

13. HUMEAN MINDS AND MORAL THEORY

SHELDON WEIN

SAINT MARY'S UNIVERSITY

ABSTRACT. Grant that Hume is a contractarian. Justice then arises from more basic features of humans and their circumstances. Among these more basic features from which justice arises Hume includes (in addition to self-interest narrowly construed) the widely held passions of benevolence and sympathy. But it is mysterious why he included them in his contractarian theory for the derivation of justice does not need them, and may even be weaker with them included. This paper suggests that Hume's philosophy of mind, in particular his account of the imagination, forced him to include benevolence and sympathy along with self-interest as the passions on which justice is based.

Hume holds that our passion for justice is not one we simply have but one which arises from other features of our natures and the conditions in which we find ourselves. Put briefly, Hume's account of how humans came to have justice goes like this. Human beings are motivated by various desires. The rational individual is one who adopts the most efficient means available for satisfying her desires. Reason, being the slave of passions, does not tell one what ends one ought to have, but is confined to telling the agent how she may best attain that which she pre-critically desires.¹ Among the passions all (or, at any rate, most) people have are the quiet passions of benevolence and sympathy. A benevolent person is one who feels good upon learning of the success of others. A sympathetic person is one who is pained by their suffering.² Put another way, people are made happy when they see others succeed and are pained when they see others suffer. Now, there are at least two reasons why it is difficult for people to satisfy their benevolent and sympathetic passions. First, these passions are "quiet"; they are apt to be overruled or outweighed by other passions which, being "louder", are more likely to excite people to action. These louder passions, particularly the self-regarding passions, will constantly move people to act in ways which will not fulfill their benevolent and sympathetic impulses. Second, the human situation is such that, in the circumstances of justice, uncoordinated human interaction often has unforeseen negative consequences.³ Thus, even when our benevolent and sympathetic passions are the ones which move us to action, the collective disutilities which result from uncoordinated human interaction conspire to make actions motivated by benevolence and/or sympathy unlikely to have their desired effect.

The result of this feature of human motivation, conjoined with these characteristics of the human situation, is that benevolence and sympathy will only rarely motivate us and that they will even more infrequently lead to the satisfaction of our desire to see others do well and avoid suffering. Most of the time we are likely to be frustrated because we want to see others do well and are seldom able to bring this about. What is needed, then, is an institutional framework to ensure that our benevolent and sympathetic impulses lead to actions which actually help others, and a commitment on the part of each to defend and promote that framework.⁴ Such institutions will only find support if they result in each person (taken severally) being better off than she would have been without those institutions.⁵ The rules of justice are simply those rules which ensure that each member of the community governed by those rules is better off than she would have been had everyone refused to govern his or her actions by those rules. It is rational for each individual to accept these rules (and the coercive measures necessary to ensure that people follow them) because by doing so she will satisfy more of her pre-critical desires than would otherwise be possible. Justice arises because rational, benevolent, and sympathetic individuals interacting in the circumstances of justice need justice as a means of attaining some of their most important goals. Without justice, universally felt passions of benevolence and sympathy will be constantly and systematically frustrated. Our shared commitment to the institutions of justice is a derivative virtue arising from the conjunction of instrumental rationality, the passions of benevolence and sympathy, and the circumstances in which human interaction occurs.⁶

Assuming that the preceding account is roughly correct, we can ask what features of Hume's general philosophic outlook led him to develop this account of justice. My answers to this question will be dogmatic. I will simply indicate which features of Hume's thought I find salient here and hint at how I think they influenced him.

Hume thinks (correctly, I believe) that reason is the slave of the passions.⁷ This is important here because, had Hume thought that reason, by itself, could move us to act, then he might have found, as have so many other philosophers, that reason simply tells us to be just and moves us to so act. With this route for explaining the widespread propensity to be just unavailable to him, Hume was left with but two alternatives. He either had to hold that we are, by nature, inclined to do what justice requires—that one of our passions is the passion to follow the rules of justice—or he had to show how justice arises from more fundamental elements of human motivation. Since Hume specifically denies that we have a passion for virtue, only the second of these alternatives was available to him. Hence, he sought to account for our propensity to do what justice requires of us by reference to some other motive(s).⁸

The second feature of Hume's philosophy that is relevant to this question is his understanding (or, as many would claim, his discovery) of the logical connection argument. Hume realized that causes must be logically independent of their effects. What is particularly important here is his realization that, in principle, anything can be the cause of anything. Thus he could allow for the possibility that the cause of our propensity to be just need not be a simple passion to do what justice requires.⁹ This

leads to another feature of Hume's philosophy, *viz.*, that humans are capable of acquiring new passions. What new motives a person might cultivate or develop as a means of ensuring that other (natural) passions are more frequently satisfied is, for Hume, a purely contingent question.¹⁰ Thus, it was open to him to argue that our propensity to be just results from the need to develop intermediate motives to ensure that more basic passions are not frustrated.¹¹

Hobbes's account of why people do what justice requires is, in some respects, similar to Hume's. Like Hume, Hobbes does not think that people follow the rules of justice simply because they want to be just. Rather, Hobbes accounts for our propensity to follow the rules of justice by arguing that everyone has an overriding desire to stay alive and that justice is needed to ensure that we do not perish. But Hobbes failed to see that humans can acquire new impulses to action. This accounts for the fact that Hobbes needed an external force, the powerful Sovereign, to move people to keep his laws of nature. Hobbes's account of human motivation was so impoverished that he failed to see what Hume was able to see, namely, that internal sanctions could do the same job as the external Sovereign (and our fear of it) does. There are, of course, other features of Hume's general philosophic outlook which influenced his account of justice, but these seem to be the most salient ones.¹²

One odd feature of this account is that Hume included benevolence and sympathy (along with self-interest, narrowly construed) as the basis of his account of the nature of justice. This is peculiar because the account sketched above would have worked just as well had Hume not thought that benevolence and sympathy were widespread human passions. As David Gauthier has impressed upon us all, justice can arise as a result of instrumentally rational individuals interacting under conditions of moderate scarcity, even if these individuals are pre-critically devoid of compassion for others.¹³ Gauthier shows that nontuistic rational maximizers—that is, individuals who take no interest in the interests of others, but are concerned solely to further their own personal welfare—will, in the circumstances Hume describes, come to adopt and support certain principles governing their mutual interaction. The principles adopted are, so Gauthier argues, the principles of justice. Gauthier shows, in effect, that justice is an emergent property which arises from the interaction of instrumentally rational individuals under conditions where cooperation is both possible and necessary (Hume's circumstances of justice) and that it is not necessary to presuppose that such individuals are wholly selfish. Since benevolence and sympathy add nothing to the force of the argument (and since, for certain purposes, they detract a great deal from it), we must ask why Hume included them in his account of justice.

I should add at this point that neither Hume nor Gauthier seems to have noticed that the argument for the emergence of justice works even if we assume that the individuals involved are malevolent and/or sadistic.¹⁴ Such individuals—those who are instrumentally rational and malevolent and/or sadistic, and who realize that cooperation is both possible and necessary—will come to adopt and support certain constraints on their actions. The rules specifying these constraints are rules which Hume and Gauthier would, I suspect, be willing to call rules of justice. Indeed, the very same argument works even if we assume that we have a mixed initial set of individuals, some of whom are benevolent and/or sympathetic, some of whom have only

selfish passions, and some of whom are malevolent and/or sadistic. All instrumentally rational agents can be shown to have a motive to accept and support principles of justice regardless of what particular passions these agents start out with. Roughly, the argument for this claim proceeds as follows.¹⁵ An instrumentally rational individual is one who: [1] perceives contingent facts and makes inferences from those facts, [2] has particular passions which move her to act, [3] feels satisfied when her passions lead to successful action and is frustrated when her actions fail to achieve their ends, [4] can formulate strategies for successful action and for avoiding frustration, [5] can, and does, act in conformity with those strategies which she perceives as most likely to lead to the satisfaction of her passions. The disposition to maximize satisfaction, and hence the disposition to maximize successful action (or to avoid frustration and unsuccessful action), is an unconditional requirement of the instrumental conception of rationality. Clearly, some courses of action are capable of producing more satisfaction than others, and if the instrumentally rational individual perceives this fact, she must pursue those courses of action. The strategies for action which the instrumentally rational individual may adopt fall into two categories. First, and most obviously, she may choose to act on the world in order to get satisfaction. Second, the rational individual may choose to act on herself in order both to cultivate those passions which are easy to satisfy and to rid herself of passions which cannot easily be satisfied.

Suppose that an instrumentally rational individual finds that she is malevolent—that is, she gets satisfaction from perceiving the frustration of others. What should this rational individual do to satisfy her passions? Obviously, she ought to act prudentially on the world to ensure that others suffer, or, if this is difficult, she ought to get rid of this passion. Clearly the second strategy, getting rid of the passion, is the better one. If the rational individual pursues the first strategy, others will adopt the strategy of isolation or fighting the malevolent individual. Since this will not maximize the malevolent individual's satisfaction, she rationally ought to adopt the strategy of eliminating the malevolent passion, a passion which is bound to lead to frustration. But should the instrumentally rational individual seek to become benevolent? If the individual is already benevolent, should she act on this passion or seek to quell or eliminate this passion? If the rational individual does become benevolent—if she learns to delight in the welfare of others—then she gets satisfaction every time others succeed. Since others, being rational, are seeking to succeed and are generally fairly good at it, the benevolent rational individual is going to get a lot of satisfaction out of the efforts of others. Being benevolent is like having unpaid servants: they work and you get some of the benefit. So, the rational individual will seek to make herself benevolent. Furthermore, she will, whenever it is feasible, act on this passion to ensure that others succeed in their chosen projects. In sum, the instrumental conception of rationality (the conception both Hume and Gauthier adopt) recommends that people take steps to rid themselves of their malicious tendencies and recommends that they adopt benevolent passions. Once this is established, one can develop an account of justice along Humean lines.

But, then, why did Hume include benevolence and sympathy in his account when the argument could have been developed just as well without them? One possible answer is that the preceding argument that all instrumentally rational agents must be or become benevolent actually occurred to Hume and that he in fact accepted it,

but found it to be so obvious and trite that he never bothered to mention it. The problems with this answer are too great for it to be acceptable. In the first place the argument is not obvious. In the second place, it certainly is not trite. If it is correct, it has enormous consequences for anyone seeking to reconstruct morality on the basis of minimal assumptions about the nature of human rationality. But, most devastating of all, the answer has, so far as I have been able to discover, no textual support whatsoever. Hume seems to have no interest at all in showing that justice can be derived from any logically possible preference set. Rather, he wanted to show that human beings, as they naturally are, can and will come to develop rules of justice to govern their mutual interaction. This obvious explanation of why Hume included benevolence and sympathy (together with self-interest) in his account of justice seems to be correct. Hume simply observed that everyone has certain tendencies to behave morally. He sought an explanation for this, rejected the idea that for each virtue people have, they have a passion to do what is required in order to have that virtue, and then showed that, given the passions of benevolence and sympathy, he could account for all the other virtues. But why are some virtues emergent and others, the natural virtues, basic? To say of a given virtue that, in fact, it has some source other than our natural passion to be virtuous in that respect does not advance the matter. The question here is not about what characteristics make a virtue derivative. Rather, what is at issue is why Hume thought that, of the virtues people have, some are really derivative virtues. Why did Hume classify the virtues the way he did? What sorting principle did he use? Why, for instance, did he not hold that justice is a natural virtue and claim that benevolence, sympathy, and narrow self-interest are derived from it? As I suggested above, Hume seems to have simply observed that people are, for example, just, but that they are not naturally just. He, as it were, simply saw that people are naturally benevolent and sympathetic and only derivatively just. Put another way, Hume thought that it was a brute fact that people are benevolent and sympathetic, and not a brute fact that they are just.¹⁶

But how does one, especially if one is David Hume, observe that people are benevolent and sympathetic *and* observe that they are naturally so? If Hume could observe that people are by nature benevolent and by nature sympathetic, then, by parity of reasoning, he should not only have been able to observe that the white billiard ball hit the red billiard ball and the red billiard ball moved but also have observed that the white billiard ball hitting the red billiard ball caused the red billiard ball to move. But, Hume holds, we cannot observe causes *qua* causes. All that we observe is certain regularities (constant conjunction, priority in time, *et cetera*) in the behavior of billiard balls. We then postulate certain causal relations to explain this behavior.¹⁷ Similarly, one would think that *all* that one could observe is that people are benevolent, sympathetic, just, and so forth. To also observe that people are naturally sympathetic, naturally benevolent, but not naturally just seems to grant people powers of observation beyond what Hume thinks we can have. Humeans can observe benevolence, but not its naturalness. They can observe people being just, but not the derivative character of justice.

So the question arises again, why did Hume begin his account of justice by presupposing that people are naturally benevolent and sympathetic? I suggest that the answer to this question is not to be found in Hume's moral theory, but in (of all

places) his philosophy of mind. It is my belief that Hume thought that benevolence and sympathy are natural, rather than derivative, because, given the truth of his philosophy of mind, it is necessarily the case that people are benevolent and sympathetic. (I do not claim—indeed I am inclined to deny—that Hume realized this. Hume seems to have thought it was just a contingent fact that people are benevolent and sympathetic and that this fact bore no special significance for his philosophy of mind.) Let us suppose for the moment that Hume's psychological theory, in particular his account of imagination, is correct. Let us further suppose that I decide to do something to you, say, to step on your toe. But first I decide to reflect on the consequences of this proposed action. One consequence is that you will feel pain. In imagining the consequences of my act, then, one thing I will imagine is you feeling pain. Now, on Hume's psychological theory, for me to imagine something is to have a faint copy (an idea) of the experience of that something (the impression of which the idea is a copy). Thus, for me to imagine pain is to be slightly pained myself.¹⁸ As Thomas Nagel observes, "The imagination of the pain of others will itself be painful".¹⁹ Hence, I will be sympathetic to your plight. This sympathy will, unless I have some other strong passion which moves me to step on your toe, be sufficient to make me abandon the strategy of stepping on your toe. Similar reasoning applies to the case of benevolence. I conclude, then, that Hume's philosophy of mind, together with the fact that we can imagine the feelings of others, entails that people are naturally benevolent and sympathetic.

Obviously, Hume thought that we can imagine the feelings of others and that his philosophy of mind was an accurate description of human psychological capacities. Consequently, if the above argument is correct, he had to think that everyone was naturally benevolent and sympathetic. Hence he had no reason for trying to derive benevolence and sympathy. A program like Gauthier's, which attempts to derive compassion for others on the basis of the psychological states that nontuistic rational maximizers would seek to cultivate in the circumstances of (potentially) cooperative human interaction, simply is not needed by Hume because he did not need to derive benevolence and sympathy from rational self-interest. For Hume, these passions are, and must be, ones that everyone has simply in virtue of their having a Humean psychological make-up.²⁰ Consequently, there is no need for Hume to show that they are derived from some more basic features of the human condition. Furthermore, the plausibility of Hume's moral theory serves, albeit only partially and very indirectly, to corroborate one of the most implausible parts of his philosophy, his account of imagination.²¹ □

END NOTES

¹ I assume here, and throughout, that a rational agent has a consistent preference set; she cannot, for instance, prefer pears to peonies and peonies to pears.

² This use of "benevolence" and "sympathy" differs from Hume's. On Hume's (inconsistent) uses of these terms, see Selby-Bigge's Introduction to *Hume's Enquiries*, especially pages xxi-xxx.

3 Not only are the consequences not always foreseen, but even when they are they are not always, to use a term of Sidgwick's, forefelt. See *The Methods of Ethics* (seventh edition), page 110.

4 This does not deny that self-interest is the motive which explains, or accounts for the origin of, justice. See *A Treatise of Human Nature*, Book III, Part II, Section II, page 492.

5 Which institutions will best do this and which institutions one ought to support are very different questions. I ignore the problem here.

6 See David Gauthier, "David Hume: Contractarian", *The Philosophical Review*, Volume LXXXVII, #1, 1979. Outside the *Treatise* and the *Inquiry Concerning the Principles of Morals* Hume frequently makes contractarian sounding remarks. Thus, in his essay "On Suicide", he claims that "All our obligations to do good to society seem to imply something reciprocal. I receive the benefits of society, and therefore ought to promote its interests."

7 *A Treatise of Human Nature*, Book II, Part III, Section III, page 415.

8 *A Treatise of Human Nature*, Book III, Part II, Section II.

9 See J. L. Mackie, *Causation: The Cement of the Universe*, Oxford: Oxford University Press, 1974. Chapter 1.

10 Of course, society (through education) will usually take care of this task.

11 I do not, of course, deny that self-interest plays an important role here. See Thomas Nagel, *The Possibility of Altruism*, Princeton University Press, 1970, Chapter 2, Section 4, for an excellent short account of the logic of Hume's theory of moral motivation.

12 A major difference between Hobbes and Hume stems from the fact Hobbes seems to have been more anxious to use his moral and political views to change the way people act. Hobbes was in the business of *prescribing*, whereas Hume was interested in *describing*.

13 *Morals by Agreement*, Oxford 1986. That Hume may be characterized as a contractarian is argued for in Gauthier's "David Hume: Contractarian" (cited in note #6 above) and in Russell Hardin's "The Strategic Structure of Moral Problems" (unpublished ms. dated November 1983).

14 I use "sadistic" here as the antonym for "sympathetic". Throughout this paper I have ignored the difference between benevolence and sympathy.

15 I am indebted to discussions with Duncan MacIntosh for the ideas in the remainder of this paragraph and in the next paragraph.

¹⁶ Hume's insightful observations (*A Treatise of Human Nature*, Book III, Part II, Section II) about the disparity between the extension of self-love and/or benevolence on the one hand and justice on the other are wholly unconvincing.

¹⁷ For a more complete statement of Hume's account, see Mackie (cited in note # 9 above).

¹⁸ A copy of something need not have all the properties of the original. Thus, my copy of Pissarro's *Boulevard des Italiens* is, unlike the original, not valuable. Similarly, an idea of pain need not be painful. This may be so. But Hume clearly holds that our idea of pain is, like the impression of pain from which it is copied, painful. I thank Randy Carter for pointing this out to me.

¹⁹ Nagel, page 11. Actually, things are more complicated than this. When I imagine you in pain, that idea is like my idea of my being in pain, which is, in turn, like my experience of pain.

²⁰ The "must" in this sentence indicates contingent necessity. Hume's psychological theory is a theory about matters of fact.

²¹ An earlier version of this paper was read to the American Philosophical Association. I thank Nathan Brett, John Deigh, Clarence Johnson, Ann Levey, and Bob Martin for helpful comments. My research was funded by grants from the Social Sciences and Humanities Research Council of Canada.