Article:
Chamberlain, J. (2026) Hume on characters, virtues, and “durable principles of the mind”. Hume Studies, 51 (1). pp. 9-33. ISSN: 0319-7336

https://doi.org/10.1353/hms.2026.a989191

Hume on Characters, Virtues, and "Durable Principles of the Mind"

James Chamberlain, University of Sheffield

Forthcoming in *Hume Studies*

Please cite final version

Correspondence: School of History, Philosophy and Digital Humanities, University of Sheffield, 45 Victoria Street, Sheffield, S3 7QB, United Kingdom. Email: j.p.chamberlain@sheffield.ac.uk

Abstract: Hume is widely understood to believe that all virtues and vices are "durable principles of the mind," and that durable principles of the mind are character traits. Several scholars therefore read him as a virtue ethicist. I argue that we should reject all such interpretations. I argue that Hume allows that some virtues and vices are simply single perceptions, such as a motivationally strong desire to help a stranger or to murder someone. Therefore, I argue, we should not read him as a virtue ethicist. Hume is, however, a reductionist about character, who understands a character trait as simply a long-lasting but interrupted succession of perceptions. On my interpretation, a durable principle of the mind is an uninterrupted, relatively short-lived succession of different perceptions within an actor's mind, of a kind that mainly interests Hume because he thinks it enables observers to associate a person's action with her mind.

Introduction

When Hume first introduces his theory of moral evaluation, he claims that to "approve of a character is to feel an original delight upon its appearance" (T 2.1.7.5).[1] He often states that we approve or disapprove of characters.[2] Hume calls the approved parts of a character "virtues," and the disapproved parts "vices" (T 2.1.7.3). He also argues that we never morally evaluate actions *as such*:

> If any *action* be either virtuous or vicious, 'tis only as a sign of some quality or character. It must depend upon durable principles of the mind, which extend over the whole conduct, and enter into the personal character. Actions themselves, not proceeding from any constant principle, have no influence on love or hatred, pride or humility; and consequently are never consider'd in morality. (T 3.3.1.4)[3]

Here, Hume is widely understood to argue that virtues and vices *are* "durable principles of the mind" (henceforth just "durable principles"). Furthermore, durable principles are widely understood as *character traits*. For example, Paul Russell claims that "when Hume is speaking of virtues and vices, he is referring to specific traits of character which are, for him, 'durable principles of the mind'."[4] On any such interpretation, "the moral qualities of an agent which arouse our moral sentiments must be 'durable or constant'."[5]

Hume is thus widely read to understand virtues and vices as character traits, where a trait is, as I will henceforth assume, a relatively long-lasting feature of a person.[6] Perhaps unsurprisingly, therefore, several scholars read Hume as a virtue ethicist.[7] In this paper, however,

I will argue against all such interpretations. To do so, I will initially set aside Hume's discussions of durable principles, and consider his accounts of characters and virtues with fresh eyes. I will argue that he understands virtues, character traits, and characters as follows.

Virtues and vices are, fundamentally, just different kinds of "perceptions" (T 1.1.1.1). Any token perception of any generally useful or agreeable type will cause approbation in anyone who contemplates it, and so constitute a Humean virtue. For example, if we encounter someone whom we believe to possess a single, short-lived benevolent desire then we will feel approbation, since benevolence is very typically useful and agreeable. The desire will therefore be a virtue.

Hume never uses the term "character trait," or even the word "trait."[8] Nevertheless, he understands what *we* would call a "character trait" as a series of perceptions. Drawing on recent work by Hsueh Qu, I will argue that a Humean character trait is nothing more than an interrupted succession of a certain type of perception in a person's mind over time.[9] As such, a trait can exist only intermittently, whenever the relevant perceptions exist. My interpretation thus rejects the claim, endorsed by Qu and some others, that Hume is committed to the thesis that a trait can have an uninterrupted existence.[10]

Like the individual perceptions of which they are composed, traits can be virtues or vices. For example, a benevolent trait will be a virtue. It will consist of an interrupted succession of benevolent desires in an individual mind over time.

What Hume calls a "moral character" is, I will argue, ultimately that set of perceptions which consists of all and only those perceptions in a person's mind that are apt for moral evaluation (T 3.1.1.12; T 3.3.3.2). However, Hume recognizes that we can, and do, morally evaluate people for just those parts of their characters that we know about.[11] He thinks we will approve of a person for (what I will call) her *synchronic moral character*, if we believe that her

mind at a certain time includes one or more perceptions of some useful or agreeable kind(s). We will approve of her for her *diachronic moral character* if we believe that her mind includes a succession of such perceptions over time. Of course, Hume allows that people experience perceptions of many kinds, so that we may evaluate a person as virtuous in some respects and vicious in others. We might approve of someone as we watch him help a friend in financial distress, while disapproving of him for his generally imprudent attitude towards money.

Even so interpreted, Hume clearly *does* understand many virtues as character traits. This fact, along with Hume's frequent discussions of "character," might explain why he is so widely believed to understand *all* virtues as character traits. He writes approvingly of Cleanthes, for example, whose diachronic character seems to consist purely of virtuous traits, like kindness, cheerfulness, and industriousness (EPM 9.2). Yet Hume also allows that some virtues are simply short-lived perceptions. In his view, *anyone* who experiences a benevolent desire thus possesses a genuine virtue, at least for a moment, even if she is otherwise utterly selfish in her desires.

One implication of this, I will argue, is that Hume's theory is significantly different from that of any typical virtue ethicist's. Another is that he cannot think that all virtues are "durable principles of the mind"; at least, not if the term means "enduring character traits."

I will therefore revisit Hume's discussions of durable principles, starting with those in Book 2 of his *Treatise*. I will argue that, *regardless* of the details of his theory of moral evaluation, Hume does not appear to understand durable principles as character traits of *any* kind. Instead, a durable principle is an uninterrupted succession of different but causally related perceptions, of a kind that anyone who performs an intentional action will experience *at and around the time of action*. Durable principles can include various sense impressions, passions, and ideas, but they *always* include the perception that caused the action.

Furthermore, despite Hume's name for them at T 3.3.1.4, I will argue that durable principles need not endure for *very* long. From start to finish, they may last for only a few minutes. Certainly, this is a peculiar usage of the term "durable." Nevertheless, by calling these successions of perceptions "durable," Hume means only that they are long-lasting *in comparison with actions*.

For example, any benevolent desire will be caused by a series of perceptions, including various feelings, particularly love, and various ideas, like a belief about the best way to please one's loved one. If the desire is acted on, then it will cause a further series of perceptions, including beliefs about its effects, feelings of sympathetic pleasure, and so on. This succession of different perceptions, which centers around and includes the benevolent desire, is a durable principle of the mind.

In Book 3, Hume argues that our forming a belief about an actor's durable principle is a *necessary condition* for us to morally evaluate that person for her action. However, I will argue, he does not believe that durable principles are the *objects* of our moral evaluations. As I read it, Hume's argument at T 3.3.1.4 can be summarized as follows: Before we can sincerely call any action "virtuous" or "vicious," we must habitually associate that action with a durable principle: a succession of causally related perceptions in the actor's mind, which will have started before she performed the action, and which will continue to exist for some time after the action. Actions occur too quickly for us to *directly* associate them with the minds of those who perform them, but we *will* associate a person's action with her mind if we infer that it was caused by a durable principle. Only then can we experience any love, hatred, approbation, or disapprobation for the actor.

So understood, T 3.3.1.4 tells us little about the objects of our moral evaluations. Hume never argues that any virtues or vices *are* durable principles, or that *all* virtues or vices are character traits. I will therefore conclude that he consistently allows that virtues and vices can be either single perceptions or character traits. His theory of moral evaluation implies that, whenever we believe that a person's mind includes at least *one* token perception of any useful or agreeable type, we will approve of the person for possessing that perception.

1. Hume on characters and virtues

For now, set aside Hume's discussions of durable principles. Consider instead his claim that the mind, or psychological self, is just a "bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" (T 1.4.6.4). As Terence Penelhum observes, Hume thinks that our idea of the self *just is* the idea of a mind, or "that succession of related ideas and impressions, of which we have an intimate memory and consciousness" (T 2.1.2.2).[12] Hume certainly recognizes that we believe ourselves to be embodied. Nevertheless, he treats a "thinking person" as identical with a "mind" (T 1.4.6.17).

Hume stresses that "nothing is ever present to the mind but its perceptions" (T 3.1.1.2). He understands a perception as a mental object, which, like any "object," can exist only so long as it remains "invariable and uninterrupted" (T 1.4.6.6). However, Hume argues, "there is no impression constant and invariable" (T 1.4.6.2). Since he believes that ideas change even *more* rapidly than impressions (T 2.3.9.12), he can confidently assert that *nothing* within the mind "remains unalterably the same, perhaps for one moment" (T 1.4.6.4).[13]

Given this theory of the mind, Qu argues, any Humean character trait must ultimately reduce to an interrupted succession of "fleeting and perishing" perceptions (T 1.4.2.20).[14] This is because Hume also believes that a causal power can be nothing more than the *exercise* of that power (T 2.1.10.4). This implies that a disposition to X can be nothing more than "X"s actual occurrences—namely, its regularly occurring."[15] Hume is therefore committed to understanding a benevolent disposition, for example, as *nothing more* than an interrupted succession of short-lived benevolent desires in a mind, which occur at frequent or regular intervals over time. Roughly following John Bricke, I will call this the "reductionist theory" of character.[16]

As Qu observes, Hume implicitly distinguishes the "synchronic self," or that bundle of perceptions which constitutes the mind at a single time, from the "diachronic self," or that "connected succession of perceptions" which constitutes the mind over time (T 2.1.2.3).[17] Hume can therefore understand (what we would call) a character trait as simply a part of a person's diachronic self. Indeed, at one point, he insists that there are "such dispositions as benevolence and generosity," but then immediately describes these as "sentiments" with "causes, effects, objects, and operations, marked by common language and observation" (EPM App2.6).[18] He thus appears to reduce these virtuous dispositions to the perceptions involved, just as Qu argues he should do.

However, Qu also understands Hume's "durable principles of the mind" as durable character traits. He therefore agrees with Bricke and Jane McIntyre that Hume *cannot* coherently be a reductionist about character, since durable character traits must "endure even when not exercised."[19]

McIntyre argues that Hume is not a reductionist about character, by arguing that his description of a calm passion as "a settled principle of action" implies that *token* calm passions

"typically endure and can exist without being exercised" (T 2.3.4.1).[20] However, as Qu and Russell both argue, Hume appears to mean just that a calm *type* of passion can be a settled principle of action.[21] Furthermore, Qu observes that Humean passions are always *felt*, so that they cannot have "any significant durability, since they are not felt for meaningfully lengthy durations."[22] I therefore agree with Qu that Hume is committed to the reductionist theory of character.

Qu concludes from this that Hume's theory of character is ultimately incoherent.[23] Yet, before accepting this conclusion, we should ask whether Hume really *does* understand durable principles as durable character traits. Indeed, we have a further reason to doubt this. As we have seen, Hume's account of durable principles is widely read to imply that *all* virtues and vices are character traits. However, there are good reasons to think that Hume would reject this conclusion.

For one thing, Hume officially defines a virtue as "*whatever mental action or quality gives to a spectator the pleasing sentiment of approbation*" (M App.1.10).[24] This definition includes no mention of characters or character traits. Instead, it explicitly states that a mental "action" can be a virtue. Hume has previously insisted that all mental actions are perceptions: "The mind can never exert itself in any action, which we may not comprehend under the term of *perception*" (T 3.1.1.2). He has also claimed that any *physical* action will be "produc'd and annihilated in a moment" (T 2.2.3.4). He presumably believes this to be true of mental actions too – indeed, he uses very similar language to describe perceptions. Therefore, Hume's *definition* of a virtue implies that a single, short-lived perception can cause us to feel approbation, such that it will be a virtue.

Hume similarly defines the closely related notion of "personal merit" without reference to character: "PERSONAL MERIT consists altogether in the possession of mental qualities, *useful* or *agreeable* to the *person himself* or to *others*" (EPM 9.1; see also T 3.3.1.30). He believes that all moral evaluation is ultimately grounded in the sentiments of approbation and disapprobation.[25] Combining the two theses, he argues that we will "approve of a person, who is possess'd of" any such "qualities" (T 3.3.1.29). Despite this, some scholars read Hume to argue that we morally evaluate characters but *not* persons.[26] However, while Hume often claims that we approve of characters, he also often states that we can "blame" or "approve" of persons.[27] During his discussion of durable principles, he describes the moral sentiments as "sentiments concerning … [a] person" (T 3.3.1.5).

It seems, therefore, that persons and mental actions are no less important than characters or character traits for Hume's moral theory. Moreover, mental actions can be virtues, and all mental actions are perceptions. Hume also sometimes claims that we approve of mental "qualities"[28] or "sentiments"[29] *as well as* "characters." It is far from obvious that "qualities" should be understood as character traits. Even more importantly, "sentiments" are presumably *passions*, and Hume understands all passions as perceptions (T 2.1.1.1).

Yet Hume's discussions of virtues and vices often *do* focus on characters. Unfortunately, he never explains what he means by a "character." We have seen, however, that a Humean character must ultimately reduce to some set or succession of perceptions. Perhaps Hume simply believes that a person's character is *identical* with her mind, and so with her as a "thinking person." However, he appears to distinguish a person from her "moral character" (T 3.1.1.12; T 3.3.3.2). He also understands a "praise-worthy character" as one "consisting of all the most amiable moral virtues" (EPM 5.10). I therefore suggest that a Humean *moral character* consists

of *just* those perceptions in a person's mind that are apt for moral evaluation. These will include various passions, for example, but presumably no sense impressions.

Furthermore, Hume's implicit distinction between the synchronic and diachronic mind suggests a similar distinction between a person's synchronic moral character and her diachronic moral character. A synchronic moral character will be a set of morally evaluable perceptions in a person's mind at a certain time. A diachronic moral character will be a succession of morally evaluable perceptions in a person's mind over time.

Hume appears to make just this distinction when he claims that a person's chances of showing "strength of mind" will vary "according to the *general* character or *present* disposition of the person" (T 2.3.3.10). In Hume's view, to exhibit strength of mind, as when we appear to act on reason rather than passion, is to act on our calm passions rather than on our violent passions. Here, I suggest that Hume is arguing that, if the present arrangement of perceptions in a person's mind – her "*present* disposition" – involves strong, calm desires that are appropriately related to other perceptions, then she will likely exhibit strength of mind.[30] Moreover, if she is in good habits, then her "*general* character," or that succession of perceptions which constitute the morally evaluable parts of her mind over time, will have made it likely that she will act on her calm desires.

Admittedly, Hume is not fully consistent in his use of these terms. For example, although he elsewhere uses the phrase "present disposition of our mind" (T 3.3.1.16), he also sometimes uses "disposition" synonymously with "general character." This seems to be his meaning, for example, when he notes that a "person of an obliging disposition" can sometimes give an uncharacteristically "peevish answer" if they have a toothache, or if they are hungry (EHU

8.15).[31] Nevertheless, at T 2.3.3.10, Hume appears to distinguish a synchronic from a diachronic character.

Hume at least sometimes appears to suggest that we can approve of a person *just* for her synchronic character. For example, he thinks that we will be "struck with any signal instance of GREATNESS of MIND or Dignity of Character," and he gives the example of "the famous silence of AJAX in the ODYSSEY, which expresses more noble disdain and resolute indignation, than any language can convey" (EPM 7.4).[32] This seems to imply that we will approve of Ajax for his dignified indignation, regardless of *any* beliefs we might have about his diachronic character.

Hume also states that "the benevolent or softer affections are ESTIMABLE; and wherever they appear, engage the approbation, and good-will of mankind" (EPM 2.1). This suggests, for example, that if we see one passing stranger cheerfully help another passing stranger, then this alone will be sufficient to engage our approbation.

Furthermore, Hume appears to think that we can disapprove of *vicious* synchronic characters. Consider his famous discussion of a "[w]ilful murder," in which he attempts to provide an exhaustive list of the morally relevant facts in a particular case (T 3.1.1.26). Aside from the murderous action, Hume claims, those matters of fact which might even *potentially* constitute a vice include "only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case." If Hume were to think that we can *only* disapprove of people for possessing vicious character traits or dispositions – or, indeed, vicious durable principles – then he surely ought to have explicitly and unambiguously referred to these here. However, he does not. *Prima facie*, this passage suggests that a single, short-lived, but motivationally strong desire to murder can constitute a vice, even if the murderer never previously wanted to harm anyone,

and never will again. It suggests that we can take someone to be vicious *just* by disapproving of her synchronic character.

Finally, if Hume were to believe that all virtues and vices are character traits, then he would be committed to the claim that *all* intentional actions of morally evaluable kinds must reveal or stem from traits of character. However, as Russell argues, any such claim would be highly implausible.[33] As A.E. Pitson observes, Hume believes that "our characters are, to some extent, 'inconstant and irregular'," and so he must allow that people will *sometimes* act out of character.[34] Hume cannot *rule out* that someone who is generally disposed not to feel a certain desire might experience that desire at some point. In Hume's view, even the most generally hard-hearted person *might* be moved to pity, under certain conditions. Even a mild-mannered person might commit a crime of passion. Moreover, we can, and do, morally evaluate people for acting on such uncharacteristic passions.

So, Hume appears to allow that, in at least some cases, a single perception, such as a motivationally strong desire to help a stranger or to murder someone, can be a virtue or a vice. He also appears to believe that we approve and disapprove *of* persons, *for* their possession of certain characters, such that we will approve, roughly, of anyone whose synchronic or diachronic character includes at least one useful or agreeable perception.

More precisely, Hume thinks that we will approve of anyone who possesses at least one token perception of any type that is *typically* useful or agreeable. This is implied by his argument that we will approve of anyone who *wants* to help her friends or country, even if she cannot do so:

> Where a person is possess'd of a character, that in its natural tendency is beneficial to society, we esteem him virtuous, and are delighted with the view of his character, even tho' particular accidents prevent its operation, and incapacitate him from being serviceable to his friends and country. (T 3.3.1.19)

We will approve of *anyone* who wants to help her friends, because any such desire will be a benevolent one, and because benevolence tends to be useful and agreeable.[35] Similarly, Hume argues, we will approve of anyone who justly repays money, even in those rare cases where doing so will cause only misery, and so be "pernicious in every respect" (T 3.2.2.22). Justice *tends* to be very useful, and so it is "a moral virtue, merely because it has that tendency to the good of mankind" (T 3.3.1.9).[36]

Hume also thinks we can morally evaluate persons for *not* possessing perceptions of kinds that people typically possess. He observes that parents typically feel affection for their children, of a kind which causes desires to care for their children. We therefore habitually expect parents to care for their children, and so, Hume argues, we will "blame a father for neglecting his child. Why? because it shews a want of natural affection, which is the duty of every parent" (T 3.2.1.5). If a father fails to feel affection for his child, and so acts neglectfully, then this may well be a *moral* failure, even if he never intended to act neglectfully, and even if he is generally affectionate. So, some Humean vices, and presumably some Humean virtues, are constituted by the *absence* of perceptions.

I therefore take Hume to be committed to the following thesis: If we believe that a person possesses at least *one* perception of any type that tends to be useful or agreeable (or lacks at least one perception of a type that people typically experience and which tends to be harmful or

disagreeable), then we will, *ceteris paribus*, feel at least some approbation for that person, even if we know nothing about her character traits.

On this interpretation, Hume's moral theory entails – somewhat trivially – that a virtue must be a part of a character, since a Humean character is just the morally evaluable part of a person. Less trivially, it entails that our moral evaluations center on persons and virtues, rather than actions, their (particular) consequences, or moral rules. Is this a sufficient reason to count Hume as a virtue ethicist? I believe not, for the following reasons.

Virtue ethics is a normative theory. As Rosalind Hursthouse observes, virtue ethicists believe that a right action is "simply, what a virtuous agent would, characteristically, do (or have done) in the circumstances," and they understand a virtuous agent as "one who has, and exercises, those character traits that are the virtues."[37] As Christine Swanton stresses, they generally understand a virtue as "a good, excellent, or admirable character trait."[38]

Hume's arguments mainly concern metaethics and moral psychology. Nevertheless, he argues that we *should* moralize as we are naturally inclined to do (T 3.3.6.3; EPM 9.15–25). He thinks that we (naturally do, and so should) evaluate a person's action as morally right if, and only if, we believe that it was caused by a perception of some typically useful or agreeable kind. A right action is, simply, one so caused. Furthermore, some Humean virtues are not character traits of *any* kind. Hume does not, therefore, argue that "the primary object of moral evaluation is a trait of character," *or* that a "good action" is "what the virtuous agent (one with this trait) would do."[39] We might well agree with Don Garrett that Hume incorporates elements of virtue ethics into his sentimentalist moral theory.[40] Nevertheless, I believe that any attempt to understand Hume *as* a virtue ethicist risks missing, or misreading, much that is interesting and distinctive about his theory.[41] His theory, as I understand it, defies simple categorization.

However, since my interpretation is incompatible with the claim that all Humean virtues are character traits, we cannot accept it without first addressing Hume's arguments concerning "durable principles of the mind." In the next section, I will argue that, *regardless* of the details of Hume's theories of virtue or character, his Book 2 account of durable principles suggests that he does not understand them as character traits.

2. Durable principles in Book 2

Hume first introduces durable principles during his discussion of love and hatred, where he addresses "some difficulties, concerning particular causes" of these passions (T 2.2.3.1). He returns to durable principles when he discusses "liberty" and "necessity" (T 2.3.2.6). In this section, I will address his aims in each of these discussions, along with some implications for our understanding of durable principles.

Hume uses "love" to mean, roughly, any positive feeling that we experience towards a person, if our feeling is caused by the person's being related to something that is independently pleasing (T 2.2.1.6). He stresses that the object of love or hatred can only be a "person or thinking being" (T 2.2.6.4; T 2.2.1.7), or a "thinking person" (T 2.2.1.6).

According to Hume, love and hatred – as well as pride and humility – are produced via a kind of process that he calls a "double relation of ideas and impressions" (T 2.1.5.5). Hume's account of this process is notoriously complicated, but the crucial point is as follows: It is *only* if our idea of a person is closely associated with an idea of something that pleases or pains us that we can feel love or hatred for that person. For example: "A prince, that is possess'd of a stately palace, commands the esteem of the people upon that account; and that *first*, by the beauty of the palace, and *secondly*, by the relation of property, which connects it with him" (T 2.2.1.5). The

palace pleases the people, the people associate the palace with its owner, and so the people feel love, or "esteem," for the prince. If the people did *not* closely associate the palace with the prince, then the palace could never make them feel any love for him, no matter its beauty.

Here, we must note that Hume sees property as "a particular species of *causation*; whether we consider the liberty it gives the proprietor to operate as he please upon the object, or the advantages, which he reaps from it" (T 2.1.10.1). To understand this, we must briefly consider his theory of causal inference.

According to Hume, we can only make causal inferences once we have come to associate our ideas of two (or more) objects by custom or habit, typically by having frequently seen one kind of object immediately followed by another. If we habitually associate any two ideas, then we will automatically form the one idea whenever we form the other idea. So, Hume thinks, if we have frequently seen people freely use and enjoy the objects that we call their "property," then we will habitually associate the practice with the term, and *vice versa*. Moreover, Hume argues, when any idea is habitually associated with any present *impression*, then some of the "force and vivacity" or "liveliness" of that impression will transfer to the idea (T 1.3.8.2). The idea will thus be enlivened, and so become a *causal belief*: "A LIVELY IDEA RELATED TO OR ASSOCIATED WITH A PRESENT IMPRESSION" (T 1.3.7.5). For example, if a newspaper article states that a certain palace is the "property" of a prince, then our visual impressions of this newspaper will cause us to habitually form the lively idea of the prince freely using and enjoying the palace. We will henceforth habitually associate our ideas of the prince and the palace. And, Hume argues, if the palace is aesthetically pleasing, then this will cause us to feel love for the prince.

Before T 2.2.3, Hume's account of love and hatred had suggested that we will simply love or hate anyone "in proportion to the pleasure or uneasiness we receive from him" (T 2.2.3.2). However, he now recognizes "some difficulties" with this claim, since we clearly hate people significantly *more* if they intentionally hurt us than if they do so by accident. Yet Hume's causal theory of love and hatred appears unable to account for this, because an accidental harm can cause as much pain or "uneasiness" as a deliberate one.

So, Hume introduces his account of durable principles to help explain *why* we feel less love or hatred for anyone who accidentally causes us pleasure or pain than for those who, we believe, *intentionally* caused that pleasure or pain. He begins by making "a distinction":

> If that quality in another, which pleases or displeases, be constant and inherent in his person and character, it will cause love or hatred independent of the intention: But otherwise a knowledge and design is requisite, in order to give rise to these passions (T 2.2.3.4).

Hume's only examples of "constant and inherent" qualities are "deformity" and "folly" (T 2.2.3.4). He does not say what he *means* by "constant and inherent." However, he typically calls a mental quality "inherent" if he thinks we cannot remove it by any effort of the will.[42] Hume therefore appears to have in mind qualities that we cannot remove at will, and which may endure over a whole lifetime. He also appears to think it a simple matter to explain why we might hate people with displeasing "constant and inherent" qualities. If we have repeatedly seen someone say foolish things, for example, then we will *strongly* associate the idea of that person with the idea of foolishness. Therefore, we will typically think of foolishness, by habit, whenever

we think of the person. Furthermore, since foolishness is a displeasing quality, we will feel at least *some* uneasiness whenever we think of it (T 2.2.8.4). We will therefore experience the kind of double relation of ideas and impressions that produces hatred.

So, for Hume, the simplest causes of love and hatred are mental or physical qualities that endure for a very long time, such that we will very closely associate the qualities with the persons who possess them. Having briefly explained this, Hume attempts to explain why we feel more hatred for someone who *intentionally* harms us than for anyone who "wounds and harms us by accident" (T 2.2.3.3). He first argues that we cannot hate anyone *just* because their action harms us. By "action," Hume appears to mean simply a physical behavior. As we have seen (in section 1), he believes that any action will be "produc'd and annihilated in a moment." Therefore, he argues, an action must lack the kind of causal "relation" that can "connect… [the] action sufficiently with the person" to produce hatred:

> '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 is as if it had never been. (T 2.2.3.4)[43]

Note that someone can be the "immediate cause and author" of an action *without* it being related to her mind, or "sensible and thinking part." Any action must be caused by a person's body, but we need not understand it to be caused by her mind. Moreover, any action, unlike a constant or inherent quality, will be insufficiently durable for an observer to *directly* associate it

with a person's mind. Unfortunately, Hume does not say what he means by "durable." However, Johnson's dictionary of 1755 has the primary definition as: "Lasting; having the quality of long continuance." This seems to reflect Hume's usage in this passage. His point seems to be that, unlike *very* long-lasting qualities such as folly and deformity, which we will closely associate with the thinking persons who possess them, an action can last only for a "moment," so that we cannot directly associate it with the person who performed it.

So, if Hume is to explain why we hate those who *intentionally* harm us, then he will need an account of some object, set of objects, or process that is involved in intentional (but not unintentional) action, and which lasts longer than an action, such that it lasts long *enough* to allow us to associate the harm so caused with the mind of the actor. This account must also be consistent with his "bundle" theory of the mind.

It is in this context that Hume introduces durable principles, which he initially describes simply as "certain qualities." He writes, "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 2.2.3.4). Donald Ainslie understands these "qualities" as *character traits*:

> People who perform the same action with the same intention might nonetheless merit different moral responses; it depends on how the intentional action fits into the person's life as a whole. And thus Hume concludes that we must look at the "mental qualities" or character traits from which our actions as well as our intentions spring.[44]

However, even setting aside the details of Hume's moral theory, I do not think that such "qualities" can be plausibly understood as character traits. My first reason for this is simple but, I think, important. In the following passage, Hume claims that "the principal part of an injury is the contempt and hatred, which it shews in the person, that injures us; and without that, the mere harm gives us a less sensible uneasiness." He appears to mean, very plausibly, that our "uneasiness" and "mortification" will be caused by the person's contempt and hatred *for us* (T 2.2.3.5). Presumably, anyone who feels enough hatred or contempt to *intentionally* injure us will continue to feel such passions "after the action is perform'd," at least for some time. Our ideas of these passions will therefore help "facilitate the transition of ideas" *from* that of the action *to* that of the person. If we think of the person's action, we will be *very* likely to think of the hatred that they feel for us, and thus of their mind. Therefore, it is simply these *occurrent feelings*, rather than any *traits* of contempt or hatred, that appear to form the most important parts of the "quality" that is shown by the actor's intention.

My second reason relates to Hume's later discussion of liberty and necessity. There, he claims, again, that we can only feel hatred (or anger) for someone who performs any "criminal or injurious actions" if their actions have some "relation to the person or connexion with him" (T 2.3.2.6). Hume now uses this claim to argue against the "doctrine of liberty or chance" (T 2.3.2.6). In his view, the doctrine of liberty asserts that all actions occur purely by "chance," without *any* cause (T 2.3.1.18). Notably, they have *no* mental causes: they "proceed not from some cause in the characters and disposition of the person, who perform'd them" (T 2.3.2.6). Therefore, Hume argues, the doctrine of liberty implies that a person who performed a wicked action must be blameless, for "as… [the action] 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" (T 2.3.2.6).

Hume thus presents his opponents with a dilemma. They can either claim that our actions have *no* mental causes, and so reject any notion of moral responsibility, or they can accept that our actions are caused by something "durable or constant," and so allow that we are morally responsible for our actions. Hume presses them to choose the second option.

For this argument to be persuasive, Hume surely cannot mean to imply that the "durable or constant" causes of our actions must all be long-lasting *character traits*. His argument relies on the implicit claim that the *only* alternative to allowing that durable principles cause our actions would be to allow that our actions have *no* mental causes. So, if he were to insist that *all* durable principles are character traits, his argument would fail. As his opponents would surely realize, he would have offered no reason to think that our actions might not be caused by short-lived psychological "principles" instead of long-lasting ones.[45]

Perhaps Hume simply failed to notice this problem with his argument. However, I think this unlikely, not least because he repeats it in the first *Enquiry*.[46] I suggest that he is better understood to allow that durable principles can endure for only a relatively brief time. Clearly, they last longer than any action. However, since Hume thinks that actions pass in a "moment," he may not think they last very long. They are merely "durable" and "constant" *enough* to approximate the kinds of "constant and inherent" qualities – like foolishness or deformity – which we very strongly associate with their possessors. They need only last long enough to make us believe that actions have mental causes.

In summary: Hume believes that we can only feel love or hatred for a "thinking person" if we associate her mind with some pleasing or displeasing object or quality. He thinks it obvious

that we will associate *very* long-lasting qualities with the persons whose qualities they are. In contrast, however, he thinks that actions cannot last long enough for us to directly associate them with actors' minds. Therefore, he argues, we will feel *no* hatred for an actor, no matter how much her action may hurt us, if it is merely her action and its effects that displease us.

Nevertheless, Hume argues, a pleasing or painful action *can* cause love or hatred, if and only if it indicates the presence of a mental "quality" that has remained "after the action is perform'd," and that is of a kind that will "connect" the action "with the person, and facilitate the transition of ideas from one to the other." Hume therefore classifies these "qualities" – durable principles – among the "durable or constant" causes of love and hatred, seemingly for two reasons. First, because he believes that they affect our passions in very similar ways to folly, deformity, and other very long-lasting qualities. Second, because he wants to contrast them with actions, which he believes to be too short-lived to have any significant effect on our passions. I have argued, however, that Hume does not believe that they are *very* long-lasting. He does not understand them as character traits. So, how *does* he understand them?

3. Hume's understanding of durable principles

I suggest that durable principles are best understood as follows:

Durable principle of the mind:	A causally related succession of different perceptions within an actor's mind, the existence of which an observer can infer from her action, and which will include the perception that proximally causes the action.

A durable principle will typically, but not necessarily, include an intention to produce the *consequences* of the action. For Hume, an intention is a particular kind of perception: it is an "actuating passion or motive, capable of producing the action" (T 3.2.5.6). Moreover, Hume understands passions as "reflective impressions," which are typically caused by identifiable, prior perceptions (T 2.1.1.1). As Elizabeth Radcliffe observes, he thinks that "we can identify passions in two ways: phenomenally, by how they feel, and structurally, by the circumstances that attend them, including their causes and effects."[47]

For example, benevolence *feels* a certain way, but it is *also* recognizably the kind of passion that is caused by an idea of a loved one's potential pleasure, such that it constitutes a desire for that pleasure to occur (T 2.2.6.5). A similar passion is pity: "a concern for… the misery of others, without any friendship … to occasion this concern" (T 2.2.7.1). Hume argues that we can "pity even strangers, and such as are perfectly indifferent to us," if we sympathize with their misery (T 2.2.7.1). In the *Treatise*, he argues that sympathy involves an idea of another person's passion becoming enlivened into "the very passion itself," such that we will typically *feel* miserable if we see a stranger looking miserable (T 2.1.11.3). This sympathetic feeling will, under certain conditions, cause a desire for the stranger's misery to be alleviated. And this desire is the passion of pity.

Generally, Hume thinks that we can often infer which perceptions cause which others, and which passions cause which actions. He thinks it "obvious" that we *want* our loved ones to be happy, for example, but also that we can never "hate any [person] without wishing his misery," at least when the *idea* of his potential misery is "presented by the imagination" (T

2.2.6.5). Therefore, he would allow that, if we see someone *act* in a seemingly hateful manner, then we will habitually associate his action with precisely these kinds of passions and ideas.

Say Jean intentionally hurt David. David will probably infer from Jean's behavior that she *wanted* to hurt him. If so, then he will infer that her mind will have contained various related passions, like "contempt" or "hatred," both before and "after the action is perform'd." He will also infer that Jean's mind will include, and have included, various *other* related perceptions, such as a belief about the most effective way of harming him, a pleasure at the fulfilment of the desire, memories of the desire, and so on. This causally related series of perceptions, or *durable principle*, will not last anything like as long as a quality of foolishness or deformity. However, it will last *long enough* to allow David to associate the harmful action with Jean's mind.

The key point is this: By inferring that Jean's action was caused by a *desire* to hurt him, David will habitually form a lively, complex idea of a durable principle *that is a part of her mind*, and that is causally related to both Jean's action and David's pain. David will, therefore, habitually associate his pain *with* Jean's mind (and *vice versa*). Therefore, David will experience the kind of double relation of ideas and impressions that will cause him to feel hatred for Jean.

According to Hume, an inference of this kind is "necessary, in order to produce some relation, and connect … [the] action sufficiently with the person" for the action to cause love or hatred (T 2.2.3.4). Furthermore, Hume thinks that we *cannot* make any similar inferences if we believe that actions were accidental, so that we will treat any accidental action as *merely* physically caused by a person. Presumably, he thinks that we cannot easily form beliefs about people's minds from behaviors that accidentally cause us pain. This seems at least somewhat plausible. For example, if I tread on your foot without realizing it, then you can infer little about my desires or other mental states, other than that I wanted to take a route that involved placing

my foot on the part of the road where your foot was. Therefore, in Hume's view, you will not closely associate your pain with my mind, and so a double relation of ideas and impressions cannot easily occur. You will think that my *body* caused you pain, but not that *I* – a "thinking being" – caused you pain.

Later, Hume will summarize this argument by claiming that people "are not blam'd for such evil actions as they perform ignorantly and casually, whatever may be their consequences … because the causes of these actions are only momentary, and terminate in them alone" (T 2.3.2.7). However, he allows that some *non-moral* hatred will occur in such cases, because there will be a "very small relation" between the pain and the actor's mind (T 2.2.3.6). If I accidentally tread on your foot, then you will associate your pain with *one* of my desires, at least; my desire to step on that part of the road where your foot was. You will, therefore, feel a weak passion of hatred. However, in Hume's view, you will not *morally* blame me, because you will not infer that my mind includes any perceptions of any generally harmful or disagreeable kinds. And, as we have seen, Hume thinks that we *only* disapprove of people if we believe that their minds include at least one such perception.

In short, Hume argues that we will only feel any strong hatred for a person who hurts us if we habitually associate their *mind* with our pain. He *also* thinks that, if a person accidentally hurts us, then we will not closely associate their mind with their own action, and so we will not associate their mind with our pain. In contrast, if we infer that an action was caused by a *motive*, then we will habitually form further beliefs about various other perceptions in the actor's mind: enough, at least, that we *will* associate this mind with our pain. Such successions of perceptions, which exist at and around the time of action, are what Hume will later call "durable principles of

the mind." In the next and final section, I will argue that his Book 3 discussion of them is merely intended to apply his Book 2 argument to the *moral* sentiments.

4. Durable principles in Book 3

Just before *and* after his discussion of "durable principles of the mind," Hume stresses the similarities between the moral sentiments and love and hatred, seemingly describing the moral sentiments as "that love or hatred, which arises from mental qualities" (T 3.3.1.6; see also T 3.3.1.3). He later describes them as "nothing but a fainter and more imperceptible love or hatred" (T 3.3.5.1). I therefore agree with Páll Árdal that Humean moral sentiments are produced via double relations of ideas and impressions.[48]

Hume claims that, "when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles in the mind and temper" (T 3.2.1.2). As elsewhere, his point is that we only approve of people for possessing certain mental qualities – in this kind of case, *motives*. At T 3.3.1.4, Hume draws on his Book 2 account of durable principles to offer a brief explanation of this point. He argues that we can *only* treat an action as "virtuous" or "vicious" if it is a "sign of some quality or character" that *included* the motive which caused the action. At least, I will now argue, this is what he means to argue for when he discusses "durable principles of the mind."

As we have seen (in section 2), Hume has already argued that, although all actions are physically caused by persons, this "relation alone is too feeble and inconstant" to make us associate actions with persons. Therefore, actions *as such* cannot cause us to feel love or hatred – or, indeed, pride or humility. Hume now implies that we should apply this argument to the moral sentiments: "Actions themselves, not proceeding from any constant principle, have no influence

on love or hatred, pride or humility; and consequently are never consider'd in morality" (T 3.3.1.4).

Hume has also argued that, for any action to cause us to feel love or hatred, something is needed to *connect* the action with the actor's mind, such that we will closely associate the two. Durable principles play precisely this role, as we have seen. Love and hatred thus *depend* on durable principles, and so too must the moral sentiments. Therefore, any evaluation of an actor *as* virtuous or vicious "must depend upon durable principles of the mind, which extend over the whole conduct, and enter into the personal character." Recall, a durable principle will exist for some time before *and* after an action (it will "extend over the whole conduct") and it will form a *part* of the actor's mind (it will "enter into the personal character").

So, Hume's point at T 3.3.1.4 is that we can *only* morally evaluate an actor for her action if the action indicates the presence of a durable principle in her mind: one that included the (pleasing or displeasing) motive behind the action. This is because we cannot directly associate her action with her mind, and because the moral sentiments – like love, hatred, pride, and humility – can only occur to us if we associate a person's mind with something pleasing or displeasing.

However, it is not immediately obvious what Hume means by "quality or character." He repeats this phrase in the following passage, where he argues that, to understand the "origin of morals," we must never consider any action as such, but "only the quality or character from which the action proceeded. These alone are *durable* enough to affect our sentiments concerning the person" (T 3.3.1.5). So, the phrase does not refer to a *single* kind of object, since Hume claims that "these" things are "*durable* enough to affect our sentiments."

I suggest that the kind of "quality" here mentioned is a durable principle: that kind of quality which, Hume argued at T 2.2.3.4, will be shown by an "intention," and which will last long enough to "connect" an action with a person. However, I take "character" to refer – in these two passages, at least – to (what we would call) a character trait. Hume's point, I suggest, is that *either* durable principles *or* character traits can enable us to associate an action with a person.

We have seen (in section 1) that the reductionist account of character entails that a character trait can be durable *only* in the sense that it can be a long-lasting but interrupted succession of perceptions. Nevertheless, actions can "proceed" from durable characters, so understood. For example, whenever Cleanthes performs a generous action, its proximal cause will be one of the many generous desires that together constitute his long-lasting trait of generosity. Furthermore, if we know Cleanthes well and we see him act generously, then – Hume thinks – we will associate the action and its effects with Cleanthes's trait of generosity, and so with *him*.

Yet durable principles play a more fundamental role than character traits: our moral evaluations of actors "depend upon durable principles of the mind." *Whenever* we infer a motive from an action, Hume thinks, we will do so by inferring the presence of a related durable principle. In some such cases, we may *also* believe that the actor possesses a relevant character trait. Yet, in other cases, we may have no such beliefs.

Say we see a stranger's hand enter her pocket, pull out a £20 note, and hand it to a homeless man. This action may certainly please us. However, to *approve* of this person for so acting, Hume thinks, we must associate the stranger's action with her mind, and infer that it was caused by a motive of some useful or agreeable kind. Perhaps we noticed her concerned expression when she saw the homeless man, and so inferred that she experienced a sympathetic

sadness. We can then infer that she experienced a *piteous desire*, and that this desire caused her action. We will habitually make further inferences about her perceptions in the recent past: that her sympathetic feeling was caused by her idea of the man's sadness, that this was caused by her visual perceptions of him, and so on. We can also make inferences about her present perceptions: that she approves of her own behavior, that she feels pleased that the homeless man is pleased, that she still feels some pity, and so on.

On my interpretation, it is this *causally related succession of perceptions* that will comprise the durable principle that will "extend over the whole conduct" of the piteous person. Once we have inferred its existence in her mind, we will be likely to think of her piteous desire, by customary association, whenever we think of her mind. We will also be likely to think of her desire, and so to form the habitually associated idea of her mind, whenever we think of her action or its consequences. We will thus associate her passion of pity *and* any pleasure that it caused us with *her*, a "thinking being." We will therefore experience the kind of double relation of ideas and impressions that produces approbation.

Indeed, a single durable principle may contain *several* virtues or vices. Consider a case where someone acts justly by repaying a loan, despite knowing that they could escape doing so. Perhaps we observe this person spot the money's owner in the street, call her over, and pay her the money, while saying how pleased he is to do so. He has, we may infer, a durable principle that includes various perceptions: visual impressions, memories, beliefs about the person's identity, and so on. It will also include several passions, including a painful feeling of *disapprobation* at the idea of keeping the borrowed money (an "abhorrence of villainy and knavery"), a pleasant feeling of *approbation* at the idea of repaying it (a "regard to justice"), and

a *desire* to repay the money (T 3.2.1.9). We will approve of his desire *and* moral sentiments, but not of the durable principle as such.

So understood, Hume's account of durable principles does not imply that all virtues and vices are character traits. Indeed, it tells us very little about his understanding of virtues, characters, or character traits. It is not, therefore, in tension with the interpretation of these that I set out in section 1. I therefore believe that we should endorse this interpretation. Hume allows that virtues and vices can be either character traits or single perceptions. He is also a reductionist about character.

On this interpretation, Hume can still explain why every character "appears to me, in practice, pretty durable and untransmutable" (M App.2.4). He need claim nothing more than that people's characters typically comprise similar kinds of perceptions over time. Somewhat similarly, he can allow that our "existence is more durable" than the kinds of short-lived causes of joy which, he thinks, can cause little pride (T 2.1.6.7).[49] Hume observes that such objects are of a "much shorter duration" than "ourselves," and that we are related to them for only a "small … part of our existence." Again, therefore, "durable" seems to mean simply "long-lasting." Hume merely claims that a person's mind will typically exist for much longer than any such short-lived objects.

My interpretation is also consistent with much of what Hume writes about moral responsibility. For example, he writes that

> Men are less blam'd for such evil actions, as they perform hastily and unpremeditately, than for such as proceed from thought and deliberation. For

> what reason? But because a hasty temper, tho' a constant cause in the mind, operates only by intervals, and infects not the whole character. (T 2.3.2.7)[50]

To act on a "hasty temper," it seems, is to act on a vicious desire that one later regrets. Clearly, if we know that someone acted in this way, then we will infer that he possessed various perceptions of which we disapprove, such as a desire to perform the harmful action, a pleasure at its performance, and so on. However, we will *also* infer that he possessed various perceptions of which we approve, such as a feeling of remorse, and so on. Whenever we think of this person, *both* sets of ideas will come to mind, so that we will be likely to feel both moral approbation *and* disapprobation. Hume believes that, in such cases, our opposing passions "become mutually destructive" (T 2.3.9.15). If we were to feel the *same* degree of approbation as of disapprobation, then our opposing sentiments would destroy each other, and our mind would be in "perfect tranquillity" (T 2.3.9.15). However, in cases where a harmful action has been hastily performed, Hume thinks, we will ultimately experience at least some disapprobation.

In the same passage, Hume argues that harmful "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 2.3.2.7). Therefore, Hume argues, "repentance wipes off every crime, especially if attended with an evident reformation of life and manners" (T 2.3.2.7).

Bricke reads this to mean that we will not disapprove of a repentant ex-criminal because we will not believe that she still possesses any "enduring conative qualities of a certain undesirable sort."[51] However, Hume says nothing about durable or enduring qualities here. He has already suggested, at T 2.2.3.4, that if a wrongdoer's behavior changes – presumably such

that she exhibits sympathetic responses to the pains she caused, desires to make amends, and so on – then there will be a corresponding "alteration" in the "transition of ideas" in our minds: "In which case the passion is likewise alter'd." We will start to think of these *pleasing* qualities instead of the earlier, painful qualities when we think of the person, and so we will start to feel love, at least if we think of this person's mind just from the time at which she repented to the present moment. Presumably, we can evaluate her diachronic character *before* repentance too, by disapproving of her succession of perceptions from young adulthood until she repented.

In summary, Hume believes that we will typically feel a moral sentiment whenever we infer that a person's mind includes or included (or, in relevant cases, lacks or lacked) at least one perception or succession of perceptions of some type that tends to be useful, agreeable, harmful, or disagreeable. Any such perception, or interrupted succession of similar perceptions, will constitute either a virtue or a vice. The object of our moral evaluation will be the person in question.

Conclusion

I have argued that Hume's "durable principles of the mind" are *not* virtues, vices, or character traits. Instead, a durable principle is a succession of different perceptions, which may only last for a few minutes, and which will exist at and around the time that a person performs an action. In Book 2, Hume refers to durable principles to help him explain why we love or hate people who intentionally cause pleasure or pain, but not those who do so unintentionally. In Book 3, he refers back to this argument, to help explain why we never morally evaluate actions as such.

Given this, we can allow that Hume understands character traits as interrupted successions of similar perceptions, just as Qu argues he should do, in accordance with his

accounts of the mind and causal powers. We can also allow that some Humean virtues and vices are simply perceptions: a thesis that is implied by Hume's official definition of a "virtue," and which is also strongly suggested by several other passages in his works. I have argued that we should not, therefore, read Hume as a virtue ethicist. We should, however, take him to consistently endorse the reductionist theory of character to which his theory of mind commits him.

NOTES

My thanks to participants at the British Society for the History of Philosophy Annual Conference 2022 and to two anonymous reviewers at *Hume Studies* for their insightful feedback on this paper. My thanks too to Donald Baxter, Danielle Charette James, Avital H. Levi, Katherina Paxman, Xiao Qi, and Aaron Alexander Zubia, for very helpful discussions. I am particularly grateful to Enrico Galvagni and Peter Millican for many patient and invaluable discussions about earlier drafts of this paper, and about Hume's account of character more generally.

[1] David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (Oxford: Clarendon Press, 2000). Cited by "T," followed by Book, part, section, and paragraph number. All emphasis is in the original.

[2] E.g.: T 3.1.1.2; T 3.1.1.12; T 3.3.1.30; EPM 5.10; EPM 5.38. (David Hume, *An Enquiry concerning the Principles of Morals,* ed. Tom L. Beauchamp [Oxford: Clarendon Press, 1998]. Cited by "EPM," followed by section and paragraph number. All emphasis is in the original.)

[3] Hume repeats this argument in T 3.3.1.5.

[4] Paul Russell, *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility* (Oxford: Oxford University Press, 1995), 95. See also, e.g.: Donald Ainslie, "Hume and Moral Motivation," in T*he Cambridge Companion to Hume's "Treatise*," eds. Donald Ainslie and Annemarie Butler (Cambridge:

Cambridge University Press, 2015), 283–300, 292; Annette Baier, *A Progress of Sentiments: Reflections on Hume's "Treatise"* (Cambridge MA: Harvard University Press, 1991), 109; John Bricke, *Mind and Morality: An Examination of Hume's Moral Psychology* (Oxford: Oxford University Press, 1996), 240–45; Rachel Cohon, *Hume's Morality: Feeling and Fabrication* (Oxford: Oxford University Press, 2008), 131; Erin Frykholm, "The Ontology of Character Traits in Hume," *Canadian Journal of Philosophy* 42 (S1) (2012): 82–97, 83; Jane McIntyre, "Character: A Humean Account," *History of Philosophy Quarterly* 7, no. 2 (1990): 193–206; A.E. Pitson, *Hume's Philosophy of the Self* (London: Routledge, 2002), 88–96; Elizabeth Radcliffe, *Hume on the Nature of Morality* (Cambridge: Cambridge University Press, 2022), 40; Constantine Sandis, *Character and Causation: Hume's Philosophy of Action* (New York: Routledge, 2019), 78; Jacqueline Taylor, *Reflecting Subjects: Passion, Sympathy, and Society in Hume's Philosophy* (Oxford: Oxford University Press, 2015), 102.

[5] *Freedom and Moral Sentiment*, 97.

[6] Several scholars emphasise this last point. E.g.: Ainslie, "Hume and Moral Motivation," 292–93; Baier, *A Progress of Sentiments*, 183; John Bricke, "Hume's Conception of Character," *The Southwestern Journal of Philosophy* 5, no. 1 (1974): 107–13; Cohon, *Hume's Morality*, 128; Frykholm, "The Ontology of Character Traits in Hume"; McIntyre, "Character: A Humean Account"; Hsueh Qu, "Hume's Dispositional Account of the Self," *Australasian Journal of Philosophy* 95, no. 4 (2017): 644–57; Radcliffe, *Hume on the Nature of Morality*, 40: Russell, *Freedom and Moral Sentiment*, 95–96; Sandis, *Character and Causation*, 77–78; Christine Swanton, *The Virtue Ethics of Hume and Nietzsche* (Oxford: Blackwell, 2015), 48; Taylor, *Reflecting Subjects*, 102.

[7] E.g.: Cohon, *Hume's Morality*, 161; Rosalind Hursthouse, "Virtue Ethics and Human Nature," *Hume Studies* 25 (1999): 67–82; Swanton, *The Virtue Ethics of Hume*; Jacqueline Taylor, "Virtue and the Evaluation of Character," in *The Blackwell Guide to Hume's* Treatise, ed. Saul Traiger (Oxford: Blackwell, 2006): 276–95.

[8] My thanks to Enrico Galvagni for pointing this out to me (personal correspondence).

[9] "Hume's Dispositional Account". Given Hume's nominalism, a Humean trait is fundamentally an interrupted succession of *resembling* perceptions in a mind over time. I address the relevant kinds of resemblance in section 3.

[10] E.g.: Bricke, "Hume's Conception"; McIntyre, "Character: A Humean Account,"; Qu, "Hume's Dispositional Account."

[11] As Baier observes, Hume uses "character" in various ways (*A Progress of Sentiments*, 188). Here, I will ignore those passages where "character" is synonymous with "reputation," for example (e.g., EPM 6.14).

[12] Terence Penelhum, "The Self in Hume's Philosophy." *Southwestern Journal of Philosophy* 7, no. 2 (1976): 9–23, 14. Hume is often understood to introduce a *further* idea of the self at T 1.4.6.5 (e.g.: Donald Ainslie, "Scepticism About Persons in Book II of Hume's *Treatise*." *Journal of the History of Philosophy* 37 [1999]: 469–92; Baier, *A Progress of Sentiments*; Pitson, *Hume's Philosophy*). I argue against this interpretation of T 1.4.6.5 elsewhere (paper under review).

[13] Hume eventually worried that his theory of personal identity was internally inconsistent (T App.21). However, this worry only occurred to him *after* he had written Book 2, and he retained hopes that the inconsistency might be resolved. Moreover, it only concerned his theory of personal *identity*: he still maintained that, if he thinks of *himself*, then he cannot "perceive any thing but the perceptions" that appear to constitute his mind (T App.15).

[14] Qu, "Hume's Dispositional Account."

[15] Qu, "Hume's Dispositional Account," 650.

[16] Bricke, "Hume's Conception," 109.

[17] Qu, "Hume's Dispositional Account," 644n2.

[18] Hume uses "benevolence" very differently in the *Treatise* and moral *Enquiry*, but the details need not concern us here.

[19] McIntyre, "Character: A Humean Account," 197. See also: Bricke, "Hume's Conception,"109; Qu, "Hume's Dispositional Account," 647.

[20] McIntyre, "Character: A Humean Account," 200–1.

[21] Qu, "Hume's Dispositional Account," 649; Russell, *Freedom and Moral Sentiment*, 106n3. See also Pitson, *Hume's Philosophy*, 90.

[22] Qu, "Hume's Dispositional Account," 648.

[23] According to Qu, Hume thinks we create a "fictional notion" of the self as composed of durable dispositions, and that such fictitious selves are the objects of all moral judgments ("Hume's Dispositional Account," 656). However, so interpreted, Hume not only believes that all moral judgments "rest on a philosophically erroneous conception of the self," but he also frequently conflates this erroneous view with his own, official conception of the self (ibid).

[24] I argue elsewhere that there are some significant differences between the theories of moral evaluation in Hume's *Treatise* and moral *Enquiry* (James Chamberlain, "Hume on Calm Passions, Moral Sentiments, and the 'Common Point of View'," *Hume Studies* 47, no. 1 [2022]: 79–101). However, I think his understanding of characters, virtues, and vices remain essentially unchanged. He consistently understands all virtues and vices as "mental qualities" that cause approbation or disapprobation (T 3.3.1.3; T 3.3.1.6; T 3.3.1.9; T 3.3.4.1; EPM 1.10; EPM 9.1).

[25] Elsewhere, I argue that Hume understands *all* moral judgments as sentiments (James Chamberlain, "Hume's Emotivist Theory of Moral Judgements," *European Journal of Philosophy* 28, no. 4 [2019]: 1058-72; "Hume on Calm Passions"). However, several scholars argue that some Humean moral judgements are beliefs (e.g., Cohon, *Hume's Morality*; Don Garrett, *Hume* [London: Routledge, 2015]). Here, I claim only that Hume's arguments about durable principles concern the moral sentiments, and that these play an important role in his theory of moral judgment.

[26] Baier, *A Progress of Sentiments*, 134; Åsa Carlson, "The Moral Sentiments in Hume's *Treatise*: A Classificatory Problem," *Hume Studies* 40 (2004): 73–94, 74.

[27] E.g.: T 3.2.1.3; T 3.2.2.8; T 3.3.1.16; T 3.3.1.19; T 3.3.1.29; EPM 5.41n24.1; EPM 7.22.

[28] E.g.: T 3.3.1.14; T 3.3.1.15; T 3.3.1.17.

[29] E.g.: T 3.1.2.3; T 3.1.2.11; T 3.2.8.8.

[30] Hume uses "disposition" and cognate terms somewhat similarly elsewhere, as at T 2.3.7.5.

[31] David Hume, *An Enquiry concerning Human Understanding*, ed. Tom L. Beauchamp (Oxford: Clarendon Press, 2000). Cited by "EHU," followed by section and paragraph number.

[32] The adjective "signal" is defined in Johnson's dictionary of 1755 as, "Eminent; memorable; remarkable."

[33] Russell, *Freedom and Moral Sentiment*, 102.

[34] Pitson, *Hume's Philosophy*, 99. Pitson's quotation of Hume is from EHU 8.15.

[35] I argue for this interpretation, with reference to Hume's accounts of sympathy and the "common point of view," in my "Hume on Calm Passions." I also argue that Hume thinks we can feel *violent* love for a person even as we approve of her for having a virtuous character. I take this to be his (poorly phrased) point when he claims, of someone who helps his "narrow circle," that we "approve of his character, and love his person" (T 3.3.3.2).

[36] I argue for this interpretation in James Chamberlain, "Justice and the Tendency towards Good: The Role of Custom in Hume's Theory of Moral Motivation." *Hume Studies* 43, no. 1 (2017): 117-37.

[37] Hursthouse, "Virtue Ethics and Human Nature," 68.

[38] Swanton, *The Virtue Ethics of Hume*, 48.

[39] Cohon, *Hume's Morality*, 161.

[40] Garrett, *Hume*, 247.

[41] Nevertheless, I think Hume closer to a virtue ethicist than, e.g., James Harris allows ("Hume on the Moral Obligation to Justice," *Hume Studies* 36, no. 1 (2010): 25–50). Harris argues that Hume sees *no* moral significance in the motives or character traits that lie behind the virtue of justice. I disagree, as we will see in section 4.

[42] Hume typically has in mind qualities that, he believes, all humans *innately* possess. E.g.: T 3.2.5.9; T 3.2.9.3; T 3.3.1.17; T 3.3.3.7; T 3.3.6.3; EHU 10.5.

[43] See also EHU 8.29.

[44] Ainslie, "Hume and Moral Motivation," 292.

[45] This problem generalizes to other understandings of Humean character traits. Say that a Humean trait is, e.g., a short-lived but deeply entrenched or unchangeable part of a character. Hume *still* cannot plausibly stipulate that actions must either be caused by such traits or have *no* psychological causes.

[46] Hume is more conciliatory in the later work, claiming to oppose only "the principle … which denies necessity" (EHU 8.29). However, the essential argument is the same.

[47] Elizabeth Radcliffe, *Hume, Passion, and Action* (Oxford: Oxford University Press, 2018), 102.

[48] Páll Árdal, *Passion and Value in Hume's "Treatise"* (2nd edition) (Edinburgh: Edinburgh University Press, 1989), 112–15. I understand the moral sentiments as literally calm passions of love and hatred (Chamberlain, "Hume on Calm Passions"). All that matters here, however, is that they resemble love and hatred. Of course, there are opposing arguments (e.g., Carlson, "The Moral Sentiments," 82–85).

[49] Ainslie suggests that Hume thinks the causes of pride must be similarly "durable" ("Scepticism About Persons," 484). However, Hume merely claims that they cannot be "casual and inconstant" (T 2.1.6.7).

[50] See also EHU 8.30.

[51] Bricke, *Mind and Morality*, 240.

WORKS CITED

Primary sources:

Hume, David. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. Oxford: Clarendon Press, 2000 [1739/40].

Hume, David. *An Enquiry concerning Human Understanding*. Edited by T. L. Beauchamp. Oxford: Clarendon Press, 2000 [1748].

Hume, David. *An Enquiry concerning the Principles of Morals*. Edited by T. L. Beauchamp. Oxford: Clarendon Press, 1998 [1751].

Johnson, Samuel. *A Dictionary of the English Language*. London: W. Strahan, 1755.

Secondary sources:

Ainslie, Donald C. "Scepticism About Persons in Book II of Hume's *Treatise*." *Journal of the History of Philosophy* 37, no. 3 (1999): 469–92

Ainslie, Donald C. "Hume and Moral Motivation." In *The Cambridge Companion to Hume's "Treatise"*. Edited by Donald Ainslie & Annemarie Butler, 283–300. Cambridge: Cambridge University Press, 2015.

Árdal, Páll. *Passion and Value in Hume's* "*Treatise"* (2nd edition). Edinburgh: Edinburgh University Press, 1989.

Baier, Annette. *A Progress of Sentiments: Reflections on Hume's "Treatise."* Cambridge MA: Harvard University Press, 1991.

Bricke, John. "Hume's Conception of Character." *The Southwestern Journal of Philosophy* 5, no. 1 (1974): 107–13.

Bricke, John. *Mind and Morality: An Examination of Hume's Moral Psychology.* Oxford: Oxford University Press, 1996.

Carlson, Åsa. "The Moral Sentiments in Hume's *Treatise*: A Classificatory Problem." *Hume Studies* 40 (2004): 73–94.

Chamberlain, James. "Justice and the Tendency towards Good: The Role of Custom in Hume's Theory of Moral Motivation." *Hume Studies* 43, no. 1 (2017): 117-37.

Chamberlain, James. "Hume's Emotivist Theory of Moral Judgements." *European Journal of Philosophy* 28, no. 4 (2019): 1058-72.

Chamberlain, James. "Hume on Calm Passions, Moral Sentiments, and the 'Common Point of View'." *Hume Studies* 47, no. 1 (2022): 79–101.

Cohon, Rachel. *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press, 2008.

Frykholm, Erin. "The Ontology of Character Traits in Hume." *Canadian Journal of Philosophy* 42 (S1) (2012): 82–97.

Garrett, Don. *Hume*. London: Routledge, 2015.

Harris, James. "Hume on the Moral Obligation to Justice." *Hume Studies* 36, no. 1 (2010): 25–50.

Hursthouse, Rosalind. "Virtue Ethics and Human Nature." *Hume Studies* 25 (1999): 67–82.

McIntyre, Jane. "Character: A Humean Account." *History of Philosophy Quarterly* 7, no. 2 (1990): 193–206.

Penelhum, Terence. "The Self in Hume's Philosophy." *Southwestern Journal of Philosophy* 7, no. 2 (1976): 9–23.

Pitson, A.E. *Hume's Philosophy of the Self*. London: Routledge, 2002.

Qu, Hsueh. “Hume's Dispositional Account of the Self.” *Australasian Journal of Philosophy* 95, no. 4 (2017): 644–57.

Radcliffe, Elizabeth. *Hume, Passion, and Action*. Oxford: Oxford University Press, 2018.

Radcliffe, Elizabeth. *Hume on the Nature of Morality*. Cambridge: Cambridge University Press, 2022.

Russell, Paul. *Freedom and Moral Sentiment: Hume’s Way of Naturalizing Responsibility*. Oxford: Oxford University Press, 1995.

Sandis, Constantine. *Character and Causation: Hume’s Philosophy of Action.* New York: Routledge, 2019.

Swanton, Christine. *The Virtue Ethics of Hume and Nietzsche*. Oxford: Blackwell, 2015.

Taylor, Jacqueline. “Virtue and the Evaluation of Character.” In *The Blackwell Guide to Hume’s “Treatise*.” Edited by Saul Traiger, 276–95. Oxford: Blackwell, 2006.

Taylor, Jacqueline. *Reflecting Subjects: Passion, Sympathy, and Society in Hume's Philosophy*. Oxford: Oxford University Press, 2015.