Hume

Peter Millican

Hume's theory of morals is widely misunderstood and often unwittingly caricatured, partly because it is subtle and fits uneasily into popular taxonomies and partly owing to Hume's talent for the eloquent aphorism, which is easily remembered while the subtleties are overlooked. Most notorious is the famous statement in his youthful *Treatise of Human Nature* that 'Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them' (*T* 2.3.3.4, 415). This is commonly read as an insistence on the total impotence of human reason, apparently implying its lack of any jurisdiction over the principles of action and a consequent moral scepticism or at least irrationalism: 'I have prov'd,[1] that reason is perfectly inert, and can never either prevent or produce any action or affection.... Moral distinctions, therefore, are not the offspring of reason. Reason is wholly inactive, and can never be the source of so active a principle as conscience, or a sense of morals' (*T* 3.1.1.8–10, 458).

But Hume is no moral sceptic or irrationalist, and there is plenty else in the *Treatise* to indicate that he sees reason as playing a major role in determining moral principles. To remove any doubt, both of these points were emphasized very clearly when he later composed *An Enquiry Concerning the Principles of Morals*, his favourite work and one that he explicitly insisted should be taken as his authoritative voice[2]:

> Those who have denied the reality of moral distinctions, may be ranked among the disingenuous disputants; nor is it conceivable, that any human creature could ever seriously believe, that all characters and actions were alike entitled to the affection and regard of every one.
>
> *M* 1.2, 169–70

> 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....
>
> *M* App. 1.2, 285

The overall moral theory of the two works is broadly similar, but their approach and presentation is very different. Book 3 of the *Treatise*, entitled 'Of Morals', aims to place morality within a general theory of the human mind, starting from an analysis of the mind's faculties and contents – notably the passions that drive us – and strongly highlighting theoretical arguments about human motivation. The *Enquiry*, by contrast, starts from an analysis of the moral judgements that Hume observes to be generally accepted, aiming to identify what is common to them. Only then does it turn to morality's place within the human mind, and accordingly the famous theoretical arguments that had appeared in the first section of *Treatise* Book 3 – having been significantly edited and shortened – are relegated to the first 'Appendix' of the later work. Since these arguments are very famous and controversial, giving plenty of scope for interesting critical examination, Hume's moral theory has tended to be discussed overwhelmingly with reference to the *Treatise*. The *Enquiry* has been largely neglected until recently, an unfortunate irony given that Hume's relegation of the famous arguments may well reflect a recognition that some of those arguments, at least as presented in the *Treatise*, are fundamentally defective. To get a faithful overall picture of Hume's moral theory, therefore, we must take account of both works, embellishing the broad and clear strokes of the mature *Enquiry* with the theoretical detail supplied by the *Treatise*, but with a keen eye for differences between the two that apparently indicate changes of mind rather than of approach or emphasis.

Hume's utilitarian virtue ethics

Hume approaches morality not only as a *scientist of human nature*, aiming to understand its 'springs and principles' (*E* 1.15, 14), but also – inevitably – as himself a human being who partakes of that nature. As a human scientist, Hume observes our ubiquitous tendency to praise and censure actions and personal 'characters', and he seeks for the principles that underlie this behaviour. He finds the common thread to be that we generally judge actions according to the character they reveal, and that we judge characters according to their general tendency to be 'useful or agreeable' to the possessor or to others:

> If any *action* be either virtuous or vicious, 'tis only as a sign of some quality or character. It must depend upon durable principles of the mind, which extend over the whole conduct, and enter into the personal character. Actions themselves, not proceeding from any constant

principle, have no influence on love or hatred, pride or humility; and consequently are never consider'd in morality. . . . We are never to consider any single action in our enquiries concerning the origin of morals; but only the quality or character from which the action proceeded.

T 3.3.1.4–5, 575

PERSONAL MERIT consists altogether in the possession of mental qualities, *useful* or *agreeable* to the *person himself* or to *others*.

M 9.1, 268

That moral judgement applies primarily to *characters* or mental qualities rather than to *actions* makes this a form of *virtue ethics*. That the distinction between virtues and vices is drawn according to *usefulness* and *agreeableness* (rather than any appeal to divine or ultimate human purposes) makes it a form of *utilitarian virtue ethics*.[3]

There is of course plenty of scope for debate about both of these principles. An act-utilitarian, for instance, would insist that moral judgement applies primarily to individual acts rather than to characters, while a rule-utilitarian or Kantian would judge acts according to their guiding rule or maxim. Likewise, there is scope for disagreement over which mental qualities are appropriately to be classed as 'virtues' or 'vices' – *pride*, for example, has traditionally been viewed by Christians as the primary 'deadly sin', and *humility* as a cardinal virtue, whereas Hume sees 'pride or self-esteem' and 'vanity or the desire of reputation' (*T* 2.2.1.9, 332) as crucial spurs to moral behaviour, a point he emphasizes in the very last paragraph of the *Treatise*:

[W]ho can think any advantages of fortune a sufficient compensation for the least breach of the *social* virtues, when he considers, that not only his character with regard to others, but also his peace and inward satisfaction entirely depend upon his strict observance of them; and that a mind will never be able to bear its own survey, that has been wanting in its part to mankind and society?

T 3.3.6.6, 620

He also expresses very similar thoughts at the end of the final section of the *Enquiry* (*M* 9.21–5), where he famously confronts the challenge of the 'sensible knave' who hopes to benefit by appearing moral while secretly taking advantage of opportunities for immoral gain. It is debatable whether Hume has a fully satisfactory answer; indeed, he acknowledges (at *M* 9.23) that someone who is completely unmoved by the desire to view himself as

virtuous might well be beyond persuasion. No doubt this is disappointing for anyone who seeks a universally persuasive answer to the question 'Why be moral?' But it does not present any objection to Hume's theory, for it is a sad fact of life that someone who is overwhelmingly self-interested is indeed unlikely to appreciate the richer happiness that typically comes from the 'social virtues' – from genuinely caring about others – and from the shared affection and companionship they make possible. As many philosophers have recognized, at least since Aristotle, an appreciation of the virtues is best achieved by parental example, training and habituation, not by self-interested calculation. It is very plausible to argue that parents, wishing the best for their children, have excellent reason to inculcate *sincere* virtuous desires and affections, since the greatest satisfactions in life are thus made possible. But a person who has already grown up with a purely selfish disposition may well find it impossible to understand how this could be the case, as indeed Hume infers from his famous Copy Principle: 'A man of mild manners can form no idea of inveterate revenge or cruelty; nor can a selfish heart easily conceive the heights of friendship and generosity' (*E* 2.7, 20).[4] Perhaps the best hope for such a person is that habituation of *simulated* concern for others will open the way in time to feelings of *genuine* concern, so that even the 'sensible knave', despite himself, can ultimately develop the empathetic Humean virtues. All this should give pause for thought to those who, under the spell of crude economic theory (and often an even cruder theory of psychology),[5] have been all too ready to pronounce that 'greed is good', without regard either for the psychological health of those brought up with this message, or for the social health of a society in which so little encouragement is given even to *present an appearance* of selfless virtue.

The language of morals

It might seem that disputes about the identification of the virtues and vices would be hopelessly intractable, with philosophers simply disagreeing in ways that reflect their differing theories. But Hume begins his *Enquiry* by proposing an ingenious method of resolution, by appeal to the nature of *common language*:

> [W]e shall endeavour to follow a very simple method: We shall analyse that complication of mental qualities, which form what, in common life, we call Personal Merit: We shall consider every attribute of the mind, which renders a man an object either of esteem and affection, or of hatred and contempt; every habit or sentiment or faculty, which, if

ascribed to any person, implies either praise or blame, and may enter into any panegyric or satire of his character and manners.... a philosopher ... needs only enter into his own breast for a moment, and consider whether or not he should desire to have this or that quality ascribed to him, and whether such or such an imputation would proceed from a friend or an enemy. The very nature of language guides us almost infallibly in forming a judgment of this nature; and as every tongue possesses one set of words which are taken in a good sense, and another in the opposite, the least acquaintance with the idiom suffices, without any reasoning, to direct us in collecting and arranging the estimable or blameable qualities of men.

M 1.10, 173–4

He then begins to build his catalogue of virtues accordingly, starting with 'the benevolent or softer affections', which, 'wherever they appear, engage the approbation, and good-will of mankind', as shown by the positive colouring of the words through which they are expressed: 'The epithets *sociable*, *good-natured*, *humane*, *merciful*, *grateful*, *friendly*, *generous*, *beneficent*, or their equivalents, are known in all languages, and universally express the highest merit' (*M* 2.1, 176). Hume seems to be on fairly solid ground in his assertion that these words are universally taken as expressions of virtue, though clearly not all would agree that they reach 'the highest merit', Kant being the most conspicuous opponent of this view.[6]

Even if Hume's catalogue of terms is agreed, however, there is a risk that his method of appeal to language might sometimes fail to deliver substantial results. For as Aristotle famously taught, many virtues are associated with complementary vices, lying on a scale with the ideal character placed at a 'mean' between the two extremes. Thus, *courage* is a virtue, *cowardice* and *rashness* both complementary vices, and we can all agree on the colouring of these words, but this does not imply that we will agree on the substantial question of where the ideal 'mean' lies nor where each boundary is crossed between virtue and vice. Aristotle has often been criticized for the vacuity of his 'doctrine of the mean' for precisely this reason, and Hume at least provides a relatively determinate answer:

No quality, it is allowed, is absolutely either blameable or praise-worthy. It is all according to its degree. A due medium, say the PERIPATETICS,[7] is the characteristic of virtue. But this medium is chiefly determined by utility. A proper celerity, for instance, and dispatch in business, is commendable. When defective, no progress is ever made in the execution of any purpose: When excessive, it engages us in precipitate

and ill-concerted measures and enterprises: By such reasonings, we fix the proper and commendable mediocrity in all moral and prudential disquisitions; and never lose view of the advantages, which result from any character or habit.

M 6.2, 233

Questions remain about how the various utilities involved are to be assessed, predicted and compared, but Hume's example convincingly illustrates how the appropriate balance – for example, between speed and caution in performing some industrial process – might be judged in particular cases. However, his particular solution in terms of utility, though certainly plausible, is not implied by the agreement in language on which he hopes to base his theory. Someone could agree that *courage* is a virtue, and *rashness* a vice, without agreeing that the boundary between them is to be determined by considerations of utility.

A similar point can be made by returning to *pride*, which Hume again considers as involving something like an Aristotelian mean between extremes:[8]

We shall begin with examining the passions of *pride* and *humility*, and shall consider the vice or virtue that lies in their excesses or just proportion. An excessive pride or over-weaning conceit of ourselves is always esteem'd vicious, and is universally hated.

T 3.3.2.1, 592

But tho' an over-weaning conceit of our own merit be vicious and disagreeable, nothing can be more laudable, than to have a value for ourselves, where we really have qualities that are valuable. . . . nothing is more useful to us in the conduct of life, than a due degree of pride, which makes us sensible of our own merit, and gives us a confidence and assurance.

T 3.3.2.8, 596–7

Hume is well aware that Christian philosophers such as Aquinas take a far more negative view of pride, and he emphasizes, in a conciliatory tone, the universal consensus that an *appropriate* degree of the passion is entirely acceptable: 'The most rigid morality allows us to receive a pleasure from reflecting on a generous action' (*T* 2.1.7.8, 298–9). Aquinas would indeed agree, since he considers the sin of pride to involve 'an *excessive* desire for one's own excellence which rejects subjection to God'.[9] But this just serves to illustrate how easily agreement *in words* can mask *substantial* disagreement

between widely diverging moral systems. If we restrict the words 'pride' and 'vanity' to what we consider to be cases of *excessive* self-regard, then of course we can agree that they denominate vices rather than virtues, but we might still disagree radically about the degree of self-regard that is appropriate.[10] And thus Hume's appeal to the agreed positive (or negative) moral tone of our words for virtues (or vices) gives far less solid evidence of a genuine moral consensus than he sometimes appears to suggest.

The corruptions of religion

All this does not entirely undermine Hume's method, and of course he is well aware that there is plenty of disagreement about moral issues, notwithstanding the established moral tone of much of our language. But the crucial result that he takes from his survey of virtues and vices is that they all plausibly depend on considerations of *agreeableness* and *usefulness*, either to the person who has them or to others. He then appeals to this implicit common standard to provide a criterion for judging alleged virtues and vices, sometimes in a way that rejects the view of them taken by orthodox moralists, especially those inspired by religion:

> And as every quality, which is useful or agreeable to ourselves or others, is, in common life, allowed to be a part of personal merit; so no other will ever be received, where men judge of things by their natural, unprejudiced reason, without the delusive glosses of superstition and false religion. Celibacy, fasting, penance, mortification, self-denial, humility, silence, solitude, and the whole train of monkish virtues; for what reason are they every where rejected by men of sense, but because they serve to no manner of purpose; neither advance a man's fortune in the world, nor render him a more valuable member of society; neither qualify him for the entertainment of company, nor encrease his power of self-enjoyment?[11] We observe, on the contrary, that they cross all these desirable ends; stupify the understanding and harden the heart, obscure the fancy and sour the temper. We justly, therefore, transfer them to the opposite column, and place them in the catalogue of vices; nor has any superstition force sufficient among men of the world, to pervert entirely these natural sentiments. A gloomy, hair-brained enthusiast, after his death, may have a place in the calendar;[12] but will scarcely ever be admitted, when alive, into intimacy and society, except by those who are as delirious and dismal as himself.

> *M* 9.3, 270

Section 14 of Hume's *Natural History of Religion*, entitled 'Bad Influence of Popular Religions on Morality', sets out to explain – with examples – why religious people, even when supposedly devoted to the service of a morally perfect God, will typically attempt to win His favour 'either by frivolous observances, by intemperate zeal, by rapturous extasies, or by the belief of mysterious and absurd opinions' (*N* 14.1).[13] Hume's ingenious explanation is that the very qualities which make genuinely virtuous actions desirable in themselves – their agreeableness and usefulness – make them less attractive to the superstitious believer, who wants to find some *distinctive* way of showing devotion to God. Such a believer will therefore be more attracted towards a devotional practice which is either pointless or painful, such as fasting in Lent or Ramadan, or self-flagellation: 'It seems the more purely religious, because it proceeds from no mixture of any other motive or consideration' (*N* 14.6). And for similar reasons, such corrupted morality is likely to be encouraged by priests who see an opportunity for consolidating their influence: 'the more unaccountable the measures of acceptance required by [the divinity], the more necessary does it become to abandon our natural reason, and yield to their ghostly guidance and direction' (*N* 14.8).

Hence Hume accounts for the common observation that religious fervour is often associated not with devotion to genuine morality but rather with appalling crimes and barbarity. And he quotes historians to confirm that in the ancient world, as often in modern times (such as the religious wars of the seventeenth century, or recent terrorist atrocities), '*Those who undertake the most criminal and most dangerous enterprizes are commonly the most superstitious*' (*N* 14.7).[14] Hume points out that monotheism, in particular, is prone to zealous intolerance, enforcing religious conformity to reflect the unity of the deity (*N* 9.3). And his anti-religious animus becomes especially evident in a long footnote to the essay 'Of National Characters' (*Essays* 199– 201 n. 3), where he explains how the character and position of clergymen is especially liable to lead them into hypocrisy, ritualism, promotion of superstition and fraud, conceit, intolerance of disagreement and vengeful vindictiveness. He takes it as a commonplace 'that all prudent men are on their guard, when they meet with any extraordinary appearance of religion', while acknowledging that 'probity and superstition, or even probity and fanaticism, are not altogether incompatible'.[15]

A naturalistic account of morality

Hume's moral philosophy has become particularly influential in recent years, inspiring a wide range of thinkers, from emotivists (e.g. A. J. Ayer) and error-

theorists (e.g. J. L. Mackie) to 'quasi-realists' (e.g. Simon Blackburn and Allan Gibbard). Some of Hume's appeal derives from the specific detail of his meta-ethics, which we have yet to consider. But much is also due to his position as the greatest pioneer of the project to develop a positive moral theory within a fully naturalistic framework, explaining morality as part of a 'science of human nature' that makes no appeal to religious doctrine and which fits comfortably into the post-Darwinian worldview. Crucial to this is Hume's forthright rejection of religion as the ground of morality, a rejection made all the more emphatic by his insistence that religion – so far from providing even a valuable inducement towards moral behaviour (as then universally taken for granted by Christian apologists) – is frequently a corrupting influence. Similar themes would later be emphasized by Nietzsche, though with the very different aim of *undermining* morality, at least as it is generally understood. But Hume, as we saw earlier, is no moral sceptic, and he seeks a naturalistic explanation of morality which ultimately *vindicates* it as a crucial aspect of the good life, rather than a debunking explanation which dismisses it as a superstitious illusion or conspiracy of the weak.

Nietzsche was writing in the wake of Darwin's *Origin of Species* (1859) and *Descent of Man* (1871), but nevertheless Hume's philosophy – dating from over a century earlier – fits even more comfortably with the modern evolutionary outlook that sees humankind as just one species of animals, set within a natural order that operates according to all-embracing causal laws and without intrinsic purpose.[16] No fewer than three of the six parts of the 1739 *Treatise*,[17] respectively, giving accounts of human *reason*, *pride and humility*, and *love and hatred*, end with sections devoted to the corresponding features of animals (*T* 1.3.16, 2.1.12 and 2.2.12), while a fourth part, on *the will and direct passions*, omits such a discussion only for the sake 'of avoiding prolixity . . . since nothing is more evident, than that . . . the will and direct passions, as they appear in animals . . . are of the same nature, and excited by the same causes as in human creatures' (*T* 2.3.9.32, 448). As animals among others, we cannot expect nature to make our lives easy, a thought vividly expressed by the character Philo in Hume's posthumous *Dialogues Concerning Natural Religion*:

> Look round this universe. What an immense profusion of beings, animated and organized, sensible and active! You admire this prodigious variety and fecundity. But inspect a little more narrowly these living existences, the only beings worth regarding. How hostile and destructive to each other! How insufficient all of them for their own happiness! How contemptible or odious to the spectator! The whole presents nothing but the idea of a blind Nature, impregnated by a great vivifying principle,

and pouring forth from her lap, without discernment or parental care, her maimed and abortive children.

D 11.13

Hume sees morality as an adaptation to the situation in which we find ourselves, starting off from the affection and benevolence that we naturally feel towards those close to us. He takes this natural benevolence as an obvious fact and does not present an evolutionary account of its origin. But a great deal of what he says in building the rest of his moral theory on this foundation is extremely congenial to an evolutionary viewpoint.

In the *Treatise*, Hume draws a distinction between *natural* and *artificial* virtues (of which there are echoes in Appendix 3 of the *Enquiry*). Natural virtues, on this account,[18] are qualities of mind that we possess by a natural instinct (e.g. kindness to children, pity for the unfortunate, gratitude to benefactors) and which we also naturally approve of, because they tend to bring immediate good on each occasion of their exercise. Artificial virtues, by contrast, are those 'that produce pleasure and approbation by means of an artifice or contrivance, which arises from the circumstances and necessities of mankind' (*T* 3.2.1.1, 477). Hume's paradigm example of such an artificial virtue is *justice* – by which he means mainly property rights – while others involve promises, government, international law and chastity. It is characteristic of these that they can fail to bring good on specific occasions, and can even cause harm, because their value comes from the overall system of which they are a part:

A single act of justice is frequently contrary to *public interest*; and were it to stand alone, without being follow'd by other acts, may, in itself, be very prejudicial to society. When a man of merit, of a beneficent disposition, restores a great fortune to a miser, or a seditious bigot, he has acted justly and laudably, but the public is a real sufferer. . . . But however single acts of justice may be contrary, either to public or private interest, 'tis certain, that the whole plan or scheme is highly conducive, or indeed absolutely requisite, both to the support of society, and the well-being of every individual. 'Tis impossible to separate the good from the ill. Property must be stable, and must be fix'd by general rules. Tho' in one instance the public be a sufferer, this momentary ill is amply compensated by the steady prosecution of the rule, and by the peace and order, which it establishes in society. And even every individual person must find himself a gainer, on ballancing the account; since, without justice, society must immediately dissolve . . .

T 3.2.2.22, 497

The artificiality of justice is also revealed by the complex rules that property relations typically involve, regarding 'possession acquired by occupation, by industry, by prescription, by inheritance, by contract, *&c.* Can we think, that nature, by an original instinct, instructs us in all these methods of acquisition?' (*M* 3.41, 202). Hume is keen to insist, however, that the artificiality of such rules does not undermine either their moral significance or their essential place in human society: 'Mankind is an inventive species; and where an invention [such as justice] is obvious and absolutely necessary, it may as properly be said to be natural as any thing that proceeds immediately from original principles, without the intervention of thought or reflection' (*T* 3.2.1.19, 484).

Thus, Hume believes that morality – though an essential part of human life – is, to a significant extent, *invented* rather than *discovered* (and as we shall see, this even applies somewhat to his 'natural virtues'). Morality starts from our natural instincts but is then refined by thought and reflection into a system whose features – though *actually* dependent on human nature – can easily give the *illusion* of being an independent aspect of reality such as might be divinely created and discoverable through reason. Let us now look a bit more closely at this Humean account of the genealogy of morals.

The genealogy of morals

The *Treatise* discusses the artificial virtues before the natural,[19] whereas the *Enquiry* follows a more logical sequence, starting with Section 2 on benevolence. We have already seen (in 'The language of morals' above) how Hume draws attention to the positive colouring of the words '*sociable*, *good-natured*, *humane*, *merciful*, *grateful*, *friendly*, *generous*, *beneficent*' (*M* 2.1, 176). In pursuit of his general (quasi-)utilitarian strategy,[20] he then goes on to argue – by appeal to common human experience of, and judgement about, these 'softer affections' – that 'the UTILITY, resulting from the social virtues, forms, at least, a *part* of their merit, and is one source of that approbation and regard so universally paid to them' (*M* 2.8, 179).

In Section 3 of the *Enquiry*, Hume moves on to justice, and it is here that we see the germs of an evolutionary account of morality which explains it as starting from family affection and tribal allegiance, then moving out to society more generally. Justice is necessitated by the human situation in which we need to cooperate with other people but are greatly partial to our own interests in preference to theirs. If we had never had a need of others for any of our wants – because nature 'bestowed ... such profuse *abundance* ... that ... without any care or industry', we could obtain whatever we wanted,

then 'the cautious, jealous virtue of justice would never once have been dreamed of' (*M* 3.2–3, 183–4). Likewise, 'the divisions and barriers of property and obligation' would never have been thought of if everyone felt 'no more concern for his own interest than for that of his fellows' (*M* 3.6, 185). Families can exhibit such 'enlarged affections', mutually benevolent to such an extent that – corroborating Hume's theory – 'all distinction of property [is], in a great measure, lost and confounded among them' (*M* 3.7, 185). Another condition for the development of justice is that through cooperation, we can indeed mutually achieve the necessities of life. Thus, in dire emergencies such as 'a city besieged . . . perishing with hunger' or a shipwreck, 'the strict laws of justice are suspended . . . and give place to the stronger motives of necessity and self-preservation' (*M* 3.8, 186). Hume sums up these points by emphasizing again his key theme of utility:

> Thus, the rules of equity or justice depend entirely on the particular state and condition, in which men are placed, and owe their origin and existence to that UTILITY, which results to the public from their strict and regular observance. Reverse, in any considerable circumstance, the condition of men: Produce extreme abundance or extreme necessity: Implant in the human breast perfect moderation and humanity, or perfect rapaciousness and malice: By rendering justice totally *useless*, you thereby totally destroy its essence, and suspend its obligation upon mankind.
>
> *M* 3.12, 188

Utility explains our need for cooperation, and the need is sufficiently obvious that it is relatively straightforward to explain, in general terms, how morality as a whole – including the artificialities of justice and its rules – is likely to have arisen.

Again Hume starts from the nature of humanity, this time 'that natural appetite betwixt the sexes, which unites them together, and preserves their union, till a new tye takes place in their concern for their common offspring' (*T* 3.2.2.4, 486).[21] From this minimal foundation,

> a family immediately arises; and particular rules being found requisite for its subsistence, these are immediately embraced; though without comprehending the rest of mankind within their prescriptions. Suppose, that several families unite together into one society, which is totally disjoined from all others, the rules, which preserve peace and order, enlarge themselves to the utmost extent of that society; but becoming then entirely useless, lose their force when carried one step farther. But

again suppose, that several distinct societies maintain a kind of intercourse for mutual convenience and advantage, the boundaries of justice still grow larger, in proportion to the largeness of men's views, and the force of their mutual connexions. History, experience, reason sufficiently instruct us in this natural progress of human sentiments, and in the gradual enlargement of our regards to justice, in proportion as we become acquainted with the extensive utility of that virtue.

M 3.21, 192

What needs to be explained, in this story, is how people who have already learnt 'some rule of conduct and behaviour' within their immediate family (*M* 3.16, 190), in which their affection and benevolence towards each other is instinctive and strong,[22] can then be induced to extend this rule-respecting behaviour to a progressively wider circle where such natural bonds are far weaker (or even entirely absent). The benefits of cooperation with others are indeed obvious even in a primitive society, whether to combine in dealing with external threats, hunting animals and harvesting crops, or simply recognizing that 'it will be for my interest to leave another in the possession of his goods, *provided* he will act in the same manner with regard to me' (*T* 3.2.2.10, 490). Appealing to promises as the original basis of such cooperation is hopeless, because the 'observance of promises is itself one of the most considerable parts of justice; and we are not surely bound to keep our word, because we have given our word to keep it' (*M* App. 3.7, 306).[23] Besides, society obviously pre-dates language, and Hume's far more plausible account is in terms of the development of a tacit *convention* whereby we help each other conditionally on observing the other's cooperation: 'Thus two men pull the oars of a boat by common convention, for common interest, without any promise or contract' (*M* App. 3.8, 306; cf. *T* 3.2.2.10, 490). Each knows that the other will stop cooperating if he attempts to 'free ride' by taking advantage without reciprocating, and this mechanism can also extend to instances of cooperation that are not simultaneous, as long as the interactions are foreseen as repeating:[24]

I learn to do a service to another, without bearing him any real kindness; because I forsee, that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or with others. And accordingly, after I have serv'd him, and he is in possession of the advantage arising from my action, he is induc'd to perform his part, as foreseeing the consequences of his refusal.

T 3.2.5.9, 521

This overall Humean account is far more plausible than stories about rational intuition of moral forms or intrinsic purposes, identification of morality with knowledge of the divine or Kantian respect for law as such, while at the same time being supportive of morality as a valuable institution. Hence, it is not surprising that a variety of contemporary thinkers have seen it as the core of a correct and fruitful account, with potential for enrichment from the insights of evolution, game theory, psychology and the philosophy of language.[25]

Reason, passion and systematization

We saw in 'The language of morals' above that, in the *Enquiry*, Hume begins his investigation with a study of the *language* of virtues and vices, looking at our already established institution of moral ascription with the aim of identifying its central unifying feature, namely, the endorsement of *character traits* that are *useful* or *agreeable*.[26] He then goes on to build on this theoretical unity by inviting his readers to adjust their view of those exceptional *supposed* virtues that fail to fit the framework (notably the 'monkish virtues' of celibacy, fasting, penance, mortification, etc.). He thus appeals to systematization as a means of shifting moral perception: when we see the true shape of our overall moral framework, we can be motivated to adjust our outlying judgements to conform to it. This is a clever strategy in a contentious field, using the established consensus enshrined in our very language as a lever of persuasion (though we also saw that this verbal consensus may be less than it initially appears). The order of discussion also ultimately gives Hume a neat way of answering the 'controversy started of late … concerning the general foundation of MORALS; whether they be derived from REASON or from SENTIMENT' (*M* 1.3, 170):

> 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; since nothing but that faculty can instruct us in the tendency of qualities and actions, and point out their beneficial consequences to society and to their possessor. . . . But though reason … be sufficient to instruct us in the pernicious or useful tendency of qualities and actions; it is not alone sufficient to produce any moral blame or approbation. Utility is only a tendency to a certain end; and were the end totally indifferent to us, we should feel the same indifference towards the means. It is requisite a *sentiment* should here display itself, in order to give a preference to the useful above the pernicious tendencies. This sentiment can be no other

than a feeling for the happiness of mankind, and a resentment of their misery; since these are the different ends which virtue and vice have a tendency to promote. 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.

<div align="right">

M App. 1.2–3, 285–6

</div>

By *reason* in these contexts, Hume means simply our *cognitive* faculties, by which we discover what is true and what is false.[27] And his argument for saying that reason so understood is insufficient for morality is very straightforward and commonsensical: belief or knowledge of what is the case cannot motivate us unless we *care* about the relevant facts.

In the earlier *Treatise*, Hume follows a very different path, starting from an account of our passions (in Book 2) and then immediately setting out to prove that 'Moral Distinctions [are] not deriv'd from Reason' (title of *T* 3.1.1). Here, he most famously appeals to an argument drawn from Section 2.3.3, 'Of the Influencing Motives of the Will', to the effect that reason *cannot* motivate:

> REASON is the discovery of truth or falshood. Truth or falshood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now 'tis evident our passions, volitions, and actions, are not susceptible of any such agreement or disagreement; being original facts and realities, compleat in themselves, and implying no reference to other passions, volitions, and actions. 'Tis impossible, therefore, they can be pronounced either true or false, and be either contrary or conformable to reason.

<div align="right">

T 3.1.1.9, 458

</div>

Though this argument has been extremely influential, its logic is unclear and its upshot obscure. In particular, although it has commonly been taken to imply that the *products* of reason – namely beliefs – cannot *cause* actions, this is not something that Hume believes.[28] Passions (e.g. desires, hopes and fears) certainly motivate us, but beliefs can do likewise: 'The effect, then, of belief, is to raise up a simple idea to an equality with our impressions, and bestow on it a like influence on the passions' (*T* 1.3.10.3, 119). And Hume begins his discussion of the 'direct' passions by making clear that the prospect of pleasure and pain, and the belief that these will be the consequences of certain behaviour, are the chief driver of our actions (*T* 2.3.9.1, 438; cf. *T*

2.3.3.3, 414).[29] The famous argument is therefore probably best seen instead as merely expressing a logical taxonomy, dividing the role of *reason* – that is, the discovery of truth or falsehood – from that of *will* – that is, intentional action.[30] It is probably no coincidence that Hume dropped this misleading argument from his recasting both of Book 2 of the *Treatise* (i.e. the *Dissertation on the Passions*) and of Book 3 (i.e. the *Enquiry*).[31]

Another major difference between the *Treatise* and the *Enquiry* is that in the former, Hume provides a sophisticated (and arguably egoist) explanation of our concern for others based on the mechanism of *sympathy*, whereby we literally come to share the feelings of those we encounter by responding to their manifestation of those feelings (*T* 2.1.11.2–3, 316–17). In the *Enquiry*, he more straightforwardly identifies 'humanity' or 'a fellow-feeling with others' as a clearly observable 'principle in human nature' and implies that his previous attempt to '[resolve it] into principles more simple and universal' had been mistaken (*M* 5.17 n. 19, 219–20).

Nevertheless, the themes we have already explored – involving moral language, systematization and the benefits of cooperation – combine powerfully (especially within an evolutionary perspective) to provide such an explanation of our 'moral sentiments'. First, the establishment of morality leads naturally to our judging things from 'a general point of view', since only thus can we consistently converse with (and thus influence and negotiate with) others. Language itself provides a powerful facilitator of such impartiality:

> General language, . . . being formed for general use, must be moulded on some more general views, and must affix the epithets of praise or blame, in conformity to sentiments, which arise from the general interests of the community. . . . Sympathy, we shall allow, is much fainter than our concern for ourselves, and sympathy with persons remote from us, much fainter than that with persons near and contiguous; but for this very reason, it is necessary for us, in our calm judgments and discourse concerning the characters of men, to neglect all these differences, and render our sentiments more public and social. Besides, that we ourselves often change our situation in this particular, we every day meet with persons, who are in a situation different from us, and who could never converse with us, were we to remain constantly in that position and point of view, which is peculiar to ourselves. The intercourse of sentiments, therefore, in society and conversation, makes us form some general unalterable standard, by which we may approve or disapprove of characters and manners. And though the heart takes not part entirely with those general notions, . . . yet have [they] a considerable influence,

and being sufficient, at least, for discourse, serve all our purposes in company, in the pulpit, on the theatre, and in the schools.

M 5.42, 228–9

Since language is our medium of thought, moreover, we shall inevitably find ourselves considering matters in the same terms that we use to converse with others: as Hume points out against the 'sensible knave', it is simply not possible (at least for most of us) to maintain an outward pose that is radically different from our inner orientation. Partly for this reason, the moral outlook – whereby we attempt to take account of others' interests from a general point of view – is usually *in our own interest*. The benefits of mutual cooperation in society are so great that anything which tends to fit us better into such cooperative relationships will – most of the time – be to our advantage: useful to ourselves, as well as to others. Hence evolution, both biological and societal, will naturally lead us genuinely to *care* about others and also about moral considerations such as fairness. So we should not be at all surprised to find that nature has 'hardwired' us to do so to a significant extent.[32] Such hardwiring, together with the evident importance of morality and its enshrinement in our language, helps to explain our common tendency to *objectify* our moral judgements and to see them as part of external reality even though they are not. Hume observes that this tendency towards objectification applies even in aesthetics, where critical discussion and consideration – identification of, and agreement on, desirable and undesirable features, etc. – naturally leads us in the direction of systematization of our language and thought.[33] But in morals, practical utility provides a far stronger force in the same direction. So although morality is not part of the world – the province of objective truth and falsehood – our moral judgement makes it appear to us as though it were:

> Thus the distinct boundaries and offices of *reason* and of *taste* are easily ascertained. The former conveys the knowledge of truth and falsehood: The latter gives the sentiment of beauty and deformity, vice and virtue. The one discovers objects as they really stand in nature, without addition or diminution: The other 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.
>
> *M* App. 1.21, 294

This 'new creation' of morality is one of which Hume fully approves, and his 'anti-realism' here is constructive and not part of any debunking exercise.[34] His account of morality fits very plausibly within a scientific framework that shuns postulation of any fanciful extra-sensory realities of divine purposes,

moral forms or *a priori* duties. And it shows how morality, in a thoroughly benign form, can make *good* sense without such dubious encumbrances, as an institution of which we likewise can fully approve, as worthy of cultivation both in our own lives and those of our children.

Notes

1 Here, Hume provides a footnote reference back to *T* 2.3.3. See 'Reason, passion and systematization' below for more on this argument.

2 In his short autobiography 'My Own Life', Hume describes the *Enquiry* as 'in my own opinion . . . of all my writings, historical, philosophical, or literary, incomparably the best' (*MOL* 10). In 1775, he asked his printer William Strahan to affix an 'Advertisement' to the volume of his works containing his two *Enquiries*, *Dissertation on the Passions* and *Natural History of Religion*. In this, he renounces the *Treatise* and ends: 'Henceforth, the Author desires, that the following Pieces may alone be regarded as containing his philosophical sentiments and principles' (*E* Adv, 2).

3 The word 'utilitarian' as applied to moral theories was coined by Jeremy Bentham, but Hume frequently speaks of 'utility', as for example in the title of Section 5 of the *Enquiry*: 'Why Utility Pleases'.

4 What is widely known as Hume's Copy Principle states that 'all our ideas are nothing but copies of our impressions, or, in other words, that it is impossible for us to *think* of any thing, which we have not antecedently *felt*, either by our external or internal senses' (*E* 7.4, 62; cf. *E* 2.5, 19). In the *Treatise* (*T* 1.1.1.12, 7), Hume describes this as the 'first principle I establish in the science of human nature'.

5 Hume makes clear in the second *Enquiry* that he is completely opposed to the psychological egoism that is so frequently presumed by those who praise him – together with his younger friend Adam Smith – as heroes of 'free-market' economics (see note 29 below).

6 Indeed, one of Kant's most implausible suggestions is that an action done from pure benevolence, rather than out of respect for moral duty, lacks any moral worth. The humanity and warmth of Humean morality here seems far more attractive than the relatively puritanical legalism of Kant. Consider whether you would prefer to inhabit a world where everyone treats you with genuine benevolence, naturally empathizing with your pleasures and pains and spontaneously acting accordingly, or a world where people feel no personal concern for you, but all act morally out of undiluted respect for the Moral Law.

7 'Peripatetics' are followers of Aristotle, for example, the mediaeval schoolmen.

8 Shortly after these passages, Hume echoes the Aristotelian thought that proper pride or 'greatness of mind' is an especially central virtue, since

'Courage, intrepidity, ambition, love of glory, magnanimity, and all the other shining virtues of that kind, have plainly a strong mixture of self-esteem in them' (*T* 3.3.2.13, 599–600). Likewise in the *Enquiry*, 'A desire of fame, reputation, or a character with others, is so far from being blameable, that it seems inseparable from virtue . . . and a generous or noble disposition' (*M* 8.11, 265). He later remarks that pride 'may be either good or bad, according as it is well or ill founded, and according to the other circumstances which accompany it' (*M* App. 4.3 n. 66, 314).

9 Quoted (with my emphasis) from Eileen Sweeney, 'Vice and sin', in Pope 2002, 162: 'Pride . . . has a central place in Aquinas's account. Pride is the first sin, the source of all other sins, and the worst sin. He defines pride as an excessive desire for one's own excellence which rejects subjection to God (Ia IIae, q. 162, aa. 1, 5). . . . [E]very sin begins in turning from God and hence all sins begin in pride . . . the motive for acquiring all the lesser goods one prefers to God is pride, that through them one "may have some perfection and excellence" (Ia IIae, q. 84, a. 2). . . . In the *Secunda secundae*, Aquinas depicts pride as the original sin.'

10 Note that Hume's own usage of these words also varies, sometimes designating general characteristics and sometimes *excessive* levels of these characteristics (thus risking the false impression that he is contradicting himself). Compare 'self-satisfaction and vanity may not only be allowable, but requisite in a character' (*T* 3.3.2.10, 597) with 'VANITY . . . is so justly regarded as a fault or imperfection' (*M* 8.11, 266). At the beginning of his essay 'Of the Standard of Taste', Hume himself observes that the use of terms that carry a positive or negative flavour can mask substantial differences of judgement, both in aesthetics and morals.

11 Notice that the last four clauses state in turn that the 'monkish virtues' are not useful to the man himself, nor useful to others, nor agreeable to others, nor agreeable to himself.

12 That is, he might be made a saint.

13 Indeed Hume even jokes that 'if . . . a popular religion were found, in which it was expressly declared, that nothing but morality could gain the divine favour; if an order of priests were instituted to inculcate this opinion, in daily sermons, and with all the arts of persuasion; yet so inveterate are the people's prejudices, that . . . they would make the very attendance on these sermons the essentials of religion, rather than place them in virtue and good morals' (*N* 14.3).

14 Compare the comment in the *Dialogues* that 'If the religious spirit be ever mentioned in any historical narration, we are sure to meet afterwards with a detail of the miseries, which attend it' (*D* 12.11).

15 James Boswell reports Hume as saying on his deathbed 'that the Morality of every Religion was bad . . . that when he heard a man was religious, be concluded he was a rascal, though he had known some instances of very good men being religious'. Boswell suggests that Hume may here have been

deliberately reversing the then widely accepted principle that unbelievers, lacking fear of divine punishment, are generally immoral.

16 Hume is a *determinist*, believing that everything happens in accordance with universal causal laws; for the relevant texts, see Millican 2010. He is also a *compatibilist*, taking determinism to be entirely consistent with moral responsibility, though his view is distinctive and commonly misunderstood. For excellent discussions, see Russell 1995 and Botterill 2002. Determinism fits very comfortably with Hume's virtue ethics, which judges actions according to the *character* from which they flow. Accordingly, he sees moral judgement as requiring that actions be thus determined by character (*E* 8.28–30, 97–9). Note also that Hume's *sentimentalism* – by basing moral judgements on natural emotions – neatly steps sidesteps any metaphysical claim to the effect that determinism makes moral judgement inappropriate (*E* 8.34–5, 101–3).

17 Only Books 1 and 2 of the *Treatise* were published in late January 1739, whereas Book 3 did not appear until twenty-one months later.

18 Hume points out at that the word 'natural' can be understood differently depending on whether it is contrasted with the miraculous, the unusual or the artificial (*T* 3.1.2.7–9, 473–5; *M* App. 3.9 n. 64, 307–8).

19 Justice is discussed in *T* 3.2.1–2 and 6, property rights in *T* 3.2.3–4, promises in *T* 3.2.5, government in *T* 3.2.7–10, international law in *T* 3.2.11 and chastity and modesty in *T* 3.2.12. The natural virtues are considered – far more briefly – in Part 3, specifically 'greatness of mind' in *T* 3.3.2, goodness and benevolence in *T* 3.3.3, and natural abilities in *T* 3.3.4. Note that Hume's analysis of the virtues makes it hard to draw a clear line between so-called *natural abilities* and *moral virtues*, since both are typically useful or agreeable. In Appendix 4 of the *Enquiry*, he develops the argument of *T* 3.3.4 to maintain that this supposed distinction is 'merely verbal'.

20 Recall that Hume is not a classical utilitarian, but a virtue theorist who takes utility to be a criterion of virtue.

21 Such appetites are, of course – though Hume understandably does not say this – very easy to explain from an evolutionary perspective.

22 And, one might add, very easily explicable from a genetic viewpoint, in terms of 'inclusive fitness', whereby one would expect creatures that are capable of discriminating behaviour to evolve to favour those that share their genes. See Dawkins 1989, chapter 6.

23 Hume expands on this point in *T* 3.2.5, explaining why 'the obligation of promises' must be considered as artificial rather than natural, depending as it does on the existence of society. Once respect for property is established, the value of an institution of promising becomes clear, as in Hume's example of bargaining about mutual help in harvesting crops that ripen at different times (*T* 3.2.5.8, 520–21). But attempting to explain the historical origin of society in terms of some form of promise or social contract is hopeless, as Hume makes clear also in his classic essay 'Of the Original Contract'.

24 This point was famously illustrated by Robert Axelrod's computer experiments on the Iterated Prisoner's Dilemma, in which repetition proves to be the key factor that favours cooperative strategies over the selfish behaviour that dominates the 'one-shot' case. See Dawkins 1989, chapter 12.

25 See, for example, the books by Binmore, Blackburn, Mackie and Ridley in the 'Recommended reading' section.

26 Or, of course, both. Hume's fourfold classification of the virtues does not require that a virtue be *exclusively* useful or agreeable, *exclusively* to oneself or others. Virtues can be useful *and* agreeable, to oneself *and* others.

27 See, for example, *T* 3.1.1.9, 458; *M* App. 1.21, 294; and *P* 5.1.

28 It is a central principle of Hume's philosophy that causal relations can be known only through experience and not by any such aprioristic reasoning.

29 However, not all our behaviour is driven by the prospect of personal pleasure and pain, and Hume rejects psychological egoism (though it is arguable that in the *Treatise*, he *started from* an egoist position: see Millican 2020, 281–9). Various 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 . . . or in other words, pain and pleasure . . . and proceed not from them, like the other affections' (*T* 2.3.9.8, 439). This last point perhaps evinces a recognition that the psychological egoist puts the cart before the horse in considering all behaviour to be selfish, since it is typically through the satisfaction of *antecedent* desires (for something other than pleasure) that we derive pleasure. Hume refutes the 'selfish hypothesis' most forcefully in Appendix 2 of the second *Enquiry*, and also attacks it in his 1741 essay 'Of the Dignity or Meanness of Human Nature'.

30 This distinction is now commonly expressed in terms of 'direction of fit': what Hume calls our *reason* – our *cognitive* faculty – aims to conform our mind (i.e. our beliefs) to the world, whereas the *will* – our *conative* faculty – aims to conform the world to our mind (i.e. our desires).

31 The contrast is especially marked given how conspicuously he presented it *twice* in the *Treatise*. Hume was an acute detector of sophistry, and it seems most likely that he became well aware of the argument's problems.

32 See, for example, Prinz 2006, who usefully surveys recent evidence for a close link between emotion and moral judgement (thus undermining the Kantian claim that rationality – rather than emotion – provides the key). For more details on these issues, see the book by Churchland in the 'Recommended reading' section. The books by Binmore and Blackburn also stress how natural moral sentiments – by enhancing the sanctions associated with non-cooperation – can play a valuable role in helping to establish the reciprocal altruism that lies at the basis of much moral behaviour.

33 Hume seems rather complacent about the extent to which our judgements can be expected to converge under the pressure of this sort of

systematization (especially in his essay 'Of the Standard of Taste', where he deals with aesthetic judgement). He confronts the issue of moral relativism in 'A Dialogue' (effectively a fifth appendix to the *Enquiry*), arguing that the variation in moral attitudes between different cultures can be accounted for in a uniform manner, by appeal to the variability of utility with context.

34 There has been considerable discussion in the scholarly and philosophical literature of the extent to which Hume should be considered a 'projectivist' about morality, and in what sense(s). Such discussions have provoked much interesting philosophy, though they seem unlikely to result in any determinate conclusion given the scarcity of the relevant textual evidence, and the unlikelihood that Hume himself thought through the issues with anything like the same sophistication that we are able to bring to the issue after another 250 years of philosophical development. If Hume were alive today, I think he would be more interested in the scientific exploration of morality than in such subtle philosophical explication of ways of thinking about it.

References

References to Hume's *Treatise* (*T*), *Enquiry concerning Human Understanding* (*E*) and *Enquiry concerning the Principles of Morals* (*M*) provide section and paragraph numbers, followed by page references to the standard Oxford Selby-Bigge editions. Several recent printed editions include both means of reference, and they can also be found on the web editions at www.davidhume. org, from which the quotations are taken. In the case of Hume's *Dialogues concerning Natural Religion* (*D*), *Natural History of Religion* (*N*), *Dissertation on the Passions* (*P*) and various essays (always referenced by name), only paragraph numbers are given: these also are available at www.davidhume.org.

Botterill, G. (2002), 'Hume on Liberty and Necessity', in P. Millican (ed.), *Reading Hume on Human Understanding*. Oxford: Clarendon Press, 277–300.

Dawkins, R. (1989), *The Selfish Gene* (new edition). Oxford: Oxford University Press.

Millican, P. (2010), 'Hume's Determinism', *Canadian Journal of Philosophy* 40: 611–42.

—— (2020), 'The Relation between Hume's Two *Enquiries*', in Jacqueline Taylor (ed.), *Reading Hume on the Principles of Morals*. Oxford: Oxford University Press, 273–99.

Prinz, J. (2006), 'The Emotional Basis of Moral Judgments', *Philosophical Explorations* 9: 29–43.

Russell, P. (1995), *Freedom and Moral Sentiment*. Oxford: Oxford University Press.

Sweeney, E. (2002), 'Vice and Sin', in Stephen J. Pope (ed.), *The Ethics of Aquinas*. Georgetown: Georgetown University Press, 51–68.

Recommended reading

Baillie, James, *Hume on Morality* (London: Routledge, 2000) gives a systematic account of Hume's theories of the passions and morality, in both the *Treatise* and the *Enquiry*, in a form that is likely to be especially useful to students working through these texts.

Binmore, Ken, *Natural Justice* (Oxford: Oxford University Press, 2005), written by a prominent economist, explains morality as the solution to a game-theoretic coordination problem, with Hume given credit for finding the solution more than two centuries before others (pp. 8–9).

Blackburn, Simon, *Ruling Passions* (Oxford: Clarendon Press, 1998) is a sophisticated but accessible treatment of morality by one of the foremost moral philosophers in the world today. It covers a wide range of issues, developing a Humean theory in the direction of Blackburn's favoured 'quasi-realist' account of morality.

Churchland, Patricia, *Braintrust: What Neuroscience Teaches Us about Morality* (Princeton: Princeton University Press, 2011) gives a biological explanation of morality in a similar spirit to the Humean account sketched in my last two sections here.

Mackie, John Leslie, *Hume's Moral Theory* (London: Routledge & Kegan Paul, 1980) provides an excellent philosophical discussion of Hume's theory of morals (as presented in the *Treatise*), placing it within its historical context and stressing the 'error theory' for which Mackie is well-known through his book *Ethics: Inventing Right and Wrong* (London: Penguin, 1977).

Ridley, Matt, *The Origins of Virtue* (London: Penguin, 1997) gives a lively popular discussion of the evolution of morality, with Hume's insight from *T* 3.2.5.9 given prominence in the discussion of the Prisoner's Dilemma.