

# Hume's Decisive Turn Away from Egoism

Peter Millican, Hertford College, University of Oxford

## Abstract

Hume's moral theory appears to have started, in his *Treatise of Human Nature* (1739-40), from a basis in Lockean psychological egoism. By contrast, his *Enquiry concerning the Principles of Morals* of 1751 incorporates a full-scale attack on such egoism, originally placed very prominently in the first part of Section 2 (though ultimately moved to Appendix 2). 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's *Sermons*. 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). Many otherwise puzzling characteristics of Hume's discussions of morals – both early and late – can be understood more clearly when put into this context of Hume's early commitment to, and later renunciation, of psychological egoism.

---

## 1. The Egoist Foundations of the *Treatise*

*Treatise* Books 2 and 3 present 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).

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>1</sup>

This claim that the *Treatise* is fundamentally egoist is, however, controversial, attracting both supporters and opponents.<sup>2</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” (2000, p. 90); 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” (2015, p. 114);<sup>3</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]” (2011, p. 72). We can quickly address the first two of these points – on natural virtues and sympathy – but the third will require more detailed consideration.

Hume does indeed recognise “sociable” natural virtues, 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*, 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>4</sup>

The *Treatise* does, however, acknowledge a notable anti-egoist exception, in that – as Owens observes – we naturally feel benevolence towards those we love, and “anger” towards those we hate:

---

<sup>1</sup> This connection 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” (*Essay* 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).

<sup>2</sup> Other supporters include Rachel Cohon (2008, pp. 31-5), Stephen Darwall (1993, p. 423), and Mikael Karlsson (2006, pp. 246-7).

<sup>3</sup> 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>4</sup> McGilvary (1903, pp. 291-4) and Bricke (1996, pp. 129-35) insist that Humean sympathy can communicate others’ *desires* as well as pleasures or pains, thus potentially giving me a desire for another’s wellbeing which is 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). Rather, as Merivale points out (2014, p. 129), Hume’s actual treatment of pity at *Treatise* 2.2.7.2 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”, *DOP* 3.7, emphasis added).

“Love is always follow’d by a desire of the happiness of the person below’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)

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 clearly recognises that their association constitutes a departure from egoism, for he later remarks that resentment 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 is 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 crucial 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>5</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>6</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 an 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 App* 2.6-7, 12).<sup>7</sup> If this suggestion is correct, then the

---

<sup>5</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 significantly against McGilvary and Bricke’s thesis that Humean sympathy can operate in that way (cf. note 4 above), 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>6</sup> Kail (2007, p. 141) sees the difficulty as being that “two impressions can merge to produce another”. 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; P 1.7). Passmore (1980, p. 127) suggests instead that Hume’s anxiety concerns his desire “to show ... a precise parallel between love and pride, between hate and humility”, but there is no obvious textual evidence to back this up.

<sup>7</sup> In the following section, the relevant passages are quoted from *EPM* 2.6 and 2.12 in the original 1751 edition.

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 recognised 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 recognising 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 recognises 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>8</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 his general account “of the love of relations” in terms of sympathy, so there is clear evidence of development in his view here.<sup>9</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>10</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,

---

<sup>8</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).

<sup>9</sup> The change was permanent, as shown by *Essays* 162-3 (“The Sceptic”, para. 10), *EPM* 3.40, and *DOP* 3.3 n. 4.

<sup>10</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.

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*.

## 2. 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 that provoked his *Letter from a Gentleman to His Friend in Edinburgh* 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>11</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>12</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 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>13</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

---

<sup>11</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 Hume also has a significant principled reason to downplay the distinction between natural and artificial virtues in his later moral theory, because there they become answerable to the same general criteria through his definition of virtue in terms of usefulness and agreeableness (*M* 9.1, 12), and its function in *correcting* erroneous moral judgements such as regarding the “monkish virtues” (*M* 9.3). This is another significant and systematic change in Hume’s moral theory, consequent on his crucial transition away from egoism and towards acknowledgement of a broad “humanity” (cf. §3 below).

<sup>12</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).

<sup>13</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 capitalisation), and the paragraph numbers correspond exactly.

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>14</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>15</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 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 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>16</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

---

<sup>14</sup> An observation perhaps partly intended to acquit the *Treatise* of "sapping the Foundations of Morality".

<sup>15</sup> Again, contrast this sceptical observation with the lament for lost simplicity at *T* 2.2.6.2, quoted in §5.

<sup>16</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 confirming that Hume's theoretical anxiety at *T* 2.2.6.2 is itself associated with egoism.

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, 1.14 n.)

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 it was reading Butler which 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 key point.<sup>17</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>18</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.

### 3. From Sympathy to Humanity

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

---

<sup>17</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.” (*Essays* 85-6).

<sup>18</sup> This speculation, from Merivale (2014, §2.1), 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 17 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*.

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 85 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 *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 7 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 22 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 which characterised that theory:<sup>19</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, 2.5 n.; *EPM* 6.3). Moreover “humanity” now becomes the dominant term, mentioned in no fewer than 48 paragraphs of the second *Enquiry*.<sup>20</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 eulogised 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>21</sup> It is

<sup>19</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>20</sup> Thus while paragraphs referring to “sympathy” decline by a factor of nearly 4 times from *Treatise* Books 2 and 3 to the second *Enquiry*, those referring to “humanity” multiply by a factor of very nearly 7 times. The *Dissertation on the Passions*, incidentally, includes only three paragraphs referring to “sympathy” (none to “humanity”), with at best vague hints of the theory of the *Treatise* at *DOP* 3.4 and 3.11.

<sup>21</sup> In the *Treatise*, Hume attempts to provide a mechanism for standardisation 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. 3.3.1.30). This may be the best he can do within a predominantly egoist theory, but as Jackie Taylor points out (2002, pp. 52-6; 2015, pp. 113-6), it is morally inadequate, since such a narrow circle will commonly display partiality and prejudice. Hume himself clearly recognises this risk in the 1757 essay “Of the Standard of Taste” (paras 1, 11; *Essays* 226-7, 233).



these universal sentiments on which our language is “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 not only explained (limited) humanity as arising from sympathy, but also 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>22</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 recognises 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.

---

<sup>22</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*.

# Bibliography

## Works by David Hume

- T* (1739-40), *A Treatise of Human Nature: A Critical Edition*, eds David Fate Norton and Mary J. Norton, Oxford University Press, 2007 – references indicated by “*T*”, and given to book, part, section and paragraph number, or “*T App*” (Appendix) followed by paragraph number
- Essays* (1741-83), *Essays, Moral, Political, and Literary*, ed. Eugene F. Miller, Liberty Classics, 2nd edition 1987 – references indicated by “*Essays*”, and given to page number
- LFG* (1745), *A Letter from a Gentleman to his Friend in Edinburgh*, Edinburgh (anonymous, no named printer or bookseller); republished in facsimile ed. Ernest C. Mossner and John V. Price, Edinburgh University Press, 1967 – references indicated by “*LFG*”, and given to page number
- EHU* (1748), *An Enquiry concerning Human Understanding*, ed. Peter Millican, Oxford University Press, 2007 (1777 edition, includes textual variants) – references indicated by “*EHU*”, and given to section and paragraph number
- EPM* (1751), *An Enquiry concerning the Principles of Morals*, ed. Tom L. Beauchamp, Oxford University Press, 1998 (copytext 1772 edition) – references indicated by “*EPM*”, and given to section (or appendix, e.g. “*EPM App 1*” etc.) and paragraph number
- DOP* (1757), *A Dissertation on the Passions and The Natural History of Religion*, ed. Tom L. Beauchamp, Oxford University Press, 2007 (copytext 1772 edition) – references to *Dissertation* indicated by “*DOP*”, and given to section and paragraph number

## Other Works

- Bricke, John (1996), *Mind and Morality: An Examination of Hume’s Moral Psychology*, Oxford University Press.
- Butler, Joseph (1736), *Fifteen Sermons Preached at the Rolls Chapel*, 3<sup>rd</sup> edition, London.
- Butler, Joseph (1736), *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*, Dublin.
- Cohon, Rachel (2008), *Hume’s Morality: Feeling and Fabrication*, Oxford University Press.
- Darwall, Stephen (1993), “Motive and Obligation in Hume’s Ethics”, *Noûs* 27, pp. 415-48.
- Garrett, Don (1997), *Cognition and Commitment in Hume’s Philosophy*, New York: Oxford University Press.
- Garrett, Don (2015), *Hume*, Routledge.
- Gill, Michael (2000), “Hume’s Progressive View of Human Nature”, *Hume Studies* 26, pp. 87-108.
- Hobbes, Thomas (1651), *Leviathan: or, The matter, forme, & power of a common-wealth ecclesiasticall and civill*, London.
- Kail, P. J. E. (2007), *Projection and Realism in Hume’s Philosophy*, Oxford: Oxford University Press.
- Karlsson, Mikael M. (2006), “Reason, Passion, and the Influencing Motives of the Will”, in Saul Traiger (ed.), *The Blackwell Guide to Hume’s Treatise*, Blackwell, pp. 235-55.
- Locke, John (1690/1700), *An Essay Concerning Human Understanding*, ed. P. H. Nidditch, Clarendon Press, 1975 (“*Essay*”).
- McGilvary, Evander Bradley (1903), “Altruism in Hume’s *Treatise*”, *Philosophical Review* 12, pp. 272-98.
- Merivale, Amyas (2014), *An Enquiry concerning the Passions: A critical study of Hume’s Four Dissertations*, University of Leeds PhD thesis.
- Owens, David (2011), “The Problem with Promising”, in Hanoach Sheinman (ed.), *Promises and Agreements: Philosophical Essays*, Oxford University Press, pp. 58-79.
- Passmore, John (1980), *Hume’s Intentions*, third edition, London: Duckworth.
- Taylor, Jacqueline (2002), “Hume on the Standard of Virtue”, *The Journal of Ethics* 6, pp. 43-62