Preference and Logic

Richard Bradley

Department of Philosophy, Logic and Scientific Method

London School of Economics

r.bradley@lse.ac.uk

In her ambitious and wide-ranging book *Rationality and the Structure of the Self*, Adrian Piper seeks to establish the basic principles of what she calls transpersonal rationality, the form of rationality constitutive of the Kantian conception of the self. Transpersonal rationality is governed by principles that require us to transcend our personal preoccupations and interests and focus on those that apply to all in equal measure. In contrast the rival Humean conception of the self, the main foil for her argument, draws on an egocentric form of rationality directed at the instrumental fulfilment of the agent's desires but not at their content.

No short discussion of this book can hope to do justice to its richness, so I will focus on one important strand of her argument, that concerning the interpretation of formal decision theory and its concepts and principles. Piper's position on this question is both very interesting and unorthodox. It is fair to say that the Humean conception is the predominant one amongst economists and decision scientists, but Piper argues that formal decision theory does not depend on it and that decision theory can be given Kantian foundations. Indeed, she argues that the Kantian conception provides a better foundation for the rationality principles underpinning mainstream decision theory and that its adoption solves important difficulties for this theory.

Piper's argument centres on the concept of preference, a notion that gives content to formal decision theory and allows for explanations of choices in terms of utility maximisation. What Piper argues, correctly I believe, is that to justify the axioms of preference that underpin expected utility theory, preferences cannot be understood purely behaviouristically, but must be taken to be psychological states – judgements – susceptible to rational evaluation. In particular, the central axiom of transitivity, if it is to serve as a universal rationality constraint on preference, must express a *logical* principle. For if preference is interpreted choice-theoretically, as in Revealed Preference theory, then transitivity merely captures a behavioural regularity which may or may not be exhibited by particular agents but which is of no normative significance. On the other hand, if it is interpreted purely formalistically then the representation theorems of decision theory become vacuous; demonstrations of necessary and sufficient conditions on a primitive binary relation for numerical representation of the expected utility form that lack any normative or explanatory significance.

Piper gives formal expression to this point in Chapter III of the second volume of her book where she attempts to give a rigorous foundation to transitivity as a logical principle.

The key, she argues, is to recognise that binary preference relations are not themselves logical connectives taking propositions as arguments, but rather are constituents of judgements:

> "expressions of intentional operators that denote certain of a subject's intentional attitudes – namely, preference and weak preference respectively – toward pairs of intentional objects – namely preference alternatives." – (Vol. 2, p. 115)

This puts the judgement that one alternative is preferable to another on the same footing as the judgement that some proposition is true or the judgement that it would be good if the proposition were true. And like such categorical judgements, these relational judgements, or the propositions that encode them, can serve as constituents of more complex propositions, compounded by means of the usual Boolean connectives: and ($\wedge$), or ($\vee$) and not (¬).

We are now in a position to see how logic can get a hold on expressions of preference. Consider, for instance, the weak reflexive relation 'at least as preferable as', which I will denote by '$\precsim$' so that '$X \precsim Y$' denotes the judgement that *Y* is at least as preferable as *X*. Let $X < Y$ mean that $(X \precsim Y) \wedge (\neg(Y \precsim X))$. The logical properties of Boolean compounds of such expressions are no different from other expressions. So, in particular the propositions $X \precsim Y$ and $\neg(X \precsim Y)$ are contradictory, and hence the preference judgements they express are not co-tenable. But since preferences are expressed by choices, it follow that:

> "… choice behavior is just as much subject to the consistency requirements of classical logic as speech behavior, in both normative and descriptive systems; and so that the utility-maximizing model of rationality is similarly subject to a more inclusive, Kantian model of rationality that places classical logic at its base." - (Vol. 2, p. 115)

What light does this conclusion shed on the requirement of transitivity? Within the judgemental interpretation, transitivity is best regarded as an inferential principle, applicable *inter alia* to those binary relations encoding an ordering principle such as betterness or preference. Now someone who judges that $X \precsim Y$ and that $Y \precsim Z$ but not that $X \precsim Z$, does not make any strictly contradictory judgements, so cannot in my opinion be said to violate the law of non-contradiction. They do however fail to adopt all the logical consequences of the judgements that they make and so fall short of logical omniscience. Such failures are criticisable: one can say to such a person 'given that you judge *Y* to be at least as preferable as *X* and *Z* to be at least as preferable as *Y*, you are committed to judging that *Z* as at least as preferable as *X*. For this follows from what you judge to be true". But the failure is quite different from that of logical inconsistency.

Why then do decision theorists typically regard transitivity as a requirement of rationality? The reason is, I suggest, because they typically assume that agents' preferences are *complete* (i.e. they respect the condition that $X \precsim Y$ or $Y < X$). When this is the case the person who violates transitivity makes judgements that are *implicitly inconsistent* in the sense that, were they to derive all logical consequences of the judgements that they make they would find themselves with strictly inconsistent ones. For example, suppose you judge that $X \precsim Y$ and that $Y \precsim Z$ and that your judgements violate transitivity. Given the assumption of completeness this implies that you judge that $Z < X$ (or else, contrary to hypothesis, your judgements would respect transitivity). But that fact that you judge that $X \precsim Y$ and that $Y \precsim Z$ implicitly commits you to the judgement that $X \precsim Z$. And so, were you to derive this consequence by applying the inference rule captured by transitivity, you would end up judging both that $X \precsim Z$ and that $Z < X$. But by definition $Z < X$ implies that $\neg(X \precsim Z)$.

The fact that violations of transitivity in the presence of completeness implies that the agent's preference judgements are implicitly inconsistent is, I suspect, why Piper regards failure of transitivity as tantamount to logical inconsistency. But an agent's preferences are not rationally required to be complete. She might lack the information required to form a judgement on some matters or she might lack the time or cognitive resources to do so or, for that matter, the issue may be of so little relevance to her practical interests that is not worth her while to do so. Those who reduce preference to choice will argue that insofar as you are disposed to choose one way or another, you implicitly hold preferences over all alternatives. But this is not Piper's conception of preference and so I see no reason for her to insist on completeness. But then transitivity should not, as I have argued, be viewed as a requirement of consistency, but rather the weaker one of closure of one's judgements under logical consequence.

12 July 2018