5

HUME ON FREE WILL AND MORAL RESPONSIBILITY

Peter Millican

David Hume's views on morality – the topic of both Book 3 of his *Treatise of Human Nature* (1740), and of his *Enquiry concerning the Principles of Morals* (1751) – are widely familiar and much discussed, but his view of *moral responsibility* is far less well known, and even rather obscure. To piece this view together, we must examine his theory of intentional agency, which he expounds under the heading "Of Liberty and Necessity". But there are significant discrepancies between his discussions of this topic in Book 2 of the *Treatise of Human Nature* (1739) and the *Enquiry concerning Human Understanding* (1748). And his theory has been widely misunderstood for a number of reasons, including confusion about his theory of causal necessity and a tendency to view him through the lens of subsequent writers in the empiricist tradition. So although Hume's writings "of liberty and necessity" are amongst the most widely known in the philosophical canon, achieving a reliable interpretation of his settled opinion on these topics – and on the closely related topic of moral responsibility – requires careful analysis.

1 Hume's theory of causation

Hume famously analyses causal relations in terms of mere "constant conjunction" between types of event, where the cause is regularly followed by the effect.[1] Many of his readers down the years have interpreted this theory as implicitly denying that there is any genuine *causal necessity* between a cause and its effect, thus perhaps suggesting skepticism about causal relations themselves. But such readings are incorrect if we understand the relevant terms in accordance with his own theory. First, *Hume clearly believes in causal relations* as understood on the regularity account, and indeed frames explicit rules to identify them:

> it may be proper to fix some general rules, by which we may know when [objects] really are . . . causes or effects to each other (*T* 1.3.15.2, cf. *T* 1.3.13.11)

> all objects, which are found to be constantly conjoin'd, are upon that account only to be regarded as causes and effects (*T* 1.4.5.32)

DOI: 10.4324/9781003282242-9

Secondly, Hume insists – both before and after developing this theory – that *necessity is essential to causation*. Indeed, his long investigation into the origin (and hence nature) of our "idea of necessary connexion" takes off from this premise, when he asks what *more* is needed for the causal relation to hold between two "objects", besides (single-case) contiguity and temporal priority:

> An object may be contiguous and prior to another, without being consider'd as its cause. There is a NECESSARY CONNEXION to be taken into consideration; and that relation is of much greater importance, than any of the other two above-mention'd. (*T* 1.3.2.11)

To achieve a clearer understanding of this key idea, Hume employs what is commonly known as his Copy Principle, for which he has previously argued (in *T* 1.1.1 and *EHU* 2):

> It seems a proposition, which will not admit of much dispute, 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. (*EHU* 7.4, cf. *T* 1.3.14.1)

This is the core of his *conceptual empiricism*, whereby all the contents of our thoughts are derived either from external sense experience (sight, touch, hearing etc.), or from internal "reflection" (feelings, emotions, etc.). Hume eventually (*T* 1.3.14.20, *EHU* 7.28) tracks down the "impression of necessary connexion" to something that we understand through reflection, namely *customary transition of the mind* from observed cause *A* to expected effect *B*, in response to repeated observation of *A*s followed by *B*s – what we now call *inductive inference*. His declared success in this quest clearly indicates that he takes the corresponding idea to be legitimate, since it has a genuine – though surprising – impression-source, rather than being a confused "fiction".[2] Hume then frames on this basis two "definitions of cause", the first couched in terms of constant conjunction and the second in terms of mental inference:

> Similar objects are always conjoined with similar. Of this we have experience. Suitably to this experience, therefore, we may define a cause to be *an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. . .. The appearance of a cause always conveys the mind, by a customary transition, to the idea of the effect. Of this also we have experience. We may, therefore, suitably to this experience, form another definition of cause; and call it, *an object followed by another, and whose appearance always conveys the thought to that other*. (*EHU* 7.29, cf. T 1.3.14.31)

The prominent role of mental inference in these discussions has led many interpreters to presume that he is a *subjectivist* about causal necessity (in the sense of considering such necessity to be *observer-relative*), and hence denies it as an objective factor within causal relations. But this is to ignore that in both works – and almost immediately in the

Enquiry – Hume goes on to provide two parallel "definitions of necessity", again in terms of constant conjunction and mental inference. These serve to vindicate his claim that necessity is *essential* to causality, on *both* definitions:

> According to my definitions, necessity makes an essential part of causation (*T* 2.3.1.18, cf. *EHU* 8.25)

> Necessity may be defined two ways, conformably to the two definitions of *cause*, of which it makes an essential part. (*EHU* 8.27, cf. *T* 2.3.2.4)

So Hume is neither skeptical about causal relations nor about causal necessity, as long as these are understood in accordance with his definitions. Let us now go on to see how Hume applies these definitions in arguing for the deterministic thesis which he calls "the doctrine of necessity".

2 The "doctrine of necessity"

Hume's discussions entitled "Of liberty and necessity" – in both *Treatise* 2.3.1-2 and *Enquiry* 8 – are largely devoted to establishing what he calls *the doctrine of necessity* in the sphere of human behavior. This doctrine states that all such behavior is subject to *causal necessity*. On Hume's view as explained previously, such causal necessity is simply a matter of events' occurring in conformity with deterministic causal laws, from which, therefore, those events could be predicted at least *in principle* (i.e., by a hypothetical being who knows all the relevant initial conditions and laws, and is able to calculate their workings in detail). But previous philosophical orthodoxy, as represented by the influential Newtonian Samuel Clarke, took causal necessity to involve more than mere *conformity* with predictive laws. For Clarke, *real necessity* is to be understood on the model of *mechanical impulse* – when one body in motion smashes into another and *forces it* to move – so it is a mistake to think of intelligent agents – acting on reasons or motives – as subject to such necessity. We might *call* such agents' behavior "morally necessary" in the sense of being entirely predictable in principle (i.e., the minimal deterministic "necessity" that Hume endorses), but if so, that is "merely a *figurative Manner of Speaking*" and "not indeed any *Necessity* at all" (Clarke 1717, 15–17).[3]

One of the main aims of Hume's discussions "Of liberty and necessity" is to undermine this alleged distinction between *physical* and *moral* necessity, by arguing that *our only possible understanding of causal necessity is in terms of his two definitions*, and that these apply equally to the physical and moral (i.e., human) worlds. So Hume's primary motivation in applying his Copy Principle to track down the "impression of necessary connexion" seems to be to circumscribe the limits of *what we can possibly mean* by causal necessity. On this basis, he insists that those such as Clarke who claim that "there is something else in the operations of matter" – some form of supposed genuine *necessity* that goes beyond anything in human behavior – are using terms that have no corresponding idea, and are therefore "unintelligible" (*T* 2.3.2.4, cf. *A* 34, *EHU* 8.22).[4]

Hume's argument for the doctrine of necessity (i.e., determinism) within the human sphere aims first to establish and then build on these conceptual claims. He starts from what he takes to be the "universally" agreed predictability of "the operations of external bodies", in which "Every object is determin'd by an absolute fate to a certain degree and

direction of its motion" with "not the least traces of indifference or liberty" (*T* 2.3.1.3), and "every natural effect is so precisely determined by the energy of its cause, that no other effect, in such particular circumstances, could possibly have resulted from it" (*EHU* 8.4). Then he refers back to his two definitions of *cause*, and draws from them two parallel criteria of *causal necessity*, which he aims to prove apply to the human world as much as to the physical: "Here then are two particulars, which we are to consider as essential to necessity, viz. the constant *union* and the *inference* of the mind; and wherever we discover these we must acknowledge a necessity" (*T* 2.3.1.4). The version in the *Enquiry* is more explicit about the *conceptual* or *semantic* aspect of this argument – that it hinges on questions of *meaning* as determined by the limits of our *ideas*:

> Our idea, therefore, of necessity and causation arises entirely from the uniformity, observable in the operations of nature; where similar objects are constantly conjoined together, and the mind is determined by custom to infer the one from the appearance of the other. These two circumstances form the whole of that necessity, which we ascribe to matter. Beyond the constant *conjunction* of similar objects, and the consequent *inference* from one to the other, we have no notion of any necessity, or connexion. (*EHU* 8.5)

An entry in Hume's own index to the *Enquiry* – "Necessity, its definition" – makes explicit that he understands this last statement as *defining necessity* (citing also the passage from *EHU* 8.27 quoted earlier). The index has a quite separate entry for "Cause and Effect . . . Its Definition", referring to the two definitions of cause at *EHU* 7.29 (and the footnote at *EHU* 8.25). So Hume clearly distinguishes his definitions of necessity from these earlier and more famous two definitions of cause, though they are obviously closely related.

As we have seen, Hume casts this argument in *semantic* terms – as restricting *what we can possibly understand or mean by causal necessity* – in order to undermine the orthodox claim that "moral" and "physical" causation are of a different nature. But he also implicitly appeals to this semantic theme when attempting to dissolve the debate as having turned hitherto on a *misunderstanding*:

> If it appear, therefore, that all mankind have ever allowed, without any doubt or hesitation, that these two circumstances take place in the voluntary actions of men, and in the operations of mind; it must follow, that all mankind have ever agreed in the doctrine of necessity, and that they have hitherto disputed, merely for not understanding each other. (*EHU* 8.5-6)

This attempted dissolution might seem a bit far-fetched with regard to "the most contentious question, of metaphysics, the most contentious science" (*EHU* 8.23), but Hume suggests that the debate's very intractability is indicative of some persisting misunderstanding (*EHU* 8.1-3).

The bulk of Hume's subsequent argument, however, appeals to considerations that seem *observational* more than *semantic*, because they emphasize not so much people's assumptions about the regularity of human behavior – and the fact that they make inferences accordingly – but rather, that such regularity *actually* obtains, so that such inference is *in fact* reliable. Accordingly, the first part of his strategy is to "prove from experience that our actions have a constant union with our motives, tempers, and circumstances" (*T* 2.3.1.4, cf. *EHU* 8.7).

After giving a variety of examples to support and illustrate this claim (*T* 2.3.1.5-10, *EHU* 8.7-15), Hume concludes "that the conjunction between motives and voluntary actions is as regular and uniform, as that between the cause and effect in any part of nature" (*EHU* 8.16, cf. *T* 2.3.1.12). In the *Enquiry*, he also at this point claims to have shown "that this regular conjunction has been universally acknowledged among mankind", in which case "it may seem superfluous to prove, that this experienced uniformity in human actions is a source whence we draw *inferences* concerning them" (*EHU* 8.16), though he goes on to illustrate the latter anyway (*EHU* 8.17-20). In the *Treatise*, he separates out this last task of showing

> that as the *union* betwixt motives and actions has the same constancy, as that in any natural operations, so its influence on the understanding is also the same, in *determining* us to infer the existence of one from that of another (*T* 2.3.1.14).

His argument for this mixes illustration (*T* 2.3.1.15, 17) with appeal to his theory of necessity (*T* 2.3.1.16, 18).

3 Determinism, chance, and some reservations

Hume's argument for "the doctrine of necessity" is by no means compelling. The complexity of the human brain and associated faculties (both sensory and motor) render it hopelessly unfeasible to establish, in the sphere of human behavior, that there are "not the least traces of indifference", or that "every effect is so precisely determined, that no other effect, in such particular circumstances, could possibly have resulted" (cf. *T* 2.3.1.3 and *EHU* 8.4, as quoted previously). Even in the physical world, such claims are dubious (as has been emphasized by quantum mechanics), but at least there it is possible to test and measure isolated systems with relatively great accuracy, thus giving a plausible basis for extrapolation to more complex phenomena. But there is obviously no way of even beginning to study human thought and behavior *in isolation* from the unfathomable complexity of our bodies and our environment (both physical and social). Determinism about the mind and behavior may be plausible, but support for it is likely to come less from direct experimental confirmation than from more general considerations – and perhaps prejudices – such as the desire for explanatory completeness (thus ruling out randomness), the belief that mental phenomena are dependent on a deterministic physical substrate (as strongly suggested by evolutionary considerations), and the conceptual difficulty of envisaging a non-random alternative to causal explanation.[5]

In the 18th century, following Newton's impressive achievements, physical determinism was widely taken for granted. Those inclined to materialism, such as Hobbes and Hume, would naturally extend determinism equally to the "moral" sphere of human behavior, on the basis that humans – like everything else in the world – are purely material beings. But Christians such as Clarke, believing in a realm of immaterial spirits, typically considered mental causation to operate quite differently. Both sides, however, were equally committed to the Causal Maxim discussed in *Treatise* 1.3.3, that "*whatever begins to exist, must have a cause of existence*". And the Maxim was also commonly understood as extending to any *change*, not just new beginnings of existence. So both materialists and Christians were typically committed to *universal causation*, but differed in respect of the *kinds of causation* they countenanced, and in particular, whether human

behavior involved *purposive causation by spirits* (as opposed to mechanical causation by bodies). Neither of these sides would typically understand causation as involving *randomness*, because while materialists tended to attribute all causation to deterministic matter, Christians opposed the "Epicurean" hypothesis of *chance* as dangerously atheistic. Hume was therefore reflecting this consensus when he reported as "commonly allow'd by philosophers, that what the vulgar call chance is nothing by a secret and conceal'd cause" (*T* 1.3.12.1).[6]

All this helps to explain why Hume's case for the "doctrine of necessity" – understood as equivalent to universal determinism – can seem to us rather vague and complacent. At best it shows that determinism about human behavior *might plausibly* be true, because just as we presume that unpredictable changes in our health (*EHU* 8.14) and the weather (*EHU* 8.15) are explicable by underlying physical causes (even when these causes may be unknown, and are anyway far too complex for us to work out in detail); so "The philosopher, if he be consistent, must apply the same reasoning to the actions and volitions of intelligent agents" (*EHU* 8.15). In the case of the physical world, however, Hume is able to offer rather more, because here, not only can "the philosopher" reasonably consider it "at least possible [that] the contrariety of events [proceeds] . . . from the secret operation of contrary causes"; but also, "This possibility [may be] converted into certainty by farther observation; when . . . upon an exact scrutiny, a contrariety of effects always betrays a contrariety of causes". Hume gives the example of a clock or watch that stops unexpectedly, and an artisan who, on examining it, finds a grain of dust interfering with the mechanism. This discussion is the closest that Hume comes to giving solid *evidence* for the truth of determinism. But there is no such relative solidity to be had in the world of mind and behavior, where hidden mechanisms cannot be opened up and examined in detail. So again, Hume's case looks less than totally convincing.

The same can be said for his claim "that the conjunction between motives and voluntary actions is as regular and uniform, as that between the cause and effect in any part of nature" (*EHU* 8.16). For motives result in actions only through complex causal chains involving bodies and other physical things, with many potential obstacles on the way. Indeed, Hume had observed in just the previous paragraph how physical circumstances can impact on us, potentially affecting both our motives and our ability to act on them:[7] "A person of an obliging disposition gives a peevish answer: But he has the toothake, or has not dined" (*EHU* 8.15). Hence his specific focus on "the conjunction between motives and voluntary actions" seems to be a slip (cf. Botterill 2002: 285–286). His overall view would leave him free to acknowledge the complex interplay between mental and physical, and to expect deterministic patterns across the entire system, but not necessarily within the "moral" sphere considered separately.

4 Introducing "hypothetical liberty"

In his discussion "Of liberty and necessity" in *Treatise* 2.3.1-2, Hume mostly uses the term "liberty" not as applying specifically to human action, but as meaning simply *chance* or *absence of necessity* (e.g. *T* 2.3.2.2, 2.3.2.6-7) – something which, as we have seen, he rejects. He there refers to "free-will" only once, equating it with both "liberty" and "chance", and accordingly denying its existence also (*T* 2.3.1.18). This might suggest that Hume takes a "hard determinist" position, denying moral freedom on the basis that we are causally determined. But on the contrary, at *T* 2.3.2.5-7 he argues that his deterministic

position, so far from undermining morality, is *essential* to it; then in Book 3 he goes on to develop in detail his own positive theory of morality. When he came to write the first *Enquiry*, Hume corrected the false impression of hard determinism by *endorsing* "liberty", having defined it quite differently from his usage in the *Treatise*, and as entirely consistent with determinism.

> By liberty . . . we can only mean *a power of acting or not acting, according to the determinations of the will*; that is, if we chuse to remain at rest, we may; if we chuse to move, we also may. Now this hypothetical liberty is universally allowed to belong to every one, who is not a prisoner and in chains. (*EHU* 8.23)

The same paragraph famously starts by describing this discussion as a "reconciling project with regard to the question of liberty and necessity", so Hume's compatibilist intention is entirely explicit.

But the precise nature of his compatibilist theory is not so clear. Looking back on some occasion when I have moved or stayed at rest – to follow Hume's example – a minimal interpretation of his "hypothetical liberty" might simply require that the act I performed was *in fact* in accordance with my will:

> *Either* I willed to move, and did move; *or* I willed to stay at rest, and did stay at rest.

But a more fulsome interpretation would require *also* a counterfactual condition, such as:

> *If* I had willed differently, *then* my action would have been correspondingly different.

It might be suggested that Hume ought to favor the minimal interpretation, on the basis that his determinism would rule out the counterfactual possibility of my having willed differently, or to respect the puzzling *T* 1.3.14.34 corollary to his definitions of cause, that "The distinction, which we often make betwixt *power* and the *exercise* of it, is . . . without foundation" (cf. Penelhum 2000: 162). But determinists are generally happy with counterfactuals corresponding to causal laws, while Hume's commitment to his implausible corollary weakens when he considers our passions and probabilistic thinking (*T* 2.1.10.4, 6, 10; 2.2.5.7). Indeed the corollary is absent from the *Enquiry*, where Hume's chosen term – "*hypothetical* liberty" – strongly suggests a counterfactual interpretation. Moreover his choice of example here – moving or not moving – is suggestively reminiscent of the discussion in Locke's *Essay*, which seems explicitly to require the two-way (counterfactual) power of the more fulsome interpretation:

> so far as a Man has a power to think, or not to think; to move, or not to move, according to the preference or direction of his own mind, so far is a Man *Free*. Where-ever any performance or forbearance are not equally in a Man's power; where-ever doing or not doing, will not equally follow upon the preference of his mind directing it, there he is not *Free*, though perhaps the Action may be voluntary. (*Essay* II xxi 8)

Given Hume's Lockean background, and "the absence of any Humean signals to the contrary", Bricke suggests that "it is reasonable to read Hume's doctrine in the standard . . . way", as involving two-way "implementability of relevant alternatives" (2008: 208). This

seems right, but before confidently drawing this conclusion, we need to examine some further textual complications.

5 "Liberty of spontaneity" and a puzzling mismatch

In *Treatise* 2.3.2, Hume attempts to explain why people are *falsely* tempted to think of their actions as not determined, and argues against the concern that determinism would undermine morality. Much of this discussion is reproduced in *Enquiry* 8 (21-2 and 26-31), but the best-known passage is confined to the *Treatise*:

> Few are capable of distinguishing betwixt the liberty of *spontaneity*, as it is call'd in the schools, and the liberty of *indifference*; betwixt that which is oppos'd to violence, and that which means a negation of necessity and causes. The first is even the most common sense of the word; and as 'tis only that species of liberty, which it concerns us to preserve, our thoughts have been principally turn'd towards it (*T* 2.3.2.1)

Here Hume gives a clear impression of endorsing the "liberty of *spontaneity*" as his favored account of free will, though the term does not appear at all in the *Enquiry*. This quite strongly suggests that it is another name for his "hypothetical liberty", especially when both are contrasted with just one other sense of "liberty", namely indifference or chance, and in the light of Hume's emphatic statement at *EHU* 8.23 that if we speak of "liberty" in any reasonable sense, "we can only mean" his hypothetical liberty. But the first sentence of this initial paragraph in the *Treatise* significantly muddies the water: "I believe we may assign the three following reasons for the prevalence of the doctrine of liberty, however absurd it may be in one sense, and unintelligible in any other" (*T* 2.3.2.1). This seems to be describing the doctrine that we have "liberty" as *absurd* if that is understood as indifference, and *unintelligible* otherwise. So Hume apparently considers "liberty of *spontaneity*" to be unintelligible, even though he then goes on to refer to it as "the most common sense of the word; and . . . that species of liberty, which it concerns us to preserve". But how can it concern us to preserve something that is unintelligible? And if, on the other hand, "liberty of *spontaneity*" is indeed intelligible, then isn't it perverse that elsewhere in the *Treatise* Hume prefers to use "liberty" in something other than its "most common sense", and indeed another sense that he himself calls "absurd"?

But putting aside any such perversity, and assuming that the *liberty of spontaneity* of the *Treatise* is indeed intelligible, does this really equate with the *hypothetical liberty* of the *Enquiry*?[8] Most previous scholars have indeed affirmed this identification, but it seems hard to square with Hume's statements that his hypothetical liberty is "universally allowed to belong to every one, who is not a prisoner and in chains", whereas by contrast spontaneity is "that species of liberty, which it concerns us to preserve" and which "our thoughts [are] principally turn'd towards" (thus suggesting that it is easily lost, and very far from universal). People anxious about their freedom are most commonly thinking about such things as political, religious, social or family pressures that constrain what they can do for fear of punishment or opprobrium; but they are relatively rarely in danger of being chained up in prison. If, for example, a gang boss threatens that my factory will be burned down if I continue to compete with his business, or a community elder threatens me with social ostracism if I continue my relationship with some outsider, then I am likely to be seriously concerned by this impairment of my freedom. But neither of these apparently deprives me

of the minimal "hypothetical liberty" explicated in the *Enquiry*, which seems to amount to little more than a capacity for *voluntary agency*: to apply my will in whatever situation I am placed, with all its implied consequences and constraints. For in these situations, if I choose to do as I have been ordered, I may; and if I choose to defy the instruction (and take the consequences), I also may. So apparently I have "hypothetical liberty", but not the sort of freedom "which it concerns me to preserve". Which, then, is the genuine Humean notion?

6 A traditional misreading, and a resolution of the puzzle

I suspect that previous scholars have often viewed Hume through the lens of later thinkers in the empiricist compatibilist tradition, such as A. J. Ayer in his classic essay "Freedom and Necessity":

> it is not, I think, causality that freedom is to be contrasted with, but constraint. . .. If I am constrained, I do not act freely. . .. An obvious instance is the case in which I am compelled by another person to do what he wants . . . the compulsion need not be such as to deprive one of the power of choice. . .. [But] if . . . no reasonable person would be expected to choose the other alternative, then the action that I am made to do is not one for which I am held to be morally responsible. (Ayer 1954: 278-9)

Ayer's identification of freedom with *lack of constraint* (rather than *lack of causality*) seems to echo Hume:

> the idea of necessity [seems] to imply something of force, and violence, and constraint, of which we are not sensible (*T* 2.3.2.1)

> if the definition [of cause] above mentioned be admitted; liberty, when opposed to necessity, not to constraint, is the same thing with chance; which is universally allowed to have no existence (*EHU* 8.25)

The first of these passages immediately precedes Hume's mention of *liberty of spontaneity* in the *Treatise*, which as we saw earlier, talks of such liberty as "oppos'd to violence". The second passage appears, in a similar spirit, to be distinguishing between *liberty of indifference* – as "opposed to necessity" – and Hume's *hypothetical liberty* – as "opposed . . . to constraint".

There is, however, a very different way of interpreting Hume's language, which avoids the puzzling mismatch between "spontaneity" understood in Ayer's way and "hypothetical liberty" as described in the *Enquiry*. This alternative interpretation is strongly supported by a later passage in the *Enquiry*, where Hume explains why the latter notion is appropriate to moral appraisal:

> It will be equally easy to prove, and from the same arguments, that liberty, according to that definition above mentioned, in which all men agree, is also essential to morality, and that no human actions, where it is wanting, are susceptible of any moral qualities, or can be the objects either of approbation or dislike. For as actions are objects of our moral sentiment, so far only as they are indications of the internal character, passions, and affections; it is impossible that they can give rise either to praise or

> blame, where they proceed not from these principles, but are derived altogether from external violence. (*EHU* 8.31)

An action that is "derived altogether from external violence" (e.g., being thrown downstairs by thugs) is quite different from an action performed *from fear of violence* (e.g., complying with a gang boss's commands for fear of being thrown downstairs). And if Hume's talk of "violence" in these contexts is thus understood, then this in turn suggests that when he speaks of "force, and violence, and constraint" in the *Treatise* (*T* 2.3.2.1), he has in mind *physical* force, violence, and constraint – which fits precisely with his reference in the *Enquiry* to the prisoner in chains. I conclude, therefore, that whereas Ayer's paradigm of "constraint" is *coercion by threat*, Hume's paradigm is *physical restraint* such as by chains, or a straitjacket, or physical force. Understood in this way, Hume's "liberty of spontaneity" as "oppos'd to violence" (*T* 2.3.2.1) fits closely with "hypothetical liberty" as "universally allowed to belong to every one, who is not a prisoner and in chains" (*EHU* 8.23).

7 Two-way liberty and what we seek to preserve

If the sorts of constraints that Hume has in mind as *thwarting* hypothetical liberty are indeed *physical* constraints such as prison walls and chains, then this gives extra support to the two-way or counterfactual interpretation of Hume's "hypothetical liberty" that was tentatively supported in section 4. When thus constrained, details about the causal operations of our decision-making mental processes seem beside the point: our liberty is undermined in a far more straightforward way. So I think Hume would agree with Locke (*Essay* II xxi 10) that a man who is *locked* in a room where he wishes to stay (owing to the "desirable Company"), may be staying *voluntarily* – at least if he is unaware of the locked door – but he is not *free* with regard to staying or leaving. Regardless of the operation of his will, or whatever might determine it, he does not physically have the power to leave.

Of course the man in the locked room is still free with regard to *other* choices (e.g., whether or not to speak, to move, or to attempt to leave). So this brings to light an issue implicit in Hume's definition of *hypothetical liberty*, but not explicitly noted by him (nor by most commentators); namely, that the "power of acting or not acting" can be understood as *relative to the choice in question*.[9] That being so, it might not after all be unreasonable to see this kind of liberty as one that we are anxious to preserve, not just minimally (e.g., moving our finger or not), but with as much scope as possible, so that we have the power to do, or not to do, a multitude of different things. In this way, our desire to remain out of prison and unchained can indeed be seen as concerned with maximizing our "power of acting or not acting". If Hume was thinking along these lines, then he might after all have had the *hypothetical liberty* of the *Enquiry* in mind when, in the *Treatise*, he described *liberty of spontaneity* as "that species of liberty, which it concerns us to preserve". This would yield an interpretation that neatly reconciles Hume's favored notions of liberty in the two works.

8 Moral responsibility

Just one passage in Hume's philosophical writings – at *T* 2.3.2.6 – talks about the conditions for *responsibility* using that very term.[10] But the points it makes are echoed elsewhere (notably *EHU* 8.29-30), and follow familiar themes in his theory of moral

appraisal, which focuses on enduring *qualities of mind* in the agent that may be judged as either *virtues* or *vices*:

> The constant and universal object of hatred or anger is a person or creature endow'd with thought and consciousness; and when any criminal or injurious actions excite that passion, 'tis only by their relation to the person or connexion with him. But according to the doctrine of liberty or chance, this connexion is reduc'd to nothing. . . . Actions are by their very nature temporary and perishing; and where they proceed not from some cause in the characters and disposition of the person, who perform'd them, they . . . can neither redound to his honour, if good, nor infamy, if evil. The action itself may be blameable. . . . But *the person is not responsible for it*; and as it proceeded from nothing in him, that is durable or constant . . .'tis impossible he can, upon its account, become the object of punishment or vengeance. 'Tis only upon the principles of necessity, that a person acquires any merit or demerit from his actions, however the common opinion may incline to the contrary.[11] (*T* 2.3.2.6, emphasis added)

According to Hume's virtue-ethical theory, we judge behavior morally in terms of the *qualities of mind or character* that the relevant actions evince, rather than the *type of action* (*T* 3.3.1.4, *EPM* 1.10). And we judge these qualities of mind by their *general tendencies* rather than *specific consequences* (*T* 3.3.1.30, *EPM* 9.12). In its ultimate form in the second *Enquiry*, this constitutes something like a *utilitarian virtue ethics*,[12] with *reason* playing a crucial role in assessing these general tendencies of characters, while *sentiment* generates our approval of beneficial outcomes and disapproval of bad (*EPM Appx* 1.2-3). In the *Treatise*, by contrast, Hume downplays the role of reason, and devotes far more attention to our moral sentiments, explaining how the mechanism of *sympathy* plays a crucial role by causing the ideas that we have of other people's pains and pleasures to become enlivened into impressions that we ourselves feel (*T* 2.1.11.2-7).[13] This mechanism is non-rational – just part of our natural psychology – as is our tendency to feel distinctive *moral sentiments* of approval and disapproval when we consider certain characters of mind and their good or bad tendencies from a "*general*" view (*T* 3.3.1.23, 27-30).

Many scholars – notably Paul Russell in his influential 1995 book *Freedom & Moral Sentiment* – have accordingly seen the key to Humean moral responsibility as residing in these natural moral sentiments. But there is an important distinction between moral *responsibility* and moral *culpability*. Suppose that an intentional action exhibits some quality of mind – maybe it is *kind*, or *sadistic*, or *cautious*. If the action was *kind* or *sadistic*, then we would appropriately judge it (respectively) as virtuous or vicious, but if the action was *cautious*, then our judgment might well be morally neutral. Though some people are indeed more cautious than others, this "quality of mind" is neither an obvious virtue nor an obvious vice: in different contexts, different degrees of caution can be appropriate. But judging an action as morally neutral does not remove the agent's responsibility; it simply means that their responsibility for the action brings them neither credit nor discredit. In other words, it is one thing to be *responsible* for an action; it is quite another to be *morally praiseworthy* or *morally culpable* for doing that action. Moral assessment *of an action* presupposes responsibility, but an agent can be *responsible* without being either *praiseworthy* or *culpable*.

I suggest that Hume's account of *hypothetical liberty* is intended to explain what it is to be *responsible* for an action, whereas his account of *moral sentiments* is intended to explain what it is for a responsible action to have *a moral valence*, either positive or negative. I am

free in Hume's sense with respect to action *A* if my situation is such that, whether I choose to do *A* or to refrain from *A*, I shall be able to carry through that choice. The crucial "determination" here is in the causal chain between my volition and the action (rather than the causation of my volition), so whether *A* occurs or not is suitably determined by my decision, and faithfully reflects it. Thus I am the voluntary author of what happens, straightforwardly *responsible* for it because my volition was a determining factor in its causation. And if that volition reflects an enduring feature of my character, then the action will also reflect my enduring character, thus providing an appropriate basis for morally judging me if the character thus revealed is virtuous or vicious.[14] In such judgement, the moral sentiments indeed play a crucial role (as Russell emphasizes), but they are not crucial for assessing *whether or not* I am responsible; only whether the action would redound to my credit or my discredit.

This division of labor makes good sense of Hume's text, and explains why he says so little to connect moral sentiments with responsibility.[15] It also provides a much cleaner way of analysing some cases, as compared with Ayer's very broad appeal to "constraint" which treats *coercion* as a paradigm. For if a gangster threatens me with a gun and orders me to apply my thumb to open my employer's digital cash register (say), then this need not undermine my ability to think rationally and decide what to do in the given situation. Ayer would say that my moral freedom is lost, but a better Humean response is to say that my opening of the cash register remains morally free (as long as the gangster is not physically forcing my thumb against the detector), because it is under my voluntary control. Both Ayer and Hume can agree that I would not be morally *culpable* for this action, but whereas Ayer would attribute this to a lack of moral freedom, Hume's response is more straightforward: *in that situation*, complying with the order was the *right* (or at least *permissible*) thing to do.

In this way, Hume's account can avoid the major difficulty for classic compatibilists such as Ayer, of distinguishing those "constraints" that supposedly undermine moral freedom from the various risks, threats, and limitations that are just part of normal life. But Hume's very thin conception of moral freedom as simple voluntariness faces well-known difficulties with regard to more complex, non-rational factors that may determine our choices, such as addiction, indoctrination, and coercive control. These might be addressed in a Humean spirit by appeal to higher-order desires (as in Frankfurt 1971), but a fully adequate theory would probably have to move beyond the relatively crude "Humean" model of desire-driven behavior, and take more account of the messy realities of human psychology, in the study of which Hume himself was such a pioneer.[16,17]

Notes

1 For detailed discussion of Hume's theory of causation, see in particular Millican (2021) and (forthcoming). Both make reference to numerous earlier interpretive debates, which are mainly ignored here.

2 In contrast with the bogus notions of *substance*, *accident*, and *inherence*, for example, rejected at *T* 1.4.3.3-8 and 1.4.5.2-6.

3 Clarke appears to accept that human behavior is predictable in principle, since he appeals to this in explaining God's foreknowledge of our actions – for discussion, see Millican (2010: 619).

4 Quite apart from any doubts about Hume's Copy Principle and his determinism, one might reasonably be suspicious of his apparent implicit denial of a distinction between physical and moral causation, in so far as the former concerns events that are explicable only in terms of physical laws, whereas the latter concerns *actions* whose explanation involves reference to *motives*, *intentions*, *plans*, *strategies*, or other *forward-looking considerations*. "Moral causation" in this sense need

not, however, be confined to actions of conscious agents, as illustrated by AI systems (such as chess computers) and arguably by evolutionary phenomena.

5 For related remarks about the popular "causal closure of physics" doctrine, see Millican 2010: 636–637.

6 See also *T* 2.3.1.18, *EHU* 6.1 and 8.25, and for detailed discussion of Hume's determinism, Millican (2010).

7 Note also that Hume attacks supposed theoretical objections to the possibility of mental/physical causal interaction by appeal to his theory of causation, in an important pro-materialist argument at *T* 1.4.5.30-33.

8 Contemporary usage leaves this unclear, because some took "spontaneity" to mean simple *voluntariness* (e.g. Watts 1732: 5–6; Ramsay 1751: 107, 295), whereas others saw it as requiring more rational control (Bayle 1697, "Rorarius" note F, pp. 228–229; Voltaire 1764: 31).

9 This point is easy to overlook when the discussion is couched – as it often is – in terms of whether humans are "free agents" in an absolute sense, as opposed to being "free to do *X*".

10 It is also striking that Hume says very little about the conditions for just *punishment* except within his short and posthumously published essay "Of the Immortality of the Soul", as illuminatingly discussed by Russell (1995, ch. 10).

11 This is Hume's argument that his *doctrine of necessity* (i.e. determinism), so far from undermining morality as "the common opinion" presumes, is actually essential to it, ensuring the crucial link between motives and actions.

12 For more details and discussion, see Millican (2012).

13 This mechanism of sympathy apparently disappears from the second *Enquiry*, but is commonly assumed to remain implicit. My own view, however (see Millican 2020, §§5-7), is that it is crucial in the *Treatise* precisely because Hume is there developing his theory on a predominantly egoistic basis. Its absence from the *Enquiry*, accordingly, reflects his new awareness – evident in the second *Appendix* "Of Self-love" – that we are genuinely altruistic, and hence capable of caring about others without literally sharing in their feelings.

14 This, again, implies nothing about the causation of *my volition* or *my character*, so Hume is not forced into any regress about whether we have responsibility for either of these. Indeed, he insists that the virtues and vices for which we may be morally appraised include entirely involuntary *natural abilities* (see *T* 3.3.4 and *EPM Appx* 4). It is only in respect of *voluntary actions* that moral appraisal presupposes responsibility and hence hypothetical liberty.

15 With one very important exception noted at *E* 8.35: that these sentiments – and hence ascriptions of blame, praise, and moral responsibility – are not undermined by abstract reflections on such metaphysical issues as the deterministic nature of the universe.

16 For illuminating discussion of these sorts of complications, see Pitson (2016: 389–392).

17 I am very grateful to Don Garrett, Hsueh Qu, Paul Russell, and especially Max Kiener, for helpful comments on an earlier draft of this paper.

Further reading

Botterill (2002) gives a lively and illuminating analysis of Section 8 of the first *Enquiry*, distinguishing Hume's position sharply from that of "classic" compatibilists.

Pitson (2016) offers a sensitive and wide-ranging account of Hume's position, drawing useful links with more recent literature.

Russell (1995) presents his influential "naturalistic" interpretation of Hume's position, based in part on a distinctive understanding of Hume's view of causation. For a much briefer summary, see Russell's article "Hume on Free Will" in the *Stanford Encyclopedia of Philosophy*.

References

Ayer, A. J. (1954). "Freedom and Necessity," chapter 12 of *Philosophical Essays* (pp. 271–284). London: Macmillan.

Bayle, Pierre. (1697). *Historical and Critical Dictionary: Selections*, trans. Richard H. Popkin. Indianapolis: Bobbs-Merrill, 1965.

Botterill, George. (2002). "Hume on Liberty and Necessity," in Peter Millican (ed.), *Reading Hume on Human Understanding* (pp. 277–300). Oxford: Clarendon Press.
Bricke, John. (2008). "Hume on Liberty and Necessity," in Elizabeth S. Radcliffe (ed.), *The Blackwell Companion to Hume* (pp. 201–216). Oxford: Blackwell.
Clarke, Samuel. (1717). *Remarks Upon a Book, Entituled, A Philosophical Enquiry Concerning Human Liberty*. London: James Knapton.
Frankfurt, Harry G. (1971). "Freedom of the Will and the Concept of a Person," *The Journal of Philosophy*, 68, 5–20.
Hume, David. (1739–40, 2007). *A Treatise of Human Nature: A Critical Edition*, eds. David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.
———. (1748, 2007). *An Enquiry Concerning Human Understanding*, ed. Peter Millican. Oxford: Oxford University Press.
———. (1751, 1998). *An Enquiry Concerning the Principles of Morals*, ed. Tom L. Beauchamp. Oxford: Oxford University Press.
Locke, John. (1690/1975). *Essay Concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Clarendon Press.
Millican, Peter. (2010). "Hume's Determinism," *Canadian Journal of Philosophy*, 40, 611–142.
———. (2012). Chapter 5 "Hume," in Tom Angier (ed.), *Key Thinkers: Ethics* (pp. 105–131). London: Bloomsbury; revised in second edition 2023, pp. 89–111.
———. (2020). "The Relation Between Hume's Two *Enquiries*," in Jacqueline Taylor (ed.), *Reading Hume on the Principles of Morals* (pp. 273–299). Oxford: Oxford University Press.
———. (2021). "Hume on Causation and Causal Powers," in Benjamin Hill, Henrik Lagerlund, & Stathis Psillos (eds.), *Reconsidering Causal Powers: Historical and Conceptual Perspectives*. Oxford: Oxford University Press.
———. (forthcoming). "Hume as Regularity Theorist – After All!", *Hume Studies*, 49, (2024).
Penelhum, Terence. (2000). "Hume and the Freedom of the Will," in T. Penelhum (ed.), *Themes in Hume: The Self, the Will, Religion* (pp. 156–176). Oxford: Clarendon Press.
Pitson, Tony. (2016). "Hume, Free Will, and Moral Responsibility," in Paul Russell (ed.), *The Oxford Handbook of Hume* (pp. 380–400). Oxford: Oxford University Press.
Ramsay, Andrew Michael [Chevalier]. (1751). *The Philosophical Principles of Natural and Revealed Religion* (2nd ed.). London.
Russell, Paul. (1995). *Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility*. Oxford: Oxford University Press.
Voltaire [François-Marie Arouet]. (1764). *The Newtonian Philosophy Compared with That of Leibnitz, Translated from the French of M. de Voltaire*. Glasgow.
Watts, Issac. (1732). *An Essay on the Freedom of Will in God and in Creatures*. London (originally published anonymously, but included in volume 6 of Watts' *Works*, 1753).