

# 14

## The Relation between Hume's Two *Enquiries*

Peter Millican

The relation between Hume's *Enquiries* has many strands, running in both directions. Most of these involve similarities between the two works, often contrasting with the *Treatise of Human Nature* from which they both trace their descent. Perhaps most obvious in this respect is their elegant and engaging presentational style, and their relatively narrow focus compared with the more theoretically systematic and comprehensive *Treatise* (see §1 below). The treatment of religion (§2) provides another obvious contrast: the *Enquiries* are more explicitly hostile to religious metaphysics and morality than the *Treatise*, though this too seems to be more a presentational than philosophical difference, indicating not so much a change in Hume's attitude towards religion, as in his willingness to display it.

A deeper parallel between the *Enquiries* is their forthright rejection of extreme positions that were apparently endorsed in the *Treatise*. In two prominent cases (see §3 below), this toning down is presented as a *reconciliation* between superficially opposed positions: in Section 8 of the first *Enquiry*, between the doctrines of liberty and necessity; and in the second *Enquiry*, between the moral claims of reason and sentiment. Here the *Enquiry* discussions seem to involve *clarifications* of the *Treatise* positions rather than abandonments, tidying up terminology and avoiding the extravagant statements that had previously given a misleadingly extreme impression of Hume's philosophy.<sup>1</sup> In other cases—most notably the treatment of epistemological and moral scepticism—the relation between Hume's earlier and later views is more controversial, usually owing to the relative obscurity of the *Treatise*. In Book 1 Part 4, Hume sometimes seems pulled into intensely upsetting extremes of scepticism, whereas the first *Enquiry* urbanely expounds a far less threatening "mitigated scepticism." Some scholars see this as yet another example of clarification rather than revision, but in §4 I shall sketch my own view that Hume's attitude to epistemological scepticism changed significantly between the two works.

Given the focus of this volume on the second *Enquiry*, the issue of *moral* scepticism demands more detailed consideration. In §5 and §6 below, we shall see evidence

<sup>1</sup> Such statements include "the doctrine of liberty is . . . absurd . . . in one sense, and unintelligible in any other" (T 2.3.2.1); "Reason is, and ought only to be the slave of the passions" (T 2.3.3.4).

of a very significant shift in Hume's position, not only presentationally—as manifested by his forceful denial of moral scepticism in the second *Enquiry*—but also substantially, in his formulation of powerful arguments that refute the Lockean psychological egoism from which his thinking appears to have begun. The *Treatise* itself displays an awareness that egoism is not the whole story, and much of Hume's later case against it builds on points already anticipated there. But his eventual utter rejection of egoism may have been inspired—and was at least cemented—by his discovery of the arguments of Joseph Butler (to whom due credit is given in the original first *Enquiry*). Following this change, Hume reinterprets the notions of “sympathy” and “humanity,” and is able to abandon the sometimes convoluted associationist explanations that he had previously relied on “to resolve [unselfish tendencies] . . . by a philosophical chymistry, . . . into . . . self-love” (EPM App 2.4). As explained in §7, this withdrawal from confident associationism also has a parallel in the first *Enquiry*, where Hume's modified theory of belief and probability likewise builds on points anticipated in the *Treatise* (especially the *Appendix*). Realization of these difficulties in his earlier positions may have contributed to Hume's tendency in the *Enquiries* to minimize his dependence on associationist explanations, even where he still finds them plausible.

Finally, we turn to what is perhaps the most interesting—though highly controversial—issue regarding the philosophical relationship between the two *Enquiries*, namely, the extent to which there is a genuine parallel between Hume's causal and moral metaphysics. In his mature treatment of these issues, we see some significant methodological connections that were not present in the *Treatise*: most notably, a strong emphasis on linguistic clarification as a means of resolution, culminating in the derivation of paired definitions of the key concepts. §8 below brings these discussions together, showing how Hume's use of his two definitions of “personal merit” in the second *Enquiry* can help us to interpret the two definitions of causal necessity in the first *Enquiry* (and inherited from *Treatise* 1.3.14.31). It then turns to the fashionable but somewhat anachronistic question whether Hume can appropriately be seen as a “projectivist” about causation and morality. In the case of causation, the textual evidence is weak *unless* we treat as “projection” Hume's empiricist insistence that causal attributions are to be understood in terms of our natural tendency to draw causal inferences. There is stronger textual evidence in the case of morality, with Hume's suggestion that moral taste “raises . . . a new creation,” but this comes from a paragraph whose point is to insist on “the distinct boundaries and offices of *reason* and of *taste*” (EPM App 1.21), attributing causal judgements to the one, and moral judgements to the other. So although there are suggestive parallels between them, Hume himself clearly saw fundamental differences here.

## 1. Titles and Styles

Hume's most obvious motive for the major change in presentation between the *Treatise* and the *Enquiries* was the much greater popular success of his *Essays*—meanwhile published in 1741 and 1742—compared with the disappointing reception of the *Treatise* of 1739–40. So when he “cast anew” Book 1 of the *Treatise* in 1748 (MOL, 8), he published his new work as *Philosophical Essays concerning Human*

*Understanding*, prominently subtitled “By the AUTHOR of the Essays Moral and Political.” This title, however, was somewhat misleading, because the philosophical discussion throughout is integrated and continuous, with later “essays” building on earlier ones and the work as a whole presenting a unified picture. The same applied even more to the “essays” that he initially wrote to recast his moral theory, and his new publication appropriately appeared under a title that signalled a sustained and unified investigation of a limited topic area: *An Enquiry concerning the Principles of Morals*.<sup>2</sup>

The two recastings were brought together in 1753, as volumes 2 and 3 of the four-volume *Essays and Treatises on Several Subjects*, and then closer still within the 1758 single-volume edition. Here they were joined by *A Dissertation on the Passions*—the recast *Treatise* Book 2—with the order of the three works reflecting that of the original *Treatise* (i.e. understanding, then passions, then morals). It was in this 1758 edition that what we now know as the first *Enquiry* was finally given its enduring title: *An Enquiry concerning Human Understanding*. Through this sequence of events, we see the presentation of the moral *Enquiry* apparently influencing—at least to some extent—that of its older sibling. In written style, however, the influence goes the other way, and the second *Enquiry* broadly follows the first: less academic and technical than the *Treatise*, but still clearly “accurate and abstract” philosophy rather than “easy and humane,” in the terms that Hume himself uses at EHU 1.8.<sup>3</sup> Both works aim to present Hume’s philosophy in a rigorous but relatively approachable way, “reconciling profound enquiry with clearness” (EHU 1.17), and in this both succeed admirably. Admittedly there is some contrast here, with the second *Enquiry* coming across as less technical and more “literary,” partly because its most theoretical discussions are relegated to appendices, and partly because it includes numerous allusions to history and classical writers. But neither contrast seems to have any deeper philosophical significance: perhaps Hume simply took the view that the subject of morals was better suited than “metaphysics” to appeal to a general audience, and also found the subject matter more conducive to illustration from literary sources.<sup>4</sup>

## 2. Religion and Irreligious Ethics

Hume had an intense critical interest in religion throughout his adult life, as manifested in many of his philosophical writings. The 1739–40 *Treatise*, however,

<sup>2</sup> Two passages corrected in the errata to the first edition (1751, 2.5n., now EPM App. 2.5, n. 60; and EPM 3.34, n. 12) show that at least some of the second *Enquiry* was composed on the assumption that it would appear as a set of essays.

<sup>3</sup> At EHU 1.7 Hume begins an extended defence of “what is commonly called *metaphysics*,” and the second *Enquiry* identifies itself with this “speculative” study of human nature at EPM 2.5, 6.22, App. 2.5, and App. 4.21.

<sup>4</sup> Hume had also recently read a great number of classical works, many of these in preparation for his longest and most scholarly essay, “Of the Populousness of Ancient Nations.” Though not published until 1752, Hume stated in a letter of 18 April 1750 to John Clephane that it was “The last thing I took my hand from” (HL i 139).

makes relatively little explicit mention of it, apparently mainly for prudential reasons.<sup>5</sup> In December 1737, Hume wrote to Henry Home (later Lord Kames) that he was “at present castrating my Work, that is, cutting off its noble Parts, that is, endeavouring it shall give as little Offence as possible” to make it suitable for presentation to Bishop Joseph Butler, whose good opinion he sought (NHL 2). This caution did not prevent the implicit “downright Atheism” of the *Treatise* being held against him when he applied for a Chair at Edinburgh in 1745 (LFG 15), and by 1747 he seems to have come to the conclusion that he might as well be more candid, remarking to James Oswald: “I see not what bad consequences follow, in the present age, from the character of an infidel; especially if a man’s conduct be in other respects irreproachable” (HL i 106). In 1748 he wrote to Kames, who had tried to dissuade him from publishing the first *Enquiry*: “I won’t justify the prudence of this step, any other way than by expressing my indifference about all the consequences that may follow” (HL i 111). The *Enquiry* duly appeared, with its incisively critical discussions of the Free Will Defence (EHU 8.32–6), miracle reports (Section 10), the Design Argument (Section 11), and various other topics in religious epistemology. Three years later, in 1751, the concluding section of the second *Enquiry* added to this critique an entertainingly withering attack on religious morality and its “monkish virtues” (EPM 9.3, cf. §8 below).

The main anti-religious impact of the second *Enquiry*, however—like that of *Treatise* Book 3—is in providing an account of morality which is entirely naturalistic, building on the basis of the human condition and human sentiments, with absolutely no resort to “superstition” in the form of supposed divine commands or posthumous retribution etc. This reflects back on at least two discussions in the first *Enquiry*, namely of “liberty and necessity” in Section 8, and of “a future state” in Section 11. In the former, Hume explains that “remote considerations” about the divinely ordained ordering of the universe (or, by extension, about determinism in general) cannot undermine morality, since this is founded on sentiment rather than metaphysics (EHU 8.35). Hume takes our moral sentiments and judgements to be responsive to “qualities of the mind,”<sup>6</sup> and this provides another link with his discussions of liberty and necessity (in both the *Treatise* and the first *Enquiry*), where he argues that causal necessity, so far from undermining morality, is essential to it:

The only proper object of hatred or vengeance, is a person or creature, endowed with thought and consciousness; . . . Actions are, by their very nature, temporary and perishing; and where they proceed not from some cause in the character and disposition of the person who

<sup>5</sup> The doctrines of the *Treatise* do, however, have quite significant anti-theistic implications that Hume does not point out, as explored extensively by Paul Russell, *The Riddle of Hume’s Treatise: Skepticism, Naturalism, and Irreligion* (New York: Oxford University Press, 2008). For a brief survey of Hume’s attitude to religion and his relevant works, including discussion of his use of irony and dialogue to disguise his more incendiary views, see Peter Millican, *Reading Hume on Human Understanding: Essays on the First Enquiry* (Oxford: Clarendon Press, 2002), pp. 34–40. Note that Section 11 of the first *Enquiry* prudently discusses the Design Argument in dialogue form, and the second *Enquiry* discusses moral relativism in “A Dialogue,” effectively a third appendix in the 1751 edition (and from 1777 a fifth). The *Dialogues concerning Natural Religion*, though published only posthumously in 1779, were apparently written mainly in 1751–52, being started shortly after completion of the second *Enquiry*.

<sup>6</sup> See for example T 3.2.5.4, 3.3.1.3, 3.3.1.30, 3.3.5.1, EPM 1.10, 8.0, n. 50, 9.12.

performed them, they can neither redound to his honour, if good; nor infamy, if evil... actions render a person criminal, merely as they are proofs of criminal principles in the mind... But, except upon the doctrine of necessity, they never were just proofs, and consequently never were criminal. (EHU 8.29–30, cf. T 2.3.2.5–7)

This argument sharpens the difficulty of the Problem of Evil, which Hume raises a few paragraphs later (albeit with prudent obliqueness). If morality *requires* that our actions be determined by our character, then there is no prospect for any satisfactory explanation of “how the Deity can be the mediate cause of all the actions of men, without being the author of sin and moral turpitude” (EHU 8.36). “These are mysteries,” Hume continues, which inevitably involve us in “difficulties, and even contradictions”—making his anti-theistic conclusion fairly obvious, though unstated.

The connection between Hume's naturalistic moral theory and Section 11 of the first *Enquiry* is related, but even less explicit. Here the relevant question (EHU 11.4, 19–22, 27–8) is whether atheism and the denial of a future state pose a threat to morality. Hume's “sceptical friend” argues that they cannot do so, because any inference to the supposed deity's moral qualities (such as might be thought to determine what awaits us in a future state) must be drawn from the world as we find it, and hence cannot “give us measures of conduct and behaviour different from those which are furnished by reflections on common life” (EHU 11.27). In his own voice, Hume offers the response (11.28) that even if theories about divine morality are fanciful, nevertheless people may be influenced by them; hence undermining belief in divine punishment could be morally dangerous even if logically justified. But he quickly moves on to defend religious tolerance on other grounds (EHU 11.29), and the main question is rather left hanging.<sup>7</sup> Hume's real answer becomes apparent not in the first *Enquiry* but the second (and also *The Natural History of Religion*, Sections 9–10 and 13–14), where as noted above he reveals that so far from viewing religion as a potential support for morality, he sees it rather as a corrupting influence. Moreover, he considers morality quite capable of naturalistic defence, being fully explicable in terms of the human situation (EPM 3.1–21) and personally recommendable as conducive to mutual cooperation, social harmony, self-respect, and peace of mind (EPM 9.14–25).

### 3. Presentational Clarification and Compromise

Although the two *Enquiries* are relatively forthright in their opposition to religion, in other respects they present Hume's philosophy as being significantly less sceptical and corrosive of orthodox positions than the impression given by the *Treatise*. Whether these changes are superficial and *merely* presentational, or whether on the other hand they reflect deeper developments in Hume's own views, are in some cases controversial questions, and we shall come back to the most important of these shortly (in §§4–7 below). But in at least some respects, the move from *Treatise* to *Enquiries* does seem to be mainly presentational, involving clarification and

<sup>7</sup> For a more detailed account of the arguments of this section, see J. C. A. Gaskin, “Religion: The Useless Hypothesis.” In Millican, *Reading Hume*, pp. 349–69; see especially pp. 365–8.

refinement of Hume's position to tone down its misleadingly paradoxical appearance, rather than fundamental philosophical change.

Perhaps the most obvious instance of this comes in Section 8 of the first *Enquiry*, where Hume's theory of "liberty and necessity" is essentially unchanged from that of the *Treatise*, though presented in very different terms. In the *Treatise*, Hume frequently equates "liberty" with *chance* (T 2.3.1.15, 2.3.2.2, 2.3.2.6, 2.3.2.7, 2.3.2.8) or *indifference* (T 2.3.1.3, 2.3.2.2), and he explicitly attacks "the doctrine of liberty" so understood as "absurd" or "unintelligible" (T 2.3.2.1, cf. 2.3.1.18). This forthright denial of human liberty can come across as very radical, but it is also somewhat anomalous, given Hume's own statement that "liberty," in "the most common sense of the word," is to be understood not as "liberty of *indifference*" but rather "liberty of *spontaneity*," characterized as "that which is oppos'd to violence" and as "that species of liberty, which it concerns us to preserve" (T 2.3.2.1). The *Enquiry* clears up this confusion and moderates the radical language, by instead presenting a "reconciling project" (EHU 8.23) between the doctrines of liberty and necessity, arguing in favour of both of them on the basis of appropriate definitions. "Liberty" is now to be understood as "*a power of acting or not acting, according to the determinations of the will*; that is, if we chuse to remain at rest, we may; if we chuse to move, we also may" (EHU 8.23).<sup>8</sup> By renouncing any sceptical denial of human freedom, Hume is better able to argue for the point that really concerns him, namely the application of deterministic causal understanding to human behaviour.

Another instance of presentational compromise concerns the role of reason within morality, the subject of some of the *Treatise*'s most notoriously extravagant rhetoric: "Reason is, and ought only to be the slave of the passions" (T 2.3.3.4); "'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger" (T 2.3.3.6); "Reason is wholly inactive, and can never be the source of so active a principle as . . . a sense of morals" (T 3.1.1.10). Such passages might seem to imply that reason is irrelevant to motivation or morality, and have often been read as stating at least that the *products* of reason—namely beliefs—cannot *cause* actions. But as we shall discuss more fully in §5 below, Hume in the *Treatise* takes the view that "the prospect of pain or pleasure"—which is surely a reference to belief about likely outcomes—is what most typically motivates our actions (T 2.3.3.3, 2.3.9.1), so there seems to be a mismatch here. A related problem concerns the famous "Representation Argument" of T 2.3.3.5 and 3.1.1.9, which gives at least some appearance of attempting to settle a causal claim (about the influence of reason) by a priori considerations (about the nature of our passions), something which should be

<sup>8</sup> Most commentators have identified this "hypothetical liberty" (EHU 8.23) with the "liberty of spontaneity" that Hume alludes to in the *Treatise*. But as George Botterill points out, the mere "power of acting or not acting"—which is "universally allowed to belong to every one, who is not a prisoner and in chains" (EHU 8.23)—fits very poorly as "the most common sense of the word" or as "that species of liberty, which it concerns us to preserve." Botterill absolves Hume of any confusion, but I suspect that in the *Treatise* Hume was himself muddled in his understanding of "spontaneity," which had historically been interpreted in various ways, sometimes involving reflective judgement and sometimes mere voluntariness (making it a poor basis from which to attempt to clarify Hume's own considered view). Fortunately, any such confusion was decisively remedied in the *Enquiry*. See George Botterill, "Hume on Liberty and Necessity." In Millican, *Reading Hume*, pp. 277–300.

quite impossible on Humean principles.<sup>9</sup> By the time he came to recast his work, Hume had apparently recognized these problems, and he entirely omitted the argument, and the misleading rhetoric, from *both* the second *Enquiry* and the *Dissertation on the Passions*.<sup>10</sup> Indeed the *Enquiry* makes a point of reversing any irrationalist impression, setting out a far more balanced view in the introductory section, and clearly asserting the relevance of reason right at the start of the first Appendix, which is devoted to moral metaphysics:

I am apt to suspect, . . . that reason and sentiment concur in almost all moral determinations and conclusions. (EPM 1.9)

One principal foundation of moral praise being supposed to lie in the usefulness of any quality or action; it is evident, that *reason* must enter for a considerable share in all decisions of this kind. (EPM App 1.2)

In this case the unclarity of the *Treatise* account makes it hard to assess how far Hume has changed his view. But certainly the way in which he presents it is radically different, in line with his general desire to play down—apparently sincerely—the radical negativity of his earlier work. Indeed, the rhetoric of Section 1 of the second *Enquiry* seems expressly designed to portray Hume's position as occupying a reasonable middle ground and pursuing another “reconciling project,” this time between the claims of reason and sentiment.

#### 4. Rejecting Extreme Scepticism

Another important presentational difference between the *Treatise* and the *Enquiries*—and arguably also a major philosophical development—is the change in Hume's apparent attitude to scepticism, both epistemological and moral. In the *Treatise*, Hume does not explicitly address sceptical themes until Part 4 of Book 1, but when he does so, the sceptical arguments he expounds are extreme and corrosive, apparently undermining even his own philosophical endeavours.<sup>11</sup> The interpretation of his position here is highly controversial, but at least the superficial impression left by the final section of Book 1—entitled “Conclusion of this book”—is that scepticism carries all before it:

I have already shewn, that the understanding, when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any

<sup>9</sup> For clear statements that no causal claim can be a priori, see T 1.3.15.1, 1.4.5.30; EHU 12.29.

<sup>10</sup> It is striking how much attention interpreters have paid to the Representation Argument—presented so prominently in two different books of the *Treatise*—usually without any apparent awareness that Hume dropped it from his recasting of *both* books. Here, as often elsewhere, Hume shows better judgement than many of his commentators.

<sup>11</sup> For much more on the issues to be discussed here, see Peter Millican, “Hume's Chief Argument.” In *The Oxford Handbook of David Hume*, ed. Paul Russell (New York: Oxford University Press, 2016). There I argue that the scepticism of the *Treatise* is indeed corrosive, in the sense of undermining Hume's own scientific ambitions, and that he finds himself unable to control it at the end of Book 1, resorting to a largely ad hoc and unconvincing palliative to justify ignoring the problem. The first *Enquiry*, while not entirely settling the sceptical concerns, provides a far more convincing and satisfactory response to them, and this—I believe—represents a significant change in Hume's philosophy, as well as in his presentation of it.

proposition, either in philosophy or common life.... I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another.

(T 1.4.7.7–8)

The highly critical 1745 pamphlet to which Hume's *Letter from a Gentleman to his friend in Edinburgh* was a response found ample material in this concluding section to press its first major charge of "Universal Scepticism."<sup>12</sup> But that section is not alone, and reflects much of the tone and content of the preceding discussions in Part 4. The first sentence quoted above faithfully sums up the main argument of Section 1.4.1 on "scepticism with regard to reason," perhaps the most corrosively sceptical of any of Hume's arguments. The following section, on "scepticism with regard to the senses," starts more optimistically with a complacent confidence in "the existence of body" (1.4.2.1), but by the end that confidence has completely evaporated, with Hume now being "more inclin'd to repose no faith at all in my senses, or rather imagination, than to place in it such an implicit confidence." He continues in apparent despair:

I cannot conceive how such trivial qualities of the fancy, conducted by such false suppositions, can ever lead to any solid and rational system.... What then can we look for from this confusion of groundless and extraordinary opinions but error and falshood?... 'Tis impossible upon any system to defend either our understanding or senses;... Carelessness and inattention alone can afford us any remedy. (T 1.4.2.56–7).

Yet more scepticism follows in Hume's discussion of "the modern philosophy" and its distinction between primary and secondary qualities, where he identifies "a direct and total opposition betwixt our reason and our senses;... betwixt those conclusions we form from cause and effect, and those that persuade us of the continu'd and independent existence of body" (T 1.4.4.15). And although his account of the "intellectual world" is initially portrayed as free of "such contradictions" as he has "discover'd in the natural" (T 1.4.5.1), Hume's discussion of personal identity in Section 1.4.6 is both radically subversive of standard views and—notoriously—later confessed to be inadequate in the 1740 *Appendix* published with *Treatise* Book 3.<sup>13</sup> There Hume admits his inability to renounce or render consistent the principles that underlie his position (T App. 21), and although the precise nature of his difficulty is radically unclear, there is no doubt that here, at least, his sceptical arguments have become seriously corrosive of his own philosophy.

<sup>12</sup> This pamphlet (or possibly manuscript, since we know its form and content only from Hume's quotation of it) was composed to undermine Hume's candidacy for the Chair of Ethics and Pneumatical Philosophy at Edinburgh, which could suggest strong prudential reasons for the later toning down of his scepticism, both epistemological and moral (the final major charge accused the *Treatise* of "sapping the Foundations of Morality"). But in fact, as we shall see, there is plenty of evidence that Hume's modifications in this direction had a solid philosophical motivation, and his highly critical treatment of religious topics in the *Enquiries* seems to indicate that prudential caution was a relatively minor consideration in their composition.

<sup>13</sup> At T App.10, Hume apparently alludes to T 1.4.5.1 in mentioning his earlier "hopes, that however deficient our theory of the intellectual world might be, it wou'd be free from those contradictions, and absurdities, which seem to attend every explication, that human reason can give of the material world."

The first *Enquiry's* treatment of scepticism is very different, both in tone and content. Initially it is more overt in its sceptical nature, for example in the title given to the foundational Section 4 which expounds at length Hume's "Sceptical doubts concerning the operations of the understanding." But he later devotes the concluding Section 12 to spelling out his considered attitude to this and other sceptical arguments, rejecting their more extreme or "Pyrrhonian" consequences and embracing instead "a more *mitigated* scepticism or *academical* philosophy" (EHU 12.24). Most importantly, there is little trace in the first *Enquiry* of the corrosive scepticism that so bedevilled the end of *Treatise* Book 1. In particular, the argument of "scepticism with regard to reason" is entirely dropped, and Hume seems implicitly to dismiss the kind of reflexive doubts that underlie it when he attacks "antecedent" scepticism at EHU 12.3. "Scepticism with regard to the senses" is still represented, but toned down by omitting any worries about the supposed incoherence of a changing object retaining its identity through time, and also by Hume's statement that "It is a question of fact, whether the perceptions of the senses be produced by external objects, resembling them" (EHU 12.12), implying that he sees no *contradiction* in the belief in body (though at EHU 12.15 he still finds the primary/secondary quality distinction seriously problematic). Personal identity is not discussed in the *Enquiry*, but both Hume's apparently more relaxed view of identity over time, and his omission of the Separability Principle—a major source of the implausible metaphysics of *Treatise* 1.4.5–6—suggest that this could now be less of a problem for him.<sup>14</sup> Assessing how far these changes are merely cosmetic, and how far they indicate a fundamental change in Hume's position, is unsurprisingly controversial, since he himself does not give any very clear assessment of the matter.<sup>15</sup> Personally I do see fundamental change, but for present purposes we can put this question aside, and simply focus on the presentational issue which is very clear indeed: the first *Enquiry* seems to have been intended, at least in part, to dispel the impression of Hume's epistemology and metaphysics given by *Treatise* Book 1, as being overwhelmingly negative and sceptical.

## 5. The Egoist Foundations of the Treatise

*Treatise* Books 2 and 3 are sceptical in a different way, presenting a picture of moral motivation that is primarily egoist and hedonist, thus precluding any more noble account of morality. Here Hume follows in the footsteps of John Locke, who wrote that "happiness and that alone... moves *desire*" (*Essay* II xxi 41), understanding happiness itself in terms of pleasure. Accordingly, "what has an aptness to produce Pleasure in us, is that we call *Good*, and what is apt to produce Pain in us, we call *Evil*, for no other reason, but for its aptness to produce Pleasure and Pain in us, wherein consists our *Happiness* and *Misery*" (*Essay* II xxi 42).

<sup>14</sup> The Separability Principle is stated most fully at T 1.1.7.3, and applied at T 1.4.5.5 and 1.4.6.3, for example.

<sup>15</sup> Though Hume famously renounced the *Treatise* in favour of the first *Enquiry* on numerous occasions, both in private letters and publicly (cf. §9 below), his words in doing so are open to varying interpretations.

The hedonism of the *Treatise*—expressed in terms reminiscent of Locke—is very evident in Hume’s theory of what he calls the *direct* passions. Explicitly and repeatedly, he equates *good* with *pleasure*, and *evil* with *pain* (e.g. T 2.1.1.4, 2.3.1.1, 2.3.9.1–8), explaining our motivation accordingly:

’Tis *obvious*, that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, . . . 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, emphasis added)

’Tis easy to observe, that the passions, both direct and indirect, are founded on pain and pleasure . . . Upon the removal of pain and pleasure there immediately follows a removal of love and hatred, pride and humility, desire and aversion, and of most of our reflective or secondary impressions. (T 2.3.9.1)

Hume’s account here seems to be egoist as well as hedonist, for although “pain or pleasure” could in principle refer to *somebody else’s* pain or pleasure, it is *our own* that is commonly thought to have the “obvious” and intimate connection with motivation that he is clearly assuming, such that “Upon [their] removal . . . there immediately follows a removal of . . . desire and aversion.”<sup>16</sup>

This claim that the *Treatise* is fundamentally egoist is, however, controversial, attracting both supporters and opponents.<sup>17</sup> Those insisting that “Hume is no egoist” (*verbatim*) include Michael Gill, who appeals to Hume’s belief “that humans possess ‘natural’ virtues, many of which are inherently sociable”;<sup>18</sup> Don Garrett, who points out that “the operation of sympathy guarantees that human beings are concerned for the pleasures and pains of others as well as their own”;<sup>19</sup> and David Owens, who says Hume “allows that human beings care for family and friends as well as for themselves [T 3.2.2.5].”<sup>20</sup> We can quickly address the first two of these points—on natural virtues and sympathy—but the third will require more detailed consideration.

<sup>16</sup> See also T 1.3.10.2, which seems clearly to have our own “pain and pleasure” in mind when describing these as our “chief spring and moving principle”. The same intimate connection with motivation is clearly asserted by Locke: if the “*perception of Delight*” that God “has been pleased to join” to various of our thoughts and sensations “were wholly separated from all our outward Sensations, and inward Thoughts, we should have no reason to prefer one Thought or Action, to another; . . . And so we should neither stir our Bodies, nor employ our Minds”; see John Locke, *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon Press, 1975 [1690/1700]), II vii 3 (the following section extends the same point to pain). Locke was himself following Thomas Hobbes, who identified *delight* with *pleasure* (*Leviathan* I vi 11) and was explicitly egoist: “of all voluntary acts, the object is to every man his own good” (*Leviathan* I xv 16); *Leviathan: or, The matter, forme, & power of a common-wealth ecclesiasticall and civill* (London, 1651).

<sup>17</sup> Other supporters include Rachel Cohon, *Hume’s Morality: Feeling and Fabrication* (Oxford: Oxford University Press, 2008), see pp. 31–5; Stephen Darwall, “Motive and Obligation in Hume’s Ethics.” *Noûs* 27 (1993): 423; and Mikael M. Karlsson, “Reason, Passion, and the Influencing Motives of the Will.” In *The Blackwell Guide to Hume’s Treatise*, ed. Saul Traiger (Oxford: Blackwell Publishing Co., 2006), pp. 246–7.

<sup>18</sup> Michael Gill, “Hume’s Progressive View of Human Nature.” *Hume Studies* 26 (2000): 87–108: 90.

<sup>19</sup> Don Garrett, *Hume* (Abingdon: Routledge, 2015), p. 114. Garrett’s sentence continues “. . . , and he recognizes other basic instinctual desires and aversions besides the desire for pleasure and aversion to pain.” This important additional point will also be addressed below.

<sup>20</sup> David Owens, “The Problem of Promising.” In *Promises and Agreements: Philosophical Essays*, ed. Hanoch Scheinman (Oxford: Oxford University Press, 2011), p. 72.

Hume does indeed recognize “sociable” natural virtues in the *Treatise*, but his explanation of them fits well with egoism, for it is based on reducing concern about society to concern about ourselves:

we have no . . . extensive concern for society but from sympathy; and consequently 'tis that principle, which takes us so far out of ourselves, as to give us the same pleasure or uneasiness in characters which are useful or pernicious to society, as if they had a tendency to our own advantage or loss. (T 3.3.1.11)

The effect of Humean sympathy is to replicate other people's perceived emotions in our own: when we see someone happy or sad, the operation of sympathy makes us feel a similar happiness or sadness. Sympathy thus makes us care about other people's pain or pleasure because these are *causes* of *our own* pain or pleasure (cf. T 2.3.3.3 above). But without the underlying assumption of egoism, this mechanism could be unnecessary: the prospect of others' pain or pleasure might motivate us *directly* (as Francis Hutcheson, for example, maintained), without first requiring that they be echoed in our own first-personal feelings through the operation of sympathy. Hence Hume's emphasis on sympathy as a central mechanism of the moral theory of the *Treatise*, so far from telling against the suggestion that this theory is fundamentally egoist, actually tells significantly in its favour.<sup>21</sup>

The *Treatise* does, however, acknowledge what looks like a notable anti-egoist exception, in that—as we saw Owens observing—we naturally feel benevolence towards those we love, and “anger” towards those we hate: “Love is always follow'd by a desire of the happiness of the person belov'd, and an aversion to his misery: As hatred produces a desire of the misery and an aversion to the happiness of the person hated.” (T 2.2.6.3).<sup>22</sup> So marked is this tendency that the passions of love and benevolence naturally *blend together*, “forming a compound” and generating a “uniform impression” (T 2.2.6.1). The same is true of hatred and anger. Conceptually, however, they remain distinct, and Hume himself seems to recognize that their association constitutes a departure from egoism, for he later remarks that resentment

<sup>21</sup> A different view is suggested by E. B. McGilvary, “Altruism in Hume's *Treatise*.” *Philosophical Review* (1903): 291–4 and John Bricke, *Mind and Morality: An Examination of Hume's Moral Psychology* (Oxford: Oxford University Press, 1996), pp. 129–35. They insist that Humean sympathy can communicate others' *desires* as well as pleasures or pains, thus potentially giving me a desire for another's well-being unmediated by my own pleasure and pain. Some of Hume's text indeed gives this impression, as when he talks about the sympathetic communication of “inclinations and sentiments” or “opinions and affections” (T 2.1.11.2, 7). These terms are vague, and could potentially cover many things, but in practice, Hume's explanations almost always seem to be in terms of the communication of pleasures and pains, and McGilvary is simply mistaken in claiming that “Hume accounts for pity [as] sympathy with another's pain or with his desire to rid himself of that pain” (p. 291, echoed by Bricke, p. 129). Hume's actual treatment of pity at *Treatise* 2.2.7.2–5 talks of “affliction and sorrow” being communicated by sympathy, not desire (and likewise the 1757 *Dissertation* says “compassion is an uneasiness in the *sufferings* of another,” DP 3.7, emphasis added). For more on this, see Amyas Merivale's excellent and groundbreaking recent study of Hume's mature theory of the passions, *Hume on Art, Emotion, and Superstition: A Critical Study of the Four Dissertations* (Abingdon: Routledge, 2019), pp. 58–9.

<sup>22</sup> The apparent contradiction between benevolence and egoism is not quite as straightforward as it might seem, however, since as Merivale points out (*Hume on Art, Emotion, and Superstition*, pp. 64–5), Hume at T 2.2.9.15 describes benevolence as “an original pleasure arising from the pleasure of the person belov'd,” thus providing a potential egoist explanation of benevolent behaviour.

against someone who injures me “often . . . makes me desire his evil and punishment, *independent of all considerations of pleasure and advantage to myself*” (T 2.3.3.9, emphasis added).

It seems fairly clear, then, that in respect of benevolence and hatred at least, the *Treatise* is not *consistently* egoist, and the question is whether this exception is sufficient to undermine my claim that it is *fundamentally* egoist. Potentially significant contextual evidence here is given by the paragraph immediately preceding the quotation above:

... I begin to be sensible . . . of a misfortune, that has attended every system of philosophy, with which the world has been yet acquainted. 'Tis commonly found, that in accounting for the operations of nature by any particular hypothesis; . . . there is always some phaenomenon, which is more stubborn, . . . the difficulty, which I have at present in my eye, is no-wise contrary to my system; but only departs a little from that simplicity, which has been hitherto its principal force and beauty. (T 2.2.6.2)

The rest of the section is devoted to arguing a theoretical point, that our tendency towards benevolence and anger, proportioned to our love and hatred rather than to any egoistic calculation, must be generated “by the original constitution of the mind” (T 2.2.6.6). It cannot apparently be accounted for by sympathy,<sup>23</sup> and “abstractedly consider'd,” there is no necessity to our minds’ working in this way. It is therefore just a brute fact of human nature; one that Hume apparently sees as being inconveniently out of line with his general theory.

The remarkable passage at T 2.2.6.2 is surprisingly little discussed in the literature, and its significance is debatable.<sup>24</sup> Read most naturally, however, Hume is expressing disappointment at the complication of his theory from having to add “original” and “unaccountable” principles connecting love with benevolence and hatred with anger (T 2.2.6.6, 2.3.9.8). One plausible explanation for such disappointment would be that he has hitherto been attempting to build his account on a simple egoist basis, a suggestion circumstantially supported by the striking fact that the inconsistency between egoism and benevolence provides the focus of *the only other topic-specific references to theoretical “simplicity” in Hume’s entire philosophical corpus* (at EPM

<sup>23</sup> Given the reluctance with which Hume acknowledges this complication in his theory, his failure even to consider whether it might be explicable through the sympathetic communication of others’ *desires* counts against McGilvary and Bricke’s thesis that Humean sympathy can operate in that way, especially since sympathy is immediately invoked in the following section to explain *pity* and *malice* in cases where there is no “friendship or enmity to occasion this concern or joy” (T 2.2.7.1–2).

<sup>24</sup> P. J. E. Kail, *Projection and Realism in Hume’s Philosophy* (Oxford: Oxford University Press, 2007) sees the difficulty as being that “two impressions can merge to produce another” (p. 141). But Hume expresses his worry as arising *after* he has explained blending, when he turns to examine “those ingredients, which are capable of uniting with love and hatred.” Moreover so far from being an unfortunate complication, the blending of emotions turns out to be an important unifying feature of his theory which explains hope and fear on a close analogy with his theory of probability (T 2.3.9.9–17; DP 1.7–10). John Passmore, *Hume’s Intentions* (London: Duckworth, 1980) instead suggests that Hume’s anxiety concerns his desire “to show . . . a precise parallel between love and pride, between hate and humility” (p. 127). Hume does indeed remark on the difference between these pairs (T 2.2.6.3), but without any obvious concern about the asymmetry, and this is not the main focus of the section.

App 2.6–7, 12).<sup>25</sup> If this suggestion is correct, then the incompatibility of his account of benevolence with pure egoism serves to confirm, rather than undermine, the straightforward view that the theory of the *Treatise* starts from an egoist basis.

Having recognized benevolence and anger as exceptions to his general theory, Hume seems to have edited his account of the direct passions to reflect this, though apparently rather clumsily:

Beside good and evil, or in other words, pain and pleasure, the direct passions frequently arise from a natural impulse or instinct, which is perfectly unaccountable. Of this kind is the desire of punishment to our enemies, and of happiness to our friends; hunger, lust, and a few other bodily appetites. These passions, properly speaking, produce good and evil, and proceed not from them, like the other affections. (T 2.3.9.8)

This passage starts with a bald assertion of hedonism—that “good” and “evil” are to be identified with “pleasure” and “pain” respectively—but then goes on to observe that we sometimes desire other things owing to an “unaccountable” natural instinct (presumably alluding here to the discussion at T 2.2.6 that we have just examined). These instinctive desires can “produce good and evil” (i.e. pleasure and pain), presumably through the satisfaction or frustration that we experience when they are fulfilled or fail. But such an account is strained: an expectation of pleasure and pain is not essential to these desires (e.g. we can strive towards a future goal—such as the eventual defeat of an invading enemy—whose fulfilment we do not expect to see), and indeed if such an expectation were essential, then they would not be “perfectly unaccountable.” So recognizing these “unaccountable” desires is in serious tension with the crude hedonism of treating “pleasure” and “pain” as “other words” for good and evil.

The passage just quoted acknowledges that we have a natural instinct not only to desire “punishment to our enemies” and “happiness to our friends,” but also “hunger, lust, and a few other bodily appetites.” It is tempting to speculate that Hume’s acknowledgement of these further instincts was prompted by his work on Section 2.2.11 of the *Treatise*, “Of the amorous passion, or love betwixt the sexes,” which recognizes the phenomenon of bodily *appetite*, and describes hunger as a “primary inclination of the soul” (T 2.2.11.3). The absence of parental affection from the list may seem surprising, given Hume’s statement in the immediately following section that “The affection of parents to their young proceeds from a peculiar instinct in animals, as well as in our species” (T 2.2.12.5).<sup>26</sup> But the isolation of this statement—as a paragraph in itself—rather suggests that it was a last-minute insertion. Hume’s initial explanation of such affection at T 2.2.4.2 had subsumed it under

<sup>25</sup> In §6 below, the relevant passages are quoted from EPM 2.6 and 2.12 in the original 1751 edition. More general comments on theoretical “simplicity” are at T 1.3.16.3 and EHU 4.12 (cf. T 2.1.3.6–7, 3.1.2.6; A 1; EPM 1.10). Hume does elsewhere remark on how some of his own theories are simple (e.g. T 2.1.12.1; EPM 5.45, 9.1), particularly when emphasising their ease of application to the thoughts and feelings of animals (e.g. T 2.1.12.9, 2.2.12.1, cf. 1.3.16.8).

<sup>26</sup> This section “Of the love and hatred of animals” also notes that animals—like humans—feel affection or enmity according to how they are treated (T 2.2.12.3). The analogy between animals and humans is a strong theme in Hume’s thought, both in the *Treatise* and the two *Enquiries* (cf. also T 1.3.16, 2.1.12, 2.3.9.32, EHU 9, EPM App. 2.8–9, 13).

his general account “of the love of relations” owing to closeness of connexion, so there is clear evidence of development in his view here.<sup>27</sup> When he later describes the “calm passions” that result from original instincts, he lists “the love of life, and kindness to children” alongside “benevolence and resentment” (T 2.3.3.8).<sup>28</sup> But nowhere in the *Treatise* do we find Hume systematically bringing together these various lists (presumably a symptom of the hasty publication that he later regretted). We must look to the second *Enquiry* for the most comprehensive catalogue, citing “hunger, thirst, and other appetites, resentment, love of life, attachment to offspring, and other passions” as all arising “from a simple original instinct in the human breast, which nature has implanted for like salutary purposes” (EPM 3.40). But by then, as we shall see, Hume had completely abandoned the initial egoism of the *Treatise*.

## 6. Rejecting Moral Scepticism and Egoism

Although Hume himself displays no such subversive intention, it is not surprising that the fundamentally secular and egoist theory of the *Treatise* was viewed by his critics as being seriously dangerous to morality. The hostile 1745 pamphlet mentioned in §4 above charged him “With sapping the Foundations of Morality, by denying the natural and essential Difference betwixt Right and Wrong, Good and Evil, Justice and Injustice; making the Difference only artificial, and to arise from human Conventions and Compacts” (LFG 19).<sup>29</sup> Prominent amongst the passages cited to press this charge are one attributing the origin of justice to “*the Selfishness and confined Generosity of Men*” (LFG 10, cf. T 3.2.2.18) and another affirming “that there is no such Passion in human Minds, as *the Love of Mankind* merely as such, independent of personal Qualities, of Service or of Relation to ourself” (LFG 10, cf. T 3.2.1.12). Both have an obvious basis in the egoism of the *Treatise* and the limited boundaries of Humean sympathy.<sup>30</sup>

In stark contrast with the *Treatise*, the second *Enquiry* of 1751 highlights from the start Hume’s determination to repudiate any accusation of moral scepticism or egoism. In the first two paragraphs of Section 1, he insists that “Those who have refused the Reality of moral Distinctions . . . really do not believe at all the Opinion they defend” (1751, 1.1–2). Then Section 2 starts by outlining two types of egoist

<sup>27</sup> The change was permanent, as shown by *Essays* 162–3 (“The Sceptic,” para. 10), EPM 3.40, and DP 3.3, n. 4.

<sup>28</sup> He also mentions a different form of calm passion, “the general appetite to good, and aversion to evil, consider’d merely as such,” though this leaves inexplicit what understanding of “good” and “evil” he has in mind here.

<sup>29</sup> This may partly explain why the artificiality of justice—so prominent in the *Treatise*—is mentioned as such just once in the second *Enquiry*, and merely in a footnote to an appendix (EPM App. 3.9, n. 64). But see also note 45 below for a principled reason why the distinction between natural and artificial virtues becomes philosophically less significant in Hume’s later moral theory.

<sup>30</sup> The second *Treatise* passage continues “Tis true, there is no human, and indeed no sensible, creature, whose happiness or misery does not, in some measure, affect us, when brought near to us, and represented in lively colours: But this proceeds merely from sympathy, and is no proof of such an universal affection to mankind, . . .” (T 3.2.1.12).

view, one (commonly associated with Mandeville) that cynically dismisses all would-be moral concern as hypocrisy, and the other—less extreme—

much insisted on by Philosophers, . . . that whatever Affection one may feel, or imagine he feels for others, no Passion is, or can be disinterested; that the most generous Friendship, however sincere, is a Modification of Self-love; and, that even unknown to Ourselves, we seek only our Gratification, while we appear the most deeply engag'd in Schemes for the Liberty and Happiness of Mankind. (1751, 2.2).<sup>31</sup>

Hume includes Epicurus, Hobbes, and Locke amongst those professing this “selfish System of Morals” (1751, 2.3), and goes on to explain how such a theory need not imply a denial of “the Reality of moral Distinctions” (which he has already dismissed as incredible):

AN *Epicurean* or a *Hobbist* readily allows, that there is such a Thing as Friendship in the World, . . . tho' he may attempt, by a philosophical Chymistry, to resolve the Elements of this Passion . . . into those of another, and explain every Affection to be Self-love, twisted and moulded into a Variety of . . . Appearances. But . . . even according to the selfish System, . . . I esteem the Man, whose Self-love . . . is so directed as to give him a Concern for others, and render him serviceable to Society: As I hate or despise him, who has no Regard to any Thing beyond his own pitiful Gratifications and Enjoyments. In vain would you suggest, that these Characters . . . are, at the Bottom, the same. (1751, 2.4)

Having thus observed that the status of the selfish hypothesis is “not so material, as is usually imagin'd, to Morality and Practice” (1751, 2.5),<sup>32</sup> Hume goes on nevertheless to advance a battery of objections to it, some of which we have seen anticipated in the *Treatise*.

The most obvious objection to the egoist theory is its mismatch with common experience of human affections, many of which seem plainly unselfish, and have only been supposed otherwise “from that Love of *Simplicity*, which has been the Source of much false Reasoning in Philosophy” (1751, 2.6, cf. §5 above).<sup>33</sup> Other strong considerations come from the analogy with animals, both their naïve kindness which cannot plausibly be attributed to “refin'd Deductions of Self-interest” (2.8), and the manifest “Tenderness to their Offspring” observed “in all sensible Beings” (2.9), our own experience of which is again manifestly contrary to the selfish theory: “What Interest can a fond Mother have in View, who loses her Health by assiduous Attendance on her sick Child, and afterwards languishes, and dies for Grief, when freed, by its Death, from the Slavery of that Attendance?” (1751, 2.9). Moving on to a more fundamental objection, Hume observes that certain “bodily Wants or Appetites,” and also “mental passions” such as the desire for fame, power, or vengeance, are directed immediately toward specific objects, rather than involving a merely

<sup>31</sup> This and the following quotations are taken from the 1751 first edition of the second *Enquiry*, where the discussion of egoism occupies Part 1 of Section 2. For the 1777 posthumous edition, Hume moved this discussion into Appendix 2, which is where the final versions of these passages are now to be found. The text is largely unchanged (apart from capitalization), and the paragraph numbers correspond exactly.

<sup>32</sup> An observation perhaps partly intended to acquit the *Treatise* of “sapping the Foundations of Morality.”

<sup>33</sup> Again, contrast this sceptical observation with the lament for lost simplicity at T 2.2.6.2, quoted in §5.

indirect desire motivated by a quest for pleasure. In such cases, the directed desire precedes—and its satisfaction explains—the pleasure, rather than the anticipated pleasure generating the desire. Thus egoism puts the cart before the horse:

Nature must, by the internal Frame and Constitution of the Mind, give an original Propensity to Fame, 'ere we can reap any Pleasure from it, or pursue it from . . . Self-love, and a Desire of Happiness. If I have no Vanity, I take no Delight in Praise: If I be void of Ambition, Power gives no Enjoyment: If I be not angry, the Punishment of an Adversary is totally indifferent to me. (1751, 2.12)

Once this key point has been accepted—that we can, and do, directly desire some things for themselves, rather than merely as means to pleasure—it becomes clear that there is no “Difficulty of conceiving, that this may likewise be the Case with Benevolence and Friendship, and that, from the original Frame of our Temper, we may feel a Desire of another’s Happiness or Good” (1751, 2.13).<sup>34</sup> Having broken the spell of the egoist assumption that we can only be motivated by our own pleasure and pain, we are free to open our eyes to the obvious empirical evidence that we do in fact desire the good of others (most obviously family and friends), and not merely for selfish reasons. Indeed, from this perspective it becomes clear that “if we consider rightly of the Matter, . . . the Hypothesis, which allows of a disinterested Benevolence, distinct from Self-love, has really more *Simplicity* in it, and is more conformable to the Analogy of Nature, than that which pretends to resolve all Friendship and Humanity into this latter Principle” (1751, 2.12). Even theoretical simplicity ultimately fails to tell in favour of the selfish theory, which has by now been comprehensively refuted.

Although most of these objections can be seen as developing from points first made in the *Treatise*, it seems that Hume’s decisive turn away from egoism had a distinct and specific inspiration which is identified in the first *Enquiry* (rather than the second). A footnote appended to EHU 1.14 in the 1748 and 1750 editions clearly acknowledges the “cart before the horse” point above:

It has been prov’d, beyond all Controversy, that even the Passions, commonly esteem’d selfish, carry the Mind beyond Self, directly to the Object; that tho’ the Satisfaction of these Passions gives us Enjoyment, yet the Prospect of this Enjoyment is not the Cause of the Passion, but on the contrary the Passion is antecedent to the Enjoyment (1748, 1.14n.).

In the previous sentence there is a further note “See *Butler’s Sermons*,” thus crediting Joseph Butler with this point, and indeed the argument of his Sermon XI “Upon the Love of our Neighbour” has long been considered the classic refutation of naïve egoism. Here we have strong evidence that reading Butler ultimately persuaded Hume that egoism was not only wrong in detail but fundamentally misguided. If so, that reading had presumably occurred by the time he wrote the essay “Of the Dignity or Meanness of Human Nature,” published in 1741, which makes the same

<sup>34</sup> Or indeed another’s harm, in the case of “Vengeance, [which] may be so eagerly pursued, as to make us knowingly neglect every Consideration of Ease, Interest or Safety” (1751, 2.13). Note that this discussion—focused on attacking egoism—has culminated in the same conceptual nexus as T 2.2.6.3, again somewhat suggesting that Hume’s theoretical anxiety at T 2.2.6.2 is itself associated with egoism.

key point.<sup>35</sup> And it is even possible that a preliminary reading of the sermons prior to 1739—perhaps stimulated by Hume's high opinion of Butler's 1736 *Analogy of Religion*—lay behind the clumsily edited paragraph at T 2.3.9.8 which talks of natural passions that “properly speaking, produce good and evil, and proceed not from them, like the other affections.”<sup>36</sup>

However, this might be, it seems clear that once Hume had fully absorbed Butler's point, he not only accepted it but reinforced it with further arguments, turning completely against the initial egoism of the *Treatise* and thus switching sides on one of the most central debates in moral philosophy. Having declared his position on the matter clearly and prominently, at the beginning of Section 2 of his 1751 moral *Enquiry* (now Appendix 2), he removed from the first *Enquiry* the footnote that provides such a useful clue regarding a key source of this fundamental change of mind.

## 7. Sympathy, Humanity, and Associationist Explanation

Hume's abandonment of egoism, not surprisingly, leaves significant traces in his treatment of *sympathy* and *humanity*, terms that are initially quite distinct but which he ultimately blurs. In the *Treatise*, sympathy is an association-based mechanism whereby the “inclinations and sentiments” of others are communicated to us (T 2.1.11.2). Its explanatory basis is that our lively ideas of others' “passions and sentiments” are close to impressions, and they can be enhanced through association so that “the ideas of the affections of others are converted into the very impressions they represent” (T 2.1.11.7–8; cf. also 2.1.11.14–18, 2.2.4.6–7, 2.3.6.8). This mechanism of sympathy plays a huge role in Books 2 and 3 of the *Treatise*—being mentioned in no fewer than eighty-five paragraphs—and amongst many other functions it explains our “natural humanity” or general concern for “the happiness of [our] fellow-creatures” (T 3.3.1.12). Any such humanity is imperfect and partial, for as the author of the hostile 1745 pamphlet pointed out (see §6 above), Hume insists at

<sup>35</sup> “The virtuous sentiment or passion produces the pleasure, and does not arise from it. I feel a pleasure in doing good to my friend, because I love him; but do not love him for the sake of that pleasure.” (E-DM, 85–6).

<sup>36</sup> See Merivale, *Hume on Art, Emotion, and Superstition*, pp. 71–2. This speculation from Merivale is chronologically plausible. But it is unclear whether Hume's wording genuinely manifests a grasp of Butler's “cart before the horse” point (as the 1741 quotation in footnote 35 above somewhat corroborates), or is instead a relatively ad hoc attempt to force the recalcitrant phenomena of benevolence, resentment, and bodily appetites (whose independent recognition in the *Treatise* we have traced in §5 above) into the theoretical framework of direct passions that are supposedly “founded on pain and pleasure” (T 2.3.9.1). Merivale's speculation also requires that Hume could have become dimly aware of Butler's point without grasping it sufficiently to appreciate its potentially devastating impact on the more fundamental egoism of the *Treatise*. Though unconvinced on this particular issue, I should like to acknowledge with thanks many discussions with Merivale that convinced me of a fundamental change in Hume's view, contrary to the spirit of my paper “Hume on Ethics.” In Tom Angier ed., *Key Thinkers: Ethics* (London: Bloomsbury, 2012), which—while focusing mostly on the second *Enquiry*—stated that “The overall moral theory of the two works is the same” (p. 106) and cited the T 2.3.9.8 paragraph as evidence that even in the *Treatise*, “Hume is no psychological egoist” (p. 128, n. 29).

T 3.2.1.12 “that there is no such passion in human minds, as the love of mankind, merely as such, independent of personal qualities, of services, or of relation to ourself.” And our imperfect humanity, such as it is, plays no very significant role in the *Treatise*, being mentioned in only seven paragraphs, apparently just one virtue amongst others (T 3.3.1.24, 3.3.3.3–4) and doubly unreliable as a moral criterion since it is “often . . . contrary to the laws of justice” (T 3.3.1.12).

In the second *Enquiry*, by contrast, “sympathy” is mentioned in only twenty-two paragraphs, and the term seems generally to be used in a more everyday sense—for a sentiment of fellow-feeling rather than a mechanism—and independently of the specific theory of the *Treatise*. Only the following passage comes close to speaking of the literal vivacity transfer characteristic of that theory:<sup>37</sup> “Bring . . . virtue nearer, by our acquaintance or connexion with the persons, or even by an eloquent recital of the case; our hearts are immediately caught, our sympathy enlivened, and our cool approbation converted into the warmest sentiments of friendship and regard” (EPM 5.43). Soon after, and elsewhere, “sympathy” seems to be closely linked and even virtually equated with “humanity,” involving “not . . . any regards to self-interest, but . . . a tendency to public good” (EPM 5.45, cf. 9.12). Sometimes, indeed, they seem to be identified as a single sentiment (1751, 2.5n.; EPM 6.3). Moreover “humanity” now becomes the dominant term, mentioned in no fewer than 48 paragraphs of the second *Enquiry*.<sup>38</sup> This stylometric change might be thought merely presentational, given Hume’s continuing emphasis in the central Section 5 (entitled “Why Utility Pleases”) on feelings of a broadly sympathetic nature. But there are at least three reasons to suggest that it runs much deeper, reflecting his theoretical turn against egoism. First, where the *Treatise* had denied any “*Love of Mankind* merely as such,” Hume now speaks of “a generous concern for our kind and species” (EPM 2.5), “a warm concern for the interests of our species,” “a general approbation of what is useful to society,” and “our natural philanthropy” (EPM 5.39–40). Secondly, in the Conclusion of the *Enquiry* (EPM 9.6–9, cf. also 6.5), this broad humanity is eulogized as the basis of moral sentiments and judgements, being “the same principle” in “all mankind” and giving rise to “universal sentiments of censure or approbation.”<sup>39</sup> It is these universal sentiments on which our language is

<sup>37</sup> At EPM 7.2 and 7.21, Hume speaks of sympathy as a “contagion,” which is suggestive of communication of emotion from one person to another, but silent regarding the mechanism. There are likewise various other passages that are reminiscent of Hume’s theory of sympathy from the *Treatise*, but none so specific as to warrant any strong interpretative conclusions regarding how much of that theory he retained.

<sup>38</sup> Thus, while paragraphs referring to “sympathy” decline by a factor of nearly four times from *Treatise* Books 2 and 3 to the second *Enquiry*, those referring to “humanity” multiply by a factor of very nearly seven times. The *Dissertation on the Passions*, incidentally, includes only three paragraphs referring to “sympathy” (none to “humanity”), with only one reminiscent of the specific theory of the *Treatise* (DP 3.4, cf. 2.33 and 3.11).

<sup>39</sup> In the *Treatise*, Hume attempts to provide a mechanism for standardization of moral sentiments—and thus reliable moral judgement—based on sympathy with the agent’s “narrow circle” of acquaintance (T 3.3.3.2, cf. 3.3.1.18, 3.3.1.30). This may be the best he can do within a predominantly egoist theory, but as Jacqueline Taylor points out, it is morally inadequate, since such a narrow circle will commonly display partiality and prejudice. See Jacqueline Taylor, “Hume on the Standard of Virtue.” *The Journal of Ethics* (2002): 52–6; and *Reflecting Subjects: Passion, Sympathy, & Society in Hume’s Philosophy* (Oxford: Oxford University Press, 2015), pp. 113–16. Hume himself clearly recognizes this risk in his 1757 essay “Of the Standard of Taste” (paras 1, 11; E-ST, 226–7, 233).

“moulded” (EPM 9.8), thus establishing the linguistically agreed standard of moral thought and communication from which the *Enquiry's* investigation began (at EPM 1.10). Thirdly, Hume explicitly contradicts the theoretical aspirations of his earlier account, which as we saw above both explained (limited) humanity as arising from sympathy, and attempted to explain sympathy itself, whereas now Hume dismisses such attempts as unnecessary and probably futile:

It is needless to push our researches so far as to ask, why we have humanity or a fellow-feeling with others.... We must stop somewhere in our examination of causes;...No man is absolutely indifferent to the happiness and misery of others. The first has a natural tendency to give pleasure; the second, pain.... *It is not probable, that these principles can be resolved into principles more simple and universal, whatever attempts may have been made to that purpose.*

(EPM 5.17, n. 19, emphasis added)

So an associationist explanation that had been postulated in the *Treatise* is disowned in the second *Enquiry*, not just implicitly by omission, but through explicit denial. Presumably this change reflects, at least in part, Hume's awareness that sympathy—as interpreted in the *Treatise*—is unable to do the theoretical job required.<sup>40</sup> In the *Treatise*, starting from an egoist foundation, he had based morality on sympathy and argued (very reasonably) that this would be unable to ground any genuinely universal humanity. Now in the *Enquiry*, he recognizes the manifest fact of universal humanity (for example, in his own feelings), identifies this as the true basis of moral thought and language, and consistently draws the conclusion that the mechanism of sympathy cannot provide the requisite basis for morals.

Although quite distinct from Hume's moral theory, there is a strikingly close parallel to this last development—the abandonment of an associationist explanation for good theoretical reasons—within the first *Enquiry*. There, Sections 5 and 6 sketch his theory of belief and probability as founded on custom, but seem significantly less ambitious than the corresponding (and much more extensive) sections of the *Treatise*. No attempt is now made to *define* belief as the *Treatise* had done; indeed, the possibility of any such definition is explicitly denied (EHU 5.12, cf. T 1.3.7.5). In Part 2 of Section 5, Hume sets out modestly to find “some explications and analogies, that will give satisfaction,” but he prefaces the discussion by remarking that it can be entirely ignored by “readers of a different taste” since “the following enquiries may well be understood, though it be neglected” (EHU 5.9). Its upshot is the relatively vague claim that “belief, where it reaches beyond the memory or senses, is of a similar nature, and arises from similar causes, with the transition of thought and vivacity” that characterizes association of ideas—we thus have “some analogies, by which [belief] may be explained” (EHU 5.20). All this is in contrast, especially as applied to

<sup>40</sup> See also EPM 6.3, which argues that sympathetic identification cannot explain our approval of qualities useful to another person by attempting to reduce this to “*self-love*,” and EPM 5.13, which describes as a “weak subterfuge” the appeal to imaginative identification with “distant ages and countries.” If the interpretation above is correct—that such reductive explanation was precisely the intended role of sympathy in the *Treatise*—then these paragraphs may amount to an explicit renunciation of that account, replacing it with “a quite different principle, which... interests us in the felicity of the person whom we contemplate” (EPM 6.3). Hume ends this paragraph talking of “a pleasing sentiment of sympathy and humanity,” but this cannot be sympathy as understood in the *Treatise*.

probability, with the more literally hydraulic *Treatise* account, in which force and vivacity seem to be conveyed from the initial “impulse” given by the relevant impression either to the believed idea (T 1.3.8.2) or to be divided amongst the various associated ideas (before potentially being recombined according to the similarity of the outcomes, as at T 1.3.11.12 and 1.3.12.10–11). Indeed, the *Enquiry* explicitly shuns any attempt to explain the precise mechanism, talking of “an inexplicable contrivance of nature” (EHU 6.3). And the *Appendix* to the *Treatise* indicates that even in 1740 Hume already had serious doubts about his earlier theory: the inserted paragraph T 1.3.7.7, and App. 2–9, prefigure the discussion in the *Enquiry* which inherits text from them. In other respects, also, the *Enquiry* account manifests positive development rather than mere abridgement: for example, in contrast to the corresponding discussions in the *Treatise* (T 1.3.8 and 1.3.9), Hume is careful at EHU 5.18–20 to distinguish quite clearly between *association of ideas* through the relation of causation (e.g. I see a picture and think of its painter) and *customary inference* (e.g. I see a fire and infer that it will be hot). The latter is *analogous* to association, but is not straightforwardly an instance of it.

It is a familiar observation that the recast versions of the *Treatise*—the two *Enquiries* and the *Dissertation on the Passions*—are more elegantly written and less densely systematic than the original, treating issues without attempting to place them all within a comprehensive account of the mind. It is tempting to link these stylistic and systematic changes: to assume that the omission of theoretical detail is primarily for presentational reasons—to provide a more attractive and popular exposition—rather than representing any significant change of view. But in the two cases we have briefly discussed here, as with his more fundamental sceptical concerns, we can trace a convincing *philosophical* source of Hume’s dissatisfaction with his earlier view: in the case of belief, through the reservations apparent in the 1740 *Appendix*; and in the case of sympathy and humanity, through the various texts that reveal his increasing realization that general human benevolence is genuine and not merely an artefact of refined self-interest. This is not the place to argue the point in detail, but my own view is that in most other cases too, the development from the *Treatise* to the *Enquiries* represents philosophical progress and is rarely a matter of mere abridgement.<sup>41</sup> When Hume omits earlier material, he usually does so for very good reason. He also, of course, reworks much of the material he preserves, and here too—as we have already seen and will soon see further—there are clear indications of philosophical as well as stylistic change.

## 8. Causation and Morality: Methodology, Definitions, “Projection”

We come finally to examine a particularly interesting—but controversial—philosophical link between the two *Enquiries*, concerning the relationship between

<sup>41</sup> Regarding the *Dissertation on the Passions*, see Amyas Merivale, “Hume’s Mature Account of the Indirect Passions.” *Hume Studies* 35 (2009): 185–210 and *Hume on Art, Emotion and Superstition*. Merivale demonstrates that even the *Dissertation*, which is commonly thought of as a lightweight summary of *Treatise* Book 2, displays significant philosophical development and improvement.

Hume's causal metaphysics (as developed in Section 7 of the first *Enquiry* and applied to the question of liberty and necessity in Section 8), and his moral metaphysics (as addressed in Sections 1 and 9 of the second *Enquiry*, but especially Appendix 1). There are parallels between Hume's discussions of these issues at various expository and methodological levels, but not all of these are straightforward or obviously intentional. Starting at the level of superficial presentation, we see the rhetoric of reconciliation—between liberty and necessity, and between the claims of reason and sentiment—as already discussed in §3 above. Hume's route to reconciliation in each case involves focusing on the relevant language and terms, on which basis he claims to find underlying agreement amongst all parties once matters are correctly understood (EHU 8.3; EPM 1.10). In the second *Enquiry*, Hume goes on to explain that by building his theory on a catalogue of virtues and vices as manifested empirically in our language and the attitudes it enshrines—"a question of fact, not of abstract science"—he is "following the experimental method, and deducing general maxims from a comparison of particular instances."<sup>42</sup> His aim is "to discover the circumstances on both sides, which are common" to the relevant qualities, those "in which the estimable qualities agree on the one hand, and the blameable on the other; and thence to reach the foundation of ethics, and find those universal principles, from which all censure or approbation is ultimately derived."<sup>43</sup> Hume's enthusiasm for the experimental method here might perhaps intend some contrast with the aprioristic argument of *Treatise* 2.3.3.5 and 3.1.1.9 (briefly discussed in §3 above), which seems to draw its conclusion from "a general abstract principle," a method which Hume now remarks "is a common source of illusion and mistake" (EPM 1.10). But the *Treatise* had advertised itself (in its subtitle) as "an attempt to introduce the experimental method of reasoning into moral subjects," so in this respect we have at most a correction rather than a fundamental change of course. As one might expect, the first *Enquiry*—with its very explicit distinction between *relations of ideas* and *matters of fact*—places moral investigation in the latter category, as requiring empirical study of such facts as "the general taste of mankind" (EHU 12.33). Perhaps when he wrote this Hume had already envisaged the compilation of his catalogue of virtues and vices, using language as his guide. If so, then again we find in the first *Enquiry* a clue regarding the development of his moral theory between the *Treatise* and the second *Enquiry*.

The deepest methodological parallel between Hume's treatment of necessity in the first *Enquiry*, and morality in the second, involves his appeal to definition, and in particular, his distinctive use of *paired* definitions. The two definitions of cause at EHU 7.29 are very familiar, although their interpretation is notoriously

<sup>42</sup> See also EPM App. 1.10, in which Hume emphasizes that his method is "to examine a plain matter of fact, to wit, what actions have this influence" [of eliciting approval or disapproval].

<sup>43</sup> Hume's method of appeal to linguistic consensus is more problematic than he seems to realize here, since agreement on the colouring of moral terms can hide disagreement regarding their appropriate application, e.g. one person's "courage" might be another's "rashness." See my "Hume on Ethics," pp. 109–12 for more on this, and note Hume's later apparent appreciation of this point in the first five paragraphs of his 1757 essay "Of the Standard of Taste."

controversial.<sup>44</sup> Hume's main motive for uncovering these definitions is revealed in Section 8, where they provide the key to understanding the notion of necessity and thus defending "the doctrine of necessity" in the human sphere (EHU 8.5, 8.27). In the second *Enquiry* (but not in the *Treatise*), Hume provides two definitions of "personal merit" or "virtue," though these occur in different places rather than explicitly paired.<sup>45</sup> Ordering these to match with the two definitions of cause, the "second" definition (EPM 8.0, n. 50, App. 1.10) focuses on *the mental phenomenon that unifies and confers significance on the relevant circumstances* (in the case of causal necessity, this is the *customary inference of the mind*; in the case of virtue, the *sentiment of approbation*). The first definition (EPM 9.1, 9.12) is derived from the second by discovering what is common to those circumstances—*what observations do in fact generate the mental phenomenon in question* (in the case of causal necessity, this is the observation of *constant conjunction*). The methodological route from the second definition to the first is spelled out in the following passage:

[Our hypothesis] . . . defines virtue to be *whatever mental action or quality gives to a spectator the pleasing sentiment of approbation*; and vice the contrary. We then proceed to examine a plain matter of fact, to wit, what actions have this influence: We consider all the circumstances, in which these actions agree: And thence endeavour to extract some general observations with regard to these sentiments. (EPM App. 1.10)

The resulting "general observations" are, of course, encapsulated in the first "delineation or definition of PERSONAL MERIT": "every quality of the mind, which is *useful* or *agreeable* to the *person himself* or to *others*, communicates a pleasure to the spectator, engages his esteem, and is admitted under the honourable denomination of virtue or merit" (EPM 9.12).

Much of the controversy over Hume's two definitions of cause arises from the mismatch between them: some constant conjunctions are never observed (and hence give rise to no inference of the mind); some causal inferences are made in

<sup>44</sup> A useful summary of the issues and the variety of scholarly views is given by Don Garrett, *Cognition and Commitment in Hume's Philosophy*, (New York: Oxford University Press, 1997), pp. 96–101. For my own preferred account, see Peter Millican, "Hume, Causal Realism, and Causal Science." *Mind* 118 (2009): 647–712 and especially §4 (but with the caveat expressed in note 52 below), and *Hume on Causation and Free Will* (Oxford: Oxford University Press, expected 2021–22).

<sup>45</sup> The nearest parallel in the *Treatise* is at T 3.3.1.30, where Hume offers "a general review of the present hypothesis [concerning *natural* virtues and vices]: Every quality of the mind is denominated *virtuous*, which gives pleasure by the mere survey . . . This pleasure . . . may arise from four different sources. For we reap a pleasure from the view of a character, which is naturally fitted to be useful to others, or to the person himself, or which is agreeable to others, or to the person himself." He goes on to stress the importance of "some common point of view . . . in judging of characters," which turns out to be founded on sympathy with what T 3.3.3.2 will call the "narrow circle" of the agent's acquaintance (cf. note 39 above). But perhaps because he is looking for an account that can yield (indirectly self-interested) motivation via sympathetically induced impressions, Hume seems not to have realized at this stage that the four criteria can be applied more globally—and used as a definitional standard for the *correction* of moral judgements—since their satisfaction potentially provides an objective basis which escapes from the need for genuine sympathetic engagement (such as might indeed plausibly be limited to such a narrow circle). This move also enables integration between Hume's treatment of the "natural" and "artificial" virtues, since both then turn out to be answerable to the same general criteria. So here again we see significant and systematic changes in Hume's moral theory, consequent on his crucial transition away from egoism and towards acknowledgment of a broad "humanity."

circumstances that do not correspond to any genuine constant conjunction. The parallel with his two definitions of virtue provides valuable illumination here, because there is a famous passage in the second *Enquiry* where Hume explicitly faces up to a situation in which those two definitions come apart. Many people approve of “monkish virtues” such as celibacy, fasting, penance, mortification, self-denial, humility, silence, and solitude. But these all fail the test of Hume’s first definition, and so he advocates classifying them instead as vices (EPM 9.3). In other words, having derived his “general observations” regarding the circumstances that standardly generate approbation in the spectator, Hume advocates *revising* our sensitivity and judgements of approbation in such a way as to conform to those general rules. His first definition of “personal merit,” once it has been derived, thus comes to overrule the second,<sup>46</sup> a pattern that indeed seems to hold also in respect of causation, where Hume repeatedly treats constant conjunction as the key criterion (e.g. T 1.3.15.1, 1.4.5.32) and even as “the very essence” of causal necessity (T 1.3.14.16, 1.4.5.33, 2.3.1.10, 2.3.2.2; EHU 8.22n., 8.25n.).

Moving now from methodology to metaphysics, Hume’s theories of causation and morality are often said to involve “projection,” in that causal and moral judgements attribute to objects or events properties whose nature is fundamentally mental or mind-derived. The two passages most commonly quoted in this connection are T 1.3.14.25, which observes that “the mind has a great propensity to spread itself on external objects,” and EPM App. 1.21, which says that moral *taste* “has a productive faculty, and gilding or staining all natural objects with the colours, borrowed from internal sentiment, raises, in a manner, a new creation.” But putting these two passages together as exemplifying some consistent form of “Humean projection” is very questionable, for Hume’s own attitude to them seems to be quite different. The sentence before the *Treatise* passage—“This contrary bias is easily accounted for”—makes clear that he is here *explaining away an erroneous objection to his theory of necessary connexion*, rather than presenting a positive “projectivist” account.<sup>47</sup> In so far as people “project” their inferential “feeling” onto objects, he sees them as making an error which biases them against accepting his correct account. The first *Enquiry* likewise treats as erroneous any tendency that we might have to “transfer . . . to the objects” an internal feeling such as the “sentiment of a *nisus* or endeavour,” which “enters very much into [the] vulgar, inaccurate idea” of power (EHU 7.15, n.13, 7.29, n. 17).<sup>48</sup> This deprecation of causal “projection” contrasts with the passage about moral taste in the second *Enquiry*, which appears

<sup>46</sup> Yet again we find a contrast with the *Treatise*, which says: “The distinction of moral good and evil is founded on the pleasure or pain, which results from the view of any sentiment, or character; and as that pleasure or pain cannot be unknown to the person who feels it, it follows, that there is just so much vice or virtue in any character, as every one places in it, and that ’tis impossible in this particular we can ever be mistaken” (T 3.2.8.8).

<sup>47</sup> Likewise, the following sentences in T 1.3.14.25 make clear that such “projection” is an *error*, comparable with attributing spatial location to sounds and smells (Hume alludes here to T 1.4.5.11–14).

<sup>48</sup> The latter footnote gets closest to T 1.3.14.25, saying that in the case of “energies, which are exerted, without our annexing to them any idea of communicated motion” (and which therefore involve no *nisus*), “we consider only the constant experienced conjunction of the events; and as we feel a customary connexion between the ideas, we transfer that feeling to the objects; as nothing is more usual than to apply to external bodies every internal sensation, which they occasion.” It is clear in the *Enquiry* that Hume

to endorse and celebrate its remarkable power to raise “a new creation” through the projection of “internal sentiment.”

The difference in Hume’s attitude here might seem puzzling, given the striking similarities between his treatment of causation and of morality. In both cases we have distinctive “impressions”—of “power or necessary connexion” (EHU 7.5 ff.) and “approbation or blame” (EPM App. 1.11) respectively. In both cases we have two definitions, one referring explicitly to the occurrence of such impressions, the other apparently aiming to systematize what is in common to the circumstances in which those impressions typically arise. In both cases we have a policy of conforming our judgements to the relevant systematized definition, through procedures such as the “rules by which to judge of causes and effects” (T 1.3.15), “resolv[ing] the many particular effects into a few general causes” (EHU 4.12), assessing consequences from a “common point of view” (T 3.3.1.30, EPM 9.6, cf. T 3.3.1.15), applying principles of justice that are artificially designed to foster “general utility” (EPM App. 3.11; cf. T 3.2.2.22; EPM 3.12, 3.47–8), and so forth. Hume’s description of the careful analysis required for proper moral judgement could apply equally well to causal judgement: “in order to . . . give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained” (EPM 1.9). But in fact such similarity is only to be expected, because Hume is here outlining the role of *reason* in moral judgement, in the context of *distinguishing* it from the role of *sentiment*.<sup>49</sup> As we saw in §3 above, this explicit recognition of the key role of reason in morality involves a significant change—at least presentationally—from the *Treatise*. The second *Enquiry* still seems keen to give sentiment a prominent role, in delivering the final verdict regarding “honour or infamy, approbation or censure” (EPM 1.9), *after* reason has established the facts of the case. Following this initial sketch of their respective roles, Hume expresses the hope that his investigation will reveal “how far either sentiment or reason enters into [moral] determinations” (EPM 1.10). But by the end of the *Enquiry*, it turns out that reason does nearly all of the work in establishing “the tendency of qualities and actions” (EPM App. 1.2), so almost the only task remaining for our natural sentiment of humanity is to declare an obvious preference for the useful and agreeable over their contraries: “Here, therefore, *reason* instructs us in the several tendencies of actions, and *humanity* makes a distinction in favour of those, which are useful and beneficial” (EPM App. 1.3). Moral “gilding or staining” might take its colouring from internal sentiment, but it seems to be causal reasoning that is erecting the structure and working out where to apply the paint.

aims to *replace* such inaccurate ideas (cf. EHU 7.3–5), not to endorse them: there is no suggestion here that such vulgar projection is somehow inevitable for us.

<sup>49</sup> *Reason* here is our *cognitive* faculty, “by which we discern Truth and Falshood” (1748, 1.4n.), whether of relations of ideas or matters of fact (e.g. T 3.1.1.9; EPM App. 1.6). This appears to be Hume’s standard use of the term, in common with most of his contemporaries (see Peter Millican, “Hume’s ‘Scepticism’ about Induction.” In *The Continuum Companion to Hume*, eds. Alan Bailey and Dan O’Brien (London: Continuum, 2012), pp. 79–87), and he describes reason accordingly in all of his main works (“Reason is the discovery of truth or falshood,” T 3.1.1.9; “reason . . . conveys the knowledge of truth and falsehood,” EPM App. 1.21; “reason, in a strict sense, as meaning the judgment of truth and falsehood,” DP 5.1).

Hume's main discussion of moral metaphysics in the second *Enquiry* ends with a paragraph aiming to summarise "the distinct boundaries and offices of *reason* and of *taste*" (EPM App. 1.21). Reason "discovers objects, as they really stand in nature, without addition or diminution," while taste is famously "productive," "gilding or staining... natural objects with the colours, borrowed from internal sentiment." Causal judgements regarding the "tendencies of actions" (etc.) clearly fall within the domain of reason,<sup>50</sup> despite the fact that these seem on Hume's theory to be coloured by the internal "sentiment or impression" of "customary transition" or "connexion" (EHU 7.28, 30). Why, then, cannot moral judgements equally count as operations of reason, when apparently the crucial factor preventing them from being similarly objective is a parallel kind of sentimental gilding? Hume does not address this question directly, but in both the *Treatise* and the second *Enquiry* his emphasis seems to be on the *action-guiding* nature of ethics (e.g. T 3.1.1.6) and its association with human *desires* (e.g. EPM App. 1.18–20). The "sentiment of approbation" in positive moral judgement is pleasurable and action-guiding, whereas the "sentiment or impression" of necessary connexion (associated with causal inference) is apparently "cool and disengaged" (EPM App. 1.21). Thus Hume's distinction here seems to come down to the familiar divide between the *conative* and the *cognitive*.

Much more could be said on these matters, but we have already seen enough to cast doubt on the tendency to classify Hume's causal and moral theories together under the vague and potentially anachronistic banner of "projectivism." Certainly there is material in his texts that could inspire the development of more closely parallel theories, for example Blackburn's *quasi-realism* that ties both causal and moral judgements to the distinctive *attitudes* manifested by the corresponding "impressions" (of customary inference, and of approval), and views both as "projective" for that reason.<sup>51</sup> It can even be argued—with some plausibility—that such accounts provide the best way of articulating a *Humean* theory in contemporary terms.<sup>52</sup> We must be wary, however, of over-interpreting Hume's own texts, for example by seeing metaphysical "projection" in contexts where his own aims might be more modest (e.g. merely to identify an impression of reflection to satisfy the empiricist requirements of his Copy Principle). Moreover, the famous "gilding or staining" passage—the one that is most overtly indicative of genuine (as opposed to fallacious) projection—can be read as simply attempting to characterise how *taste* differs from *reason*, which "discovers objects, as they really stand in nature, *without addition or diminution*" (EPM App. 1.21, my emphasis). Taste *adds something*, in the form of a positive emotive colouring, which Hume evidently sees as driving it *away*

<sup>50</sup> As quoted above from EPM App. 1.3, and see also for example T 1.3.15.2, 2.3.3.3, 3.1.1.12, and EPM App. 1.2.

<sup>51</sup> See Simon Blackburn, "Hume and Thick Connexions," in Millican, *Reading Hume*, pp. 259–76, especially p. 269, n. 19.

<sup>52</sup> For such an argument, see §3.5 of Peter Millican, "Against the 'New Hume'." In *The New Hume Debate, Revised Edition*, eds. Rupert Read and Kenneth Richman (London: Routledge, 2007), pp. 211–52. Likewise, in "Hume, Causal Realism, and Causal Science" (pp. 664–5), I drew on T 1.3.9.3–5 to suggest that Hume sees causal judgements, like moral judgements, as raising "a new creation." I would now be more cautious about attributing these positions to Hume himself.

from the realm of reason, and thus from causal judgement.<sup>53</sup> On Blackburn's quasi-realist account, by contrast, the projective construction of sentiment-gilded quasi-facts is supposed to provide a basis for objectivized descriptions, which thus enable truth and reason to come into play regarding *both* causation and morality. This contrast suggests that although quasi-realist and other "projectivist" theories that draw very close parallels between causation and morality may seem *Humean* in spirit, and might even perhaps represent a plausible ultimate destination for philosophical developments that we see taking place in Hume's thought between the *Treatise* and the *Enquiries*, they are unlikely to match in detail any position that he himself occupied.

## 9. Conclusion

The overall outcome of our investigation should not be at all surprising: we have found that the main philosophical publications of Hume's maturity, the *Enquiry concerning Human Understanding* and the *Enquiry concerning the Principles of Morals*—written at leisure and subsequently refined through numerous editions—exhibit many parallel developments that represent significant improvements, both presentational and philosophical, over the youthful and hastily published *Treatise*. Most philosophers, if they are good and remain active, will develop and improve their thoughts over time, and there is no reason to think of Hume as exceptional in this regard. More specifically, we have excellent evidence of his growing dissatisfaction with the *Treatise*, starting barely four months after the publication of Books 1 and 2 with a letter of 1 June 1739 to Henry Home (HL i 31), continuing with a letter of 16 March 1740 to Francis Hutcheson (HL i. 38–9), then later that year the doubts expressed in the *Treatise Appendix*, followed by further consistently negative comments in 1745 (LFG 41), 1751 (HL i. 158, to Gilbert Elliot) and 1754 (HL i. 187, to John Stewart). Most of these concern issues in Book 1, the source also of most of the criticism (e.g. from Reid and Beattie) that apparently prompted Hume to compose the famous "Advertisement" sent to his printer William Strahan in 1775 (HL ii 301), describing the *Treatise* as a "juvenile work" and asking that his later writing be taken as authoritative. Perhaps for this reason, the "Advertisement" has generally been affixed only to subsequent editions of the first *Enquiry*—the recast Book 1—but this risks giving a false impression, because "the following Pieces" which it endorsed included also the *Dissertation on the Passions*, the second *Enquiry*, and the *Natural History of Religion*, combined within the second (and final) volume of Hume's *Essays and Treatises on Several Subjects*. Modern editions have separated these works, but it is clear that Hume wished them to be seen as together "containing his philosophical

<sup>53</sup> Another potential factor here is the observer-relativity of judgements of taste, since the pleasure or pain they involve can depend on the constitution or situation of the observer (or indeed on *custom* or *caprice*, T 2.1.8.2), though we commonly attempt to adjust for such distorting factors in order to reach more stable judgements (T 3.3.1.15, EPM 1.9). Hume's exploration of moral relativity in "A Dialogue" bears interesting comparison with the discussion of aesthetic judgement in "Of the Standard of Taste" 28–31. But observer-relativity does not feature prominently in the passages where Hume is distinguishing between taste and reason.

sentiments and principles.”<sup>54</sup> Moreover, we have seen plenty of evidence that he himself would have been just as keen to insist on the second *Enquiry* as the authoritative statement of his moral philosophy, as he was to insist on the first *Enquiry* as the authoritative statement of his “metaphysics.”<sup>55</sup>

Hume's own preferences notwithstanding, the tendency of most philosophical scholarship for many years has been to focus overwhelmingly on the *Treatise*, and to consider his two *Enquiries* as providing relatively lightweight summaries of a few key discussions for more popular consumption, written to pursue “literary fame” (and perhaps to provoke the “zealots”).<sup>56</sup> Over the last couple of decades this consensus has weakened quite significantly, at least in respect of the first *Enquiry*, which has been increasingly discussed as a serious work of philosophy in its own right. I like to think that my own edited volume of 2002, *Reading Hume on Human Understanding*, played a useful role in stimulating and contributing to this growing trend. It is very much to be hoped that the current companion volume, *Reading Hume on the Principles of Morals*, will likewise help to undermine the traditional dismissive perspective on Hume's second *Enquiry*, which equally deserves to be taken very seriously as an independent, distinctive, and major work of philosophy.<sup>57</sup> Perhaps in time we shall even see a sustained focus on Hume's mature philosophy *as a whole*, viewing the later works more in relation to each other—as parts of an integrated and carefully considered philosophical system—than as writings whose interest is mainly to be understood by reference to the “juvenile” *Treatise*.<sup>58</sup>

<sup>54</sup> Hence Lorne Falkenstein and Neil McArthur's recent edition that combines the works is to be applauded: David Hume, *Essays and Treatises on Philosophical Subjects*, eds. Lorne Falkenstein and Neil McArthur (Peterborough, ON: Broadview Editions, 2013).

<sup>55</sup> Indeed, perhaps if anything more so, judging from the fundamental changes discussed in §§5–7 above, which also might help to explain why Hume so pointedly recommended the second *Enquiry* to posterity as “of all my writings, historical, philosophical, or literary, incomparably the best” (MOL, 10).

<sup>56</sup> In “My Own Life,” Hume famously describes “love of literary fame” as his “ruling passion” (21) and expresses his disappointment that the *Treatise* failed “even to excite a murmur among the zealots” (6). Building on such comments, Sir Amhurst Selby-Bigge influentially alleged, in his standard edition of the *Enquiries*, that the religious topics were included primarily for the sake of fame and notoriety (*ESB* viii). But few scholars would now endorse this cynical view, given our increased awareness of the range, quality, seriousness, and sheer volume of Hume's writings on religion. Indeed, as explained in §2 above, it is the *Treatise* that was artificially doctored in respect of the treatment of religion (and hence fails to reflect Hume's deepest commitments), having been “castrated” to remove the most forthrightly offending parts.

<sup>57</sup> Happily, there have been signs of a similar positive trend in respect of the second *Enquiry* (thus facilitating the current volume), and Jacqueline Taylor has been a particularly significant stimulant of and contributor to this trend.

<sup>58</sup> I am very grateful to Amyas Merivale for numerous discussions relating to the development of Hume's philosophy, especially in respect of the theory of the passions, and likewise to James Arnold and Jackie Taylor for discussion of Hume's ethics. Thanks also to audiences at the Scottish Seminar for Early Modern Philosophy and Oxford Hume Forum (2017), the Budapest Hume Conference (2018), and the third “Recasting the *Treatise*” workshop (2019).