

## 1 Overview: Issues and Texts

David Hume's moral theory has been immensely influential in the field of ethics. Writing in the eighteenth century and working in the wake of the rise of science, Hume described the mechanisms, or principles, of human nature that account for the moral distinctions we make. His many observations produced a moral philosophy that highlights human psychology – how human feelings and ideas demarcate the content of morality. He emphasized the practical aspect of moral distinctions and appealed to the connection between references to morality and the production of attitudes and actions. A signature view is that morality is ultimately dependent on feelings or sentiments, rather than on reason alone. On his theory, certain human feelings determine which qualities of character are virtues and which are vices. Hume was also interested in describing and categorizing the various virtues and vices and drew conclusions about the common features of each, describing virtue in terms of its immediate agreeableness or its usefulness. His account of morality extended to the rules of justice, which concern institutions like promising and property – how and why we create such rules, and what motivates people to observe them. Overall, he offered a comprehensive view of morality as a naturalistic system dependent on the human mind and on ordinary practices.

Hume's arguments have left a legacy of powerful views that are still defended in current ethical discussions. In metaethics, many sentimentalists, who defend feeling as the ground of morality, trace their roots to Hume's arguments on the psychology of moral thinking (D'Arms, 2005; Prinz, 2007; and Slote, 2010). Humeans about motivation, who maintain that desire is necessary to produce action and that belief alone cannot instigate action, take their position to originate with Hume (Blackburn, 1998; Schroeder, 2007; Sinhababu, 2009, 2017; Smith, 1994; and Williams, 1979). Among virtue theorists, some appeal to Hume's theory of the virtues to defend a spectator version of virtue ethics that can stand without teleological commitments (Driver, 2004; Greco, 2013, 2015; Swanton, 2007, 2015; and Taylor, 2006). At the same time, some philosophers understand Hume as relying on a version of utilitarian theory, given both his classification of virtues as traits useful or agreeable to the self or to others and the key role he accords to the moral spectator's sympathy with the effects of actions on others (Glossip, 1976; Rawls, 1971, pp. 32–33; and Sobel, 1997). That these schools in moral philosophy appeal to Hume as their advocate is clear, but whether all of these views are actually represented in Hume's moral theory is a separate issue.

## 1.1 Framing Questions

Commentators disagree over how to understand various features of Hume's moral theory and the details of his arguments. One question readers raise about Hume's general strategy is whether his empirical method yields a normative ethics, a means of evaluating actions and characters, or whether his project is entirely descriptive. Relatedly, Hume is credited with the famous argument that no conclusion about what we ought to do can follow from claims about what is the case. Yet, at the same time, he uses observations about human psychology to support his claims about morality. Are his claims entirely descriptive? Is he inconsistent, or is the meaning of his warning about "is" and "ought" compatible somehow with his own project? Furthermore, Hume's well-known critique of the view that reason on its own can produce motives to action and serve as the source of our morality has been the subject of debate, both concerning what his argument rules out and whether it is correct.

A controversy among those who believe that Hume has a normative ethics is about what form it takes. As I have noted, his theory has features of virtue ethics, but it also emphasizes the effects of actions on ourselves as agents and on others and points to utility as an important feature of many virtues. Furthermore, Hume's account of justice portrays the rules of justice as derived from mutual self-interest, yet commentators have wondered whether Hume can answer the question why a person should follow the rules in cases when flouting them better promotes that person's individual interest. Moreover, philosophers working in contemporary ethics ask questions about how to classify Hume's metaethical views: Is he an emotivist or a cognitivist? Is he a moral internalist or a moral externalist? He insists that morality is real, but can his own theory support such a claim? What sort of a realist can he be? Among those attentive to the evolution of Hume's thought, the issue of whether he changed his views over time, from one book to the next, is a focal point of debate. Of course, the possibility of a change of heart complicates answering all of the above questions.

My discussion will address these controversies in Hume's ethical theory. I will argue, eventually, that Hume's theory does contain a normative ethics, which he derives from human practices, and does so without violating his dictum that value conclusions cannot follow from factual premises. This requires offering a distinct reading on the relation between facts and values. On that reading, Hume's message is not skepticism about founding morality on natural facts; instead, it constitutes advice about how we discover a connection between the two – namely, by feeling, not by reasoning. Further, among points I will emphasize is that Hume references "moderation" of the sentiments, which

## *Hume on the Nature of Morality*

3

allows that we can take a broad perspective in which our intense personal feelings are muted and from which we attend to generally shared sentiments, rather than to idiosyncratic ones. Thus, we can extract standards of morality from feelings that are not relative to individuals. Morality then is real, not in the sense that it exists in the world, but in the sense that it shares the formal features of reality – consistency and coherence – while it is dependent on human nature. Morality also demonstrates reality’s practical features by its influence on language, communication, and behavior. Furthermore, on any plausible theory of the mind, we follow general psychological tendencies in our responses to the world and to others. Among them are that we regard actions and their consequences, which we can observe, as signs of persons’ character traits, which we cannot observe. I also offer an understanding of Hume’s theory of justice that shows how our natural sentiments of concern for ourselves and for those close to us can be transformed into a concern for the system of justice on the whole and then to an appreciation of rule-following behavior in itself.

### 1.2 Hume’s Chief Writings on Morality

Hume’s discussion of morality first appeared in Book 3 of *A Treatise of Human Nature*. Books 1 and 2 of the *Treatise*, “Of the understanding” and “Of the passions,” were published in 1739. Book 3, “Of morals,” followed in 1740. When it was released, the *Treatise* was a flop, to Hume’s great disappointment. It was meant for public consumption, but it was largely disregarded.<sup>1</sup> Hume later wrote a reformulation of his views on morality in his *Enquiry concerning the Principles of Morals* (1751) (often called “the second *Enquiry*”).<sup>2</sup> As I mentioned in the previous section, differences between the second *Enquiry* and *Treatise* texts have led some scholars to argue that Hume changed at least some of his views in the intervening years. Since Hume was only in his twenties when he wrote the *Treatise*, some regard it as a book of Hume’s youth and see the second *Enquiry* as offering his more considered, mature theory.<sup>3</sup> But there is by no means agreement about whether Hume’s later work signaled a change in content, or whether it was just a change in style. Furthermore, even if Hume did alter some of his views, the *Treatise* is still cited as an important source of

<sup>1</sup> In Hume’s brief autobiographical essay, “My Own Life,” published after his death, he writes, “I had always entertained a notion, that my want of success in publishing the *Treatise of Human Nature* had proceeded more from the manner than the matter, and that I had been guilty of a very usual indiscretion, in going to the press too early. I, therefore, cast the first part of that work anew in the *Enquiry concerning Human Understanding*, which was published while I was at Turin” (Hume, 1777 [1987], p. xxxv, italicization of titles added).

<sup>2</sup> He also presented a revised version of his views on the understanding in an *Enquiry concerning Human Understanding* (1748), which is often called “the first *Enquiry*.”

<sup>3</sup> Taylor (2008) is a representative.

Hume's ethical theory, especially since it covers topics and arguments that are not dealt with in the second *Enquiry*. Some commentators also prefer the more analytic approach of the *Treatise* and have looked to it for immensely influential arguments concerning, among other things, motivation to action and the relation between facts and values.

Several of Hume's essays, which appeared at various times between 1741 and 1757 and are collected as *Essays: Moral, Political, and Literary* (1777) (abbreviated here as *Essays*), are also relevant to his moral theory. This is especially so, given the moral theory's relation to Hume's political views and to his views on how we judge matters of taste, two topics that Hume treats in the *Essays*.<sup>4</sup>

I draw from the essays and from both the *Treatise* and the *Enquiry concerning the Principles of Morals*, noting some relevant differences as I go along.<sup>5</sup> The disparities between the two books have mostly to do with Hume's account of the psychological mechanism of our moral approvals and disapprovals, which I discuss later (see Sections 3.1.1, 3.1.2, 3.2.1, and 3.2.2). Because of this, the *Enquiry* emphasizes the universality of morality in a way that the *Treatise* does not, and perhaps cannot.<sup>6</sup> However, the broad contours of Hume's philosophy are the same in the two works. He firmly rejects the moral rationalism and the theological systems of writers of his day like Ralph Cudworth and Samuel Clarke, the egoism of thinkers like Bernard Mandeville and Thomas Hobbes, and the social contract theory of John Locke.<sup>7</sup> He defends sentiment-based morality grounded in human nature. He characterizes virtues as traits that are either useful or agreeable to other people or to the agent, eschewing the "monkish" virtues like humility and meekness.<sup>8</sup>

## 2 Metaethics: Hume's Case against Moral Rationalism

### 2.1 The Motivation Argument

In the *Treatise*, Hume situates his ethics within his general empiricist theory of the mind, which divides mental contents into impressions (vivid, lively, forceful experiences) and ideas (the less vivid and less lively thoughts that copy the impressions). Demonstrative reason works with ideas to discern relations

<sup>4</sup> For a discussion of themes in Hume's essays, see Watkins (2019).

<sup>5</sup> I use "T" to indicate references to the *Treatise*, followed by book, part, section, and paragraph numbers. I use "EPM" to indicate references to the *Enquiry concerning the Principles of Morals*, followed by section and paragraph numbers.

<sup>6</sup> Two recent volumes that are excellent resources on the *Enquiry* are edited by Taylor (2020) and by Kroeker, Engels, and Lemmens (2021).

<sup>7</sup> Hume's view on the origin of society is also contrary to Rousseau's. However, even though they knew one another (and had what turned out to be a tempestuous relationship), there is little evidence of intellectual influence between the two. See Popkin (1978).

<sup>8</sup> For an in-depth look at Hume's life and his works, see Harris (2015).

among them, and these relations constitute universal truths. Some moral rationalists writing in the seventeenth and eighteenth centuries argued that morality could be understood by reason in its demonstrative function, since, they contended, it consisted in immutable truths applicable to all rational beings. Hume's debate with these philosophers is expressed in terms of his question whether morality is determined by ideas or by impressions. Further, Hume writes, "In order, therefore, to judge of these systems, we need only consider, whether it be possible, from reason alone, to distinguish betwixt moral good and evil" (T 3.1.1.4). Hume's famous "Motivation Argument," in reply, says,

Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason alone, as we have already prov'd, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason.

No one, I believe, will deny the justness of this inference. . . . An active principle can never be founded on an inactive; and if reason be inactive in itself, it must remain so in all its shapes and appearances, whether it exerts itself in natural or moral subjects. (T 3.1.1.6–7)

In the first section of the second *Enquiry*, Hume describes an argument for sentimentalism analogous to one he offers in the *Treatise* (although at that point, he does not commit himself to it, but later it is clear that it is his):

The end of all moral speculations is to teach us our duty; and, by proper representations of the deformity of vice and beauty of virtue, beget correspondent habits, and engage us to avoid the one, and embrace the other. But is this ever to be expected from inferences and conclusions of the understanding, which of themselves have no hold of the affections, or set in motion the active powers of men? They discover truths: But where the truths which they discover are indifferent, and beget no desire or aversion, they can have no influence on conduct and behaviour. (EPM 1.7)

The Motivation Argument, then, says:

- (1) T version: Morals influence actions and affections. That is, they excite passions and produce or prevent action. (Morality is active.)

EPM version: The end of moral speculation is to teach duty and engage us to pursue virtue and avoid vice.

- (2) T version: Reason alone never influences actions and affections. (Reason is inactive.)

EPM version: Inferences and conclusions of the understanding (reason) discover truths, but do not beget desires and aversions and so do not influence conduct.

- (3) T version: An active principle can never be founded on an inactive one. (Implicit in EPM.)
- (4) Conclusion: Therefore, morals, or the rules of morality, are not conclusions of our reason. (Implicit in EPM.)

I consider each premise in turn in the next three subsections, taking them in reverse order.

### 2.1.1 Active and Inactive Principles

Premise (3) says that an active principle can never be founded on an inactive one.<sup>9</sup> With the development of the mechanical philosophy of the seventeenth century, philosophers debated how to account for motion. Descartes argued that while movement can be imparted from one body to another, the quantity of motion in the universe, initiated by God, was constant. Other philosophers argued that motion was increased by or originated in such events as chemical reactions and the exercise of the human will. Pre-Newtonian scientists and experimental philosophers, as well as Newton himself, appealed to active principles or “occult” qualities to explain these events and make them compatible with voluntarist theology (Henry, 1986, pp. 358–66).<sup>10</sup> Such theology accords to God the power to create anything, including value. (In fact, if God can do anything, the only way to understand changes in the world is experimentally.) Hume argues, contrary to the Cartesians, that if we cannot conceive an active power in matter, then we cannot conceive it in our idea of deity, since all ideas have the same source – namely, impressions (T 1.3.14.10). Hume also avoids reference to occult qualities, however (e.g., T 1.4.3.8; *Dialogues* 4.12).<sup>11</sup> So, active principles, for Hume, are states that effect changes in other mental or

<sup>9</sup> While Hume writes that “Any thing may produce any thing” (T 1.3.15.1), he does not mean that anything *does* cause anything. His point is that we need experience to correlate effects with their causes. So premise (3) is established empirically.

<sup>10</sup> Among the seventeenth-century natural scientists Henry names as believing in the necessary existence of active principles in matter are the English thinker Walter Warner, who represents the first attempt to develop a thoroughly mechanical philosophy, Walter Charleton (following Pierre Gassendi), and Matthew Hale (Henry, 1986, pp. 340–43). Robert Boyle, Robert Hooke, Marie Boas Hall, William Petty, and Isaac Newton are among many early moderns who appealed to various sorts of occult qualities to explain motion in an atomistic universe (Henry, 1986, pp. 344–51).

<sup>11</sup> I use “*Dialogues*” followed by part and paragraph number, to refer to Hume’s *Dialogues concerning Natural Religion* (1779).

physical states, but they are not explained by reference to hidden qualities or powers. They are manifested in our experience.<sup>12</sup>

We know from experience that some passions, such as gratitude and envy, often push us to action, and these we call “motives.” Passions, for Hume, are impressions, not of sensation, but of reflection. They are lively and forceful experiences created when persons have thoughts of, or their mind returns back to, ideas of objects that were pleasurable or painful. When I look through an arch of tree branches at a green pasture at sunset, I experience, among other things, impressions of greenness and of pleasure (impressions of sensation). When I think about the pleasurable vista, I might experience a feeling of peacefulness or a desire to return to that spot (impressions of reflection or passions). Not all passions provide impetuses to action, on Hume’s view; for example, pride does not, since it is a reactive feeling toward something pleasant connected with the agent, “a pure emotion in the soul, unattended with any desire” (T 2.2.6.3), but benevolence does, since it involves a desire for the well-being of another (T 2.2.9.3).

According to Hume’s premise (3) above, the passions that are motives are either motivating in themselves or derived from other active states. “Instincts,” like benevolence and kindness to children, are nonderived passions, active in themselves, while other motives, he says, are derived from the prospect of pleasure or pain: “’Tis from the prospect of pain or pleasure that the aversion or propensity arises toward any object: And these emotions extend themselves to the causes and effects of that object, as they are pointed out to us by reason and experience” (T 2.3.3.3). He implies, then, that the prospect of pleasure or pain has the power to motivate, which has prompted some commentators to infer that Hume’s view is that beliefs about future objects’ effects on an agent will produce motivating states. Thus, on their view, beliefs are active.

There is an alternative understanding of Hume’s claim about the origin of motivating passions, which I think is better for a number of reasons that will become apparent when I discuss premise (2). The reading I have in mind says that the active force of derived passions comes not from belief, but from certain sentiments of pleasure and pain felt in the past. My thinking of the delight of a cool drink on a hot summer day inspires desire because I have had an experience of the pleasure of drinking a cool liquid on a hot day before.

<sup>12</sup> “The New Hume” debate centers on the question whether Hume thought that unperceivable qualities of nature underlie the empirical phenomena that cause us to believe in causal connections (see Reid and Richman, 2014.) My assertion that Hume does not posit occult active powers assumes a side on this issue, but one that is well-defended by several scholars, among them Millican (2009) and Winkler (1991). Among those who defend a “realist” Hume who posits underlying properties are Strawson (2011), Kail (2007), and Wright (1983).



I desire the features of the *object* (the drink, not just the pleasure) due to having experienced its pleasantness in the past. Observation and inference then produce the belief that another object, which I have not experienced, has similar features. Since I desire those features, this object holds the prospect of pleasure, and I develop a desire for it. Thus, desire (or some other passion) originates from the impression of pleasure I've experienced in the past and gets transferred to future objects via reasoning. The desire, however, isn't arising from a belief by itself. My previous impression of pleasure generates a desire for things that are like the one that gave me the original pleasure. When I discover objects with such characteristics, my desire is directed to them. Thus, on my interpretation of Hume on motivation, pleasure and pain – rather than beliefs about prospective pleasures and pains – are active (or activating) sentiments that produce motivating passions (Radcliffe, 2018a, pp. 51–53). I say more about this reading in the next section.

### 2.1.2 *Inert Reason*

Hume's premise (2), the thesis that reason alone does not influence actions and affections, has been impactful both in the history of philosophy and in contemporary philosophy. Ironically, some philosophers think its influence in current philosophy has been based on a mistaken reading of Hume's intent. Contemporary Humeans about motivation hold the view that both belief and desire are necessary to produce an action and take Hume's argument that reason alone does not motivate to be the ground for their theory. Whether Hume is advancing their view depends on what Hume means by "reason."

In defending premise (2) of the Motivation Argument, Hume refers readers back to the *Treatise* Book 2 discussion of motivation, where he originally defended his claim about the impotence of reason. He argues that reason manifests itself in two functions: (1) demonstration, which concerns relations of ideas, and (2) causal reasoning, which concerns objects of experience and the comparison of ideas representing those objects. Demonstration involves chains of relations, each of which is recognized intuitively; Hume says that since demonstration concerns ideas, and the will concerns reality, the two are "totally remov'd from each other." Mathematics, which Hume recognizes as a demonstrative science, is useful in "almost every art and profession," but it does not by itself have any influence on the will and action.

Mechanics are the art of regulating the motions of bodies *to some design'd end or purpose*; and the reason why we employ arithmetic in fixing the proportion of numbers is only that we may discover the proportions of their influence and operation. . . . Abstract or demonstrative reasoning, therefore,



*Hume on the Nature of Morality*

9

never influences any of our actions, but only as it directs our judgment concerning causes and effects. (T 2.3.3.2)

Thus, without an established end, mathematical operations of the understanding have no practical purpose. For example, math is useful to a chemist in determining the dose of a drug needed to treat, but to do no harm to, a person of a certain weight. The mathematical calculations do not produce action, however, without the goal of treating a patient, which is supplied by a desire-like state. Although Hume's Motivation Argument is initiated to rebut the rationalist idea that demonstrative reason accesses the content of morality, he extends the premise that reason alone does not influence actions and affections to causal or probabilistic reasoning as well. He argues that the process of identifying causes and effects in the world is not sufficient to produce action. The impulse toward an object does not arise from causal reasoning; rather, we are guided by it to reach an end we have an interest in. "It can never in the least concern us to know, that such objects are causes, and such others effects, if both the causes and effects be indifferent to us" (T 2.3.3.3).

The dispute over the meaning of Hume's argument centers on whether "reason" should be treated as equivalent to "belief." A belief for Hume is an idea that has acquired force and vivacity in such a way that it feels different from ideas that are only imagined. Beliefs, on a standard reading of Hume, are generally the products of causal reasoning. Such reasoning, which has custom at its heart, uses experiences that we associate with each other by habit so forcefully that our idea that the first object produced the second is boosted to the level of belief.<sup>13</sup> This is the nature of beliefs. So, the traditional way of understanding the thesis about the impotence of reason is to say that beliefs alone do not produce actions or affections. This makes sense, because if beliefs are object-oriented, in the sense that they are about things (in the mind or the world) we take to exist independently of our thinking of them (T 1.4.2; T 1.1.1.11; T 1.3.8.15–17), then beliefs without a passion for something would leave us unmotivated. Thus, it looks as though Hume is advancing a theory of motivation whereby both desires (a subset of the passions) and beliefs about how to fulfill them are necessary to move us.

Many recent commentators have rejected that reading, arguing that nothing in Hume's case about the impotence of reason commits him to the thesis that beliefs do not motivate. Prominent among the considerations that lead to their saying this is Hume's reference to the prospect of pain or pleasure as the source

<sup>13</sup> I say that beliefs are generally the product of causal reasoning, because sometimes beliefs are produced in other ways. For instance, if I am disposed in certain ways to be excited by the idea of fantastic phenomena, I may acquire the belief that there are space alien visitors among us, when such ideas and images are presented to me in a lecture.

of our passions, which they take to refer to beliefs. The thesis that beliefs produce motives (passions) and actions also seems to be supported by Hume's discussion of the influence of belief on behavior in Book 1 of the *Treatise*, in a section called "Of the influence of belief" (T 1.3.10). I have, however, given an account of the formation of motives by prospective pleasure (Section 2.1.1) that does not attribute the force to belief. And, in other places, I have argued that the influence of belief on behavior is always for Hume in the presence of passions or desires (Radcliffe, 1999, pp. 104–12). The proponents of the view that beliefs, for Hume, do motivate, have varying readings of the second premise of the Motivation Argument, among them: (1) beliefs about good and evil in objects motivate because they contain representations of pleasure and pain (Kail, 2007, pp. 181–82); (2) the process of reasoning does not motivate, but beliefs – the outcome of the process – can (Cohon 2008b, pp. 73–77); (3) non-inferential beliefs can motivate, but reason cannot (Pigden 2009, p. 97); (4) belief is a sentiment, not a representation, so it can motivate action, even if reason, understood in some other way, cannot (Sandis 2012, pp. 206–8; Stroud 1977, pp. 158–61); (5) the faculty of reason does not motivate, while beliefs can (Owen 2016).

However, a large consideration that militates against all of these readings has to do with the fact that Hume argues that reason "alone" does not influence passions and actions. He agrees that reason does contribute to motivation, presumably in conjunction with a motivating passion, which serves as the initiative to action. So, the sense in which Hume thinks reason alone does not influence actions or affections must allow that it can nonetheless contribute to or assist in motivation. It is hard to see a way in which reason as a faculty or as a process contributes to motivation when it is not understood as belief. If we distinguish reason as a capacity or faculty – reason as the process of reasoning – and reason as a belief or product, the one sense in which reason alone does not motivate, but can contribute to motivation, is reason as belief. Otherwise, Hume is guilty of equivocation between senses of reason when he says that it cannot produce motives or actions, and yet allows that it can contribute to their production. Furthermore, the EPM version of the argument refers to "inferences and conclusions of the understanding," which "discover truths," rather than produce motivating states. The focus is on the outcome of reasoning, rather than on the capacity or on reasoning as a process. Moreover, there are good grounds for understanding beliefs, not as sentiments, but as ideas with a sentimental aspect, but I cannot offer all of the arguments for that reading here.<sup>14</sup>

<sup>14</sup> See Radcliffe, 2018a, pp. 65–88; for in-depth arguments against the various interpretations of Hume's thesis that reason alone does not motivate, see pp. 29–64.