Economics and Philosophy, **22** (2006) 365–391
doi:10.1017/S0266267106001027

HUME'S NON-INSTRUMENTAL AND NON-PROPOSITIONAL DECISION THEORY

Robert Sugden
University of East Anglia

Hume is often read as proposing an instrumental theory of decision, in which an agent's choices are rational if they maximally satisfy her desires, given her beliefs. In fact, Hume denies that rationality can be attributed to actions. I argue that this is not a gap needing to be filled. Hume's theory provides a coherent and self-contained understanding of action, compatible with current developments in experimental psychology and behavioural economics. On Hume's account, desires are primitive psychological motivations which do not have propositional content, and so are not subject to the criteria of rational consistency which apply to propositions.

In the Appendix of the *Treatise of Human Nature*, David Hume 1978 [1739–40] describes an experiment which investigates the nature of perception:

> Suppose I see the legs and thighs of a person in motion, while some interpos'd object conceals the rest of his body. Here 'tis certain, the imagination spreads out the whole figure. I give him a head and shoulders, and breast and neck. These members I conceive and believe him to be possess'd of. Nothing can be more evident, than that this whole operation is perform'd by the thought or imagination alone. (1978: 626)

This paper was written as part of a research project on the methodology of experimental economics, supported by the Leverhulme Trust. It is a sign of how long I have been thinking about this topic that my greatest debts are to my late friends Jean Hampton and Martin Hollis. For more recent assistance, I am grateful to Shepley Orr, an editor and two anonymous referees.

As a starting point for my paper, this example serves two different purposes. The more direct purpose is to illustrate Hume's sophistication as a psychologist; the experiment he describes, and the conclusions he draws from it, anticipate the findings of Gestalt psychologists more than a century and a half later. However, I also wish to use it as a metaphor. According to Hume, when our sensory data are incomplete, the imagination tends to fill in the gaps by using 'customary connexions': we unconsciously assume that we are seeing the things we are accustomed to seeing. I shall suggest that something similar has happened in scholarly interpretations of Hume's theory of decision.

The passages in the *Treatise* which deal with decision-making include some of the most famous sentences in philosophy, but they are surprisingly brief and are capable of alternative readings. My suggestion is that modern readers of Hume have filled in the apparent gaps in his presentation by assuming that he intends something similar to currently received theories of rational choice. I offer an alternative reading which I believe is more faithful to Hume's intentions. At the very least, my reading reflects a different set of customary connections. These are connections that come to mind for an author who has contributed to what is now called *behavioural economics* – the empirical, and often experimental, investigation of how the behaviour of economic agents is affected by psychological mechanisms. In this sense, I shall argue, Hume's decision theory is behavioural rather than rational.

Of those commentators who have recognised the absence of rationality in Hume's decision theory, most have seen this absence as a fundamental flaw or gap, which modern philosophy needs to correct or fill. In contrast, I shall argue that Hume's theory is coherent and self-contained. It gives us a philosophical understanding of action that is compatible with current developments in behavioural economics.

1. HUME AND HUMEANISM

When discussing theories of practical reason, philosophers often use the term 'Humean' as a synonym for 'instrumental': a Humean theory is one in which an action is rational if and only if it maximally satisfies the actor's preferences, given her beliefs, and those preferences and beliefs satisfy certain conditions of internal coherence. For example, Robert Nozick (1993: 138–40) uses Hume's famous remark about its not being contrary to reason to prefer the destruction of the world to the scratching of one's finger as justification for calling instrumental rationality 'Humean'; he then introduces some standard conditions on the internal consistency of preferences as '[o]ne tiny step beyond Hume, not something he need resist'. David Gauthier (1986: 21, 25) quotes the same passage from Hume in support of his own formulation of 'parametric' (that is, non-strategic)

instrumental rationality. Edward McClennen (1990: 4) proposes the following 'pragmatic test' for determining whether an action is rational or irrational: an action is irrational if, through choosing it, 'the agent will fail to achieve his intended objective or will fail to maximise with regard to his own preferences with respect to outcomes'. McClennen claims Hume as an 'early and unusually clear' exponent of this approach. James Dreier (1996: 249) defines 'Humean' practical rationality as having two components, instrumentality ('we may properly reason about means, but not about ends') and the consistency of preferences ('practical rationality is a matter of coherence'). And so on.

Recently, however, a number of writers have questioned whether Hume really did propose a theory of instrumental rationality (for example, Sugden 1991; Millgram 1995; Korsgaard 1997; Blackburn 1998: 238–43; Hampton 1998: 136–40, 142–51). These writers have pointed out that, in the passages of the *Treatise* that have traditionally been read as endorsements of instrumental rationality, Hume inserts crucial qualifications. In particular, consider the paragraph in Book II that includes the example of scratching one's finger. Early in the paragraph, Hume claims:

> 'tis only in two senses, that any affection can be call'd unreasonable. First, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the supposition of the existence of objects, which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects. Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. (416)

This passage certainly seems to imply that an action can be criticised as irrational if it is not directed towards achieving the agent's ends, given his beliefs. But, when summarising this paragraph, Hume adds a qualification:

> In short, a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment. (416)

When Hume returns to this topic in Book III, the point of the qualification becomes clear. He gives the following explanation of why, according to his theory, neither passions nor actions can be called reasonable or unreasonable:

> Reason is the discovery of truth or falshood. Truth or falshood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now 'tis evident our passions, volitions, and actions, are not susceptible of any such agreement or disagreement;

> being original facts and realities, compleat in themselves, and implying no reference to other passions, volitions, and actions. 'Tis impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason. (458)

Reason, he goes on to say, can influence our conduct only in two ways. It can excite a passion 'by informing us of the existence of something which is a proper object of it'; and it can discover relationships of cause and effect which 'afford us the means of exercising [a] passion'. In both cases, the role of reason is to provide judgements of fact. These judgements may be reasonable or unreasonable, but the passions and volitions that arise in response to them are not further judgements; they are psychological facts in their own right. Even if the judgement which prompts an action is unreasonable, it is only 'in a figurative and improper way of speaking' that the action itself can be called unreasonable (459).

On the most natural reading of these passages, Hume is not endorsing an instrumental form of practical reason: he is denying the existence of practical reason altogether. It seems that Hume is not, as he is often said to be, a progenitor of modern rational choice theory. To the contrary, his theoretical framework is incompatible with the most fundamental presupposition of that theory, namely, that there is such a thing as rational action.

Although there is a growing recognition that Hume was not a Humean in the modern philosophical sense of the word, most of the commentators who have favoured this reading of the *Treatise* have presented Hume's position as an extreme form of scepticism which a modern reader should reject. This is true of three of the writers I have cited as denying that Hume was a Humean (the exceptions are Simon Blackburn and myself). Elijah Millgram suggests that we should ignore Hume's 'naïve or antiquated empirical psychology' as 'counterintuitive and apparently unmotivated' and instead understand his scepticism as a corollary of his theory of semantics. That theory is 'so alien, and so thoroughly discredited' that modern readers have failed to notice it. Having noticed it, we must conclude that Hume's arguments about practical reason cannot be adapted to the uses of contemporary philosophy (1995: 81, 86–7). Christine Korsgaard interprets Hume as denying 'the instrumental principle' and, more generally, as denying that there is any such thing as practical reason; but her own concerns are to show 'what is both still necessary and possible in the theory of practical reason', and to answer the question: '[W]hat gives the instrumental principle its normativity?' (1997: 222, 253–4). On her account, Hume rejects what that question presupposes. Jean Hampton's discussion of instrumental rationality is part of an attempt to defend moral objectivity and to establish the 'authority of reason'. She is trying to show that it is self-defeating to reject moral objectivity in favour of scientific

naturalism, since the natural and social sciences presuppose the objectivity of certain norms. She treats Hume's scepticism as the *reductio ad absurdum* of the strategy of eliminating normativity from decision theory. Apparently taking it as self-evident that *some* forms of behaviour can authoritatively be criticised as irrational, she concludes: 'The problem with [the Humean] theory of reason is that it can never convict someone of acting irrationally!' (1998: 145).

John Broome (1999) takes a somewhat similar line to Hampton's. He sets out a position which he calls 'extreme Humeanism', and which corresponds with the reading of Hume as denying the existence of practical reason. An extreme Humean 'believes that no preference can be irrational [and] leaves it at that', without requiring that preferences are internally consistent, or that preferences over actions are determined by preferences over the outcomes that those actions will produce. Broome concludes that this position is 'unappealing': it 'implies that reason cannot guide people even through the most ordinary business of living' (69).

My aim in this paper is to retrieve a defensible theory from Hume's account of decision-making – a theory that is compatible with his rejection of practical reason. Let me say straight away that this theory of decision-making will not provide what Korsgaard, Hampton and Broome want. It will not assert – still less explain – the 'normativity' of the mental processes it attributes to decision-makers. It will not rest on a concept of rationality that can be used to convict people of acting irrationally, or to provide normative guidance in the business of living. Nevertheless, it will be useful for the purpose that Hume intends it to serve: explaining the mental processes that lie behind actual human decisions. I will also show that Hume offers a theoretical explanation of how, as a matter of empirical fact, people come to make the kinds of normative judgements about decision-making that Korsgaard, Hampton and Broome see as being justified by theories of practical reason. However, Hume's explanation of these judgements is, in important respects, independent of his decision theory.

My starting point is a hunch about why present-day readers find it so hard to take Hume's position seriously. The problem, I think, is that such readers – or at least, such readers as are inclined to be sympathetic with Hume – presuppose the validity of some variant or other of the modern theory of rational choice. They may entertain the possibility that there is *more* to rationality than the internal consistency of preferences, but they find it hard to make sense of the idea that there could be *less*.

In modern decision theory, the concept of preference is primitive. In most versions, the existence of preferences is axiomatic: it is simply assumed that each individual agent can rank any two options in order of preference or indifference. Rationality is then construed as requiring that choices are consistent with preferences and that preferences themselves

are internally consistent. 'Internal consistency' for preferences is defined by specifying formal properties, such as reflexivity and transitivity, that the relation 'is at least as preferred as' must satisfy. These properties are justified as principles of logic, on the analogue of the principles of standard propositional logic, applicable to particular classes of propositions.[1] In the standard subjectivist interpretation of decision theory, propositions about preference are non-cognitive – that is, they do not make claims about objective reality, but only about subjective perceptions or attitudes. Nevertheless, they *are* propositions: that is what makes them subject to tests of logical consistency. I shall say that a theory of mental states is *propositional* if it treats mental states as representing, expressing or implying propositions, and if the logical or conceptual consistency of such propositions can be used to test the rationality of the corresponding mental states. Viewed from within the conceptual framework of modern decision theory, Hume's rejection of practical reason seems to amount to denying that it is irrational for a person to maintain or to act on inconsistent propositions. It is then easy to think that Hume's position is a form of scepticism which, however challenging from a philosophical point of view, is ultimately absurd.

As an illustration of this way of thinking, consider the following passage, in which Dreier defends his preferred form of 'Humeanism'. Dreier is challenging Hume's claim that desires are not subject to rational assessment. After allowing that a person may have inconsistent first-order desires, he says:

> What's important, rather, is that these conflicting desires be resolvable into a decision. I want a deep tan, but on the other hand I don't relish the prospect of skin cancer. All things considered, I prefer to stay out of the sun. There would be real trouble if all things considered I preferred staying out of the sun to basking in it, and basking in it to a short exposure, and a short exposure to staying out of the sun. If Humeanism is committed to saying that the combination of those three preferences is perfectly rational, then Humeanism is certainly not worth defending. (1996: 250)

Notice how Dreier has translated what for Hume are *passions* or *volitions* into propositions about preference. In the domain of propositions, it

[1] This interpretation of the axioms of decision theory is explicit in the work of two of the founding fathers of that theory, Frank Ramsey and Leonard Savage. Each presents a version of expected utility theory which provides an integrated analysis of rational choice and rational belief. Ramsey (1931: 166) presents his work as an enquiry into 'the logic of partial belief'. Savage (1954: 6, 20) sets himself the task of investigating 'whether logic cannot be extended, by principles as acceptable as those of logic itself, to bear more fully on uncertainty'; he claims that his theory of decision-making has a normative status analogous with that of logic. More recently, Broome (1991: 11) argues that reflexivity and transitivity, applied to any relation of the form 'is at least as – as', are 'truths of logic'.

does seem absurd to treat inconsistency as reasonable.[2] But is Dreier's translation legitimate? If we are to understand the account of human decision-making in the *Treatise*, we need to take it on its own terms. That involves clearing our minds of the concepts and assumptions of modern decision theory. I want to suggest that Hume's decision theory is not propositional.

2. HUME AS AN EXPERIMENTAL PSYCHOLOGIST

If we are to reconstruct a theory of decision from the *Treatise*, we need to read the few passages which deal explicitly with this topic in the context of Hume's underlying conceptual framework and methodology. In this section, I explore some of the main ways in which Hume's approach differs both from modern decision theory and from modern analytical philosophy.

The first point to notice is that Hume presents his theory of mind *as experimental psychology*. This is made clear in the full title of his book: *A Treatise of Human Nature: being an attempt to introduce the experimental method of reasoning into moral subjects*. In the preface, Hume locates his work as a contribution to the 'science of man'. It is a study of 'the extent and force of human understanding ... the nature of the ideas we employ, and of the operations we perform in our reasonings'. The methodology is to be modelled on that of the natural sciences: since 'the only solid foundation we can give to this science itself must be laid on experience and observation', Hume undertakes to investigate the powers and qualities of the mind by means of 'careful and exact experiments, and the observation of those particular effects, which result from its different circumstances and situations'. He argues that the science of man cannot use the kinds of controlled experiments that the natural sciences use, because the act of placing a person in an experimental environment would induce forms of 'reflection and premeditation' that would 'disturb the operation of [the experimental subject's] natural principles'. Thus: 'We must glean up our experiments in this science from a cautious observation of human life, and take them as they appear in the common course of the world, by men's behaviour in company, in affairs, and in their pleasures' (xv–xix).

Hume delivers on these promises. He is not just – as, say, Adam Smith is – a shrewd and psychologically acute observer of human life; he thinks in genuinely experimental terms. The *Treatise* is full of reports of experiments that investigate the workings of the human mind. In many of these cases, Hume describes an experiment which the reader can carry out for herself, using herself as the subject and recording her own psychological responses to some external stimulus. Hume reports what he has found from his

[2] However, it is not self-evident that transitivity is an appropriate consistency property for propositions about preference. On this, see Sugden (1991).

own observations, and invites the reader to replicate the experiment. The experiment described in the passage I quoted in the opening paragraph of this paper is a typical example. Hume ends his account of this experiment by saying: 'Let any one examine his own mind, and he will evidently find this to be the truth' (626). Similar forms of words recur throughout the *Treatise*. In other cases, Hume refers to what he takes to be common observations of human life, but whose significance as evidence for the science of man have been overlooked. A typical example (the significance of which I will discuss later) is the claim that 'men often fall into a violent anger for injuries, which they themselves must own to be entirely involuntary and accidental' (350).

Of course, the *Treatise* is philosophy as well as empirical psychology. But, according to Hume, the contribution to philosophy is mediated by the psychology. It is by understanding how the human mind actually works that Hume hopes to provide a solid foundation for the other sciences, among which he includes mathematics, natural philosophy, natural religion, logic and morals (xv–xvi). It is a misrepresentation of the *Treatise* to say, as Millgram does, that its psychology is 'unmotivated' or to suggest that it is some sort of optional extra. Hume's psychology is grounded in observation; his theory is an attempt to organise those observations. Millgram may be right in saying that this psychology is counterintuitive. Empirical science often is: think of Gallileo's observation that all falling objects accelerate at the same rate, irrespective of their mass. But if it is naïve, it is naïve only in the same sense that Gallileo's physics is: we now know more than he did.

Since Hume's methodology is that of natural science, it is hardly surprising that he does not find any source of authority for human reason. All he can hope to do is to discover how, in fact, the mind conducts the operations that we call 'thinking' or 'reasoning'. He proposes a theory which purports to explain a wide range of such operations. He chooses to dignify two of these with the name of 'reason'.[3] One is the operation of 'demonstration', which is concerned with the discovery of what could or could not possibly exist, with what is and is not conceivable. The other is inductive inference – the discovery of relations of cause and effect. But in each case, the operation itself is described in entirely naturalistic terms. According to Hume, we discover that something is conceivable simply by conceiving it as a mental picture; our minds cannot form any representation

[3] Of course, it suits Hume's purposes to give this special status to the two mental operations that the reader needs to use in order to accept the conclusions of the *Treatise*. I do not want to get diverted into in the familiar problem of whether every empirical theory of human reasoning necessarily involves a self-defeating claim to the authority of the reasoning used by its author. My inclination is to think that Hume can acquit himself of this charge, but I shall not argue this here.

of the inconceivable (31–3). Our perception of cause and effect is simply a perception of the constant conjunction of one thing and another; which is called the 'cause' and which the 'effect' is determined by priority in time (73–84). These forms of reasoning have no authority beyond the fact that they are built in to human psychology.

A second important feature of Hume's theory of mind is that it is *dynamic*. It is a theory not just about the content of mental items, but also about how they come into and go out of existence, the temporary existence of one mental state causing the temporary existence of another. In other words, Hume is concerned not only with what it is for us to think something, but also with *when* we think it. In rational choice theory and in analytical philosophy, the content of the mind is typically conceived as a stock of ideas and perceptions that are in some sort of equilibrium with one another. In particular, rational choice theory models an agent's preferences as a stock of propositions about the relative desirability of all potential objects of choice; facing any particular choice problem, the agent consults this body of preferences and reads off those that are relevant to the case in hand. We can then test the rationality of the agent by investigating whether the items in the stock are mutually consistent. In Hume's theory, in contrast, a person's thoughts are in a constant state of flux. Mental operations are understood, not as consultations of a pre-existing stock of mental items, but as *transitions* between one mental state and another.

The central concept in Hume's theory is that of *associations* between mental states. In classifying mental states, Hume distinguishes between *original impressions* (or *impressions of sensation*), *secondary impressions* (or *impressions of reflection*) and *ideas*. Original impressions are the sensations we experience directly from contact with the outside world. Ideas are what the mind constructs for itself when it represents or re-assorts the impressions it has received. Secondary impressions are sensations that arise as a result of the workings of the mind (7–8, 275–7). For example, suppose I am walking in the mountains and see a large boulder rolling down a slope towards me. The sight of the boulder is an original impression. The thought that the boulder might hit me is an idea. The sense of fear induced by that idea is a secondary impression. Hume's theory identifies causal relationships between impressions and impressions, between ideas and ideas, and between impressions and ideas. The main hypothesis is that, at any given moment, the presence of any one idea or impression tends to bring into existence, and to add intensity to, other related ideas and impressions. In the realm of ideas, 'relatedness' can be a matter of similarity, contiguity, or cause and effect. In the realm of impressions, it is a matter of similarity only (282–4).

Hume repeatedly stresses the role of what he calls 'double relations' of ideas and impressions in inducing emotional states. His first example concerns the origin of pride. Hume treats pride as a pleasurable sentiment

of approval that a person feels in relation to himself. He asks us to consider a man who is conscious that his parish is beautiful. This man's mind contains the ideas of the parish and of its beauty. The idea of its beauty is a reflection of pleasurable original impressions. Since the idea of the parish is associated by contiguity with the idea of himself, thinking about the parish tends to evoke thoughts about himself. Since pride, like the enjoyment of beauty, is pleasurable, these two impressions are associated by similarity. Thus, according to Hume, the man's consciousness of the beauty of the parish tends to induce pride in himself: the double relation between ideas and impressions 'produces an easy transition from the one emotion to the other' (277–89).

Although some features of Hume's theory of mental states may seem over-simple to the modern reader, the central hypothesis that emotional states can be transmitted by associations of ideas is undoubtedly sound; that the workings of the mind are influenced by such associations is among the fundamental principles of modern psychology. Now that science is beginning to understand the physical workings of the brain, it is becoming clear that networks of associations between mental states have physical correlates in neural architecture. It is remarkable that, through careful introspection and observation, Hume seems to have arrived at some perception, however provisional and hesitant, of how the brain actually works.

A third feature of Hume's theory, and one that modern analytical philosophers sometimes have difficulty coming to terms with, is that it treats thought and feeling as prior to language. For Hume, thought does not always take place in words. In communicating with his readers through the medium of print, he has no option but to try to describe ideas and impressions in words, but he is conscious that any such translation must be inadequate:

> 'tis very difficult to talk of the operations of the mind with perfect propriety and exactness; because common language has seldom made any very nice distinctions among them, but has generally call'd by the same term all such as nearly resemble one another. (105)

Thus, there are distinctions between feelings 'of which 'tis impossible to give any definition or description, but which everyone sufficiently understands' (106). Hume sees it as a 'most fertile source of error' that metaphysicians 'use words for ideas, and ... talk instead of thinking in their reasonings'. Because ideas and the words that purport to represent them are closely connected, the mind easily mistakes the words for the ideas (61–2).

For Hume, language is a system of conventions (490). More particularly, the language we use to describe mental states is a set of conventions that have emerged to resolve the problems that human beings

face in trying to communicate about their states of mind. Because each person's mental states are in constant flux, and because each person's perception of an external object depends on his position in relation to it, ''twere impossible we cou'd ever make use of language, or communicate our sentiments to one another, did we not correct the momentary appearances of things, and overlook our present situation' (582). Thus, the language of mental states is more regular than our perceptions of the states themselves. For example, we say that a trait of character is virtuous if its exercise has a general tendency to induce sentiments of approval, even though each of us, because of his particular standpoint in relation to particular acts, sometimes fails to *feel* approval for what we all *call* virtuous (577–87). It follows from this view of language that feelings are conceptually prior to the language in which they are described. The structure that our language imposes on our feelings is analogous with the structure that a theory imposes when it organises empirical observations. In making sense of our observations, we must be alert to the possibility that the theory we are currently using is flawed. Similarly, in a scientific investigation of feelings, we must be alert to the possibility of error in the folk psychology that is embedded in our language.

In support of this position on the priority of thought to language, Hume appeals to the many observable similarities between human beings and other intelligent animals in terms both of physiology and of behavioural and affective responses to stimuli. Since it would be contrary to 'all our principles of reason and probability' not to attribute like effects to like causes, we should assume that the fundamental workings of the human mind are not very different from those of the minds of other intelligent animals. In explaining behaviour and affective states that are common to human beings and other animals, we should appeal only to those mental capacities that are similarly common:

> The common defect of those systems, which philosophers have employ'd to account for the actions of the mind, is, that they suppose such a subtility and refinement of thought, as not only exceeds the capacity of mere animals, but even of children and the common people in our own species; who are notwithstanding susceptible of the same emotions and affections as persons of the most accomplish'd genius and understanding. (177)

In particular, when explaining phenomena that are common to human beings and to animals which lack language, we should not assume the capacity to use language.[4]

Just as, for Hume, emotions and volitions are pre-linguistic, they are also non-propositional. This seems to be an immediate implication of the

[4] Consistently with this position, Hume maintains that non-human animals lack the sense of virtue and vice which, in his theory, is a by-product of language (326).

claim that each such mental state is an original fact and reality, complete in itself. However, many philosophers have doubted whether Hume can really mean this – claiming that, if he does, his position is untenable. For example, Donald Davidson (1980) offers what he calls 'Hume's cognitive theory of pride' as a reconstruction of 'what Hume *should* have meant'. According to Davidson, 'Hume's account of pride is best suited to what may be called *propositional pride* – pride described by sentences like, 'She was proud that she had been elected president'' (277). Taking the case of a man's pride in the beauty of his parish and using first-person terms, Davidson construes pride in the following *propositional form*:[5]

P1. I am proud that my parish is beautiful.

This formulation combines a report of a feeling (I have a feeling of pride) with a statement of a belief or judgement about the external world (my parish is beautiful). According to Davidson, that belief or judgement is *both* the cause of my feeling *and* my reason for so feeling: the theory 'explains the pride in two ways; it provides a causal explanation for it, and it gives the person's reasons for being proud'. Thus: 'The theory of propositional pride that I have extracted from the *Treatise* shows that someone who is proud always has his reasons' (285).

Generalising from this example, Davidson seems to be proposing that statements about certain types of emotional and volitional states, of which pride is the exemplar, can be analysed as:

P is S in virtue of P's belief that X,

where P is a person, S is an emotional or volitional state experienced by that person, and X is a proposition that P believes to be true. On Davidson's analysis, P's belief that X is both the cause of, and the reason for, S. This analysis of S is propositional in the sense I defined in Section 1.

By construing pride propositionally, Davidson opens up the possibility that a person's feelings can be subjected to rational appraisal, using criteria of logical or conceptual inconsistency. For example, consider the proposition 'All things considered, my parish is not beautiful.' On a natural reading, it would be inconsistent for me to assert this proposition in conjunction with P1. Thus, rationality imposes constraints of consistency on 'propositional pride': what one can rationally feel is constrained by what one believes or judges to be the case about the external world.

[5] Davidson uses a slightly different example, taken from the same passage in the *Treatise*: a man's pride in the beauty of his house. I prefer the case of the parish because it is more effective in prompting questions about the *justification* for the pride.

I suggest that Hume's understanding of pride is better represented by:

N1. Thinking about my parish as beautiful, I have an associated feeling of pride.

Notice that N1 does not treat pride as propositional. It does not refer to any proposition about the beauty of the parish, but merely records a thought and a feeling that coexist in my mind at the current moment. The thought is not a settled *belief* to which I assent, but an *idea* that is passing through my mind. Contemporaneously, I have a pleasurable feeling about myself. I am conscious that the feeling is linked to the idea (that is what is meant by saying that the feeling is 'associated'), but the feeling does not imply a belief in the truth of any proposition about the world. It's just a feeling. As a matter of empirical psychology, the feeling is caused by the thought.

How are we to decide whether pride is better described by P1 or N1? Given Hume's methodological stance, which is the better description is an empirical question, not a conceptual or linguistic one. Since this is an empirical question, the answer cannot take the form of a proof. What is at issue is a choice between theoretical models. We need to decide whether, in explaining the feelings of pride that people actually experience, it is more useful to model pride as suggested by P1 or N1. For P1 to provide a useful model, we would need a theory of how propositions like 'my parish is beautiful' provide reasons for pride, and it would have to be an empirical truth that feelings of pride are associated with beliefs in propositions which, according to that theory, provide reasons for pride.

One way of making progress towards resolving this issue is to investigate, by controlled experiments, whether feelings of pride are reliably associated with beliefs in credibly reason-giving propositions. The example of the parish is, I think, intended as just such an experiment. While the beauty of a man's house (Hume's leading example of a source of pride) might be thought to reveal his taste and wealth, the beauty of his parish seems to be a matter of pure luck. (I think we can take Hume to intend 'his parish' to be 'the parish of his birth'.) It is significant that when Hume refers to pride in the parish, he calls it *vanity*: 'Men are vain of the beauty of their country, of their county, of their parish' (306). The suggestion is that the beauty of one's parish does not give one a *reason* for pride; yet we find that it works as a *cause* of pride all the same.

An advocate of Davidson's analysis might reply that this experiment merely shows that there can be irrational pride as well as rational pride; P1 is intended only as an analysis of the latter. But remember that Davidson's analysis has been offered as a reconstruction of what Hume should have meant. If, as I have argued, Hume's aim is to explain our actual affective experiences, a distinction between 'rational' and 'irrational' emotions is redundant unless it plays some useful explanatory role. If the felt

experiences of rational and irrational pride are the same, and if both can be explained by the same theory, why distinguish between them? To criticise Hume's theory of emotion for not using the concept 'rational' would be like criticising a botanical classification for not using the concept 'weed'.[6]

If one wanted to pursue further the investigation of pride, an obvious experimental strategy would be to examine situations in which a person has the thought that is referred to in N1 (he is thinking of his parish as beautiful) but does not have the belief that is asserted by P1 (his settled belief is that his parish is *not* beautiful). Does such a person experience a feeling with the same affective qualities as in the more normal case in which he has both the thought and the belief? For example, suppose it is my belief that, all things considered, my parish is not at all beautiful, but there are one or two viewpoints from which, on a good day, it doesn't look too bad. I am passing one of these viewpoints with an impressionable visitor, who exclaims, 'What a beautiful parish!' The view and the exclamation bring to mind for me the idea of my parish as beautiful: briefly, the ideas of beauty and parish are associated together in my mind. Do I experience a correspondingly brief feeling of pride? Introspection suggests to me that I do.

Although Hume does not consider this particular experiment, the *Treatise* is full of experiments which have exactly this structure. For example:

> . . . let us consider the case of a man, who being hung out from a high tower in a cage of iron cannot forbear trembling, when he surveys the precipice below him, tho' he knows himself to be perfectly secure from falling, by his experience of the solidity of the iron, which supports him; and tho' the ideas of fall and descent, and harm and death, be deriv'd solely from custom and experience. (148)

In the spirit of Davidson's analysis of pride, one might propose a theory in which fear is represented in propositional form as:

> P2. I am afraid that I might fall to my death.

P2 can be construed as combining a report of a feeling (I feel fear) with a statement of a belief (there is a danger that I might fall to my death); the belief, it might be said, is both the cause of and the reason for the feeling. A corresponding non-propositional representation of fear is:

> N2. Thinking about falling to my death, I have an associated feeling of fear.

[6] The *Concise English Dictionary* defines 'weed' as 'wild herb growing where it is not wanted'. Roughly, what makes a plant a weed is its being the object of disapproval. The distinction between weed and non-weed is clearly meaningful, but it is not useful in botany.

Hume's experiment tests whether a person has a feeling of fear when he lacks the *belief that* there is a danger of falling, but is put in a situation which is designed to evoke vivid *thoughts about* falling. On Hume's account of what we discover in such an experiment, the person experiences a feeling with the same affective qualities – and the same observable correlate, namely trembling – as are normally experienced in situations of real danger. This suggests that fear may be better represented non-propositionally.

Here is another example, which I have mentioned in passing already. Hume is considering the passion of hatred, which in his theory has a similar status to pride: both are 'indirect passions' – passions that do not arise directly from the experience of, or the thought of, pleasure or pain, but take effect through associations of ideas (438–9). Hume claims that we tend to feel hatred for people who injure us. Since 'the principal part of an injury, is the contempt and hatred, which it shews, in the person that injures us', this tendency is particularly strong when the injury is intentional, but:

> . . . I ask, if the removal of design be able entirely to remove the passions of love and hatred? Experience, I am sure, informs us of the contrary, nor is there any thing more certain, than that men often fall into a violent anger for injuries, which they themselves must own to be entirely involuntary and accidental. This emotion, indeed, cannot be of long continuance; but still is sufficient to shew, that there is a natural connexion between uneasiness and anger, and that the relation of impressions will operate upon a very small relation of ideas. (349–50)

If we represent hatred in propositional form, we have something like:

> P3. I hate you for causing my injury.

P3 combines a report of a feeling (I feel hatred towards you) with a statement of a belief (you are the cause of my injury); the belief is both the cause of, and the reason for, the feeling. But, in the case described by Hume, this formulation seems inappropriate. My belief is not that you intentionally brought about my injury, but only that my injury resulted from some accidental and involuntary action of yours. Such a belief does not seem to provide an adequate *reason* for hatred. Even so, it remains a psychological fact that the belief can *cause* a feeling of hatred. Again, this problem is eliminated if we use the non-propositional form:

> N3. Thinking about you as the cause of my injury, I have an associated feeling of hatred.

The general conclusion I draw from these examples is that feelings are not always accompanied by the sorts of reasons that, in common language,

would be treated as justifications. In the most straightforward cases, feelings and reasons do go together in this way; the normal causes of feelings are typically treated as justifications too. But careful observation allows us to tease apart causes and reasons.

So far, I have been considering objections to the propositional analysis of emotions. Of course, Hume's non-propositional analysis confronts problems too. I now consider what I think are the two most serious possible objections. These are closely related.

First, it is implicit in Hume's approach that generic feelings such as pride, fear and hatred can be recognised as such, independently of what one feels proud *of*, afraid *of*, or hatred *for*, and independently of whether (in terms of a Davidsonian analysis) these feelings are rational or irrational. Felt experience has to be independent of reasons in this way if we are to have a genuinely empirical investigation of the causes of a generic feeling. For example, if Hume's experiment with the iron cage is to work, we need to be able to identify fear as a distinct feeling, so that the question 'Does the subject of the experiment feel fear?' is empirical and not conceptual. Are felt experiences really independent of reasons in this way? However odd this form of independence may seem in the perspective of some traditions of analytical philosophy, it is entirely credible as empirical psychology. The felt experience of fear has a distinct affective quality, independently of what the fear is about; we even know what it means to feel fear without being able to articulate what we are afraid of.

The second problem is this. Take the case of fear. If fear is a generic feeling, what does it mean to say that I am afraid *of falling*? The idea of falling may be the cause of my fear, but if that does not affect the felt experience of fear, how do I experience the association between fear and falling? This problem may seem to suggest that the fear of falling must somehow *contain* the idea of falling, as in the propositional analysis. Here, I think, we are dealing with one of those features of mental experience of which, as Hume puts it, it is impossible to give any definition or description, but which everyone sufficiently understands. If the mind works by associations of ideas, it is not surprising that we can sometimes be conscious of those associations; but it is hard to put into words exactly what this consciousness is. For the man in the cage, thinking about falling induces the sensation of fear by an association of ideas. He feels fear, and he also feels an association between this fear and the idea of falling. But the affective quality of the fear itself is just fear; the idea of falling is not part of its content.

Consider another example. People who have recently suffered episodes of food poisoning often find that they have feelings of nausea towards particular foods. Generalising across persons and episodes, the foods that take on this quality are not distinguished by any particular type or taste or propensity to transmit sickness, but simply by their having

been eaten just before the onset of illness. Think of the felt experience of nausea which wells up when one is confronted by one of these foods. In my recollection, the affective quality of this nausea does not take on the qualities of the relevant food; but there is a clear consciousness that the nausea is directed towards the food.

If aversion can work in this way, can desire do the same? If we think propositionally, it seems obvious to say that objects are desired *for* properties that they possess. Is it credible to treat desire as a generic feeling that can be associated with many different ideas and which can be recognised as a felt experience independently of what the desire is for? Perhaps surprisingly, some psychological and neurological evidence seems to suggest exactly this. Studies of the effects of addictive drugs suggest that they work by directly stimulating neural mechanisms that are associated with desiring or 'wanting'. In normal cases, wanting is associated with 'liking' – we feel desire for things that also give us pleasure – and so we are not conscious of the independent nature of desire. But addictive drugs can short-circuit the mechanisms of liking. They can create feelings of desire which, by associative learning, attach to mental representations of drug consumption. According to this model, the nicotine addict does not desire to smoke for the pleasurable feelings that smoking induces, or to avoid the painful feelings of nicotine deficiency: she simply has an intense feeling of desire which is directed towards smoking (Robinson and Berridge 1993).

In the light of this kind of evidence, it is not at all obvious that Hume's psychology is naïve. To the contrary, in so far as they purport to be analyses of feelings, propositional theories may be guilty of sophistication in the pejorative sense ('depriving person or thing of natural simplicity, making artificial by worldly experience'). Their artificiality comes from trying to impose the conceptual structure of a refined form of human language on to the natural facts of affective experience.

3. HUME'S THEORY OF DECISION

Having argued that Hume's analysis of the passions of pride, fear and resentment is non-propositional, I shall now suggest that the same is true of his analysis of the feelings that underlie decision-making. Just as Davidson's analysis of propositional pride fails to capture the fundamental logic of Hume's theory of pride, so a propositional analysis of preference fails to capture the logic of Hume's theory of decision.

Although Hume sometimes uses the word 'preference', the central concepts in his theory of decision are *desire* and *volition*. Desire is an 'emotion of propensity' which 'unites us to' the idea of some object (414, 439). That is, it is a passion which focuses on the idea of some object and induces us to approach, possess or consume it. Volition is the felt

experience of intentional action, 'the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind' (399).

On first reading, Hume's account of the relationship between pleasure, desire and volition seems similar to the classical utilitarianism of nineteenth-century economics – that is, the theory that an individual chooses those actions that can be expected to maximise his net pleasure. Defining 'good' as pleasure and 'evil' as pain, Hume says:

> 'Tis easy to observe, that the passions, both direct and indirect, are founded on pain and pleasure, and that in order to produce an affection of any kind, 'tis only requisite to present some good or evil. . . .
>
> DESIRE arises from good consider'd simply, and AVERSION is deriv'd from evil. The WILL exerts itself, when either the good or the absence of the evil may be attain'd by any action of the mind or body. (438–9)

However, Hume is not saying that the desire to perform an action is induced by the belief that that action will bring about pleasurable consequences, still less that the strength of desire is correlated with the corresponding degree of pleasure. His hypothesis is that the idea of pleasure tends to induce the feeling of desire, which in turn activates the will. I now discuss four ways in which this hypothesis is consistent with decisions which cannot be represented either as maximising net pleasure or as revealing preferences that satisfy the standard consistency axioms of rational choice theory.

3.1 The influence of contiguity on strength of desire

Hume's theory implies that the strength of our desire for something is affected by the vivacity of our mental representation of it. The vivacity of the mental representation of an object can be influenced by factors that do not affect our settled beliefs about its capacity to generate pleasure. In particular, 'every thing contiguous to us, either in space or time, [is] conceiv'd with a peculiar force and vivacity' (427). Thus, the strength of our desire for something increases with its contiguity. In a passage which anticipates recent findings of behavioural economics,[7] Hume argues that this leads to apparently irrational patterns of decision-making:

> In reflecting on any action, which I am to perform a twelve-month hence, I always resolve to prefer the greater good, whether at that time it will be more contiguous or remote . . . But on my nearer approach, those circumstances, which I at first over-look'd, begin to appear, and have an influence on my conduct and affections. A new inclination to the present good springs up,

[7] For a survey of current knowledge about temporal inconsistency, see Frederick, Loewenstein and O'Donoghue (2002).

https://doi.org/10.1017/S0266267106001027 Published online by Cambridge University Press

> and makes it difficult for me to adhere inflexibly to my first purpose and resolution. (536)

Take a concrete example. Suppose it is 1 January. I am told that I need a minor but painful surgical operation, which cannot be carried out until December. I can choose to have the operation either on 18 or 19 December. I consult my diary and find that 18 December is slightly more convenient, so I choose that. But on the morning of 18 December, if I am allowed to revise my choice, I feel a strong inclination to postpone the operation by a day. Viewed in the perspective of rational choice theory, I am revealing a temporal inconsistency in my preferences. On 1 January, I strictly prefer 'operation on 18 December' to 'operation on 19 December'; on 18 December, I have the opposite preference. This reversal of preference is not a response to new information: everything I know on 18 December, including the fact that the closeness of an operation induces a desire to postpone it, was known to me on 1 January. Nor is it a response to changes in my underlying time preferences: my desire to postpone closely approaching evils is a constant property of my psychology. (On 1 January, I would have chosen 2 January rather than 1 January as the date for an operation.) Given exactly the same information about exactly the same two states of affairs, and with no change in my underlying tastes or dispositions, my preference between them depends on the apparently irrelevant factor of the point in time at which I make the comparison. Clearly, it cannot be the case that both of these preferences are governed by the maximisation of net pleasure.

If we represent all this in terms of propositions about preferences, we have:

> P4a. I prefer 18 December to 19 December as the date for the operation in virtue of its being 1 January today.
>
> P4b. I prefer 19 December to 18 December as the date for the operation in virtue of its being 18 December today.

Comparing these two propositions, the question of their mutual consistency immediately comes to mind. It is natural to say that P4a and P4b are consistent only if the difference between the dates at which the propositions are stated provides a *reason* for the reversal of preference.[8] In rational choice theory, standard conditions of temporal consistency express the view that this kind of difference in dates is not an acceptable reason for a difference in preference.

[8] Broome (1999) takes this position, which he calls 'a non-Humean response' to an example in which a person appears to have non-transitive preferences. Broome argues that there are 'rational principles of indifference' which 'determine which specific differences between alternatives are not enough to justify a preference' (75).

On the analogy of the non-propositional representation of pride, fear and hatred, Hume's account of temporal inconsistency is better captured by:

> N4a. Having thought about the operation as a distant prospect, and now thinking about postponing it by one day, I have no associated feeling of desire.
>
> N4b. Having thought about having the operation today, and now thinking about postponing it by one day, I have an associated feeling of desire.

In this formulation, desires are matters of feeling and nothing else. The strength of a person's desire for something (or, in the case of the operation, the strength of his desire to postpone it) is affected by its closeness to him; but closeness is not the reason for the strength of the desire, it is only the cause.

3.2 The influence of associations of ideas on desire

According to Hume, desire, like other passions, is governed by associations of ideas. Indeed, it is only by appeal to associations of ideas that he is able to explain how we come desire the means to desired ends:

> 'Tis obvious, that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction. 'Tis also obvious, that this emotion rests not here, but making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and effect. (414)

In Hume's theory, when two mental items are associated with one another, consciousness of one tends to increase the vivacity of the other. For example (another of Hume's experiments), looking at a picture of an absent friend increases the force and vigour of any feelings of joy or sorrow that one has in relation to that friend (99). By exactly the same mechanism, a desire for one object becomes stronger when we think about other related objects as pleasurable. Hume offers the example of the effect on the appetite of the beauty of the manner in which food is presented. When a dish looks beautiful, we feel more appetite for it (394–5). This is yet another example of the double relation of impressions and ideas; pleasure in eating and pleasure in beauty are similar impressions, while the idea of the dish is common to both.

Although Hume does not quite do this, it is easy to construct cases in which, if this effect is at work, there can be reversals of preference. Consider the classic experimental design used by Jack Knetsch (1989) to

test whether indifference curves are 'reversible'. Each member of one group of subjects is first given a bar of chocolate and then offered the opportunity to exchange it for a coffee mug. Each member of another group is first given the coffee mug and then offered the opportunity to exchange it for the bar of chocolate. Knetsch's finding, which has been replicated many times, is that there is an 'endowment effect': other things being equal, a person is more likely to choose one object x over another object y if she has first been given x and than if she has first been given y. This effect is entirely consistent with Hume's theory of associations of impressions and ideas. Indeed, he recognises the existence of the endowment effect, even though he does not offer an immediate explanation for it: 'Men generally fix their affections more on what they are possess'd of, than on what they never enjoy'd' (482).

One possible association-based explanation treats the endowment effect as a phenomenon closely related to pride. Suppose I have been given the chocolate. Because I own the chocolate, the idea of the chocolate is associated in my mind with the idea of myself. My natural love for myself is associated by similarity with the impression of desire. Thus, my love for myself is transmitted to the perception that objects that are associated with me are desirable. (An alternative explanation, also compatible with Hume's theory, is that, having been given the chocolate, I begin to think about eating it; the idea of eating it, and the idea of this as pleasurable, have vivacity in my mind, making my desire for the chocolate more intense. If I am then unexpectedly offered the alternative of a coffee mug, the idea of using the mug has less vivacity, and my desire for the mug is correspondingly less intense.)

Again, we can compare propositional and non-propositional representations:

> P5. I prefer the chocolate to the mug in virtue of my owning the chocolate.
>
> N5. Having thought about the chocolate as mine, and now thinking about exchanging it for the mug, I have an associated feeling of aversion.

P5 offers what seems to be an inadequate reason for preferring the chocolate, while N5 simply reports a feeling and its cause.

3.3 Unresolved conflicts of desire

Recall Dreier's argument that a Humean theory of decision can allow conflicting desires, but that 'what's important' is that conflicting desires can be resolved into a decision. For Dreier, 'resolving' a conflict of desires seems to mean the formation of an all-things-considered preference. I may

have a first-order desire for a suntan and a first-order desire to avoid skin cancer, but what's important is that, when it comes to choosing whether or not to sunbathe, *either* I prefer sunbathing *or* I prefer not sunbathing *or* I am indifferent between sunbathing and not sunbathing. But, one might ask, important for what?

Rational choice theory implicitly assumes that conflicts of desires can always be resolved in this sense. (This assumption appears in the standard theory as the axiom that preferences are complete.) Certainly it is important for that theory that the assumption is true; and perhaps it would be convenient for us if we never experienced unresolved conflicts of desires. But Hume is concerned with the desires we do have, not the ones that some theory tells us we should have, or the ones it would be convenient for us to have. Far from assuming that conflicts of desire can always be resolved, he explicitly analyses cases in which such conflicts remain *un*resolved; and he explains the phenomenon of unresolved conflict as yet another implication of his theory of the association of ideas and impressions.

Hume investigates what happens 'where the objects of contrary passions are presented at once'. If the mind is simultaneously presented with two impressions or ideas, one of which is pleasurable and the other painful, do the corresponding passions 'mingle with each other [and] become mutually destructive', as an acid and an alkali do, or do they 'never perfectly unite and incorporate', like oil and vinegar? Hume's conclusion is that the answer depends on how closely the relevant ideas are associated with one another.

If there is a close association between them, the two passions tend to combine and cancel one another out. Hume claims that this is what normally occurs when the two passions arise from a single event 'of a mixt nature'. He does not give a concrete example, but I think the following case captures what he has in mind. Suppose I am thinking of going out walking. The weather is sunny but there is a cold wind. I am conscious of a mixture of pleasurable and painful ideas, both of which are associated with the weather. The pleasurable and painful ideas tend to cancel out, and I perceive the weather as, on balance, neither good nor bad (or perhaps as mildly good or mildly bad). As an example of the opposite case, Hume presents the case of a man who has just heard both of the birth of a son and of the loss of a lawsuit. The man's mind keeps running backwards and forwards between the 'agreeable' idea of the one event and the 'calamitous' idea of the other, never fixing on either, and never arriving at a representation of the combination of the two pieces of news as, on balance, good, bad, or indifferent. Hume then suggests that when we think about a prospect that offers some probability of a good outcome and some probability of a bad, we find particular difficulty in combining the ideas of good and bad because they are associated with mutually exclusive events. If the bad thoughts predominate, the resulting turmoil of emotions

is what we call fear. It the good thoughts predominate, it is what we call hope (441–3).

Hume's analysis of conflicts of emotion highlights an important difference between a static and a dynamic theory of mind. In a dynamic theory, the temporary existence of one mental state can cause the temporary existence of another. It is thus an empirical question whether, in response to a given stimulus, the mind arrives at an equilibrium state or remains in disequilibrium, constantly moving from one state to another without settling on any. Hume's theory allows the second possibility.

Although Hume does not say so explicitly, it seems to be an implication of his analysis that an individual can face a choice problem without having any settled preferences or all-things-considered desires with respect to the options between which the choice has to be made. For example, consider the man who has just heard about the birth of his son and the loss of his lawsuit. Suppose he has to choose between rushing home to see the new baby and rushing to his lawyer's office to learn the consequences of having lost the case. Hume's theory seems to imply that the man could be in a state of indecision, with no settled sense of which of the two actions he desired more.

3.4 Impulsive desires

Although Hume claims that the idea of pleasure normally induces feelings of desire, and that '[t]he chief spring or actuating principle of the human mind is pleasure or pain' (574), he does not claim that *all* desires and volitions arise in this way. Immediately after the paragraph in which he says that desire arises from good considered simply, he says:

> Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil and proceed not from them, like the other affections. (439)

Take the case of resentment, that is, the desire to return injuries. This is one of Hume's preferred examples of the 'violent emotions':

> When I receive any injury from another, I often feel a violent passion of resentment, which makes me desire his evil and punishment, independent of all considerations of pleasure and advantage to myself. (418)

On Hume's analysis, the desire to punish is a primitive response to the consciousness of injury. (Recall that it is not absolutely necessary that the original injury is perceived as intentional; even accidental injuries can cause resentment.) This desire is not derived from considerations of future pleasures and pains. Admittedly, given the existence of the desire,

satisfying it will have some positive affective quality, but the desire is not caused by thinking about the pleasure to be had from satisfying it. (I take this to be what Hume means when he says that passions of this kind produce good and evil, rather than proceeding from them.) Although the pleasure of returning injuries is real enough, one may have an intense desire to punish even when one knows that, on balance, the consequences of acting on this desire will be painful.[9]

More generally, impulsive desires cannot be rationalised as attempts to maximise net pleasure. To the contrary, they are the cause of what Hume sees as a general property of human behaviour, that men 'often act knowingly against their interests' (418). And there seems to be no reason to expect that behaviour induced by impulsive desires reveals consistent preferences.

4. NORMATIVITY

I have argued that the theory of decision that Hume presents in the *Treatise* is neither instrumental nor propositional. It is an empirical theory of desires and volitions, understood as feelings that have causes but not reasons. If my account is right, Hume intended the *Treatise* to be read as (among other things) experimental psychology. Read in this way, his theory of decision is both coherent and remarkably ahead of its time. Still, a philosophical reader may be inclined to make the same objection as Korsgaard, Hampton and Broome: we need a normative theory of practical reason but, if the *Treatise* is read as presenting a non-instrumental and non-propositional decision theory, Hume does not give us one.

The most direct answer to this objection is the one I gave in Section 1: Hume's explanatory project does not require a theory about how we ought to reason, but only a theory about how, in fact, our minds work when we engage in what we call 'reasoning'. Perhaps *we* need a normative theory of practical reason, but it is not a legitimate criticism of Hume that he has chosen to give us a theory of something else.

However, an additional answer can be given to the objection made by Korsgaard, Hampton and Broome. Although Hume's decision theory does not use normative concepts, he does offer some explanation of how human beings arrive at normative judgements about decisions. Alongside his theory of the psychological mechanisms that lie behind actual human behaviour, he provides a theory of the mechanisms that lie

[9] Once again, Hume's theory of decision anticipates recent work in behavioural economics. There is now a good deal of experimental evidence which suggests that people are motivated to punish behaviour that they perceive as unfair, even if the act of punishing is costly, and that this motivation plays an important part in stabilising cooperative practices. For an overview of this research, see Fehr and Fischbacher (2002). See also the discussion of resentment in Sugden (1986).

behind our judgements of virtue and vice – of what is praiseworthy and what is blameworthy.[10] On Hume's analysis, a trait of character comes to be regarded as a virtue if it tends to be useful either to society as a whole or to the particular person who has it (574–87). Thus, it is not surprising that Hume's list of virtues includes many of the traits that one would expect of a person who is particularly capable of engaging in and acting on instrumental reasoning, for example, perseverance, patience, activity, constancy and resolution (610–11). His discussion of temporal inconsistency is prefaced by the following remarks:

> When we consider any objects at a distance, all their minute distinctions vanish, and we always give the preference to whatever is in itself preferable, without considering its situation and circumstances. This gives rise to what in an improper sense we call *reason*, which is a principle, that is often contradictory to those propensities that display themselves upon the approach of the object. (536)

The implication seems to be that objects can be more or less preferable *in themselves*, independently of the associations of ideas that are evoked by 'situation and circumstances', and that to be capable of pursuing what is preferable in itself is a virtue – a virtue that, in ordinary language, is called 'rationality'. Given Hume's equation between 'good' and 'pleasure', it is hard to see what else he can mean by 'preferable in itself' than 'producing a favourable balance of pleasure and pain'.

So there is a sense in which Hume's theory *of virtue* is 'Humean': it includes elements of what would now be called instrumental rationality and utilitarianism. But this is not his theory of what people in fact desire, nor of how people in fact make decisions. Our ideas of virtue are matters of *taste*, as contrasted with *passion*. On Hume's account, our decisions are normally driven by our passions:

> A house may displease me by being ill-contriv'd for the convenience of the owner; and yet I may refuse to give a shilling towards the rebuilding of it. Sentiments must touch the heart, to make them controul our passions: But they need not extend beyond the imagination, to make them influence our taste. (586)

In Hume's theoretical system, judgements about the rationality of action – or, as he puts it, about what is improperly called 'reason' in the context of action – are distinct from the principles that in fact govern decisions. The elements of rationality that appear in his theory of virtue are not used in his decision theory.

At the level of methodological principle, it seems entirely reasonable to look for explanatory theories of decision-making which do not invoke

[10] My reading of Hume's theory of virtue as usefulness is similar to that of Sayre-McCord (1996), who shows how it differs from utilitarianism.

principles of rationality. That was surely true when Hume wrote the *Treatise*, when decision theory as we now know it did not exist. But, one might ask, is such an approach to decision theory still viable today? For the last 50 years, it must be said, most explanatory work in economics has assumed that agents act instrumentally with respect to preferences that satisfy standard conditions of consistency. Apparently, economists have been confident that the behaviour of economic agents is accurately and parsimoniously explained by theories of this kind. Some economists maintain that there are empirical reasons to expect rationality assumptions to work well in many of the environments that economics investigates.[11] Still, we may yet find that the best explanations of decision-making are, as Hume's theory is, based on psychological assumptions which make no reference to rationality. Current work in behavioural economics is premised on the credibility of the latter possibility.

Although my sympathies with the behavioural approach will have been obvious, I have not tried to argue that it is self-evidently superior to the rational-choice approach. It is sufficient for my argument that, more than 250 years after Hume wrote the *Treatise*, his behavioural approach to decision theory remains a viable option. That approach is not naïve, antiquated or discredited. I hope I have persuaded the reader that Hume's account of decision-making is better understood as an anticipation of experimental psychology and behavioural economics than as an anticipation of the theory of rational choice.

REFERENCES

Blackburn, S. 1998. *Ruling passions: a theory of practical reasoning*. Clarendon Press.

Broome, J. 1991. *Weighing goods*. Blackwell.

Broome, J. 1999. Can a Humean be moderate? In *Ethics out of economics*, ed. J. Broome, 68–87. Cambridge University Press.

Davidson, D. 1980. Hume's cognitive theory of pride. In *Essays on actions and events*, ed. D. Davidson, 277–90. Clarendon Press.

Dreier, J. 1996. Rational preference: decision theory as a theory of practical rationality. *Theory and Decision* 40:249–76.

Fehr, E., and U. Fischbacher, 2002. Why social preferences matter – the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal* 112:C1–C33.

Frederick, S., G. Loewenstein and T. O'Donoghue, 2002. Time discounting and time preference: a critical review. *Journal of Economic Literature* 40:351–401.

Gauthier, D. 1986. *Morals by agreement*. Clarendon Press.

[11] See, for example, Plott's (1996) argument that, with repeated experience of participation in markets, economic agents 'discover' and learn to act on internally consistent preferences. Plott's approach draws a sharp distinction between rational behaviour, which conventional economic theory is designed to explain, and non-rational behaviour, which is the preserve of psychology. Kahneman's (1996) response to Plott expresses a psychologist's scepticism about the usefulness of this distinction.

Hampton, J. 1998. *The authority of reason*. Cambridge University Press.
Hume, D. 1978 [1739–40]. *A treatise of human nature*. Clarendon Press.
Kahneman, D. 1996. Comment [on Plott, 1996]. In *The rational foundations of economic behaviour*, ed. K. J. Arrow, E. Colombatto, M. Perlman and C. Schmidt, 251–4. International Economic Association and Macmillan.
Knetsch, J. 1989. The endowment effect and evidence of nonreversible indifference curves. *American Economic Review* 79:1277–84.
Korsgaard, C. 1997. The normativity of instrumental reason. In *Ethics and practical reason*, ed. G. Cullity and B. Gant, 215–54. Clarendon Press.
McClennen, E. 1990. *Rationality and dynamic choice: foundational explorations*. Cambridge University Press.
Millgram, E. 1995. Was Hume a Humean? *Hume Studies* 21:75–93.
Nozick, R. 1993. *The nature of rationality*. Princeton University Press.
Plott, C. 1996. Rational individual behaviour in markets and social choice processes: the discovered preference hypothesis. In *The rational foundations of economic behaviour*, ed. K. J. Arrow, E. Colombatto, M. Perlman and C. Schmidt, 225–50. International Economic Association and Macmillan.
Ramsey, F. 1931. Truth and probability. In *The foundations of mathematics and other logical essays*, 156–98. Routledge and Kegan Paul.
Robinson, T. and K. Berridge (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Reviews* 18:247–91.
Savage, L. 1954. *The foundations of statistics*. Wiley.
Sayre-McCord, G. 1996. Hume and the Bauhaus theory of ethics. *Midwest Studies* 20:289–98.
Sugden, R. 1986. *The economics of rights, co-operation and welfare*. Blackwell.
Sugden, R. 1991. Rational choice: a survey of contributions from economics and philosophy. *Economic Journal* 101:751–85.