



## **Hume on Responsibility**

Lloyd Fields

*Hume Studies* Volume XIV, Number 1 (April, 1988) 161-175.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at

<http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use.

Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact [humestudies-info@humesociety.org](mailto:humestudies-info@humesociety.org)

<http://www.humesociety.org/hs/>

## HUME ON RESPONSIBILITY

## I

For Hume, to hold a person morally responsible for an action is morally to approve of him or to blame him in virtue of the action. Moreover, as he says in the Treatise of Human Nature, "approbation or blame ... is nothing but a fainter and more imperceptible love or hatred."<sup>1</sup>

How must an action be related to a person in order for the person to be held morally responsible for the action? Since to hold a person morally responsible for an action is morally to approve of him or to blame him in virtue of the action, and since moral approval or blame is nothing but a fainter and more imperceptible love or hatred, the first stage in answering this question is to attempt to answer the following more general question: How must an action be related to a person in order for the person to be loved or hated because of the action? This question, in turn, requires one to consider some of the main features of Hume's account of the passions of love and hatred.

One main feature of Hume's account of these passions is that he distinguishes between the object of love and hatred and their causes. Hume says:

As the immediate object of pride and humility is self or that identical person, of whose thoughts, actions, and sensations we are intimately conscious; so the object of love and hatred is some other person, of whose thoughts, actions, and sensations we are not conscious (T 329).

Though the object of love and hatred be some person other than one-self, the object is not, strictly speaking, the cause, that is, the sole and

sufficient cause. For if the other person were the cause, then he would cause both passions in oneself; and since love and hatred nullify one another, neither could ever exist.

The causes of love and hatred are complex and diversified. The ingredients of such causes are as follows: First, there are certain general qualities or circumstances. These may be subdivided as follows: (a) mental qualities: qualities of a person such as virtue, knowledge, wit, tend to produce love; those such as vice, ignorance, dullness, tend to produce hatred; (b) bodily features: beauty, strength, swiftness are conducive to love; their opposites, to hatred; (c) external advantages and disadvantages, such as family position, possessions, clothing, etc., can contribute to the production of love or hatred. But, secondly, these qualities or circumstances, considered in abstraction, cannot cause love or hatred. They must be related to the person who is the object of love or hatred. For example, knowing of a beautiful palace cannot lead one to esteem a particular prince unless the palace is related to him by, for instance, being his property. Thirdly, in order for mental qualities, bodily features, or external advantages or disadvantages to produce love or hatred, they must not only be related to the object of these passions; they must also induce pleasure or displeasure. Without pleasure or displeasure, neither love nor hatred could arise.

How, then, do actions figure in the causation of love or hatred? An action can cause pleasure or pain. But, for Hume, an action's production of pleasure or pain is a necessary but not a sufficient condition for the production of love or hatred towards the person whose action it is. In addition,

the action must be related to the person. But an action is of relatively short duration. Thus, Hume argues:

'Tis not enough, that the action arise from the person, and have him for its immediate cause and author. This relation alone is too feeble and inconstant to be a foundation for these passions. It reaches not the sensible and thinking part, and neither proceeds from any thing durable in him, nor leaves any thing behind it; but passes in a moment, and as if it had never been. On the other hand, an intention shews certain qualities, which, remaining after the action is perform'd, connect it with the person, and facilitate the transition of ideas from one to the other (T 349).

Now, as we have seen, Hume holds that "approbation or blame ... is nothing but a fainter and more imperceptible love or hatred..." (T 614). We should expect, therefore, that Hume's account of how an action must be related to a person in order to excite love or hatred applies to the question of how an action must be related to a person in order to excite approbation or blame. And since to hold a person morally responsible for an action is morally to approve of him or to blame him in virtue of the action, the answer to the question of how an action must be related to a person to excite approbation or blame should provide an answer to the question of the nature of moral responsibility.

Suppose, then, that we take the words 'love' and 'hatred' each to signify a genus of passions and that we regard approbation as a species of love, and blame as a species of hatred. Then we may ask: What, according to Hume, distinguishes the species, approbation, within the genus, love, and what distinguishes the species, blame, within the genus,

hatred? The answer to this question is bound up with Hume's view of the nature of virtue and of vice.

For Hume, it is a mistake to assume that the virtuous or vicious character of a mental quality or trait belongs to it independently of the minds of persons who consider or contemplate the quality or trait. Rather, the virtuous or vicious character of a mental trait, that which makes the trait a virtue or a vice, consists only in the power<sup>2</sup> of that trait to produce 'perceptions' of a certain sort in the minds of persons who consider or contemplate that trait. (I am here using the word 'perceptions' in the broad sense in which Hume uses it, that is, to signify data of immediate or non-inferential consciousness.) However, my general statement of Hume's theory of virtue and vice requires an addendum. According to Hume, sometimes we regard as a vice not a positive trait but the absence of a trait that we consider to be natural or normal. Thus, if a father lacks natural affection for his child, we regard this lack of affection as a vice, even though strictly speaking it is not a positive trait of the father that is vicious but the absence in him of a trait that the normal or average father possesses.<sup>3</sup>

If the virtuous or vicious character of a mental trait consists in its power to produce perceptions of a certain sort, of what sort are the perceptions that are produced? In An Enquiry Concerning the Principles of Morals, Hume says, "It is the nature of and, indeed, the definition of virtue, that it is a quality of the mind agreeable to or approved of by every one who considers or contemplates it."<sup>4</sup>

To say that a quality of mind is agreeable is to say that it produces pleasure. But not all

pleasures are alike. What sort of pleasure does a quality deemed a virtue produce? And, correspondingly, what sort of displeasure does a quality deemed a vice produce?

The pleasure or displeasure in question is one which arises when one considers a mental quality of a person or his character, not from a point of view in which one is concerned with how the quality or character affects one's own self-interest, but from a disinterested, impartial, impersonal point of view. Thus, Hume says, "'Tis only when a character is considered in general, without reference to our particular interest, that it causes such a feeling or sentiment, as denominates it morally good or evil" (T 472).

A mental quality counts as a virtue only because it induces a certain kind of pleasure or displeasure but also because it induces a certain kind of love or hatred, namely, approbation or blame. What differentiates approbation or blame as a species of love or hatred? Approbation or blame is that sort of love or hatred which is induced by the pleasure or displeasure derived from contemplating a mental quality or a character from a disinterested point of view.

## II

I said at the beginning of this article that Hume maintains that to hold a person morally responsible for an action is to approve of him or to blame him in virtue of the action and that approbation and blame are species, respectively, of love and of hatred. I then considered how, according to Hume, an action must be related to a person in order for the action to occasion love or hatred of

that person. Then I went on to examine what differentiates approbation or blame as a species of love or hatred. We should now be able to see how Hume's view of how an action must be related to a person in order to occasion love or hatred of that person is connected with Hume's interpretation of moral responsibility.

In discussing liberty and necessity in Book II, Part III, Section 2 of the Treatise, Hume argues (1) that the doctrine of liberty or chance, that is, the doctrine that voluntary actions are uncaused, is not presupposed by morality and religion, and (2) that, indeed, the doctrine of liberty or chance is inconsistent with morality and religion. For the doctrine in question negates a condition for holding a person responsible. The passage in which this point is argued for is important for the interpretation of Hume's conception of responsibility:

The constant and universal object of hatred or anger is a person or creature endow'd with thought and consciousness; and when any criminal or injurious actions excite that passion, 'tis only by their relation to the person or connexion with him. But according to the doctrine of liberty or chance, this connexion is reduc'd to nothing, nor are men more accountable for those actions, which are design'd and premeditated, than for such as are the most causal and accidental. Actions are by their very nature temporary and perishing; and where they proceed not from some cause in the characters and disposition of the person, who perform'd them, they infix not themselves upon him, and can neither redound to his honour, if good, nor infamy, if evil. The action itself may be blameable; it may be contrary to all the rules of morality and religion: But the person is not

responsible for it; and as it proceeded from nothing in him, that is durable or constant, and leaves nothing of that nature behind it, 'tis impossible he can, upon its account, become the object of punishment or vengeance. According to the hypothesis of liberty, therefore, a man is as pure and untainted, after having committed the most horrid crimes, as at the first moment of his birth, nor is his character in any way concern'd in his actions; since they are not deriv'd from it, and the wickedness of the one can never be us'd as a proof of the depravity of the other (T 411).

Thus, a person is morally responsible for an action only if the action is related to some enduring trait of his character. The basis in Hume's philosophy for this requirement is that, first, an action must be related to an enduring mental trait in order to occasion love or hatred; secondly, approbation and blame are species, respectively, of love and hatred; and thirdly, to regard a person as morally responsible for an action is to approve of him or blame him on account of the action. So, as an action must be related to an enduring mental trait to induce love or hatred, so an action must be related to an enduring mental trait in order for the person to be regarded as morally responsible for the action.

Hume believes that his theory of responsibility is consonant with the principle that "repentance wipes off every crime, especially if attended with an evident reformation of life and manners" (T 412). He purports to explain this principle in terms of his theory of responsibility as follows: "...actions render a person criminal, merely as they are proofs of criminal passions or principles in the mind; and when by any alteration of these

principles they cease to be just proofs, they likewise cease to be criminal" (T 412).

Hume is wrong to hold that repentance and reformation wipe out a past crime, if this means that a person who has repented and reformed is no longer responsible for a past crime.

Suppose, for example, that a man rapes a married woman. In order to avoid embarrassment, the woman does not report the crime. There is clearly here a question of the man's legal responsibility, but I wish to focus on his moral responsibility; in this example the rape is to be thought of primarily as a moral as distinct from a legal offence. Several years later the man undergoes a radical conversion of character and repents of the rape that he committed. One may now think better of the rapist than one did previously. But that does not relieve him of moral responsibility for the offence. For we believe that he ought, if possible, to 'make amends.' Making amends may take the form of offering to help the woman and her family in some way; or, if that is not feasible, the man could try to make amends by contributing time, effort or money to some organization that seeks to assist women who have been victims of rape. But if he were no longer responsible for this offence just because he had reformed his life and manners and repented of his offence, then why should we think that he ought to try to make amends?

If the reformed rapist is still responsible, then this point provides the basis of an objection to Hume's conception of responsibility. I have argued that, for Hume, to hold a person responsible for a negative action is to disapprove of him or blame him in virtue of the action. But in the preceding example, the rapist is now a reformed character. So

I do not disapprove of him now despite the fact that he committed rape in the past and the offence was indicative of his character at the time he committed it. But even though I do not disapprove of him now, I regard him as morally responsible now for what he did then. This point, however, is inconsistent with Hume's conception of responsibility and therefore constitutes an objection to it.

### III

It remains to consider how Hume's view of responsibility is related to his view of personal identity. Hume's discussion of the topic of personal identity in the Treatise, Book I, Part III, Section 6, indicates that he uses the term 'person' in two ways. According to the first use, the word 'person' signifies a succession of related perceptions. Relevant to this use of the word 'person' is Hume's declaration, "I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (T 252). Also relevant here is his statement, "They are the successive perceptions only, that constitute the mind" (T 253). According to the second use, a person is supposedly a thinking, sensible being, who remains numerically the same through time. To use the word 'person' in this way is to apply it to a mere fiction. The identity that we attribute to a whole train of perceptions "is nothing really belonging to these different perceptions, and uniting them together; but is merely a quality, which we attribute to them, because of the union of their ideas in the imagination" (T 260). Belief in the fiction of a

persisting self originates from the resemblance, and consequent confusion, between two actions of the imagination, viz., an association of ideas of distinct but related perceptions, and the contemplation of a continued object.

These two uses of the term 'person' correspond to two different frames of mind in which one may regard oneself and others. The one frame of mind is that of the accurate, reflective philosopher in his study who is concerned with the anatomy and physiology of the mind and not merely with its appearance; the other frame of mind is that of common human nature, in which even the philosopher shares when he has left his study and is attending, for example, to the practical affairs of daily life. In the former frame of mind, one is enlightened, one is able to unmask the fraud that human nature perpetrates on itself. One can discern that oneself and others are nothing but bundles of perceptions. In the latter frame of mind one is a prisoner of illusion. It is in this latter frame of mind that the mirage of a persisting identical self or person hovers before one.

In accordance with these two uses of the term 'person' and these two frames of mind, there are two ways of relating Hume's account of responsibility to his view of personal identity. Reversing the order, let us consider, first, the use of the term 'person' in which the term fails to express an idea but signifies rather the illusory notion of a thinking, sensible being that remains numerically the same through time. In this sense of the term 'person,' it is possible to hold a person responsible for an action, even though the so-called 'person' is only a fiction. For, insofar as one is a prisoner of illusion, one takes the fiction to be a reality.

Now, it is a familiar point that the intentional object of an emotion need not really exist. For example, a traveller in the desert may be overjoyed to see what he takes to be a pool of water but what is, in fact, a mirage. Analogously, an object of love or hatred may be fictional although believed to be real.

Let us leave aside for the moment Hume's discussion of personal identity and consider the following example from the point of view of ordinary language. Suppose that there are two beautiful young women who are identical twins, Alice and Beatrice. I meet Alice, who tells me that her name is Catherine; I ask her for a date, she accepts, we go out together. I meet Beatrice, who knows about my date with Alice, and Beatrice tells me that she, Beatrice, is Catherine. I ask this woman, whom I take to be Catherine, if we could go out on a date again. She accepts, and we go out together. I alternately date Alice and Beatrice over an extended period of time, thinking that I am dating the same young woman. Although, in fact, there are differences in personality between Alice and Beatrice, I take these differences to be different aspects of the same personality. Thus, I find that 'Catherine' is sometimes introspective, sometimes light-hearted. After this relationship has gone on for a while, I find that I have fallen madly in love with Catherine. Now, there is no doubt that I am in love. But, unfortunately, the object of my love is a fiction that I take to be a reality. Note, it is not the case that I love both Alice and Beatrice, for I love only one, and Alice and Beatrice are two. I love Catherine. But Catherine doesn't exist. Catherine is an illusion that has been created by the two sisters.

For Hume, as we have seen, the notion of a person as a thinking, sensible being who endures as the same being over a period of time is based on a mistake and a confusion. Nevertheless, as the previous example shows, a fiction, if it is believed to exist, can be the object of love or hatred. As a prisoner of illusion, I believe that others as well as myself are enduring conscious beings, and these intentional fictional objects, which I take to be real, can be the objects of my love or hatred. Now, approbation or blame is a species of love or hatred. Thus, such intentional fictional objects can be objects of approbation or blame. But to hold a person responsible for an action is to approve of him or blame him on account of the action. Thus, I can hold a fictional object responsible for an action.

Considering now Hume's other use of the term 'person,' one finds that to relate this use to Hume's account of responsibility is problematic. According to this other use, a person is a succession of perceptions some of which are related by the relation of resemblance and all of which are related by the relation of cause and effect. Now in Book II of the Treatise, in discussing the passions of pride and humility, Hume refers to the object of these passions, the self, as a "connected succession of perceptions" (T 277). It would be reasonable to infer, therefore, that the object of love or hatred, another person, is also a connected succession of perceptions. But as my previous example of loving Catherine shows, one can distinguish between the object of love as the lover conceives of it and what the 'object' actually is. To the lover, Catherine is a real person; in reality, nothing corresponds to Catherine except various aspects of Alice and Beatrice. Analogously, in ordinary cases of loving

or hating another person, what the lover or hater believes the object of his love or hate to be is a persisting identical person. And this point is consistent with Hume's contention that the object of such passions, from the enlightened point of view of the philosopher, is a succession of related perceptions. I maintain, however, that a lover cannot himself believe that his beloved is nothing but a succession of connected perceptions.

Suppose that the universe contains nothing but oneself. Then there occurs a perception that does not belong to the 'bundle' that constitutes oneself. This perception is followed by another, and another, and another, and so on, and these perceptions are related by resemblance and/or cause and effect. Suppose one could somehow cognize this series of perceptions. How could love (or hate) of the series of perceptions arise? What is there to love (or hate) about a series of related perceptions?

Now, as I said, for Hume, to hold a person morally responsible is morally to approve of him or to blame him in virtue of the action, and "approbation or blame ... is nothing but a fainter and more imperceptible love or hatred" (T 614). Suppose then that a particular human body exhibited behaviour in the past which caused pain in another person, that this behaviour was caused by an act of will, and that the act of will issued from an intention to cause pain. The intention and the act of will belong to a succession of related perceptions. This succession of perceptions possesses a certain character in that over an extended period of time there recur patterns of perceptions, somewhat as motifs may recur in the performance of a piece of music. The intention to cause pain was indicative of the character in

question in the sense that the intention figured in a recurring pattern of perceptions. But what would it be to disapprove of that which one conceives of as a succession of perceptions on account of an action that issued from an intention and an act of will belonging to a past segment of that succession? It appears to make no sense to hold responsible a succession of related perceptions in virtue of an action that issued from an earlier segment of that succession.

One can regard another person as responsible, i.e., disapprove of him on account of what he has done, insofar as one regards others as well as oneself from the point of view of a prisoner of illusion who believes in a persisting identical self. But disapproval cannot survive the dispelling of such an illusion. Disapproval is a passion. Now, as Hume says, since a passion does not have reference to something else, it cannot be said to be true or false, according as it corresponds, or fails to correspond, to something else (T 415). But a passion can be called unreasonable if it "is founded on the supposition of the existence of objects, which really do not exist" (T 416). The dispelling of the illusion of self is the realization that an enduring self does not exist. Therefore, from the point of view of the enlightened Humean philosopher, it must be unreasonable to disapprove of another (supposed) 'person' on account of what 'he' has done, i.e., to regard 'him' as responsible.

Lloyd Fields  
The University  
Dundee, Scotland

1. David Hume, A Treatise of Human Nature, ed. L.A. Selby-Bigge (Oxford: Clarendon Press, 1960), p. 614. Further references to the Treatise will be cited as 'T' followed by the relevant page number(s).
2. What is meant here by 'power' is, of course, to be explained in terms of Hume's analysis of cause and effect.
3. For the sake of simplicity I shall confine my attention in the following discussion to positive mental traits.
4. David Hume, Enquiries Concerning the Human Understanding and Concerning the Principles of Morals, Second Edition, ed. L.A. Selby-Bigge (Oxford: Clarendon Press, 1961), p. 261n.