

Foundations of Decision Theory

Issues and Advances

Edited by
Michael Bacharach
and
Susan Hurley

Basil Blackwell

- 14 The most general theorem is in Gorman (1968). It is perhaps worth my explaining briefly how the theorems work. The question, remember, is how good at the different locations aggregates together to make up general good. Because there are three dimensions, the aggregation has three stages. Take a person and a time. An act gives to that person at that time some good in each of the states of nature. The good in all the states aggregates together to determine the good of the person at that time. That is the first stage. Then – the second stage – the good of the person at all times aggregates together to determine the good of the person. Finally – the third stage – the good of all the people aggregates together to determine general good. But we did not have to take the stages in that order. We might have started instead, say, with a given person in a given state of nature. The person's good in that state at all times aggregates together to determine her good in that state. Then all the people's goods in that state aggregate together to determine general good in that state. Finally, good in all the states aggregates together to determine general good. These are merely two different routes to determining overall good. Therefore they must give the same answer. This means that aggregation in one dimension cannot be independent of aggregation in another. That is the basis of the theorems.
- 15 I have examined the two dimensions of uncertainty and people, and particularly Harsanyi's argument, in Broome (1987). Broome (1991d) tries to take the third dimension into account too. Broome (1991c) is a fuller investigation.

4

Decision Theory and Folk Psychology

Philip Pettit

I INTRODUCTION

The standard view of how Bayesian decision theory relates to folk psychology is that it provides an explication, under idealization, of the central sound core of that psychology. David Lewis gives expression to this explication thesis as follows.

Decision theory (at least if we omit the frills) is not esoteric science, however unfamiliar it may seem to an outsider. Rather it is a systematic exposition of the consequences of certain well-chosen platitudes about belief, desire, preference and choice. It is the very core of our common-sense theory of persons, dissected out and elegantly systematised. (Lewis, 1983a, p. 114)

But if decision theory explicates certain well-chosen platitudes of folk psychology, the alleged core of our commonsense theory of persons, does it ignore any others? In this paper I identify one neglected platitude and argue for its importance. I assume that the explication thesis is sound but I complement it with an abstraction thesis: a thesis that although decision theory explicates certain platitudes, it abstracts away from others.

The paper is in five sections. The second rehearses the main assumptions of Bayesian theory in its different versions. The third presents the explication thesis and the fourth argues for the abstraction thesis. Finally the fifth section looks at the significance of the abstraction alleged. I argue that the abstraction means that decision theory is incomplete, nonautonomous, and nonpractical. The nonautonomy result may be the most interesting for decision theorists, connecting with pressing concerns about the individuation of options.

2 DECISION THEORY

There are a number of different versions of Bayesian decision theory but we need not concern ourselves with their distinguishing features. All that we need to appreciate is the hard core of propositions that those versions share. There are three subsidiary principles associated with the theory and one central principle of rationality. The principle of rationality has two sides to it, being first a principle of rational preference and second a principle of rational choice.

The first subsidiary principle is that for any chooser to whom the theory applies we can identify a suitable domain of items over which he can have preferences and a suitable domain of items to which he can attach probabilities. In some versions these domains are the same; in others they are different. One of the things that makes the domain of preferences suitable is that it includes items that can be equated with the options facing the agent in any situation of choice or can be used to construct something equivalent; an option is usually identified as an exhaustive and exclusive disjunction of outcomes and the domain of preferences will include such disjunctions or at least such outcomes. One of the things that makes the domain of probabilities suitable is that it ensures that there will be a suitable probability associated with any outcome.

The second subsidiary principle is that the choosers to whom the theory is intended to apply have an appropriate preference ordering over the items in the preference domain. The ordering must be complete in the sense that no item fails to be ranked. It must be consistent, in the sense that if A is preferred to B and B to C then C cannot be preferred to A. It must also satisfy one or more other conditions, of which the most common imposed is a form of continuity assumption.¹

The third subsidiary principle is that the agents to whom the theory is intended to apply have an appropriate probability ordering over the items in the domain of probability. If that domain is suitable, then it will constitute a Boolean algebra; this means that if it includes two propositions A and B, for example, then it also includes not A, not B, A and B, A or B, and so on. The probability ordering will be appropriate if and only if it enables us to assign to every item X in the algebra a real number $P(X)$, such that it satisfies the Kolmogorov axioms and represents the agent's probability for that item. These stipulate the following (see Skyrms, 1975, ch. 6):

- 1 $P(X) \geq 0$ for every X;
- 2 $P(X) = 1$ if X is a tautology;
- 3 $P(X \text{ or } Y) = P(X) + P(Y)$ if X and Y are mutually exclusive.

That an agent satisfies these three subsidiary principles does not mean that he will be intuitively rational in his preferences. The principles allow this sort of irrationality, for example: that someone should prefer a disjunction of A and B both to A and to B. The principle of rationality for preference rules out such possibilities. Let the items in the domain of preference be ranked according to the agent's preferences on some scale, say from 0 to 10; its place on the scale determines what is known as an item's subjective utility. The principle of rationality for preference dictates the appropriate place on the scale, the appropriate subjective utility, for any disjunctive item such as A-or-B, where the alternatives are exclusive and exhaustive. It says that the place ought to be determined by the sum of the scale figures for A and B, each figure being discounted by a number which represents the appropriate probability, in the mind of the agent, of that item's being realized rather than the other. If A is scaled at 2 and B at 7, and if A has a probability of $\frac{3}{4}$ and B of $\frac{1}{4}$ - the numbers must add to 1 - then the appropriate place on the scale for A-or-B is $2 \times \frac{3}{4} + 7 \times \frac{1}{4} = 3\frac{1}{4}$. When subjective utility is so understood that it can be determined in this way as well as more directly, we speak of subjective expected utility (SEU).

Any theory which ascribes rational preferences makes for a decision theory, so far as each option in any situation of choice can be equated with an item in the domain of preference or with an exhaustive and exclusive disjunction of such items. The principle of rationality for choice says that the choice of an option will be rational as long as it maximizes subjective expected utility. The choice of an option O_1 over an option O_2 will be rational if and only if $SEU(O_1) > SEU(O_2)$. If O_1 and O_2 are simple items in the domain of preference, then O_1 must be ranked above O_2 . If they are disjunctions the appropriate sum for O_1 must be higher than that for O_2 .

Different versions of Bayesian decision theory differ in a number of ways.² They differ in ontology, taking the items in the domains of preference and probability to be different sorts of things, they differ in their views of the sorts of probabilities which it is appropriate to introduce, and they differ in how precisely they axiomatize the theory. But such differences still allow them to give a common endorsement to the sorts of principles presented. The presentation of those principles is sufficient for our purposes in this paper, but some may find it useful to see how the concepts involved in the theory can be given operational sense. Probably the best way to do this is to look at the approach suggested by Frank Ramsey which is described in the appendix.

3 THE EXPLICATION THESIS

The explication thesis requires two points to be established: first, that folk psychology involves a certain incontestable core of theory; second, that decision theory explicates that core. In practice the first stage in defending the thesis comes to a defense of the assumption of intentional agency, and the second to an argument for identifying subjective probabilities and utilities respectively with the desires and beliefs postulated under that assumption.

The assumption of intentional agency involves three components.

- 1 Every action issues from the agent's beliefs and desires.
- 2 Those beliefs and desires constitute a reason for the agent as to why the action should have been performed; they mean that he desired an action of a certain sort and that he believed that he would bring one about by doing what he did.
- 3 The beliefs and desires cause the action to occur in virtue of rationalizing it in this way, and not by a deviant route: not, for example, because their presence produces a temporary failure – say embarrassment – which has the fortuitous result of engendering the appropriate response.³

I shall not argue here either that the intentional assumption is implicit in folk psychology or that it is sound. Both points are generally, if not universally, granted among contemporary philosophers and I am happy to go along.⁴

The second stage in defending the explication thesis requires an argument that decision theory explicates the assumption of intentional agency. The argument might go like this. If an agent has the subjective probabilities and utilities postulated, then provided that their contents are suitable, those states serve, like the beliefs and desires assigned under the assumption of intentional agency, to give the agent reasons for choosing as he does; they do so, at least, provided that they are taken as more than fictions.⁵ The most economical way of viewing such an agent will then be to identify the posits respectively of the theory and the assumption – to equate beliefs with subjective probabilities, and desires with subjective utilities. And that view amounts to seeing decision theory as explicating the assumption of intentional agency.

It may be thought to be an objection to the equation of the two sorts of states that only subjective probabilities and utilities come equipped with numbers. But the objection is not compelling, for the number can be seen as a way of coding the degree of strength of the corresponding

belief or desire. I think that there are no persuasive objections of this kind to the equation and so I am prepared to go along with the explication thesis. Doing so without the ceremony of full-scale argument may be excusable, given that my ultimate purpose in the paper is to show that the thesis is subject to an important and little noticed limitation.

It is important to realize that the decision theory which explicates beliefs and desires involves a great amount of idealization. What the explication thesis says is that the subjective probabilities and utilities which an agent would have under the idealized circumstances described in the relevance conditions are his beliefs and desires. It does not say that, for any agent who has beliefs and desires, those states constitute subjective probabilities and utilities, or at least the full range of subjective probabilities and utilities ascribed in decision theory. Rather, what holds is the reverse, namely that if an agent has subjective probabilities and utilities, then they are his beliefs and desires by other names.

Finally, a caution. As I have stated it, the assumption of intentional agency is silent on how an agent's beliefs and desires should change in the light of certain changes of belief – changes reporting new evidence and the like. Equally, as I have stated it, decision theory is silent on how the agent's subjective probabilities and utilities should shift in response to certain changes in probabilities: this is the topic of probability kinematics. Thus the explication thesis has quite restricted scope. I have nothing against the enriched version, however, under which the thesis is that decision theory as enriched by a suitable probability kinematics explicates the assumption of intentional agency as enriched with a suitable assumption of attitudinal rationality. Indeed, henceforth I shall write as if the explication thesis takes this richer form.

4 THE ABSTRACTION THESIS

The best way into the abstraction thesis that I wish to defend is probably to identify the folk psychological platitude which I claim that decision theory ignores. I call the platitude the assumption of *desiderative structure*. What it says is that there are two quite different sorts of object which desires may have – prospects and properties – and that the desires that we form for different prospects are determined by the properties that we think they have. Any prospects that we desire, any prospects that we prefer to the relevant alternatives, we desire for the properties they display or promise to display.

A *prospect*, in my usage, is what would more commonly be described as a state of affairs: something like the state of affairs involved in my

going to London this afternoon, in Western banks' extending the repayment period on Third World loans, or in the greenhouse effect's proving not to be a reality. It is any way that the world may be. At the limit it is any token way that the world may be, that is, any particular possible world; more usually, it is any type of way the world may be, that is, any set of possible worlds. The prospect that *p* is just the set of possible worlds at which it is the case that *p*, and so on.⁶ I shall use a description to pick out any prospect; usually I will present it via a sentence – say “*p*” – expressing the state of affairs in question. But I can think of it as a state of affairs satisfying other descriptions as well as the description used to pick it out. The prospect is coarsely individuated so that it is an *a posteriori* matter whether certain prospect-identifying descriptions pick out the same prospect or not.

Every prospect involves the realization of a certain *property* or properties, whether in a given individual or individuals, in a given domain, or after a given pattern. The property involved may be of a variety of forms; thus it may be relational or nonrelational, as in the difference between the property of equality and the property of mass, it may involve a particular, as in the property of speaking French, or it may be universal, like the property of being intelligent, and so on. A property can be seen as a distinct sort of entity which belongs to any prospect that involves it, though there is a qualification to be made in a moment about this claim; the property of traveling will belong to the prospect of my going to London this afternoon, and so on. Like a prospect, a property in this sense is a coarsely individuated entity – something such that for at least some expressions in a language it is an *a posteriori* matter whether they pick out the same property or not. I may discover that what I prize as the elegance of certain paintings is just the classical quality for which you admire them or that what I took to be the cruelty of certain actions is what you described as their brutality. There are not as many properties as property-expressions: properties are independent entities to which the expressions help us refer.

But though properties can be seen in this way as distinct sorts of entities which belong to the prospects that involve them, there is a qualification to be made about that representation. This is that, for all we need to say, a property can equally be represented as itself a type of prospect. Ignoring some complications, we can represent the property of being an *F* as equivalent in all significant respects, for example, to the prospect that there is something which is *F*: that is, equivalent to the set of possible worlds at which there are *F*s. If this representation is preferred, then it will affect the formulation of the assumption of desiderative structure, but it will not alter the substance. The assumption will not be that we

desire properties as well as prospects and desire prospects for the properties that we think they have. Rather, it will be that among the set of prospects there is a special class – those corresponding, for appropriate properties, to sentences such as “there is something which is *F*” – and that, for any prospect we desire, we desire it because we see it as involving the realization of one such privileged sort of prospect. This principle will apply to prospects within the special class as well as to prospects outside it. It is important to recognize this possibility, for fear of misunderstanding, but in what follows I shall set it aside. I shall assume that properties are distinct sorts of entities from prospects and that they belong to the prospects that involve them.

The assumption of desiderative structure is that, not only do we desire prospects, we also desire properties, and that we always desire prospects for the properties we think they have. Take the notion of prospect-preference as given – the notion of ranking prospects in a preference-ordering.⁷ This enables us to identify what it is to desire a prospect and what it is to desire a property. To desire a prospect is to prefer it to the prospects that you think of as the alternatives. To desire a property is to be disposed to prefer a prospect that has it, assuming that there is only one, among a set of prospects that otherwise leave you indifferent. More intuitively, to desire a prospect is to opt for it, or to form the intention of opting for it, among the set of available alternatives; to desire a property is to value it, being disposed, if other things are equal, to desire any prospect that displays the property. Notice that under these definitions we can think of desire as being involved in the same sense in each case. If there are two sorts of desire, that is not because there are two senses of the term; it is only because there are two sorts of objects for desire in the one and only sense of “desire” available. But enough of abstract definition. It is time to introduce our distinction with examples.

Consider the self-ascription of desire involved in my saying that I desire that *p*: for example, that I desire to have a teaching job, or to live in a warmer climate, or to be moral.⁸ One context in which I may make such an ascription is where the proposition “*p*” picks out a particular prospect or state of affairs from among a set of fixed alternatives. I have a choice between staying in research and going to a teaching job, between living in Canberra and moving to Queensland, between doing something of questionable ethics and being more punctiliously moral. In such a context I shall indicate a desire for a particular prospect by reporting that I desire to teach or move nearer the tropics or be moral.

But now imagine that the context of ascription is different and that it is not assumed that “*p*” picks out one among a fixed set of alternatives. Suppose for example that I have been asked about the things that I would

like in life and that I say that I have a desire to have a teaching job, to live in a warmer climate, and to be moral. Here these ascriptions clearly do not pick out desires for particular states of affairs or prospects. So what are the objects I claim to desire? The ready answer is that I desire the properties in question in the ascriptions. The presence of one of those properties will make any prospect the more attractive to me; at the limit its presence in one of a number of alternatives between which I am otherwise indifferent will lead me to prefer that alternative. Among a set of career prospects I will tend to prefer a teaching job, at least if other things are equal; among a set of residential alternatives, one near the tropics; among a set of behavioural options, one that is morally permissible. And so on. I may prefer a nonteaching job if presented with a certain set of alternatives, for it may be that the job scores very well in respect of other attractive properties. But that the job is not in teaching will still count against it; compared with the abstractly possible job which is identical except in this respect – a job not in fact available as an alternative – it will look inferior (see Jackson, 1985a).

The assumption of desiderative structure marks this sort of distinction between desires for prospects and properties. But it also goes further. It says that whenever an agent desires a prospect, he does so because of the properties it displays: he desires it for the desired properties that it promises to realize. In scholastic terms, the prospect is the material object of his desire, the properties the formal object. The observation is common to the Aristotelian way of thinking about these matters.⁹ *Quidquid appetitur sub specie boni appetitur*: whatever is desired is desired for being good. This extra element in the assumption of desiderative structure ought not to be surprising. If there are independent desires for properties as well as prospects, and if they are not idle wheels in our psychology, the only obvious role that they can play is in the determination of prospect-desires.

There are three points that I maintain about the assumption of desiderative structure: first, that decision theory neglects it; second, that it is recognized in our folk psychology; third, that it is a reasonable assumption to make.

It ought to be clear, I think, that decision theory neglects the assumption.¹⁰ In Bayesian theory we assume a domain of items – in effect, different states of affairs or prospects – over which the agent has preferences. We suppose that every option that the agent faces – the prospect of doing A, doing B, or whatever – appears among those items or can be equated with a suitable disjunction of items: the disjunct items will be subprospects of the prospect constituted by the disjunction. Then we postulate that if a rational agent ranks certain outcomes in a particular

way in his preference ordering then he will rank disjunctions of those outcomes in a corresponding manner, and that as he ranks items or disjunctions of items which correspond to options so will he choose among them: he will choose so as to maximize expected utility. The outcomes and options invoked here are all particular states of affairs or prospects. What the theory does then is formulate a constraint of consistency which the rational agent will satisfy in the preferences and choices that he forms over such prospects.

The assumption of desiderative structure postulates a different constraint of consistency on the rational agent. It is not a constraint of consistency between a prospect-desire and the agent's preference-ranking over its subprospects and the subprospects of its alternatives. Rather, it is a constraint of consistency between a prospect-desire and the agent's property-desires, in particular, his desires for the properties exhibited by the prospect and its alternatives. In concentrating on the first constraint of consistency, decision theory neglects the second. It asks after what an agent's preferences over relevant outcomes rationally require of him in his decision between certain options. It ignores the other question, of what an agent's values – the properties he cherishes – require of him in the decision.

But while decision theory neglects the assumption of desiderative structure, and the constraint of rationality that it involves, nothing in the theory is strictly inconsistent with the assumption. The theory tells us what prospect an agent ought to desire, given his preferences over the relevant subprospects. Perhaps the assumption only tells us how he ought to form preferences over those subprospects, given his desires for the different properties they display. Perhaps it only bears on the original prospect-desire indirectly, so that there is no potential conflict between decision theory and the assumption. These matters will come up again in the next section, particularly in the discussion of the nonautonomy claim.

We have seen that decision theory neglects desiderative structure, though it does not rule it out. What then of folk psychology? Is it clear that in our everyday habits of thinking about desire we distinguish between property-desires and prospect-desires, and see the former as serving to determine the latter? I believe that this is clear and that little more needs to be said in defense of the claim than is already implicit in the remarks about the examples used in introducing the distinction. After all, those remarks appeal to what we all find familiar, and that they introduce the assumption of desiderative structure shows that this is part of our common lore.

But in case you are not persuaded, here is a thought which may

convince you that we folk psychologists, even those of us schooled in decision theory, postulate property-desires at the source of desires for prospects.¹¹ There is a family of paradoxes acknowledged by decision-theorists which ought not to be paradoxes in the absence of desiderative structure; that we find them paradoxical therefore shows that we endorse that assumption. For example, consider someone who prefers to go to the other side of town to buy a bicycle for \$50 less than he can buy it on this side – for \$150 rather than \$200 – but who does not prefer to go to the same trouble in order to save \$50 on the price of a car – to buy it for \$15,000 rather than \$15,050. Thinking strictly in terms of prospect-desires, there is nothing even slightly irrational, and nothing therefore paradoxical, about such a pair of preferences. Yet most of us do feel some tension. Obviously the explanation is that most of us assume that property-desires drive prospect-desires and are therefore ill at ease with the notion that the property-desire that is apparently relevant in the first case – the desire to save \$50 – is irrelevant in the second. If we regain our ease, it will probably be through coming to assume that the relevant property in the first case is not that feature but one which is absent in the second: say, if this makes economic sense, the feature of buying the commodity at 25 percent less.

Let us agree that as decision theory neglects desiderative structure, folk psychology recognizes it. The final question is whether it is in fact reasonable to postulate such a structure in our desires. I believe that it is, on the grounds that the structure offers the best explanation of a variety of phenomena. Consider first the ambiguity, noted above, in ascriptions of desire. In ascribing to myself or to any other agent the desire that *p*, I may be attributing a property-desire or a prospect-desire. That ambiguity may be capable of being otherwise explained, but the most natural explanation is the account in terms of desiderative structure which is assumed in my very characterization of it. This phenomenon of ambiguity is not the only one that can be nicely explained by positing desiderative structure. I shall offer four other examples here.

The first is the phenomenon of internal conflict in desires; in a recent paper Frank Jackson (1985b) has argued that this is best explained by a distinction like ours. Consider the conflict that I feel when confronted with a choice between, say, attending an important departmental meeting and seeing my son perform in the school play. Consider, more particularly, the conflict that I may continue to feel even after deciding for the meeting; that is, even after forming a preference or desire for that prospect rather than the other. How do we explain the continuing conflict, given that the first prospect has triumphed – given, in other words, that there is only one prospect-desire present, the desire for that

prospect? An attractive explanation is offered by the assumption of desiderative structure, for it enables us to say that after coming to desire the meeting rather than the play I can continue to feel the pull of the property which put the play in the running, even if it was not enough to earn it victory – the property of enabling me to see my son on stage.

A second phenomenon which desiderative structure enables us to explain is the distinction, common in many quarters, between desire *simpliciter* for a state of affairs and *prima facie* desire. That distinction is often taken as a primitive, but the assumption of desiderative structure lets us see how it can be derived from more basic considerations. Under that assumption I come to form a desire for a prospect only so far as I identify it as the bearer of certain properties which I already desire. I come to desire that *p*, period, only so far as I desire that *p*, *qua* *F*. But then it is natural to say that the prospect-desire is the desire *simpliciter* that *p*, and the other state – the desire that *p*, *qua* *F* – the *prima facie* desire. Furthermore this goes with the fact that *prima facie* desire is sometimes also cast as desire *pro tanto* or desire *secundum quid* – desire in so far as something is true, desire in a certain respect.¹²

A third phenomenon explained by the assumption of desiderative structure is closely related to the last. It is the apparent fact that linguistic desire-contexts are not extensional: that even if *A* desires that *p*, and *p* if and only if *q* – so that the *p*-object of desire just is the *q*-object – still we cannot say that *A* desires that *q* without being misleading. John desires to go to the movies, and will disappoint his mother tonight if and only if he goes – so that going to the movies just is disappointing his mother – but he may not desire to disappoint his mother, or so we regularly say. The assumption of desiderative structure suggests a straightforward explanation of this phenomenon. When we use a sentence “*p*” to ascribe a prospect-desire, we naturally pick a sentence that serves to indicate the relevant property-desire also – a sentence involving predicates that alert us to the property in question. John desires to go to the movies and desires that prospect for the property of its involving him in going to the movies, or for a closely related property. Thus if we replace “*p*” in the original desire-ascription with a sentence that picks out the same prospect but under a different property, we run the risk of misleading our audience about the property desired. John desires the prospect which involves disappointing his mother but he does not desire it for the property of disappointing his mother; he may not even have realized that it would disappoint her. Hence it is misleading to say without qualification that John desires to disappoint his mother. The failure of extensionality is unsurprising.

The fourth phenomenon which desiderative structure lets us explain is

the practice among human agents of seeking and giving certain sorts of reasons for choice. Consistently with the decision-theoretic picture, the picture involving prospect-desires only, an agent will have only one sort of reason to offer in explanation or justification of what he does: the fact, however elliptically expressed, that the option chosen best served his desires for prospects in general, according to his beliefs. But this sort of consideration will not do to answer challenges such as these. "How could you want anything so cruel?" "How could you desire such a comparatively unfair outcome?" "How could you ignore the self-destructive aspects of your decision?" To respond to these questions with a suitable reason, the agent will have to point to properties of the option, or of its potential outcomes, which made it attractive to him – ideally, to properties such that his interrogator can understand how someone might be moved as he was by desires for their instantiation. The availability of such answers, however, is explicable only under the assumption of desiderative structure.

Finally, an objection. The decision-theoretic picture, it may be said, can make room for the way in which properties are invoked in explanation of these four phenomena; there is no need to posit property-desires as independent of prospect-desires. The idea will be that although the only desires I have are desires for prospects, still I may often identify properties that are common as a matter of fact among the prospects I desire. Identifying such properties in an option forgone, I may experience conflict; identifying them in any option at all I may represent myself as having a *prima facie* desire for that option; identifying them in an ascription of desire, I may see the ascription as nonextensional; identifying them in an option chosen, I may succeed in making sense of the choice for another.

The objection comes of confusion. If I am oriented only to prospects, if properties are not objects of desire independently, then the trick suggested works in none of these cases. Consider the property of the option forgone in the earlier example: the property of enabling me to see my son on stage. If prospects are all that concern me then, even if this is generally a property found only in prospects desired, its presence in the option forgone will not explain any lingering desire. Why should it cause desire to linger given in this case that if the property had been realized then the undesired prospect would have eventuated in place of the one desired?

The point to which we are directed shows also why the trick will fail in the other cases. Identifying in an option before me a property present in prospects frequently desired in the past will not furnish me with a *prima facie* desire; at best it will provide a *prima facie* reason for predicting that

I will come to desire that option. Identifying a reference to a commonly featured property in an ascription of desire will not explain why we cannot substitute reference to another property of the prospect desired, at least if the other property is equally recognized as a property of the prospect; nothing distinctive will be signaled by reference to that property, if prospects are the only things we desire. Identifying in an option chosen a property present in prospects often desired by others will not justify the choice to them; at most it will show them that in one respect, as no doubt in countless others of equal irrelevance, the choice is like choices they make. I conclude that if phenomena like those cited are to be explained, then we need a robust distinction between prospect-desires and property-desires, a distinction such as the assumption of desiderative structure posits.

5 THE SIGNIFICANCE OF THE ABSTRACTION THESIS

The explication thesis has it that decision theory explicates, in David Lewis's words, certain well-chosen platitudes about belief, desire, preference, and choice. What we have seen is something not inconsistent with this, that equally decision theory ignores certain other platitudes about such intentional states and acts. That this is so may not be found particularly interesting, however, unless it has some lesson for the significance of decision theory. In this final section I shall try to show that our abstraction thesis does have such a lesson. I shall argue that the thesis implies that decision theory is first incomplete, second nonautonomous, and third nonpractical. The incompleteness claim bears on decision theory as a descriptive device; the nonpracticality claim bears on its status as a normative instrument. The nonautonomy claim is relevant to decision theory in both its roles.

5.1 Decision theory is incomplete

The incompleteness claim follows fairly directly from the considerations mentioned in discussion of desiderative structure. But it is probably worth spelling out in a little detail. The claim is that decision theory is an incomplete account of the matters relevant to decision-making. In decision-making the agent's preference-ordering over a certain set of prospects gets to be determined, whether consciously or unconsciously, deliberately or mechanically. Decision theory charts some of the elements that play a role in such preference formation but if the incompleteness claim is correct then it systematically misses others.

Decision theory does not allege its own completeness. But it is natural, if only because of the name given to the theory, to think of it as a complete account of preference formation.¹³ Decision theory is generally committed to the following principle of the co-determination of prospect-preferences and that principle is easily rendered as a principle of completeness.

For every prospect, the place of that prospect in a rational agent's preference-ordering is determined – given his probability function – simultaneously with the places occupied by its subprospects; the place of each subprospect is co-determined in the same way with the places occupied by each of its subprospects; and so on down to the ultimate atomic subprospects, if there are any.¹⁴

If “p” expresses a possible state of affairs or prospect, and if there are two different ways in which it may be realized relative to another proposition “q” – the two ways will be expressed respectively by “p-and-q” and “p-and-not-q” – then they are subprospects of the original prospect. Any prospect can be partitioned into a variety of sets of subprospects, but most theorists suppose that there is a single set of maximally determinate subprospects on the basis of which all others are constructed. These are the different ways that things might be at the finest level of discrimination at which the agent works; they are, from his point of view, the different possible worlds.

The principle of the co-determination of preferences easily goes over into the principle of completeness. Two attractive assumptions are sufficient to generate the shift: first, that there are indeed ultimate atomic subprospects; second, that the rational agent's preferences for non-ultimate prospects are determined by his preferences for subprospects, so that his preferences for the ultimate subprospects come out then as basic. The completeness principle holds that the rational agent assigns a place in his preference-ordering to every possible world and that the place of every other prospect – in effect, every set of possible worlds – is determined by those rankings combined with the agent's probability function. The place of “p” may be fixed by the places of “p-and-q” and “p-and-not-q” or by the places of “p-and-r” and “p-and-not-r,” and the places of each of those may be fixed in turn by further subprospects, but, however it goes, the place of everything will be fixed eventually by the agent's preferences over the ultimate subprospects he distinguishes, his subjectively different possible worlds.

The picture projected by the completeness principle, the picture usually associated with decision theory, is distinctively instrumentalist. The different possible worlds represent different possible outcomes or ends

of action. The different desires which the rational agent has *vis-à-vis* those ends are given and beyond debate. The only job in rational decision-making then is for the agent to form his preferences over the relevant options on the basis of how, given his probability function, they promise to do by his desires for those ends.

The abstraction thesis defended in the last section gives the lie to the completeness principle and to this instrumentalist picture which it projects. It means that the picture is inadequate in at least two respects. First of all, an agent need not have determinate preferences over all the relevant subprospects before he can rationally form a preference over certain prospects. Second, even if he does have such preferences, even indeed if he has determinate preferences over the ultimate possible worlds involved, those preferences cannot be seen as basic, as the unmoved movers of the system.

This second point follows from the claim that in forming preferences over prospects, even prospects as specific as possible worlds, we are moved by our preferences over the properties which we see those prospects as displaying. That claim, in effect the assumption of desiderative structure, means that the rational agent's ultimate points of reference, his ultimate motivational bearings, must be given by abstract properties rather than by concrete outcomes. In more everyday language they must be given by the agent's values rather than by his ends. Certainly he may take his guidance in decision-making from the ends to which different choices are likely to lead, but how desirable he finds those ends will depend on his values.

The first point mentioned is that not only are values more basic than ends in rational decision-making, they may serve to determine a decision without the agent's preference-ordering over relevant ends becoming determinate. Consider a situation where an agent has to form a preference between two options, A and B, where the different outcomes that are relevant to him – they may or may not be as specific as possible worlds – are A_1 and A_2 , B_1 and B_2 . Suppose now that both A_1 and A_2 are certain to realize a certain property which is of supreme importance to the agent – in this circumstance, it is an overriding value – while at least one of the B options is certain to fail in this regard. In such a situation the rational agent will form a preference for A over B and may do so without having a determinate preference-ordering between A_1 and A_2 or between them and the B outcome, if there is one, which promises to realize the value in question. The decision may be rationally determined by values in a way that abstracts from concrete ends.

We have been documenting the incompleteness of decision theory, in particular the incompleteness which follows from the abstraction thesis

of the last section. The incompleteness, in a slogan, is that decision theory looks only at ends, oblivious of the fact that values are more basic than ends and may even determine rational choice in abstraction from ends. But this incompleteness charge will not pass unchallenged. There is at least one line of objection that I can envisage.

According to a familiar style of behaviorism, preferring one state of affairs or prospect to another is simply being reliably disposed to choose it rather than the other, if you are given the choice; being indifferent between the two will then be failing to meet this condition in both directions. This approach means that you may prefer one prospect to another, or be indifferent between them, even if you have never considered them before. More generally, it means that preference and indifference between prospects are almost always given: they come on the cheap. The approach suggests two things: first that far from values being more basic than ends, an agent will have preferences over ends which outrun the control of values; second, that far from values determining choice in abstraction from ends, an agent will always have determinate preferences over the ends that are relevant to any choice.

But the assumption of desiderative structure dictates a clear response to this objection. If the rational agent forms preferences between prospects in the light of his preferences over the properties of those prospects, then preference comes to something significantly more than the disposition simply to choose. It may be that each of us has defined dispositions of this kind across the range of all prospects that we can distinguish – all our possible worlds – but the idea is that such dispositions should not be dignified with the title of “preferences.” For a disposition to choose to count as a preference, it must be a disposition to choose with reason – a disposition to choose on the basis of the properties displayed by the alternatives.

This line of response is not unreasonable. The behaviorist explication must equate preference with the disposition, already determined by how the agent is constituted, to select one of the relevant alternatives as soon as they are presented. It is that explication which suggests that I may already have a preference between two prospects, even though I have never considered their properties, never weighed up the values that they realize. But the equation of preferences with such brute dispositions is bound to seem inappropriate under the assumption of desiderative structure. And rightly so. After all, even if a person is disposed to choose one unconsidered prospect rather than another, he will equally be disposed, if possible, to consider the properties of the two before making his choice. It is not unreasonable for someone who believes in desiderative structure to refuse to equate preference with the brute disposition,

claiming that preference proper appears only after the consideration of properties or values.¹⁵

5.2 Decision theory is nonautonomous¹⁶

The second interesting result which follows from the abstraction thesis of section 4 is that decision theory is not as autonomous in relation to folk psychology as might otherwise be thought. The decision-theorist who wishes to apply his theory to determine what an agent will do, or ought rationally to do, must rely on folk psychological insight in identifying the required prediction or prescription. The best way to introduce this result will be by means of some examples.

Consider first an agent who is to be offered each of the following choices, perhaps on different days.¹⁷

- 1 Here is a (large) apple and an orange. Take your pick; I will have the other.
- 2 Here is an orange and a (small) apple. Take your pick; I will have the other.
- 3 Here is large apple and a small apple. Take your pick; I will have the other.

Imagine now that the agent has chosen the large apple over the orange and the orange over the small apple. If he is rational, what should we expect him to do in 3? By consistency or transitivity we ought, it seems, to expect him to take the large apple. But of course we know that being a well-bred young man he will not. So what do we say?

The obvious thing to say is that taking a large apple when the alternative left for another person is an orange is quite a different thing from taking a large apple when the alternative left is a small apple. Thus the three choices are not an instance of the intransitive sequence: A rather than B; B rather than C; C rather than A. The sequence, properly represented, is A¹ rather than B; B rather than C; C rather than A². A¹ is taking a large apple and leaving an orange for the other person; A² is taking a large apple and leaving a small apple for the other person.

Consider next an example which challenges, not transitivity of preference, but something called Independence or, in one version, the Sure-thing Principle. The Independence assumption is implicit in the principle of rational preference mentioned in section 1. It says that if you prefer an option involving a certain probability of A to an option involving the same probability of B, where otherwise they are the same, then for any two options differing in such a way only, you ought to prefer the one involving A – this, no matter what the probability shared by A and B,

and no matter what the alternative to A and B in each case. Thus you ought to prefer 3 to 4, if you prefer 1 to 2.

- 1 10 percent chance of A, 90 percent chance of C.
- 2 10 percent chance of B, 90 percent chance of C.
- 3 50 percent chance of A, 50 percent chance of D.
- 4 50 percent chance of B, 50 percent chance of D.

This principle is basic to traditional decision theory, since what it amounts to is the claim that an agent's preferences over the outcomes involved ought to impact on his choices homogeneously, without any difference being made by those differences in probabilities and alternatives that cancel out between options.

But this principle is subject to the same sorts of apparent counterexamples as transitivity. The most famous is the Allais Paradox but the one I prefer is this (derived from Diamond, 1967). I have to decide whether I am to give a candy to little Mary or little John, both of whom have equal claims. I prefer tossing a coin and giving the prize to Mary if the coin comes up heads and to John if it comes up tails, than giving it to John in either case. That is to say, I prefer 1 to 2.

- 1 Heads Mary wins; tails John wins.
- 2 Heads John wins; tails John wins.

But if my choice goes this way, then by Independence I ought to prefer 3 to 4.

- 3 Heads Mary wins; tails Mary wins.
- 4 Heads John wins; tails Mary wins.

But of course, being concerned with fairness, I shall prefer 4 to 3, not 3 to 4. So what are we to say?

Again, the obvious thing to say is that as the options differ in the case that challenged transitivity, so the outcomes differ here. The outcome under which Mary wins is different depending on whether or not the alternative possible outcome is that John wins. In the one case it is a fairly generated outcome; in the other it is not. Thus the example is not an instance of the scheme to which the Independence principle applies.

I hope that the two examples considered are sufficient to make clear that, if decision theory is to be applied to predictive or normative purpose, then it must be able to borrow from somewhere a principle for determining of two apparently equivalent options or outcomes, whether or not they really are relevantly similar. Of course decision theory might go to the extreme of saying that no two options or outcomes X and Y are ever suitably equivalent if they occur in the context of different alter-

natives. But that would be to make the theory useless in practice, since nothing would then follow from one choice as to what a rational agent would pick in a second.¹⁸

Clearly what is needed is a principle of equivalence for options and outcomes which is moderate in its effects; that is, which makes sufficient distinctions to handle the sort of counterexamples mentioned without making so many distinctions that no choice bears on any other.¹⁹ What type of principle would do the trick? John Broome points us in the right direction, I believe, with this principle for individuating possibilities (for example, options or outcomes): "worlds should be classified as different possibilities if and only if they differ in a way that can justify a preference" (Broome, 1990a, p. 11). Taking the large apple is a different option when the alternative is an orange from what it is when the alternative is a small apple, because it differs in a way that can justify apparently intransitive preferences. For similar reasons, Mary's winning is a different outcome when the alternative is John's winning from what it is when the alternative also gives victory to Mary.

If we take up Broome's suggestion here, and it has overwhelming attractions to my mind, then we must agree that it is a sufficient condition for one option's being non-equivalent to another option, one outcome to another outcome, that they differ in regard to properties that are desired or undesired by the agent. This proposition will enable us to say, at least in principle, when contexts (of alternatives) affect what would otherwise count as a single option or outcome X in such a way that equivalence fails across those contexts: in one context we have X¹, say, and in another X². It enables us to tell whether the relational property of having such and such alternatives – and, in the outcome case, at such and such probabilities – is sufficient to affect the identity of the option or outcome. The property will suffice to make a difference just in case it is or it involves a property desired or undesired by the agent. The relational property of leaving a small apple rather than an orange means that taking the large apple is impolite and since politeness is likely to be a property that matters to the agent, it makes a difference to the identity of the option. The relational property of there having been an equal chance of John's winning means that Mary's victory is fair and since fairness is likely to count with any agent, it makes a difference to the identity of the outcome.

The upshot is that while decision theory abstracts from the assumption of desiderative structure in regimenting folk psychology, the application of the theory requires us to practice folk psychology in a way that relies on that very assumption. Does our reliance on the assumption jeopardize the status of decision theory, making it radically indeterminate in many

cases whether an agent is faced with one or two options, one or two outcomes? I do not think so, since a property only has to matter in some measure, however miniscule, to make a difference. While there is probably great variety in how agents weight different properties, there seems to be less variation in which sorts of property they desire. In any case, what I wish to stress here is that, however this affects its status, decision theory does have to rely on the practice of a part of folk psychology from which it abstracts. Decision theory is a nonautonomous discipline.²⁰

5.3 Decision theory is nonpractical

This is enough on the incompleteness and nonautonomy of decision theory. I turn now to the final proposition that I derive from the abstraction thesis: the claim that decision theory is nonpractical. Decision theory claims to identify the pattern of choice which a rational agent will make: it spells out an ideal of rational choice. But there are two ways in which such an ideal may relate to practice:²¹ first as a *calculus* or procedure for getting the practice right; second as a *canon* or test for determining of any practice whether it is right.²² An ideal may be a calculus for successful practice without being a canon, as when it offers a generally reliable way of achieving a certain desired standard, where the achievement of that standard is the canon of success. And an ideal may be a canon of successful practice without being a practical calculus for achieving success. When a stockmarket adviser tells you to buy low and sell high, you are certainly offered a relevant canon of success, but equally certainly you are not offered an effective recipe or calculus.

Whether decision theory is practical or not depends on whether it offers just a canon of rational choice – I assume that it offers that – or something that serves also as a calculus. When we are told that the rational agent maximizes expected utility, are we told how we manage, or perhaps should manage, to be rational, namely by attending to the task of maximizing expected utility? Or are we just told that, however the rational agent does it, what makes him distinctively rational is the fact that he maximizes expected utility? Imagine that we are informed that what makes for excellence in long-distance running is the efficient use of oxygen, where there is nothing we can do to affect this feature in ourselves. The issue is whether decision theory gives us the same sort of nonpractical information: a fine analysis perhaps, but useless advice.

The habit in business schools and the like may be to treat decision theory as a calculus, but it is not an uncommon view among philosophers that at most the theory serves as a canon of rationality. One ground on

which that view has been defended is that human agents do not have the sort of access to their own utilities and probabilities that would be required for using decision theory as a calculus (see Harman, 1986, ch. 9). I believe that the abstraction thesis defended in this paper provides another reason for holding by this view, casting decision theory as a nonpractical normative device. I argue that the assumption of desiderative structure means that were an agent to try to use decision theory as a calculus, then he would be departing in a fundamental way from ordinary procedure; he would be changing the basis of his decision-making. I take it that this will be seen as a consideration against such a calculative employment of the theory, since no one admits the sort of transformative effect I allege.

If someone uses decision theory as a calculus, then he forms his preference over options on the basis of considerations such as “My subjective probability for ‘p’ is 3/4” and “My subjective utility for ‘q’ is 7.” This means, given the explication thesis which equates such probabilities and utilities with beliefs and desires, that he forms his preference on the basis of self-ascriptions of beliefs and desires. The considerations deployed can equally well be cast as follows: “I believe to degree 3/4 that p” and “I desire that q with an intensity of degree 7.” Why can they not be cast in a less subject-centered way as “It is 3/4 probable that p” and “It is desirable to degree 7 that q”? Because under the explication thesis subjective probability measures degree of belief, subjective utility degree of desire. The numbers given register the strength with which states like the belief that p and the desire that q are held, not aspects of the content of other states – states like the belief that it is probable that p or desirable that q.²³

If the agent who uses decision theory as a calculus forms preferences over options on the basis of self-ascriptions of belief and desire, then this means that the property in virtue of which he comes to prefer one option to another ultimately refers back to his own desire-satisfaction. He prefers the option for the property of promising the most satisfaction of his preferences over relevant prospects. Suppose that I use decision theory as a calculus to decide between A and B, where the relevant possible outcomes are A₁ and A₂, B₁ and B₂. The lesson is that I will then come to prefer one or the other option for its property of answering in the most satisfactory way to my preferences over A₁, A₂, B₁, and B₂.

These considerations counsel against the calculative use of decision theory, for when in ordinary practice an agent comes to prefer one option to alternatives, he often prefers it for a property other than that of answering in a certain way to his prospect-desires; say, for the property of being an obligation of etiquette, being amusing, or being in the public

interest. The agent's actual practice of decision-making is radically different from what it would be if he were to use decision theory as a calculus. Were he to apply decision theory in this way, then he would submit himself to the control of different property-desires from those which normally operate. He would be concerned in every decision with realizing the property of best satisfying his prospect-preferences, rather than with realizing independent properties like those of being mannerly, having fun, or advancing the common good.

But it may seem that I am overlooking an obvious objection. Suppose that I have a certain preference-ordering over the prospects A_1 , A_2 , B_1 , and B_2 , an ordering driven by my preferences over properties of those prospects: say, F_1 , F_2 , G_1 , and G_2 respectively. If I make my decision between A and B in the ordinary way I will make it in the light of how the options do by F_1 - G_2 properties; if I make it by using decision theory as a calculus, I shall make it in the light of how they answer to my A_1 - B_2 preferences. But since the F_1 - G_2 properties determine my A_1 - B_2 preferences, this means that either way I shall make the same choice. And does that not entail that there is no significant difference between the two practices?

No, it does not. When I make a decision, forming a preference over certain options, desiderative structure means that two things become fixed: first the material object of my desire and second its formal object. Even if the two practices in our example generate a desire or preference for the same material object, the same option, the desire has a different formal object in each case; it brings a different value into play. In the one case the option is preferred for answering in a certain way to the agent's prospect-preferences over A_1 , A_2 , B_1 , and B_2 . In the other it is preferred for answering in a certain way to the properties F_1 , F_2 , G_1 , and G_2 .

The variation in formal object makes for a difference in the scope across times and modalities of the preference formed.²⁴ If I prefer an option for property K, then I prefer it here and now for any time or situation at which it retains that property; if I prefer it for property L, then I prefer it here and now for the different range of times and situations at which it retains that other property. If I prefer the option in our example for answering to my preferences over A_1 - B_2 , then assuming that I am concerned with my preferences at the time of action rather than now at the time of decision, I prefer it now only for times and situations at which I retain those preferences. If I prefer the option for how it answers to the F_1 - G_2 properties, I prefer it now for times and situations where it continues to answer in that way. The scope is different in each case and that means that there is a sense of preference-satisfaction

such that the preferences will have different satisfaction conditions. It will be possible to satisfy in that sense the preference with the one sort of formal object without satisfying the preference with the other.

Suppose, however, that in making my decision with regard to how the option answers to my preferences over A_1 - B_2 , I have my current actual prospect-preferences in view. Does that not mean, since these are driven by properties F_1 - G_2 that the option-preference formed will have the same scope across times and modalities, regardless of difference in formal object? Yes, but there remains a difference in its scope across persons; I assume that the preference extends to different persons, given that it is driven by properties which abstract from the identity of the agent. If we ask in the two cases what course is preferred for the arbitrary agent facing exactly the same decision, then we are given different answers. In the one case the preference is that any agent with such and such prospect-preferences over A_1 , A_2 , B_1 , and B_2 should choose the favored option. In the other case the preference is that any agent facing options that answer in the required way to properties F_1 , F_2 , G_1 , and G_2 should do so. The difference in scope can scarcely be described as making for a difference in the satisfaction conditions of the preferences, since the notion of satisfaction suggests that one and the same agent is involved. What it makes for, more colloquially, is a difference in the universalization conditions of the two preferences (see Pettit 1987b, 1988a).

We have seen that the shift to using decision theory as a calculus, even if it has no effect on choice of option, will affect the values that drive an agent's decisions, the properties for which he prefers certain options over others. We have also seen that this change of value focus will mean a change in the satisfaction conditions, or at least the universalization conditions, of the agent's option-preferences. All of this said, however, an advocate of using decision theory as a calculus may still protest that it is unclear why we should worry about such a shift of value focus. The final question before us then is whether such a change is likely to make any practical difference to an agent.

The question, more generally, is whether it will make any practical difference for an agent to switch from one value set to another, if the sets yield the same output, supporting the same decisions. The answer is that the switch will make a difference if the value sets differ, as they are likely to do, on the input rather than on the output side: if they differ in the sorts of considerations that tend to support or subvert them. The answer in the general case dictates a negative response to the specific issue about using decision theory as a calculus, for the values that go with such an employment of the theory have a different profile on the input side from the values that more ordinarily move us.

One input to which it is generally appropriate to submit a set of values is the universalizability test. The challenge raised by this test is whether the agent can happily endorse, not just the decision dictated in his own case by the values he adopts, but also the decision dictated by those values for any other agent, including any agent whose action would affect him negatively. I submit that the values endorsed under a calculative use of decision theory are vulnerable to this test in a way in which the values endorsed in ordinary modes of decision-making are not.

Suppose that you are in the position of Socrates and that you are about to drink the hemlock, much to the distress of your friends, when you are challenged to put the values motivating that decision to the universalizability test. Imagine that you reached your decision out of concern that the option taken should be the honorable choice. In that case the universalizability test will hardly have any effect. You may acknowledge that you would probably feel differently about the option chosen were you in the position of your friends, but that need not cause you to weaken in your resolve; you will probably reckon that you would feel differently only because of being emotionally blinded, as they are surely blinded, to the importance of being honorable.

Imagine now that you reached your decision, using decision theory as a calculus, on the ground that the option taken promised the greatest satisfaction of your prospect-desires. It promised that satisfaction, of course, because it seemed to be the honorable choice, but the important thing in your calculations was not the source of the promised satisfaction, only the satisfaction itself. Consider then how this value orientation will fare under the universalizability test. Acknowledging that you would feel differently about the option chosen if you were in the position of any of your friends, acknowledging in other words that in their position you would not see the option as promising the greatest satisfaction of your prospect-desires, will you be as unshaken as before in your resolve to take the hemlock? Arguably not. You could think before that the feelings that you would have in the position of your friends should be discounted, being the product of blind emotion. Now you cannot take this view and you have to face the question: "Why should the satisfaction of the prospect-desires I have in the role of agent matter so much more than the satisfaction of the prospect-desires I would have in the position of my friends?"²⁵

This line of thought shows that the values adopted in the switch to using decision theory as a calculus are more vulnerable to at least one sort of challenge than the values invoked in more regular decision-making. Even if the switch does not immediately mean that different options will be chosen, then, it is likely to be of practical significance. It is likely over

the longer haul to generate a shift in the choices that the agent will make. In jargon that will ring a chord in decision-theorists, the switch may not directly alter the utilities that an agent attaches to various prospects but it will probably alter them indirectly, for it subjects the agent to a different utility kinematics.

We must conclude then that not only is decision theory an incomplete account of decision-making and a nonautonomous theory, it is also an impractical instrument for making decisions. This does not mean, of course, that decision theory has no value. On the contrary, it has the great merit of clearly explicating certain important features of our folk psychology. The lesson is only that the value of decision theory has to be carefully judged. While it explicates one part of folk psychology, it abstracts from another, and the abstraction places limitations on its utility as a descriptive and normative device.

APPENDIX

In "Truth and probability" Frank Ramsey begins with the assumption that, given his preferences, we can identify for each person an ethically neutral proposition with a probability of a half; a plausible example might be the proposition that a coin which is tossed in the air will come up heads.²⁶ Let this proposition be represented by "h." The proposition will be ethically neutral for the agent if and only if he is indifferent between two situations that differ only in whether it is true or false. It will have a probability of 1/2 for the agent if, for two possibilities A and B which leave the realization of "h" open and which are such that he prefers A to B, he is indifferent between the following gambles: A if h, B if not, and B if h, A if not. For short, he is indifferent between (AhB) and (BhA).

With this assumption in place, Ramsey suggests that we can operationalize a decision-theoretic view of the agent in roughly the following steps.

- 1 Among suitable items, find those items most preferred and least preferred by the agent, say A and Z, and assign these to the endpoints of an arbitrary scale, say between 0 and 1.
- 2 By a version of the Bayesian theory – the principle of rational preference – the gamble (AhZ) ought to appear at the midpoint of the scale, with a preference ranking or utility of 1/2. Find an item M such that the agent is indifferent between the gamble (AhZ) and M and assign this to the midpoint. (Alternatively, though this breaks with Ramsey, call the gamble "M.")
- 3 By the theory, the gambles (AhM) and (MhZ) ought to appear respectively on the quarter points of the scale, with utilities of 3/4 and 1/4. Find items F and R such that the agent is indifferent between (AhM) and F, (MhZ) and R, and assign these to the quarter points. (Alternatively, call these gambles "F" and "R" respectively.)

- 4 Continue filling in the scale by successive applications of this procedure.
- 5 If an item AA appears which the agent prefers to A, then find an item C on the scale such that he is indifferent between the gamble (AAhC) and A and assign a utility figure to AA on the basis of the theory. Use a similar procedure to determine the utility of any item that appears to which Z is preferred.
- 6 Having established the utilities of relevant items, determine the probabilities of propositions other than h by the following procedure. For any proposition p, find items B, K, and V such that the agent prefers B to K to V but is indifferent between K and (BpV). Given the utilities for B, K, and V, the theory dictates the appropriate probability for p.
- 7 In order to establish the subjective expected utility of any option, find a gamble with which it can be equated and determine the expected utility that belongs to it.

The Ramsey procedure just described is useful, not just in showing how Bayesian decision theory might be made operational, but also in revealing that under this version of that theory an agent's subjective probabilities and utilities are all fixed by a relatively austere base – an appropriate preference ordering over appropriate items. That preference ordering enables us to identify an ethically neutral event with a probability of 1/2 and, as the seven steps show, it serves in principle to determine all the agent's subjective utilities and probabilities. Instead of our three relevance principles we might have postulated just that the agent has such an ordering. Equip an agent with that ordering and you will have done all that is required to equip him with a full set of subjective probabilities and utilities. Indeed, more than that, you will also have done all that is required to equip him with rational preferences.

But though this base is relatively austere, there is one way in which it is not so austere as it may seem. It assumes that there is no problem in determining the content of an agent's preference – in determining for example that in choosing this object over that he is expressing a preference for “a chair given he has a table” over “a table given he has a table” rather than for “a chair over a table,” “something to sit on over something to write on,” “something made of plastic over something made of wood,” and so on. This feature of the procedure is worth noting, since it explains why decision theory does not get into trouble with the problem of determining the contents of subjective probabilities and utilities.²⁷ Decision theory assumes that that problem is solved. The point connects with the discussion of nonautonomy in section 4.

NOTES

I was helped by comments received when versions of this paper were read at seminars in the Australian National University and Oxford University. I was particularly helped by conversations with and comments from Paul Anand, Michael Bacharach, John Broome, Peter Gärdenfors, Susan Hurley, Frank

Jackson, David Lewis, Peter Menzies, Huw Price, and Michael Smith. I am also grateful for useful comments received from Edward Craig, Lloyd Humberstone, and Fred Schick.

- 1 The continuity assumption requires, roughly, that if P is preferred to Q, and there is a continuum of increasingly less preferred points as we move from P to Q, then for any S such that the agent prefers P to S and S to Q, there must be some point on the P–Q continuum such that he is indifferent between S and that point; there must not be a discontinuous leap from points preferred to S to points to which S is preferred. Continuity is broken by someone with strictly lexical preferences over the components of the packages on the P–Q continuum, someone for example who prefers any increase in X – any increase in the amount or chance of X – to any increase in Y, if each package has X and Y as components. If a person takes rights seriously then he may seem to have lexical preferences, for he will not sacrifice any increase, for example, in his own innocence or cleanness of hands (X), for an increase in certain other goods (Y) (see Pettit, 1987a, pp. 8–14). But rights may not introduce the offending sort of discontinuity, for even the most rights-respecting of people are willing to run a certain risk of imposing the ills against which rights are a protection for the attainment of goods that rights trump. I risk the taking of innocent life by driving to work and I do so, presumably, only for the convenience thereby attained.
- 2 For reviews, the first informal and the second technical, see Eells (1982, ch. 3) and Fishburn (1981, pp. 139–99).
- 3 This assumption is well characterized by Davidson (1980).
- 4 For a defence see Jackson and Pettit (forthcoming).
- 5 Notice, however, that to take them as more than fictions, in particular to identify utilities with desires – where desires are taken as more than fictions – need not be to equate utility with something felt, like pleasure. Desire satisfaction need not be something felt.
- 6 I ignore the complications required in virtue of the considerations raised in Lewis (1983a), Essay 10.
- 7 There is some further discussion of how to understand prospect-preference in section 4.
- 8 On this ambiguity see Jackson (1985b).
- 9 See the relevant essays in Raz (1978) and see Milligan (1980, ch. 3). The Gorman–Lancaster translation of commodities into characteristics offers a parallel in economics. See Sen (1982a, p. 30). The development of multi-attribute utility theory stems also from a recognition of the role of property-desires. See Keeney and Raiffa (1976) and Farquhar (1980, pp. 381–94).
- 10 Regular decision theory. I except the multi-attribute utility theory mentioned in note 9.
- 11 I owe the point to Frank Jackson. It is supported by the examples considered in the next section in the discussion of the nonautonomy claim.
- 12 For related points see Jackson (1985a). On desire *pro tanto*, and a difference between it and another sense of desire *prima facie*, see Hurley (1985–6). See

- also her recent book (Hurley, 1989), which appeared after this paper was in near final draft.
- 13 Notice though that many decision theorists explicitly deny completeness (see, for example, Tversky, 1975, pp. 163–73). Notice too that multi-attribute utility theory, which was mentioned in note 9, stems from a recognition of the incompleteness of regular decision theory.
 - 14 Some axiomizations of decision theory, most notably the Jeffrey–Bolker axiomatization, involve atomless algebras (see Jeffrey, 1983). While the principle of the co-determination of preferences is supported by decision theory generally, it should be noticed that it need not be fatal for decision theory if the rational agent's preference-ordering does not extend to all prospects (see Broome, 1990b).
 - 15 Should he be reluctant not just to equate prospect-preference with brute disposition but also property-preference? Perhaps, for it is not clear that property-preference has to be brute in the same sense. After all, only some properties are taken by most of us to be desirable or undesirable and, aspiring as we do to attain agreement about which are which, most of us seem to think that if we only understand what is involved we shall agree about whether a given property makes for good or ill; of course, this is compatible with our weighting the properties in different ways.
 - 16 I have been aided greatly in developing my thoughts here by conversations with Peter Gärdenfors and by reading Broome (1990a, b) and Hurley (1989).
 - 17 The example is not mine but I do not know its provenance.
 - 18 See Broome (1990a, b, ch. 4). For a different approach to the problem raised see Schick (1987).
 - 19 Alternatively, as Broome notes, we might individuate options and outcomes to the finest level possible and then look for a principle to tell us which differences ought not rationally to matter.
 - 20 A similar result follows on a different treatment of the counterexamples. We might say that the lesson of the Diamond example is that when inter-outcome properties, such as that related to fairness, serve to undermine Independence, then we reach a limit beyond which decision theory does not apply. But if we say this, then decision theory is nonautonomous in the sense of applying within a boundary which it itself does not have the resources to discern.
 - 21 This parallels a more familiar distinction in ethics (see Pettit and Brennan, 1986; Pettit, 1988b).
 - 22 Of course there is a further looser sense in which an ideal like that offered by decision theory may relate to practice: by stimulating agents to be more systematic and rigorous in their thinking. I do not doubt but that decision theory often usefully plays this role.
 - 23 Of course, this line is consistent with the claim that ordinary ascriptions of probability express degrees of belief, and that ordinary ascriptions of desirability degrees of desire.
 - 24 Here I am indebted to collaborative work with Michael Smith; see Pettit and Smith (forthcoming).
See Pettit (1987b, 1988a).

- 25 The general lesson in the offing is that universalizing decisions based on reasons of desire-satisfaction tends to generate utilitarianism (see Pettit, 1987b).
- 26 Reprinted in Ramsey (1978).
- 27 On that problem see, for example, Dretske (1986).