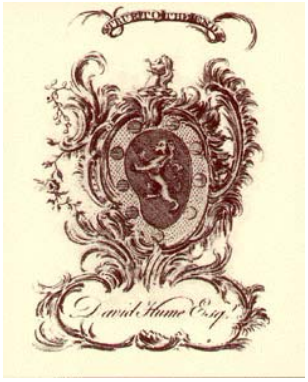

Hume's Moral Sentimentalism

Daniel Shaw

Hume Studies Volume XIX, Number 1 (April, 1993) 31-54.

Hume's Moral Sentimentalism

Daniel Shaw

In chapter 7 of his book, *Hume*, Barry Stroud considers and rejects a number of standard interpretations of Hume's sentimentalism and then argues for his own 'projectionist' interpretation.[1] In this paper I shall comment briefly on all these readings, raise objections to Stroud's proposal, and, finally, argue in favour of what I shall call the 'power' interpretation of Hume's sentimentalism.

Hume maintains that the vice or virtue of an action is not a matter of fact about the action that can be inferred from anything, by causal or inductive reasoning alone, nor a fact that can be discovered by direct perception of an action, but that, rather, vice and virtue are matters of moral sentiment. Hume writes,

> Take any action allow'd to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation which arises in you, towards this action. Here is a matter of fact; but 'tis the object of feeling, not of reason. It lies in yourself, not in the object. So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it. Vice and virtue, therefore, may be compar'd to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in objects, but perceptions in the mind.[2]

Stroud gives the following explanation of this passage:

> [W]hat we believe or 'pronounce' when we regard an action as vicious is different from and more than anything we can discover by perception of the action or by inference from its observed characteristics to other matters of fact about it. Hume grants that there are certain observable characteristics an action can be known to possess such that when we know

applied to Hume's moral theory, has to be understood in terms of Hume's own concept of contingent constant conjunction. On this view, cause and effect consists in nothing more than a constant conjunction between one type of event and another type of event.)[17]

One of the main attractions of this power interpretation is that it provides a general framework into which the many and various elements of Hume's moral theory fit neatly into place. The characteristics in actions which explain their power to cause moral feelings are the four kinds of utilitarian properties: immediate agreeableness to self, immediate agreeableness to others, utility to self and utility to others. The laws and facts of human nature which explain the power of virtue are the facts about the sentiment of humanity and the laws of sympathy: namely, the fact that (on Hume's account) contemplating the pleasures of others and therefore contemplating those actions and motives which give pleasure to others, causes the normal contemplator himself to experience a pleasurable feeling of approval, or at least causes him to be disposed to experience such feelings.

The standard condition necessary for making a correct moral judgement, a condition which corresponds to the standard viewing conditions needed to judge the colour of an object correctly, is the condition that the judgement be made from what Hume calls the common standpoint, namely, from the standpoint from which we consider only the type of character or action in general, neglecting any peculiar individuating features of the particular case—peculiarities such as how the action in question affects our private interests and its particular position in time and space; whether, for example, it happened today or a thousand years ago, whether next door or in another country. Just as, in the case of visual perception, we do not judge that an object grows smaller as it recedes into the distance just because the image it causes in our visual field grows smaller, but, rather, we learn automatically to allow for the distance-factor in making our perceptual judgement; so too, when we judge the morality of an action from the common standpoint, we learn automatically to make allowance for the action's spatio-temporal position, and we do not judge an action to be, for example, less virtuous just because, owing to its remoteness in time and/or space, or because of its special effects on our private interests, it happens to give us a smaller, that is, less intense, feeling of approval when we contemplate it.

In the case of visual perception, our shared concept of standard viewing conditions and our shared ability to correct for variations in distance bring enough uniformity, generality and consistency to our perceptual judgements for us to be able to communicate coherently with each other concerning them. This generality has two aspects:

> that the action has them we inevitably regard it as vicious. (H 178)

> In a case of wilful murder, for example, we can discover by (roughly) causal reasoning that one man deliberately and unnecessarily destroyed a human life, and caused great suffering, pain and hardship both to the victim and to others. (H 177)

And Hume would grant that when we know that the man's action has these characteristics we inevitably regard it as vicious. But,

> he quite rightly insists that that does not imply that regarding that action as vicious is simply believing that it has those observable [and inferrable] characteristics. He thinks that pronouncing an action to be vicious is something different and that is why he says that the vice entirely escapes you as long as you consider only the object thought to be vicious. (H 178)

Why does Hume think that, for example, believing an act of wilful murder to be vicious cannot consist simply and entirely in believing it to possess the observable and inferrable characteristics of being a deliberate and unnecessary destruction of human life which caused great suffering, pain, etc.? Surely, one might think, to believe all that of an action is all there need be for believing the action to be a morally vicious action. But, of course, Hume would argue that since the above belief is itself merely a conclusion of perception and causal reasoning, which in itself has no motivating power, and since, by contrast, the judgement that the act is morally vicious, being a moral judgement, must in itself be capable of motivating us to act, that is, to refrain from murder and do all we can to prevent murder, the latter active motivational moral judgement must involve something over and above the former inert conclusion of reason.

I would agree with Stroud that Hume is right about this. Imagine a case of a psychopath whose ability to argue and draw valid inferences is in perfectly good working order but whose motivational and emotional derangement is so severe that he has no conscientious aversion whatsoever to the idea of committing wilful murder. Now it is quite possible for this man to validly perform all the perceptual and causal reasoning needed to arrive at the conclusion that a given act involves the deliberate and unnecessary destruction of a human life and the causing of great pain and suffering. But, by the terms of the case, since he is psychopathic, arriving at that belief about the action gives this person no aversion towards such actions and no inclination

to refrain from or to try to prevent such actions when it is in his power to do so.

Hume would surely be right to argue that although the psychopath has reasoned to all the relevant facts about the action, he has not made the sincere moral judgement that the act is a morally vicious action. The psychopath could of course use that form of words of the action, that is, he could be taught to apply the predicate 'is vicious' to all actions of that kind; but given his motivational state he could at most be using these words in parrot fashion, that is, imitating the verbal distinctions made by normal moral observers without sincerely making the moral judgements about it that they make. The ingredient that is missing from the psychopath's judgement which prevents it from counting as a sincere moral condemnation of the act of murder is some aversion to the action, that is, some desire to avoid it which could actually motivate him to refrain from and to oppose such actions whenever it was in his power to do so.

Since Hume believes that all desires necessarily involve feelings,[3] Hume would say that what is missing in the case of the psychopath's judgement is an introspectible sentiment of disapproval.

Does Hume think that whenever anyone makes a sincere moral judgement he must always, at the time of making the judgement, actually experience a feeling of approval or disapproval? It is not entirely clear from his writings whether he would require this.[4] But if he did, his claim would be open to the objection that we sometimes do make sincere moral judgements without actually experiencing a feeling of approval at the very time of making the judgement. Think of reading a page of the daily newspaper and reading about cases of murder, theft, bribery and corruption. Does one necessarily have to have a distinct feeling of disapproval as one's eye goes down the page in order for it to be true of one that one regards these as vicious actions? Obviously not.

As in the case of the debate about feelings as motives of actions, the Humean who is trying to give a defensible interpretation which satisfies as many of Hume's main aims as possible would, I think, do well to invoke the idea of a 'calm' feeling of approval or disapproval, and to interpret 'calm feeling' dispositionally,[5] that is, as a disposition to feel approval of some act that one judges to be virtuous, a disposition which the sincere maker of the judgement that x is virtuous necessarily has at the time of making the judgement but which may, on occasion, be present in an unactualized form at the time of making the judgement. This would amount to saying that although the maker of the judgement may not actually feel approval as he sincerely thinks or utters the thought that act x is virtuous, nevertheless, it is true of him, at that very time, that, were he to seriously contemplate the prospect of doing act x when the opportunity arose, or failing to do act x when

the opportunity arose, he would actually experience a favourable sentiment towards the former option and a sentiment of aversion towards the latter; or at least that he would be consciously aware of approving of the former and disapproving of the latter.

In his encyclopedia article on Hume, D. G. C. MacNabb gives what seems to me a quite useful summary of Hume's sentimentalist theory of moral judgement:

> Hume's theory of moral judgments is that to consider a character trait or an act which springs from it as virtuous or vicious is to have a special sort of feeling of pleasure or displeasure toward it. The distinctive character of this feeling is that it is aroused only by human characters and actions, that it is aroused only when the type of the character or action is considered in general, neglecting any individuating features of a particular case, and that the feeling is affected by no features of the character or action other than its pleasantness or unpleasantness, its usefulness or harmfulness, either to its possessor or to others affected by it.[6]

MacNabb's summary is useful just because it brings together in a single paragraph the main elements of Hume's positive theory; firstly the sentimentalism, that is, the view that the making of a moral judgement involves the having of a special feeling; secondly, Hume's view that, in the final analysis, human character traits, that is, virtuous and vicious motives, are the only proper object of moral judgements (actions being relevant only as a sign of the motive behind them). Thirdly, the doctrine of the common standpoint and the correction of sentiment by reason, that is, Hume's idea that reason modifies, regulates and adjusts the way we interpret our feelings of approval and disapproval in order to give our moral judgements the generality and consistency that intelligible moral discourse and communication requires. This correction of feeling by reason enables us, as MacNabb puts it, to consider the type of character or action in general neglecting individuating features of the particular case. When, according to Hume, we adopt the common standpoint, from which moral judgements are made, reason enables us to ignore such peculiarities as variations in distance, time and space and the bearing the action has on our private interests; and finally, Hume's utilitarianism, Hume's view that the feeling of approval or disapproval is affected by no features of the character or action other than its pleasantness or unpleasantness, its usefulness or harmfulness, either to its possessor or to others affected by it.

But, although MacNabb's summary can serve as a good reminder of the main outlines of Hume's ethical theory, it is obviously an oversimplification. As Barry Stroud points out, it really isn't that easy to say "exactly what role ... feelings play in Hume's theory and exactly what the relation is between the sentiment I feel in my own breast and the moral pronouncements that I make." As Stroud says, "there are several different possible answers to this question, all of which are connected fairly closely to what Hume actually says" (H 180). It is this question I would like to take up now.

Perhaps the simplest answer we could give to this question is to interpret Hume's sentimentalism as a kind of simple subjectivism.

On this interpretation, when I sincerely say or believe that a particular action is vicious, I'm really making a report on some feature of my own psychological state, that is, I am saying or believing that I have a feeling of disapprobation towards the action. In Stroud's words:

> On this view ... moral talk is autobiography ... [All] "moral judgements" I make are about me, and when you say "x is vicious," you are saying that you have a certain sentiment towards x. (H 180)

As Stroud points out, the simple subjectivist interpretation of Hume's sentimentalism does have two main attractions. Firstly, it emphasizes the importance of feeling in moral judgements, and, secondly, it supports Hume's view that moral judgements are not arrived at by reason alone. One can report one's feelings without having to make any rational inferences, but, as Stroud also observes, the simple subjectivist view has two fatal weaknesses which make it totally unacceptable as an account of moral judgements. The first weakness is that it changes the topic of moral judgements from apparently being judgements about actions and agents into really being judgements about the speaker, that is, into being autobiography. This is a weakness because it flies in the face of all appearances. As Stroud says,

> in saying "That that was a vicious act done by an evil man" we certainly seem to be saying something about an action and an agent, not just something about our own feelings. (H 180)

The second weakness of simple subjectivism, which really makes the first one fatal, is that, while it denies that the apparent subject of moral judgements is the real subject of such judgements, it gives no explanation whatever of why actions and agents should ever have appeared to be the real subject of moral judgements. As Stroud puts it,

> appearances are sometimes deceiving and it might be that all we are saying when we say that someone evil did something vicious is something about our feelings, but if so we need at the very least some understanding of how it can seem to us that we are attributing some characteristic to the action itself, even if we are not. (H 181)

These are the main defects of the simple subjectivist reading.

Another way of interpreting Hume's sentimentalism would be to take quite literally his claim[7] that vice and virtue are not qualities in objects, but perceptions, that is, feelings in the mind. Taken literally, this means that the vice of an action is identical with a feeling I get when contemplating the action. But, as Stroud points out, that has the absurd implication that

> in saying that I get a certain feeling from contemplating x, I would be saying that I get vice from contemplating x. (H 181)

That is an absurd claim. It is doubtful that Hume intended that and therefore doubtful that Hume literally meant that vice and virtue were identical with certain feelings had by contemplators of vicious and virtuous actions.

Another interpretation of Hume's view about the relation between feeling and moral judgements would be to understand him to be saying that, although vice and virtue are not identical with certain feelings, and although vice and virtue are not really in the action, it is only because I get a certain feeling from contemplating a vicious act that I judge the action to be vicious.

On this 'explanatory' interpretation, the feeling that we get from contemplating the action is a necessary part of what explains the fact that we pass a moral judgement on the action; that is, if the feeling were not present I would not speak or judge as I do. But my utterance of the judgement is not thereby shown to be a simple report of the presence of the feeling in my mind.

The explanatory interpretation, as so far stated, is incomplete, for an answer still needs to be given to the following question: if the moral judgement 'x is vicious' is not simply a report of a feeling, but is rather explained by or caused by the feeling, what then does the judgement itself mean? To say that it is explained or caused by a feeling does not yet tell us what it means in itself.

One way of answering this question would be to give an emotivist account of the meaning of moral judgements. On this emotivist view,[8] when I say that x is good, I am not *saying that* I have a certain feeling

but rather I am simply expressing, in a linguistic way, a certain feeling or emotion that I have towards x. As Stroud says,

> It would be like a cheer for x. Cheers are linguistic, but they are typically not assertions. (H 182)

But, as Stroud also points out, it would be both anachronistic[9] and inconsistent with one of Hume's main aims to give this emotivist account as part of an interpretation of Hume's sentimentalism. Not only is there "no evidence that Hume ever considered such a theory" (H 182), but also the emotivist claim that moral judgements lack truth-values conflicts with Hume's view that when we make a moral judgement we are putting something forward which we regard to be true.[10] Of course, as Stroud also points out,

> [Hume's] considered view is that moral judgements are not literally true of anything in the action in question ... but that does not imply that we do not tend to regard those statements as being objectively true, and to put them forward as such. It is just that behaviour that Hume wants to explain. (H 182)

The interpretation we have just considered, the explanatory interpretation of Hume's sentimentalism plus the emotivist theory of the meaning of moral judgements, has two parts to it. The first part is the idea that although moral judgements are not simple reports of the speaker's feelings, nevertheless the speaker's feelings are the *basis* of his judgement, in the sense that they are a necessary part of what explains his making of the moral judgement. If he did not get certain feelings from contemplating act x, he would never make the judgement that he does make about act x.

Stroud argues that this first part of the above interpretation is wholly acceptable: it avoids the objectionable subjectivist claim while at the same time satisfying Hume's requirement that feeling be the basis of moral judgements. It is only the second part of the above interpretation, the part which, contrary to Hume's intention, denies a truth value to moral judgements, which Stroud, for that reason, finds objectionable.

Stroud's own solution to the problem of interpreting Hume's sentimentalism is to combine the first, explanatory part of the above interpretation with a different account of the meaning of moral judgements—what I shall call the projection theory of the meaning of moral judgements. The projection theory does allow a truth-value to moral judgements but does so without turning moral judgements into something that could be arrived at by reasoning. The idea of Stroud's

projection theory is that moral judgements, as part of their very meaning, do ascribe objective moral characteristics such as goodness and rightness to agents and actions, but that these objective moral properties do not really exist as properties of the actions and agents, but are just products of our moral feelings, products which we, in making moral judgements, project on to actions and agents, regarding these projections as if they were objective features of actions and agents, when in reality they are no such thing.[11]

Stroud's own account of this projection theory runs as follows:

> I contemplate or observe an action or character and then feel a certain sentiment of approbation towards it. In saying or believing that x is virtuous I am indeed ascribing to x itself a certain objective characteristic, even though, according to Hume, there really is no such characteristic to be found 'in' x. In saying that x is virtuous I am not just making a remark about my own feeling, but I make the remark only because I have the feeling I do. (H 184)

Now Stroud admits that this projection theory "goes beyond anything explicitly stated in Hume's discussion of morality" (H 185), but he nevertheless thinks that it is a good interpretation of Hume's sentimentalism because he thinks it "coheres better than any alternative with [Hume's] general philosophical aims" (H 185) without being inconsistent or incoherent in itself.

Stroud's account of how this projection theory satisfies Hume's main aims runs as follows: firstly, according to the projection theory, "without the appropriate feelings there would be no such thing as moral judgements" (H 184). That achieves Hume's aim of making morality a matter of feeling, not reason, thereby enabling moral considerations to influence the will. Secondly, also according to the theory, "What I actually feel determines the moral judgement that I make" (H 184). Since the judgement has no source other than feeling, that satisfies Hume's aim of making a moral judgement into something that cannot be arrived at by reasoning. Thirdly, according to the theory, the moral judgement is "not a report to the effect that I have such a feeling. ... Rather it is the attribution of a certain characteristic—virtue or goodness—to an action or character" (H 184-85). For this reason, the theory avoids the mysterious change of topic which marred the subjectivist reading considered earlier and it avoids making the modern emotivist move of depriving moral judgements of their truth-values, a move which there is no reason to think that Hume ever intended.

Although this interpretation has a lot going for it, I have my doubts about it. According to Stroud's projection theory, that which we project on to an action or character is not the feeling itself, which makes no sense, but rather, what we project on to the action is an idea of virtue or goodness, an idea which arises directly from the feeling of approval.

The first difficulty about this is that Hume says that it is by means of our impressions, not our ideas, that we distinguish between virtue and vice.[12]

A more serious difficulty is that, on Stroud's interpretation, Hume gives us no explanation of what it is we are supposed to be attributing to an action when we project this idea of virtue on to it; that is, he gives no account of what exactly this idea is an idea of, no account of what the word 'virtuous', understood as standing for some objective characteristic, means. In this connection, Stroud says (H 186) that it is not surprising that on the projection interpretation of his theory Hume gives no account of what the term 'virtuous' means, for, says Stroud, it is likely that Hume regards the notion of virtue or goodness as simple: "The *origins* of certain simple perceptions in the mind can be explained, but their meaning or content cannot be explicated further" (H 186).

But I think there is a serious objection to Stroud's suggestion that the term 'virtuous' might just stand for a simple idea. It is true, as Stroud says, that words which stand for simple ideas, words such as 'red' or 'sweet', are not verbally definable in the sense that they cannot be analysed into component parts. They cannot be so analysed because, being simple, they are not made up of component parts. However, although simple ideas cannot be defined in the sense of being analysed, there is another sense in which simple idea words can be defined; that is, they can be given what is commonly known as an ostensive definition. That is to say, it is possible to grasp what such words mean by direct awareness of what they stand for, and it is possible to indicate their meaning to somebody else by getting that person to become directly aware of the simple property for which these words stand. For example, we can remind ourselves of what the word 'red' means by making ourselves directly aware of the red surface of a ripe tomato.[13] And we can direct our attention to what the word 'sweet' stands for by tasting a lump of sugar. We could teach someone who lives in an area where salt is never found what 'salty'[14] means by putting a bit of salt on his tongue and having him taste it, having him become directly aware of the salty taste.

But have we any such direct awareness of some simple property for which the word 'virtuous' stands? Stroud's projection interpretation of Hume's sentimentalism attributes to Hume the view that we do have some simple idea of virtue, some simple idea produced by, but distinct from, our feelings of approval; which simple idea we then project on to

actions and characters. But it is not at all clear to me either that we do have such a simple idea or that there is anything in Hume to suggest that he thinks we have.[15]

Another major difficulty I find with Stroud's projection theory is that, on this interpretation, Hume is committed to the view that all our moral judgements, being projections on to reality of characteristics that do not really exist in reality, are all, strictly speaking, false judgements. Surely if Hume meant that, he would have said that, and he doesn't. What he does say, is that all pure rationalists have a false conception of the nature of moral judgements, not that all moral judgements are themselves false.

I should now like to give my own interpretation of Hume's sentimentalism, my own answer to the question, Exactly what role do feelings play in Hume's theory of moral judgements, and what is the relation between the sentiments we feel in our own breasts and the moral pronouncements we make?

I take as my starting point Hume's remark in the *Treatise*, "Vice and virtue, therefore, may be compar'd to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in objects" (T 469). I take this to be a reference to John Locke's theory of the secondary qualities as being not qualities *in* objects, like shape, size and solidity, that is, not primary qualities, but rather what Locke calls powers, powers which objects have (and which they have in virtue of certain of their primary properties) to give rise to certain perceptions in the mind of human observers. With one important qualification, that is the sort of theory which I think Hume is trying to apply to morality.[16] According to Locke's theory, a secondary quality such as redness is not simply a characteristic in the red object such as a certain configuration of its molecules, nor is it simply a perception in the mind of an observer of red objects. Rather, according to Locke, the redness of the object is a power which the object has to give rise to certain perceptions in normal perceivers under certain standard viewing conditions. On this theory, the red object possesses this power in virtue both of certain properties in the object as well as certain universal laws about the mental and physical make-up and responsiveness of normal human observers.

On the power interpretation of Hume's sentimentalism, virtue is the power that certain actions have (a) because of certain characteristics in them, as well as (b) because of certain laws of human nature; a power to give rise to feelings of approval in the minds of those (c) normal people who observe or contemplate those actions (d) under certain standard conditions. (The above-mentioned qualification is as follows: the term 'power' here must not be understood in terms of Locke's idea of natural necessary connection. Hume rejects that Lockean theory of power [at E 63-73]. The idea of power, if it is to be

there are two ways in which standard viewing conditions give generality to our perceptual judgements. Firstly, they enable us to make reasonably accurate perceptual judgements, for example, judgements of colour, concerning all objects which come within our visual field. Irrespective of how near or far away from us they may be, we are able to make allowances for the distance factor. Secondly, since this conception of standard viewing conditions is an idea which all normal perceivers share in common, it enables all such perceivers to make similar perceptual judgements.

By analogy, Hume's theory of the common standpoint from which moral judgements are made also gives to moral judgements the very same two types of generality as in the case of perceptual judgements. Firstly, similar moral judgements made from the common standpoint can be made concerning all similar actions whatever their position in time and space and whatever their relation to our private interests. At least in principle, it is possible to make moral judgements about all actions. Secondly, all normal moral beings who take up the common standpoint can, in principle, arrive at similar moral judgements, just as in the case of agreement about colour perception. Finally, just as our perceptual theory enables us to explain certain mistaken perceptual judgements as arising from certain abnormal physical conditions of the perceiver (for example, he mistakenly thinks something is yellow only because he happens to be suffering from jaundice, an abnormal organic condition), so, too, Hume's theory of the normal moral make-up, including the normal operations of sympathy and humanity, enable Hume to explain certain aberrant moral judgements, for example, approval of self-denial or self-injury or pointless ascetism as an end in itself, as arising from certain psychological abnormalities in people who make these judgements.[18]

In addition to providing a good unifying framework for bringing together the many elements of Hume's moral theory, the suggested interpretation is also supported by certain key passages in the *Treatise* and the *Enquiry*.

In the *Treatise*, Hume speaks of the impression, that is, the feeling of approval, arising from virtue. He says we "must pronounce the impression arising from virtue, to be agreeable" (T 470). Now this accords with the interpretation of virtue as a power which actions have to give rise to moral sentiments. In *An Enquiry Concerning the Principles of Morals*, Hume goes so far as to define virtue as "*whatever mental ... quality gives to a spectator the pleasing sentiment of approbation*" (E 289). That could well mean that virtue is the action's power to give feelings of approval to people who contemplate it.[19]

In *Hume*, Stroud mentions (pp. 182-83) two ways in which this definition could be taken (this definition of virtue as "whatever mental

quality gives to a spectator the pleasing sentiment of approbation"). One could take this definition to be equating, for example, viciousness, with that particular quality in actions which causes feelings of disapproval in human beings as they are presently constituted. Since Hume thinks that utility and disutility are among the causes of approval and disapproval, it would follow that, on this way of taking the definition, the viciousness of certain actions would be identical with their disutility, that is, with their tendency to cause pain, and the virtue of certain other actions would be, on this interpretation, identical with their utility, their tendency to cause pleasure.

But since utility and disutility are features of actions which we can discover by reasoning and observation alone, this way of taking Hume's definition would imply that vice and virtue were discoverable by reasoning and observation alone; and that conclusion conflicts with one of the main points of Hume's moral theory, namely his anti-rationalism. Moreover, since utility and disutility are characteristics which are objectively in the action and which would still be in actions whether or not anyone approved or disapproved of them,

> it follows, from this way of taking the 'definition', that our getting certain feelings is not essential to there being such things as virtues and vices, and that is not what Hume intends. (H 183)

Stroud discusses (H 183) a second way of taking Hume's definition of virtue as "whatever mental action or quality gives a spectator the pleasing sentiment of approbation." This second way comes nearer to my power interpretation, but isn't quite the same.

On Stroud's second way of taking the definition, 'x is vicious' means that x has that quality, whatever it might happen to be at the time, which actually causes us to have a certain feeling of disapproval towards x.

As Stroud says,

> This would not equate viciousness essentially with any particular objective property or characteristic, and it would have the virtue of emphasizing the importance of feelings for the very existence of such things as vices and virtues, as Hume intends. On this [way of taking the definition], if no feelings of disapprobation ever occur, or if they stop occurring, it follows that nothing is vicious. Similarly, if those feelings do occur but are not caused by any quality in the objects contemplated or observed, then nothing is vicious. ... The feelings themselves

> would not *be* virtue or vice, but virtue and vice would exist only in so far as we get such feelings. (H 183)

Stroud rejects this way of taking the definition for the reason that it implies that we could not arrive at moral judgements by feeling or sentiment alone; and Stroud seems to think that Hume claims that we can arrive at moral judgements by feeling or sentiment alone. As far as I can see, Hume does not think that we arrive at moral judgements by sentiment alone, and I therefore do not think that Stroud's reason for rejecting the second way of taking Hume's definition is a good reason.

But I think that there are two other good reasons for rejecting the second way. First of all, although, as Stroud points out, the second way of taking the definition does not make virtue identical, by definition, with any particular specified objective property of actions, nevertheless, it does claim that virtue is contingently, that is, as a matter of fact, identical with whatever property of actions happens as a matter of fact to cause us to have a sentiment of approval towards the action. It does, in other words, identify virtue with an objective characteristic in the action, and this is one thing that Hume explicitly denies. By contrast, the power interpretation which I have suggested, avoids this claim, that is, avoids identifying virtue with an objective characteristic in the action and with nothing over and above such an objective characteristic. A power of an action to cause a sentiment is not itself an objective characteristic residing wholly in the action any more than it is a subjective sentiment in the spectator. Rather, the power itself is a complex relationship between certain characteristics in the action, certain facts about human nature and certain subjective sentiments in the observer.[20] Virtue itself, on this interpretation, is the relationship described by the following hypothetical statement: If a normal observer assumes the common standpoint and contemplates an action which has characteristic c,[21] then that characteristic of the action, taken together with certain facts about the observer's human nature, will cause the observer to experience a sentiment of approval.

Against this power interpretation, Stroud could raise the very same objection which he made against his own rejected second way of taking Hume's definition, namely the objection that, on this power interpretation, moral judgements could not be arrived at by feeling or sentiment alone. The power interpretation of 'act x is virtuous' implies, among other things, that act x has some objective characteristic which causes us to have certain feelings and, as Stroud says,

> having a feeling towards x is not enough in itself to lead us to believe that some property of x caused us to have that feeling.

> ... [W]e will not believe that solely on the basis of the feeling alone. According to Hume's theory of causality, we will come to believe it only by making an inference, from an observed constant conjunction (H 183-84),

between, on the one hand, objects of a certain kind, namely the contemplation of actions, and, on the other hand, feelings of a certain kind, namely feelings of approval; and making that causal inference is a matter of reasoning, not of feeling alone. So, if virtue and vice were powers, we could not arrive at moral judgements about them by feeling alone.

But, in the *Enquiry*, Hume explicitly denies that we arrive at moral judgements by feeling alone. Rather, he says,

> *reason* and *sentiment* concur in almost all moral determinations and conclusions. The final sentence, it is probable, which pronounces characters and actions amiable or odious, praise-worthy or blameable; that which stamps on them the mark of honour or infamy, approbation or censure; that which renders morality an active principle and constitutes virtue our happiness, and vice our misery: it is probable, I say, that this final sentence depends on some internal sense or feeling, which nature has made universal in the whole species. For what else can have an influence of this nature? But in order to pave the way for such a sentiment, and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained. (E 172-73)

Hume might well allow that the causal reasoning which Stroud refers to, the reasoning needed to discover that some property of act x caused us to have a feeling of approval towards act x, was just one part of the rational way-paving process which Hume says is often necessary before sentiment can pass its final verdict.

Hume's remarks in the *Treatise* might be cited in objection to this interpretation. Hume there writes,

> To have the sense of virtue, is nothing but to *feel* a satisfaction of a particular kind from the contemplation of a character. The very *feeling* constitutes our praise or admiration. We go no farther; nor do we enquire into the cause of the satisfaction. We do not infer a character to be virtuous, because it pleases:

> But in feeling that it pleases after such a particular manner, we in effect feel that it is virtuous. (T 471)

But I do not think that in saying this Hume means to deny that, in judging an action virtuous, we assume that our feeling of approval is being caused by something about the action which we are contemplating. It is hard to see how Hume could reasonably deny that. If, for example, we thought that some pleasurable feeling we had at the time of contemplating an action was not in any way caused by the action, but was caused instead solely by a good mood that we happened to be in at the time of making the judgement, or by a shot of whisky that we had just swallowed, we would hardly regard the feeling as being a genuine feeling of approval towards the action. Rather, I take it that in this passage Hume is saying that we do not have to enquire into the mechanism by which the thought of the action produces the feeling in us. We do not have to discover exactly what it is about the action and what it is about our human nature that enables the action to cause the feeling in us. That sort of enquiry is, in Hume's view, a philosophical task, the very one which Hume himself is carrying out, not a task required of anyone and everyone who makes a moral judgement.

Another objection to the power interpretation of 'x is virtuous' runs as follows: it is, after all, possible for an observer truly to judge, for example, that an object is red without ever experiencing visual sensations of redness. A man who was blind from birth can be taught truly to judge, for example, that tomatoes are red without ever experiencing sensations of redness but, rather, by making rational inferences from facts about the shape and feel of the object plus the known facts that all objects that have that shape and feel are tomatoes and all tomatoes are red; that is, by reasoning concerning matters of fact, with no colour sensations coming into it. If moral virtue is a power of actions to cause feelings of approbation and to do so in a certain manner, then, by analogy with colour perception, it should also be possible for someone who is, so to speak, morally blind, that is, who, due to some psychological abnormality, is incapable of having normal moral feelings, still to judge that some act is virtuous, and to do so just by reasoning on the basis of facts which he has learned from others. For example, he could have learned that Smith wilfully killed innocent Jones, that that is the very sort of act which has the power to cause feelings of disapproval in normal observers (of which he, the maker of the judgement, is not one), and could therefore infer by reasoning from these facts that Jones performed a vicious act, without ever having any feelings of disapproval himself. But Hume denies that this is possible.[22] So, how could the powers interpretation of Hume's sentimentalism

possibly satisfy Hume's aim of making the having of moral feelings *necessary* for arriving at any and every moral judgement?

One way of replying to this objection would be to deny the first part of the analogy, that is, to deny that a man blind from birth really could judge that some object was red and could do so by reasoning from matters of fact alone, never himself experiencing any colour sensations. Of course, the blind man could be taught to say the words, 'x is red', in certain circumstances; but, never having experienced red sensations, he could not understand, in the fullest sense of 'understand', the meaning of 'red', and therefore could not judge (in the fullest sense of 'judge', which presupposes such understanding) that anything is red. So, too, one could argue, the person with the psychological abnormality who never experiences, for example, feelings of disapproval, could be taught by others to say the words 'act x is vicious' in certain circumstances, but, never having experienced any feelings of disapproval himself, he could not fully understand those words and therefore could not judge (in the fullest sense of 'judge', which presupposes such understanding) any act to be virtuous or vicious.

I think this reply is fair enough, as far as it goes. But that would not be the end of the argument, because the original objection could easily be developed so as to avoid this reply. The objector to the power account of virtue could allow that a man who was blind *from birth* and who therefore *never* in his life experienced colour sensations could not make colour judgements in the fullest sense; but he could then go on to argue that it is possible for a man who is now blind but who did *once* have normal eyesight and who, therefore, now fully understands the meaning of colour words. It is surely possible for him now to judge, for example, that 'this is a red tomato' without now needing to have colour sensations in order to arrive at that particular judgement. If so, then, by analogy, if virtue were a power, it should be possible for someone who is now morally blind, for example, who is psychopathic at the present time, but who had *once* had normal moral feelings thanks to which he can now at least *understand* moral words, to arrive now at the sincere moral judgement that murder is wrong, without having now to experience, or even be disposed to experience, any feelings of disapproval. But surely Hume intends to rule out this possibility. According to Hume, "The vice entirely escapes you [you, the person making the moral judgement], ... till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action" (T 468-69).

I believe that this objection does show that there is something wrong with the powers interpretation of Hume's sentimentalism; not a defect in the account itself, which I think is all right as far as it goes, but rather *something missing from* the account. When criticizing the

simple autobiographical interpretation of Hume's sentimentalism, the simple equation of 'x is virtuous' with 'I the speaker have a feeling of approval towards x', I did mention one point in favour of that interpretation—namely, that it does emphasize the importance of the judgement-maker's feeling in moral judgement, since it does have the desired implication that moral judgements could not be made in the absence of the appropriate feelings or feeling-dispositions had by the maker of the judgement. The problem with the autobiographical interpretation was that, although it makes the feeling necessary for the judgement, it mysteriously changes the topic of moral judgements from the action to the speaker's feelings. By contrast with that problem in the autobiographical interpretation, the difficulty we have just discovered in the power interpretation is that, although it preserves the action as topic of the judgement (the power, though not a characteristic residing wholly in the action, is after all a power *of the action*) nevertheless, it fails to make the subject's feeling necessary to the moral judgement.

The obvious solution to both these difficulties is to get the best of both worlds by combining elements of the two interpretations, that is, by adding an autobiographical element to the powers account. On this combined autobiographical/powers version, 'x is virtuous' would be said to mean, 'act x has the power to give rise to feelings of approval both in my mind—in the mind of me, the speaker—as well as in the minds of all normal people who observe or contemplate x from the common standpoint; and it has this power both because of certain facts about the action (utilitarian facts), as well as because of certain laws of human nature as presently constituted (that is, facts about humanity and sympathy)'.

Each of the two parts of this combined interpretation—the power part as well as the autobiographical part—provides something essential which the other part on its own would leave out: the reference to virtue as a power of a type of action preserves the action as the topic of the judgement. The autobiographical reference preserves the necessity of the feelings or feeling-dispositions had by the person making the judgement.

This interpretation of Hume's sentimentalism does not accord perfectly with the letter of everything Hume says about feelings and moral judgements. No interpretation could do that if only because, taken absolutely literally, Hume's key pronouncements on the subject are inconsistent with one another and with some of his own main aims. However, I think the proposed interpretation comes at least as near as I can get to doing justice to what I understand to be the intentions of Hume's three main sentimentalist claims:

1. The claim that, "when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you [the maker of the judgement] have a feeling or sentiment of blame from the contemplation of it" (T 469).
2. The claim that, "Vice and virtue, therefore, may be compar'd to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in objects" (T 469), but rather powers of objects to affect subjects in certain ways.
3. The claim that virtue is "*whatever mental action or quality gives to a spectator the pleasing sentiment of approbation*" (E 289).

University of Aberdeen

1. Barry Stroud, *Hume* (London, 1977); hereafter cited as "H."
2. David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 2d ed., rev., ed. P. H. Nidditch (Oxford, 1987), 468-69 (hereafter cited as "T").
3. On my interpretation of the 'calm passions' even they necessarily involve feelings (dispositionally). See Shaw, "Hume's Theory of Motivation," *Hume Studies* 15, no. 1 (April 1989): 173-81.
4. Certain familiar passages strongly suggest (a) that Hume makes no distinction between moral judgements, on the one hand, and feelings of approval or disapproval on the other, regarding the former as strictly identical with the latter. For example, his claims that "[m]orality ... is more properly felt than judg'd of; tho' this feeling or sentiment is commonly so soft and gentle, that we are apt to confound it with an idea" (T 470), and that "[t]o have the sense of virtue, is nothing but to *feel* a satisfaction of a particular kind from the contemplation of a character. The very *feeling* constitutes our praise or admiration" (T 471). Other formulations suggest (b) that although feelings are *essentially involved in* the making of moral judgements, they are not strictly identical with them. For example, "The final sentence [that is, moral judgement], which pronounces characters and actions amiable or odious, praise-worthy or blameable; ... *depends* on some internal sense or feeling" in which the language of dependency presupposes non-identity (David Hume, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 3d ed., rev., ed. P. H. Nidditch [Oxford, 1975], 172-73, emphasis added [hereafter cited as "E"]).

More to the point of my "power" interpretation: if we accept Hume's definition of virtue as "*whatever mental ... quality gives to a spectator the pleasing sentiment of approbation*" (E 289), then the moral judgement 'act x is virtuous' would be (roughly) equivalent to the proposition that act x has whatever mental quality gives to the spectator the pleasing sentiment etc. Nothing of that propositional form could (on Hume's non-propositional account of feelings) be strictly identical with a Humean feeling, though it could essentially involve feelings (see above, p. 48, for my account of this involvement).

This consideration taken together with the points that (1) interpretation (b) (above) fits better with the parallel (defended below) between Humean moral judgements and Lockean judgements about secondary qualities, and that (2) interpretation (b) supports a plausible Humean reply to the counterexample discussed earlier (above, p. 33—the case of the daily newspaper), lead me to favour interpretation (b).

5. See Shaw (above, n. 3), 173-81.
6. D. G. C. MacNabb, "Hume, David," in *The Encyclopedia of Philosophy*, ed. Paul Edwards (New York, 1967), 4:85.
7. At T 416.
8. For example, the view defended by A. J. Ayer in *Language, Truth & Logic* (London, 1958), chap. 6.
9. While standard versions of emotivism (for example, Ayer's and Stevenson's) do not do justice to the richness and subtlety of Hume's account of moral judgement (for example, do not capture the variety of roles which Hume assigns to reason in the making of moral judgements), more sophisticated versions of emotivism would essentially depend on recent developments in linguistic philosophy (for example, speech-act theory, the distinction between meaning and perlocutionary force, etc.), and could therefore only support anachronistic readings of Hume's sentimentalism. See W. D. Hudson, *Modern Moral Philosophy* (London 1983), 155-64. I should add that even standard versions of emotivism seem to me too closely associated with the methods of twentieth century linguistic philosophy to be straightforwardly attributable to Hume. That said, emotivism might well be seen as a natural development (rather than an interpretation) of Hume's sentimentalism: had Hume written in the twentieth century he might well have taken the emotivist path. Or perhaps Hume's emphasis on the essential generality of moral judgements (as conferred upon them by the common standpoint from which, necessarily, they are made) as well as his insistence upon their practical, action-guiding character might point more towards

universal prescriptivism: either way leads away from cognitivist elements in Hume's own account (see above, n. 4). (Combining the proposed power account with an emotivist element—in place of the autobiographical element discussed above, p. 48—would lessen the departure.)

10. Hume's not unsympathetic summary of commonsensical arguments in favour of rationalism (in section 1 of the *Enquiry* : for example,"Truth is disputable; not taste" [E 171]; "In every criminal trial the first object of the prisoner is to disprove the facts alleged, ... the second to prove, that, even if these actions were real they might be justified as innocent" [ibid.]; and "a false relish may frequently be corrected by argument and reflection" [E 173]), arguments which he suspects may be "solid and satisfactory" (E 172), suggest that he did think that, in ordinary discourse and at least at the level of common sense, we do tend to attribute truth values to moral judgements (and not just to the factual judgements upon which they are based). The parallel, which I go on to defend, between Humean moral judgements and Lockean judgements about secondary qualities also supports this interpretation (for we do tend to regard secondary quality judgements—for example, 'grass is green'—as truth-valuable), as does Hume's definition of virtue (at T 476; see above, n. 4).
11. John Mackie has argued at length for this very analysis of moral judgements (which he calls the 'error theory') in chapter 1 of his *Ethics: Inventing Right and Wrong* (London, 1980), and has proposed it (under the name 'objectification theory') as a possible reading of Hume's account of moral judgements, concluding that "there are some hints of the objectification theory [in Hume] and it would fit well with much that he does say" (p. 74). Mackie gives a good summary of the main arguments in support of (the intrinsic merits of) the error theory on pages 48-49. A realist attack on projectivism can be found in David McNaughton, *Moral Vision* (Oxford, 1988), 91-94.
12. Stroud gives an answer to this objection (H 185-86) which I shall not go into now.
13. That is, by looking at it.
14. As in 'salty taste' (as distinct from a definition of 'salty' in terms of chemical composition).
15. It may be possible to formulate a version of projectivism based on Hume's account of causation that does not invoke Stroud's problematic simple idea of virtue. But *some* idea seems to be required at this point. For it is difficult to see, in Stroud's words, "how we employ the very feeling in formulating a 'pronouncement' or judgement." And, if it is not a simple idea, what is it that we are

attributing to an action when we say it is virtuous? What, for Hume, does the word 'virtue' mean? The account which I defend defines it as a power. It is not clear what satisfactory answer, if any, a projectivist can give. In David McNaughton's (above, n. 11) words, "desires and feelings [on the projectivist account] are not representational states and it is impossible to imagine how they could be falsely assumed to be so. We are left with no story about how the projection takes place but with the bare assertion that it must do so somehow" (p. 93).

16. In his *David Hume: Common-sense Moralist, Sceptical Metaphysician* (Princeton, 1982), David Fate Norton has traced possible connections—through Francis Hutcheson—between Locke's theory of the secondary qualities and Hume's moral sentimentalism. In Norton's words,

> Hutcheson might well have said something like this: 'Moral perception is essentially analogous to ordinary perception as I understand it. In ordinary perception (physical and anatomical details aside), we have feelings or *sensible ideas*—that is, colors, tastes, or what Locke called the ideas of secondary qualities—and along with each of these sensible ideas we also have ideas of duration and number, the *universal concomitant ideas*. And along with some of these sensible ideas we have ideas of extension, figure, and so forth, the *representative concomitant ideas* or what Locke called ideas of primary qualities. The sensible ideas are not images or representative of things, but signs or marks of them, while the concomitant ideas do represent external reality. When we experience these three kinds of ideas together (for sensible ideas are only logically prior) we have what we call the idea of an object, or knowledge of an object.
>
> 'Moral perception is much the same. It too depends upon nonrepresentative ideas, affections, or feelings which function as the signs of external reality.' (p. 85)

"From his reading of Hutcheson," Norton continues, "Hume might have been led to think that man is able to gain moral knowledge by means of feelings or purely affective states which serve as signs of objective moral reality" (p. 92). The fact that my own power interpretation was arrived at independently of this most interesting historical evidence (the present paper being based on lectures written a year before the publication of Norton's book and uninformed by knowledge of this Locke-Hutcheson-Hume connection), tells, I think, in favour of the power interpretation.

17. This is Hume's view of the outer, objective (as distinct from the inner, psychological) component of causation.
18. See E 419 and 428 for more about these moral abnormalities which Hume thinks are brought about by the influence of superstition and false religion.
19. In view of difficulties with alternative readings discussed below, that, I think, is the best way of taking it.
20. Subjective only in the sense that they are experienced by the subject who makes the moral judgement. Powers are not subjective in the sense of being variable from one normal person to another because, on Hume's theory, all normal subjects will have the same feeling-response to these general characteristics of actions.
21. The characteristic in question will, on Hume's view, involve some form of pleasure or utility (or avoidance of pain or disutility). In an illuminating discussion of Hume's sentimentalism (much of which is in the spirit of my own view), David Fate Norton (above, n. 16) argues that Hume identifies virtue and vice not with powers of actions but with the objective characteristics mentioned above: "I submit," writes Norton, "that [Hume] holds that vice and disapprobation are not identical and that moral qualities are not merely sentiments but, rather, the objective correlates of sentiments" (p. 111). But this goes against Hume's claim that, "The vice entirely escapes you, as long as you consider the object. You can never find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. ... It lies in yourself, not in the object" (T 468-69).

    More generally, while I share Norton's reluctance strictly to identify Humean moral qualities (or moral judgements) with sentiments in the observer—these readings seem to me to overstate the subjectivist, irrealist, and non-cognitivist elements in Hume's account—it seems to me Norton's own reading places undue emphasis on the objectivist, realist and cognitivist elements. The power interpretation seems to me to get the balance right. For the concept of virtue as a power bridges the divide between these conflicting elements and thereby, without giving up any of them, goes some way to reconciling some of the tensions between them. (Norton [above, n. 16, chap. 7] contains an interesting discussion of these tensions.) It also captures as much of what Norton so well describes as "Hume's elusive ontology of morals" as Norton's own emergentist proposal without 'recasting' Hume's account. (For an interesting discussion of this emergentist 'recasting', see Norton [above, n. 16], 116-17n.)

From a position opposite to Norton's, my power account might be criticized for being too even-handed as between the above-mentioned elements; that is, for not being sufficiently sentimentalist to capture Hume's emphasis on the *feeling* of approval as *the* ingredient in virtue of which a power counts as a moral virtue. That, of course, is where Hume's emphasis lies; but more, I think, as a consequence of his heuristic position (as opponent of the pure rationalist claim that reason *alone* is the basis of morals) than of any thoroughgoing subordination of reason to sentiment. (See Norton's [above, n. 16] telling critique of the subordination thesis, pp. 5-6, 9, 17-20, 96-131, 134-35, 147-53, 208-16, 219-34, 305-10.) For a power would no more count for Hume as a genuine moral virtue in the absence of (a), (c) or (d) above (p. 40) than it would in the absence of the feeling condition: that is, all these ingredients are equally necessary conditions in Hume's account of virtue.

22. He denies it by implication at T 470, 471 and 468-69.