Exploitation and Rational Choice: Transitivity as a Moral Ideal*

Loren A. King†

June 2004

These very brief remarks are a fledgling first attempt to bring together a class of philosophical debates about the foundations of decision and game theory, on the one hand, and a distinctive concern of normative political theory, on the other. I begin with a very general outline of the debate surrounding what we loosely refer to as 'rational choice approaches' in the social sciences. I then lay out – again very roughly – the beginnings of what I hope may be a plausible normative defense of transitivity, an axiom often presented as integral to rational choice approaches. This line of defense emphasizes how a unifying theme in much political theory – the avoidance of exploitation – lets us interpret the transitivity axiom in a way that has considerable normative appeal.

The term 'rational choice theory' is an informal designation encompassing several related approaches to the study of human behaviour and social phenomena, including (most obviously) game theory, public choice, and positive political economy. What these approaches share, roughly, is a sense that we can best explain social practices and institutions by looking to the motivations of individual agents, and these motivations will generally reflect the needs and preferences – broadly speaking, the *interests* – of these agents. Interests needn't be narrowly conceived, but they are, on this view, ultimately grounded in agents themselves, not in the groups they form. Certainly we may sometimes speak of a group 'having interests', but what we really mean is that group members share some interests in common. And 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, friends, community, and nation, these are nonetheless our interests, our desires, our aspirations.

Some of the more zealous advocates of the rational choice approach have hoped that we can ultimately explain most, perhaps all social outcomes in terms of personal motives, and not the other way around. The individual agent is, on this view, the explanatory primitive in the social sciences. These agents can be described as having coherent preferences over states of the world; they are assumed to be able to order these preferences, and calculate utilities associated with actions and expected consequences; and they can, it is assumed, make informed estimates about the world around them, including the motives and likely actions of other agents.

*Prepared for presentation to the 2004 annual meeting of the Canadian Political Science Association at the University of Manitoba, Winnipeg, Manitoba, Canada. Comments welcomed.

†Department of Political Science, Brown University, Providence, RI, USA. After July 1st: Department of Political Science, Wilfrid Laurier University, Waterloo, Ontario, Canada (email: lorenking@alum.mit.edu).

Critics charge that this individualistic stance, emphasizing preference satisfaction and utility maximization, is too often unrealistic and unhelpful in explaining human behaviour. Our choices are generally made within a tapestry of meanings and associations that give sense to our needs and aspirations, and that condition our beliefs and expectations. That is, our embeddedness in this rich tapestry may shape our beliefs no only about what ends are worth pursuing, but as importantly about likely outcomes in the world around us, and these latter sorts of contextual beliefs may not be satisfactorily described by the axioms and derivations of probability theory central to the decision-theoretic roots of rational choice approaches. Our preferences and judgements are, these critics suspect, irreducibly social and historical, best understood by interpreting the narratives of our lives, rather than by deriving utility functions and calculating core points in games. To imagine that our choices can be understood in terms of complete and transitive orderings – over well-defined options, and given degrees of belief about the world – is to badly misunderstand what is interesting about human behavior and social phenomena, namely its richly historical and meaningful character.

Advocates reply that if our choices do not reflect some underlying coherence to our preferences, and if our beliefs about the world do not reliably conform to the dictates of probability theory, then we are unlikely to be successful in many facets of our lives: we will make self-defeating choices and be vulnerable to exploitation by those who discover our incoherent beliefs and systematic errors in judgement. In an evolutionary setting, our beliefs and strategies for dealing with the world would be of little use to us or anyone else, and so would be unlikely to persist over time.

Critics rejoin that the plausibility of our beliefs and attractiveness of choices are not merely functions of strategic encounters among self-interested utility maximizers, but are instead dictated by the norms and traditions that we have been raised into. Our social context determines what beliefs and strategies are desirable, and so the effort to frame explanations in terms of individual desires and degrees of belief about the world inevitably depends upon understanding the social context within which those attributes and expectations were formed, and against which they have meaning for their bearers.

And so the argument goes, back and forth, with myriad iterations and variations. My chief interest here, however, is not the explanatory hopes of rational choice advocates (whether they may succeed, say, in using their assumptions and methods to account for apparently self-defeating choices, or seemingly irrational cultural norms; I myself am rather optimistic on this point, but that is another matter). Rather, I am curious about the normative promise of the rational choice approach, although I will focus on a very limited portion of the overarching framework. Even if the complaints levied against rational choice assumptions are damning, we might still find some of the classical rational choice assumptions attractive, if they lead to persuasive claims about how we *ought* to behave in some, perhaps many situations.

What would such claims look like? No doubt the general structure is familiar to most of us. If I assert a desire to achieve some goal, yet then self-consciously act in a way that is obviously contrary to my stated aim, then surely you would have grounds for condemning my actions as irrational. How could matters be otherwise?

And yet this simple consistency criterion does not seem to exhaust our intuitive sense of what rationality is. Were you confronted with the situation just described, I suspect you would look for

factors that give sense to my apparently self-defeating behavior. Perhaps I have misunderstood certain key facts about the world? I am not irrational, but insufficiently nimble of mind, and so in spite of my best efforts my reasoning is mired in error. Or perhaps I am sufficiently informed and able to reason correctly, but my will is extraordinarily weak, due to a drug addiction or some other infirmity of character (acquired or inherited). If so, then I am not an irrational actor but 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 entirely rational, informed, and competent, but you have misunderstood my utterances: I may have spoken metaphorically, and you lack familiarity with the narratives required to make sense of my worldview. 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.

Now in a sense these complications mirror the form of the debate I've just rehearsed about the explanatory limits of conventional rational choice approaches: rationality is, in our intellectual tradition, bound up with *theoretical understanding*, as Charles Taylor explains nicely: "we have a rational grasp of something when we can *articulate* it, that means, distinguish and lay out the different features of the matter in perspicuous order" (1982, 90). But prevalent rational choice approaches in the social sciences do not seem to capture this richer sense of rationality in their definition, instead referring to preference orderings, utility maximization, and probability distributions that capture our degrees of belief about features of the world that might influence the outcomes associated with our actions. Nowhere in such axiomatic formulations can we find the richly deliberative and dialogical features of how we actually reason through problems in the world we share with others. Or so critics charge.

This doesn't seem to me to be a debilitating critique, and indeed much recent work in the social sciences has been sensitive to the richness of rationality, especially how our reasoning reflects, and is importantly constrained by, historically durable norms. Of particular interest is ongoing experimental work in economics and social psychology, combined with theoretical advances in evolutionary game theory. These efforts seem to go some considerable distance in explaining how strategies may persist, even while violating some of the dictates of classical decision and game theory.

Consider, for instance, the 'ultimatum' game, in which one player will receive some considerable award, but only on the condition that she successfully negotiate with a second player to receive some share of the award. If the second player refuses the offer, then the award is withdrawn and both players leave empty-handed. The dominant strategy in this game is to offer the other player the smallest possible increment if you are player one, and to accept any positive offer if you are player two. Any other strategy is *weakly dominated*, in the sense that there is another strategy ('offer the smallest increment possible, accept any positive offer') that will perform at least as well as alternative strategies. Yet in repeated experiments the dominant strategy is almost never chosen, and when it is adopted by first players, second players will often accept considerable losses to punish what they perceive to be objectionably greedy offers. That is, in experimental settings many players prefer to leave this game empty-handed, rather than let a rational player walk away with the bulk of the prize. In a highly readable review of this experimental literature

and summary of his own findings, Bryan Skyrms (1996, ch. 2) details a computer simulation of repeated 'ultimatum' exchanges among a population of players adopting a variety of strategies. In such evolutionary models there are reproduction and transmission rules for the players, whereby strategies can be passed on to 'offspring', or adopted by other players, as a function of their success over time. In several runs of Skyrms's model, weakly dominated 'fair' strategies ('offer somewhere near half the prize if player one, accept only similar offers if player two') can persist, and indeed prosper, under not-implausible assumptions about the other strategies at play in the population, and the likelihood of 'fair' players encountering each other in repeated exchanges.[1]

As tempting as it is to go into more detail about these and other fascinating advances in evolutionary approaches to strategic behavior and collective action, I have slipped from normative to explanatory concerns, and it is the former I really want to explore. Richer formulations of decision and game theory, and especially evolutionary approaches to strategies and norms, do seem to give some real explanatory purchase to the study of strategic behavior and collective action, but they do not seem to have straightforward normative implications. It is one thing, for example, to make the case that weakly dominated 'fair' strategies may persist as viable solutions to certain classes of social interactions. But it is quite another matter to endorse 'fair' strategies as normatively appealing merely by virtue of our being able to explain them in this way – to say, that is, that one *ought* to be fair in these sorts of exchanges just because one *can* be fair without inevitably facing repeated losses at the hands of more conventionally rational players. More generally, we might plausibly explain violations of rational choice assumptions as reasonable given the historical working of certain selection and transmission mechanisms (evolved psychological responses to risk, punishment norms, formative cultural institutions), but this is not an argument for how things ought to be. To be sure, if we think that 'ought implies can' then we may be pleased to discover that we can in fact do what we ought to do, in spite of violating apparently sensible axioms of rational decision-making, but we still need to argue in favor of that 'ought' claim!

So, having laid out this debate in very general terms, the question I now want tentatively to explore is whether one of the axioms in the classical approach to rational choosing might have some normative appeal, without reference to its explanatory utility. To do this, let me sketch in somewhat more detail a small and, for my purposes, especially interesting patch of the intellectual landscape I've just outlined.

Rational choice approaches are grounded in an account of *subjective expected utility* (SEU), given axiomatic formulations by Frank Ramsey, Bruno de Finetti, John von Neumann and Oskar Mor-

[1] To see that fairness is weakly dominated, consider Skyrms's 'easy rider' strategy ('ask for roughly half the award as first player, but accept whatever is offered as second player'). This strategy always yields positive returns against itself and against fair players (in both cases getting half the award as either first or second player), and also against strictly rational players (half of the prize as first player, some small amount as second player). In contrast, fair players do as well as 'easy rider' against both 'easy rider' and other fair players, but when fair players meet rational players, the former only gain half the award as first player, and nothing as second player. Thus 'easy riders' always do at least as well as 'fair' players, and sometimes they do better (for instance, against strictly rational players). A curious strategy that turns out to be important in several of Skyrms's simulations is one he calls 'mad dog', which demands as much as possible as first player, but also expects the same amount as second player, thus punishing both fair, easygoing, and rational players for their offers!

ganstern, and Leonard Savage (for an excellent review see Anand 1996, ch. 1). When critics challenge the assumptions of rational choice approaches, it is likely unease with one or more of the axioms of SEU that is lurking behind their complaints. And not without justification: decades of experimental work has revealed that several of these axioms are at best exceedingly poor assumptions about actual human behaviour. Three axioms in particular stand out in the experimental, theoretical, and philosophical literatures on decision theory, economics, and psychology: these are completeness, transitivity, and independence. *Completeness* requires that, for options a and b, we either prefer a over b, b over a, or are indifferent between the two options. *Transitivity* requires that if we prefer option a over b and b over c, then we also prefer a over c. Independence admits of several formal definitions, but it is roughly a requirement that our decisions not respond to features of a choice situation that are irrelevant to our preferences. For instance, consider the situation in Table 1, modeled after a famous experiment by Maurice Allais in 1953 (once again, on this experiment, the Ellsberg result reported further below, and other relevant experimental findings, an able survey is Anand 1996, ch. 2). In this table, s_1, s_2, and s_3 are states of the world that will obtain with probabilities 0.01, 0.10 and 0.89 respectively. Given this, we are faced with two sets of two lotteries. In each of the two scenarios, which lottery should we prefer?

Table 1

	s_1 $p = 0.01$	s_2 $p = 0.10$	s_3 $p = 0.89$	$E(L)$
Lottery A	0	100	10	18.9
Lottery B	10	10	10	10.0
Lottery A'	0	100	0	10.0
Lottery B'	10	10	0	1.1

Notice that the choices facing us in the two scenarios offered are essentially the same in terms of expected relative gains: lotteries A and A' promise the highest expected return across the three possible states of the world that may obtain, and the only difference between the two sets of lotteries is the overwhelming likelihood (89% odds) that in the first scenario we will come away with $10, and that in the second that we will come away empty-handed. Independence requires that, if we prefer A over B, then we also prefer A' over B'.

I dwell on this third axiom to introduce one of the more interesting early challenges to SEU: in Allais's experiment, many respondents in fact claimed to prefer B over A and A' over B'. Nor, as both Anand (1996, 75) and Weber (1998) suggest, does this seem to be a ridiculous stance, in

spite of violating independence: in the first scenario, we might feel especially foolish gambling for a 1/10 shot at $100 and being unlucky enough to hit instead upon the 1/100 chance of winning nothing, this when we are certain to win $10 by playing lottery B instead. In the second scenario, there is no such certainty, so why not take the 1/10 chance, knowing that we are most likely to win nothing anyway? Why is it irrational to be sensitive to guaranteed gains, and also to the frustration (humiliation?) we would feel if, upon foregoing a sure thing, we had a stroke of bad luck and came away with nothing? Nor, for that matter, does it seem unreasonable to be similarly sensitive to profound uncertainty: another early experiment, conducted by Daniel Ellsberg, presented a similar choice situation as the Allais experiment, but for which some of the probabilities were not known by the players. Many of Ellsberg's subjects would violate independence, but apparently out of an aversion to uncertainty, as distinct from risk. (The distinction is due to Frank Knight: risk reflects objective probabilities, such as the outcome of fair coin toss. Uncertainty, in contrast, reflects lack of knowledge about the probabilities involved in a given choice situation.) Later work summarized by Anand (1996, 34-36) seems to confirm this interpretation of Ellsberg's results.[2]

What I want to take away from these results is not the point that SEU does not accurately describe choice behavior, but the implied normative puzzle: in some plausible situations, violations of SEU do not seem to warrant the charge of irrationality, because we can tell a story that makes sense of these choices, in terms that we think of as reasonable. And yet they are choices that go against what seems to be a rather objective set of facts about the world: for example, perhaps we do not want to accuse lottery players of being irrational, but do we really want to bequeath the laurel of rationality upon such ridiculously improbable aspirations? Put more generally: on the one hand, would we want a prescriptive theory of choice that counseled us to ignore our preferences for certainty and social acceptance, say, or the simple acts of whimsy and culturally acquired tastes ('but I like the casino – it's fun!') that give our lives their richness? But on the other hand, would we want a prescriptive theory of choice that, while descriptively accurate and thus sensitive to the complexities of our (socially constituted) identities, could nonetheless offer little more by way of counsel than 'do what seems appropriate, given who you are and how you feel'?

If we want an expressly normative account of rational choosing, then we may have to forgo some of the descriptive richness that we expect from an explanatory theory of rational choice. Anand (1996, chs. 8-9) offers what seems to me to be a very sensible set of alternative axioms for this purpose, which capture features we intuitively think of as essential to the idea of rationality: for instance, our deliberative faculties that attend to the *reasons* for and against a course of action, as well as our *intentional* stance with respect to what we desire. And in place of strong restrictions such as independence and transitivity, Anand offers the weaker requirement of *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. But in making his case, Anand rejects, as without useful normative content, the axiom of transitivity. By way of conclusion, I want to press him on this point, with an eye to perhaps recovering some of the intuitive appeal of this axiom by reexamining one of the

[2]Other experimental results, such as the famous preference reversals discovered by David Tversky, point similarly to the psychological vagaries of choice under risk and uncertainty, and suggest that, whatever motivates us in these situations, SEU does not seem to describe our actual choice behavior; yet again, see Anand (1996).

arguments he rejects, in light of an abiding concern in political theory.

Central to Anand's critique of transitivity is the complaint that a transitive ordering is critically sensitive to how we define the objects over which our ranking obtains, and this in turn seems to suggest that the appeal of transitivity as a maxim guiding our evaluations of rationality depends on the space over which we are ranking options. In some cases, preference really might be analogous to, say, length of height: if I prefer more oranges to fewer, then I cannot consistently rank three oranges as preferable to two, two preferable to one, and one preferable to three. This simply violates the definition I have assigned to my preference relation (i.e. 'more than'). But Anand asks us to consider this relation: 'plays better than'. It is not inconsistent to rank chess players according to whether they play better than their opponents, and yet transitivity would surely not apply to my preference to select the better chess players in the following case: suppose a player who lost out in an early round of a tournament returns to beat a contender in the final. My desire to see this 'comeback kid' crowned tournament champion is surely not irrational, although it seems arguably to violate transitivity.

But here is an argument for transitivity as a maxim that I think is more persuasive than others, at the very least for an important range of concerns. In the famous 'money pump' argument, a consumer with intransitive preferences is vulnerable to loss in the following way. Suppose I have the following preferences over three commodities: Pab, Pbc, and Pca, where P denotes a binary preference relation over objects a, b, and c. Now as Anand correctly notes, there is nothing on the face of it inconsistent with this set of binary preferences unless we add the further requirement that all of our binary preferences ought to map into consistent ternary or higher relations, i.e. that the set Pab, Pbc must entail the a ternary relation T of the following form: $Tabc$. But assuming this is really to assume a form of independence:[3] removing one item from the choice set should not alter our binary preference structures, so that removing b from consideration should leave a binary preference relation Pac, not Pca.

But now imagine that I have gone to market with my binary preferences Pab, Pbc, and Pca. A clever fellow has observed me for several months, and has inferred my preferences correctly. He now offers to trade me my c for b. I gladly accept. Later, he approaches me elsewhere in the market with another offer: he will trade me my b for an a. Again, I gladly accept. Later still, he again approaches me, this time offering his c for my a. I accept, given my preference for c over a. I am now back where I began, with c. These trades continue over the course the morning, afternoon, and into early evening, when finally I return home, content with my c. The catch? Each of these transactions involved a small fee, which, given my preferences in each case, I was happy to pay. I am at home in the same position I was in that morning, possessing c, but I am now several dollars poorer!

Now as Anand notes, we ought not to infer that, because of my binary preferences, I will inevitably enter into the sequence of exchanges just described. More likely, if I am aware of my intransitive preferences, I will simply refuse to enter into the sequence of exchanges described: I will not go to market or, being at the market, I will refuse to enter into the sorts of exchanges just

[3]More precisely, it is an assumption of *contraction consistency*, as Anand (1996, 57) notes; see also Binmore and Voorhoeves (2003, 279 n.12).

described. But this doesn't seem to me to address the normative implications of the example.

Suppose that we do not understand the intransitive form of our preferences over some options. In this case, we are vulnerable to exploitation in just the way suggested by the 'money pump' scenario. Some intriguing work is being done suggesting that, at equilibrium and with no restrictions on market entry, transaction costs will approach zero as enterprising sellers attempt to exploit unwitting bearers of intransitive preferences.[4] But if we concede this point, it seems perverse to deny the conclusion that, without intransitive preferences, we would not have to rely on any such equilibrium result to ensure our financial well-being!

Now suppose instead that we did understand the intransitive structure of our preference ordering over the hypothesized options. A similar conclusion seems to follow: certainly we could avoid exploitation by avoiding trades that leave us in the same position vis-a-vis preference satisfaction, yet cost us some positive amount of money (or time, or some other exchangeable good or service we have command over). But the maxim 'intransitive preferences over options are acceptable so long as you avoid exchanges of these options with transaction costs' seems a strange piece of advice to offer a decision-maker in need of guidance. After all, these are still our preferences, and if, upon reflection, we accept them, it seems unhelpful to prescribe that we must not enter into an institution – contractual exchange in a market – designed to facilitate satisfaction of our desires, guided by our preferences.[5]

To be sure, we might, upon reflection, suspect that intransitivity is lurking in a decision problem involving extraordinarily complex and confusing options. Perhaps we cannot find intransitivities in our preference ordering, but we would not be surprised if, upon sorting through the complexities of the decision problem and weighing the subtle costs and benefits of each option, we found intransitivity. But in such cases, concerns about intransitivity seem particularly pronounced in just those cases where the options are being presented to us by someone who stands to benefit from our involvement in the decision at hand. Think, for instance, of confidence schemes that rely on the promise of short-term gains from risky long-term investments in real estate. We might also think that similar tradeoffs between short-term satisfaction and long-term costs are behind the surge in offers of cheap personal credit in the United States. In these and similar cases, the possibility of ex-

[4]This remark is based on as-yet-unpublished work by David Laibson (Harvard and National Bureau of Economic Research) and Leaat Yariv (UCLA).

[5]Upon reflection, we might, of course, find that our *choices* suggest intransitive preferences, not by some brute fact of our needs and tastes, but rather because the different options reflect different choice contexts. For example, as Anand suggests, a child might prefer oranges to apples and larger fruit to smaller fruit, and other things being equal, she may prefer an orange to a small apple, but a large apple to a small orange. Yet surely we would not fault her reasoning if, in deference to habits of politeness taught by her parents, she accepted a small over a large apple, if these the only options available. A more difficult case arises when the goods we are ranking are difficult or impossible to compare. Anand presents the example of selecting among policies that seem to demand different evaluative criteria, such as weighing the costs of disease prevention, through either medicine or water management, against differing levels of food production. So there are contexts in which transitivity seems to be symptomatic of the choice situation, and not necessarily our preferences *per se*. But even in these cases, the maxim 'accept intransitive outcomes when contexts force them upon you' seems to demand further argument, at the very least to explain why the intransitivities in these cases are not problematic, and such arguments will appeal to the constraints themselves – for instance, the virtues of upholding politeness as a norm, or the necessity of choosing a policy involving multiple services when some good will come of whatever choice is made, regardless of intransitivity with respect to our preferences over the outcomes associated with particular services.

ploitation at the hands of another gives normative weight to the concern that transitivity may make us vulnerable to costs that, upon sufficient information and reflection, we would rather not bear. This is not to say that such tradeoffs will necessarily reveal intransitive preferences. For instance, I agree with Binmore and Voorhoeve (2003) that efforts to find intransitivity in balancing problems involving pains of strikingly different intensities and durations (Temkin 1996; Rachels 1998) are really just instances of difficult decisions of minimizing pain along two important dimensions, and do not lead to intransitivities.

But suppose such difficult tradeoffs – between unbearable pain from sudden injury, on the one hand, and modest but persistent pain over the course of one's life from lack of treatment for a chronic condition, on the other – were offered to us by an insurance salesman trying to persuade us to buy a limited policy. If there are transitivities to be found – or even if not, but if the potential for mistaken judgements and subsequent regret seems significant – then our concern is that this salesman is exploiting us by taking our money for reasons mostly his own, for gains enjoyed chiefly by him, and for reasons the terms of which we would find unpersuasive, upon having better and more complete information and the time, skills, and resources needed to reflect carefully upon it.

These sorts of concerns – that is, about exploitation, understood as costs that don't track our interests and which are not justified by reasons we'd accept – clearly echo Anand's in formulating his alternative axioms of rational choice that better capture the deliberative, intensional, and justificatory character of rationality. And yet even accepting this move, it seems that worries about exploitation lead us to value transitivity as an axiom just because it renders us less vulnerable exploitation by others.

Exploitation (broadly understood, as above) is an abiding concern of political theorists: marxists of course are unrelentingly critical of exploitation in labor markets, and liberals of various stripes worry that the lucky might exploit their good fortune, securing their favorable prospects at the expense of those less fortunate (for instance, Rawls 1971). When civic republicans (for instance, Pettit 1997) talk about freedom as the absence of domination, it is not a big stretch to conclude that what they mean by 'freedom from the arbitrary will of another' is something like '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'. And if this broad definition of exploitation holds up, then many feminists, and indeed, many Kantians are arguably on board as well: systematic biases, in attitudes and institutions, that reflect and reproduce male advantage are objectionable because they permit men to use women as means to their own ends. And the language I've used here is, of course, roughly Kantian: to use others merely as a means to one's own ends violates the categorical imperative.

The modest defense I've offered here for the transitivity axiom – that it is given not only sense, but also normative force by one of the distinctive concerns of political theory: the avoidance of exploitation – may seem a strange concession to ask of rational choice advocates, because it seems to suggest that rationality is bound up not simply with utility maximization and judgements about states of the world, but more importantly with our character, specifically, with our being the sorts of persons who are not easily exploited by others. Taking a stand on desirable character traits is not something most decision and game theorists spend much time doing! But some normative

political theorists must also make a concession here, even if their initial inclination is to criticize rational choice approaches: if you think exploitation is a bad thing, ethically speaking, then it is hard to resist the conclusion that you should also find transitivity to be an attractive maxim for governing our reflections on what we desire and how we mean to satisfy those desires. In settings where decisions are complex and uncertain, intransitivity makes us vulnerable to being used by others in ways that, upon informed and careful reflection, we would not freely accept.

References

Anand, Paul. 1996. *Foundations of Rational Choice under Risk*. Oxford: Oxford University Press.

Binmore, Ken and Alex Voorhoeve. 2003. "Defending Transitivity against Zeno's Paradox." *Philosophy and Public Affairs* 31: 272-279.

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

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

Rawls, John. 1971. *A Theory of Justice*. Second Edition. Cambridge, MA: Belnap Press, 1999.

Skyrms, Brian. 1997. *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Taylor, Charles. 1982. "Rationality," in Martin Hollis and Steven Lukes, eds. *Rationality and Relativism* (Cambridge, MA: M.I.T. Press): 87-105.

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

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