laibson, David and Leeat Yariv. 2007. "Safety in Markets: an Impossibility Theorem for Dutch Books." Manuscript, http://www.economics.harvard.edu/faculty/laibson/files/LaibsonYariv.pd (September 7, 2010).

Lane, Ruth. 1996. "Positivism, Scientific Realism, and Political Science." *Journal of Theoretical Politics* 8: 361–82

Lee, Richard. 1984. “Preference and Transitivity.” *Analysis* 44: 129–34.

Levitt, Steven D. and John A. List. 2007. “What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?” *Journal of Economic Perspectives* 21: 153–74.

Loomes, Graham, Chris Starmer and Robert Sugden. 1991. “Observing Violations of Transitivity by Experimental Methods.” *Econometrica* 59: 425–39.

Machina, Mark J. 1982. “‘Expected Utility’ Analysis without the Independence Axiom.” *Econometrica* 50: 277–323.

Mackie, Gerry. 2003. *Democracy Defended*. Cambridge: Cambridge University Press.

Maher, Patrick. 1993. *Betting on Theories*. New York: Cambridge University Press.

McClennen, Edward F. 1990. *Rationality and Dynamic Choice: Foundational Explorations*. New York: Cambridge University Press.

McDermott, Rose. 2004. “The Feeling of Rationality: the Meaning of Neuroscientific Advances for Political Science.” *Perspectives on Politics* 2: 691–706.

McKelvey, Richard D. and Thomas R. Palfrey. 1995. “Quantal Response Equilibria for Normal Form Games.” *Games and Economic Behavior* 10: 6–38.

McKelvey, Richard, D. and Thomas R. Palfrey 1998. “Quantal Response Equilibria for Extensive Form Games.” *Experimental Economics* 1: 6–38.

Mutz, Diana. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge: Cambridge University Press.

Myerson, Roger B. 1991. *Game Theory: Analysis of Conflict*. Cambridge MA: Harvard University Press.

Narveson, Jan. 2010. “The Relevanc eof Decision Theory to Ethical Theory.” *Ethical Theory and Moral Practice* 13: 497–520.

Ostrom, Elinor. 1998. “A Behavioral Approach to the Rational Choice Theory of Collective Action.” *American Political Science Review* 92: 1–22.

Ostrom, Elinor. 2000. “Collective Action and the Evolution of Fairness Norms.” *Journal of Economic Perspectives* 14: 137–58.

Ostrom, Elinor. 2006. “The Value-Added of Laboratory Experiments for the Study of Institutions and Common-Pool Resources.” *Journal of Economic Behavior and Organization* 61: 149–63.

Pettit, Philip. 1997. *Republicanism: a Theory of Freedom and Government*. Oxford: Oxford University Press.

Philips, Michael. 1989. “Must Rational Preferences be Transitive?” *Philosophical Quarterly* 39: 477–83.

Popkin, Samuel L. 1979. *The Rational Peasant: the Political Economy of Rural Society in Vietnam*. New Haven: Yale University Press.

Quinn, Warren S. 1990. “The Puzzle of the Self-Torturer.” *Philosophical Studies* 59: 79–90.

Rachels, Stuart. 1998. “Counterexamples to the Transitivity of Better Than.” *Australasian Journal of Philosophy* 76: 71–83.

Rawls, John. 1999 [1971]. *A Theory of Justice*. 2nd ed. Cambridge MA: Belnap Press.

Raz, Joseph. 1975. *Practical Reason and Norms*. London: Hutchinson.

Riker, William H. 1980. “Implications from the Disequilibrium of Majority Rule for the Study of Institutions.” *American Political Science Review* 74: 432–46.

Rubinstein, Ariel and Ran Spiegler. 2008. “Money Pumps in the Market.” *Journal of the European Economic Association* 6: 237–53.

Samuelson, Paul A. 1952. “Probability, Utility, and the Independence Axiom.” *Econometrica* 20: 670–78.

Satz, Debra and John Ferejohn. 1994. “Rational Choice and Social Theory.” *Journal of Philosophy* 91: 71–87.

Savage, Leonard J. 1954. *Foundations of Statistics*. New York: Wiley.

Scott, James C. 1976. *The Moral Economy of the Peasant: Rebellion and Subsistence in Southeast Asia*. New Haven: Yale University Press.

Scott, James C. 1985. *Weapons of the Weak: Everyday Forms of Peasant Resistance*. New Haven: Yale University Press.

Sen, Amartya K. 1985. "Rationality and Uncertainty." In *Rationality and Freedom*. Cambridge MA: Harvard Belknap.

Shapiro, Ian. 2005. *The Flight from Reality in the Human Sciences*. Princeton NJ: Princeton University Press.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93: 279–97.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69: 99–118.

Simon, Herbert A. 1985. "Human Nature in Politics: the Dialogue of Psychology with Political Science." *American Political Science Review* 79: 293–304.

Skyrms, Brian. 1996. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.

Sugden, Robert. 1985. "Why Be Consistent? A Critical Analysis of Consistency Requirements in Choice Theory." *Economica* 52: 167–83.

Taylor, Charles. 1971. "Interpretation and the Sciences of Man." *Review of Metaphysics* 25: 3–51.

Taylor, Charles. 1977. "What is Human Agency?" In *Human Agency and Language: Philosophical Papers*, vol. I. New York: Cambridge University Press.

Taylor, Charles. 1982. "Rationality." In *Rationality and Relativism*, ed. Martin Hollis and Steven Lukes. Cambridge MA: MIT Press.

Temkin, Larry. 1996. "A Continuum Argument for Intransitivity." *Philosophy and Public Affairs* 25: 175–210.

Tricomi, Elizabeth, Antonio Rangel, Colin F. Camerer and John P. O'Doherty. 2010. "Neural Evidence for Inequality-Averse Social Preferences." *Nature* 463: 1089–91.

Tversky, Amos and Daniel Kahneman. 1979. "Prospect Theory: An Examination of Decision under Risk." *Econometrica* 47: 263–291.

Tversky, Amos and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211: 453–58.

Uzan-Milofsky, Marie. 2009. "Shall We Be Resolute?" *Rationality and Society* 21: 337–57.

Voorhoeve, Alex. 2008. "Heuristics and Biases in a Purported Counter-example to the Acyclicity of 'Better Than.'" *Politics, Philosophy & Economics* 7: 285–99.

Weber, Michael. 1998. "The Resilience of the Allais Paradox." *Ethics* 109: 94–118.

Woodward, James. 2008. "Social Preferences in Experimental Economics." *Philosophy of Science* 75: 646–57.