

10 Nietzsche's Humean (All-Too-Humean!) Theory of Motivation

Neil Sinhababu

Friedrich Nietzsche sees desire as driving all of our action and practical reasoning. David Hume was an early proponent of this view, which is now called the Humean Theory of Motivation. Nietzsche and Hume never knew that they shared this theory, as Hume died a century too early to read Nietzsche, and Nietzsche never learned of Hume's practical philosophy. But their shared view led them both to appreciate the continuities between human and animal motivation and set them against a long tradition of rationalist rivals. Kant and Plato are the great historical representatives of this rationalist tradition, while contemporary Kantians like Christine Korsgaard are among the most distinguished today. As I'll argue, Nietzsche advances the Humean project by showing how desire drives the sort of practical reasoning that contemporary Kantians regard as a central case of reason's influence over action.

Section 1 outlines the Humean theory of motivation and argues that Nietzschean drives are composed of desires. Section 2 presents Nietzsche and Hume's psychological explanations in line with this theory. Section 3 explores the continuities that their theory reveals between human and animal psychology. Section 4 considers the place they see for themselves in the history of philosophy as opponents of a long rationalist tradition. Section 5 considers the phenomenon of reflective endorsement, which Korsgaard uses to defend the applicability of her neo-Kantian moral theory to humans. Section 6 describes how *Daybreak* 109 provides a Humean account of reflective endorsement that blocks the application of the neo-Kantian theory to humans.

Nietzschean drives and the Humean Theory of Motivation

This section presents the Humean view that I interpret Nietzsche as sharing, on which desire is necessary for human practical reasoning and action. Paul Katsafanas has argued that a Humean interpretation doesn't fit well with Nietzsche's conception of drives. As I'll argue, the nature of desire allows it to constitute drives as Nietzsche understood them.

Hume famously claims 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' (Hume [1739] 2000: 2.3.3). Translating this view into the language of contemporary philosophical psychology, we can understand the Humean Theory as consisting of two principles describing human action and reasoning. Here A is for 'action', E is for 'end', and M is for 'means:'

1st Proofs: Not for Distribution

162 *Neil Sinhababu*

Desire-Belief Theory of Action: One is motivated to A if and only if desire that E is combined with belief that one can raise E's probability by A-ing.

Desire-Belief Theory of Reasoning: Desire that M is created as the conclusion of reasoning if and only if the reasoning combines desire that E with belief that M would raise E's probability. It is eliminated as the conclusion of reasoning if and only if the reasoning eliminates such a combination.

(Sinhababu 2017)

On this view, desire is necessary for action, and for any reasoning that leads to action (by creating instrumental desires). This is a descriptive, non-normative view about the psychological states that explain human motivation and reasoning. Hume and Nietzsche express this view in slightly different terminology, with Hume mostly writing of passion and Nietzsche often using '*Trieb*' (drive) and '*Affekt*' (affect). But the sort of mental states they see as explaining action have a great deal in common, much of which is expressed by contemporary theories of desire.

Paul Katsafanas considers Nietzsche's response to the Kantian view that 'the will operates as a faculty independent of the affects, enjoys causal independence from the affects, and is uniquely capable of causing action' (Katsafanas 2016: 162). Katsafanas agrees with Humeans that this view is 'psychologically unrealistic', and argues that 'Nietzsche roundly rejects this triggering model of the will' (162). He suggests interpreting Nietzsche as accepting a broadly Humean picture on which all motivation forces including the will are vectors pushing in different directions. But Katsafanas hesitates to read Nietzsche as a full-fledged Humean, largely because he sees Nietzsche's psychology of drives as incompatible with a Humean psychology of desire. I'll argue that Nietzschean drives can be understood as composed of Humean desires.

Nietzsche takes drives to interpret and evaluate the world. Desires do this by causing pleasant and unpleasant feelings when we think of their objects and direct attention to things we associate with their satisfaction. Hungry people attend to food, are pleased to discover means to get food, and are displeased if these means fail. This is how the desire to eat makes us interpret food as important, seeing it as intrinsically valuable and methods for getting it as instrumentally valuable. In making us attend to particular things, desire makes us interpret them as important. In making us take pleasure in some things and not others, desire makes us evaluate them as good and bad.

Katsafanas argues that drives 'differ from desires in two crucial respects: first, drives admit an aim/object distinction; second, drives motivate us to express their aim, rather than to attain their object (which is merely adventitious)' (273). Cosima's drive to eat aims at eating and has a particular fig as its object, while her sex drive aims at sex and has Richard as its object. Here I propose regarding drives as collections of desires that channel activity towards an aim. The sex drive consists of desires for various kinds of social and bodily interaction, with the social interactions being ones that put one in position to pursue the bodily interactions constitutive of sex. Even if the drive to eat is constituted by a lone desire to eat food, it can cause a variety of actions including walking to restaurants and reading menus, when combined with belief that these actions raise the probability of eating food.

The role of means-end beliefs in the Humean Theory explains the specific and adventitious nature of drives' objects. Our initial desires can have objects as broad and general as Katsafanas takes drives' aims to be. But when they combine with means-end beliefs, more specific desires and motivation to do specific things result. Seeing a

1st Proofs: Not for Distribution

particular fig, Cosima may form a desire to eat it and be motivated to do so, as both parts of the Humean Theory suggest. This is how the generality of drives' aims is compatible with their being composed of desires, some of which have specific objects.

Katsafanas thinks drives and desires differ in that drives continue existing after satisfaction, while desires do not: 'In typical cases, a desire dissipates once its goal is attained. If I have a headache and desire to take an aspirin, the desire will be extinguished once I take the aspirin. Drives are different. The sex drive may be temporarily sated by a sexual encounter, but it will soon arise again' (111). Here he assumes a popular but false view of desire.¹ Intrinsic desires typically persist after satisfaction. They don't motivate action, because when desire is satisfied, nothing more needs to be done. I continue desiring the absence of pain after my headache is gone, but since there's no pain to get rid of, this desire stops motivating me. Intrinsic desires' continued hedonic and attentional effects after satisfaction demonstrate their persistence. Climbers who desired to reach the mountaintop may not be moved to climb again after succeeding, but they're likely to look back on reaching the mountaintop more often than they look back on events that don't figure in their desires, and with more pleasure. Some evolutionarily old intrinsic desires (like those in hunger, thirst and sexual lust) indeed vanish after satisfaction, because biological regulation systems have evolved to temporarily quell them after satisfaction affects glucose, hydration or hormone levels. Satiation of the sex drive, which Katsafanas describes, occurs when sexual desire gets turned off by the biological regulation system. Most desires aren't regulated this way. The athlete's desire to win isn't quelled by victory, and the parent's desire for the child's happiness isn't quelled by a nice birthday party. The instrumental desire to take aspirin goes away after I take it, but that's only because I lose a means-end belief. Taking even more aspirin isn't a means to anything I want. All in all, the dogma that desires generally vanish after satisfaction fails to account for the psychological data that desires explain. Most intrinsic desires persist, and instrumental desires vanish only because one ceases to believe that the means will promote the end.

Some of Katsafanas' other objections to the Humean Theory are broad enough that I can only point in the direction of an answer: 'once we attend to the diversity of mental states and processes, the Humean account does not seem rich enough. It assimilates all of this psychic complexity to two categories, belief and desire' (272). Here I'd reply that Humeans can invoke a wealth of other mental states, like sensation, imagination and attention. As outlined above, the Humean Theory is distinctive in treating belief-desire pairs as essential in causing motivation and practical reasoning. But this leaves room for all sorts of other mental states and processes to interact with our beliefs and desires in ways that don't cause action or generate new desires, giving rise to all sorts of complicated mental phenomena. This is how my forthcoming *Humean Nature* presents moral judgement, intention, willpower and the recognition of reasons, while understanding all motivation and reasoning in line with the Humean Theory.

Katsafanas also argues that 'it is *inadvisable* to pursue the Humean strategy. For suppose the Humean reduction of all mental states to beliefs and desires can succeed. Why should we care? What purpose does this reduction serve? Why think the reduction is illuminating or explanatorily fruitful?' (272). I thank him both for footnoting work of mine that suggests answers, and for inviting me to answer his question in this chapter (Sinhbabu 2009; 2013). The Humean Theory helps us understand how our minds work by showing us how their parts fit together. It reveals continuities between human psychology and that of nonhuman animals. And it reveals Kantian moral theories to

164 *Neil Sinhababu*

be inapplicable to humans for the same reason that they're inapplicable to animals – the structures of practical deliberation constituting Kantian moral agency are absent from beasts like us.

Humean and Nietzschean psychological explanations

This section considers the psychological explanations that Hume and Nietzsche build in line with the Humean Theory. Hume provides a Humean explanation of when strong feelings as we act, while Nietzsche explains how conscience operates.

Why explain psychological phenomena in line with the Humean Theory, rather than some other theory? One reason is that the Humean Theory is very simple, explaining motivation and reasoning by using a limited list of psychological states and processes. It explains all action and practical reasoning in terms of desires and means-end beliefs. Simpler theories that fit any considerable amount of our phenomenological and behavioural data are hard to develop. If a psychological picture that includes Humean Theory fits all of our data, its simplicity will likely make it better than any other overall psychological picture. Occam's Razor is a tool of scientific theory choice, and it cuts against theories that posit additional psychological states for no further explanatory gain.

Hume presents an empirical argument that motivational states that don't come with strong feelings still constitute the operations of desire rather than a desire-independent faculty of reason. Bemoaning how 'every action of the mind, which operates with the same calmness and tranquility, is confounded with reason by all those, who judge of things from the first view and appearance' (2.3.3) he distinguishes between calm and violent passions. He notes that passions generate more emotion when their objects are vividly represented by our senses or imagination, and less emotion when their objects aren't so vividly represented. By changing 'the situation of the object', we can 'change the calm and violent passions into each other'. Perceiving something with more 'force and vivacity' increases the violence of passions for it (2.3.7). Here he takes a psychological datum that seems hard for his theory to explain – motivation without strong feeling – and shows how it follows neatly from a proper understanding of the phenomenology of desire. If treating all motivation as caused by desire explains all the phenomenological and behavioural data, we have no evidence for additionally positing motivation caused by something else.

Nietzsche presents similarly insightful explanations of how desires produce psychological phenomena that are commonly regarded as the operations of reason. His historical account of the origin of the 'bad conscience' in GM 16 is an example. Many philosophers see the operations of conscience as distinct from the operation of desire. Nietzsche's historical account treats our capacity to critically evaluate our desires as a manifestation of desire itself. On his story, when primitive humans were 'enclosed within the walls of society and of peace', the aggressive impulses they weren't permitted to express against others were turned against themselves. In this process, 'Hostility, cruelty, joy in persecuting, in attacking, in change, in destruction – all this turned against the possessors of such instincts: that is the origin of the 'bad conscience''. He emphasizes that this internal redirection of aggressive desire to attack made us the psychologically complex beings we are today: 'thus it was that man first developed what was later called his 'soul'. Of course, this 'soul' isn't anything that arose through divine grace or even the activity of reason independent of desire. It's simply a part of one whose violent, furious desire is directed internally against oneself.

1st Proofs: Not for Distribution

Tom Bailey sees Nietzsche as presenting an anti-Humean account of the 'sovereign individual' discussed at the beginning of the Second Essay, on which mental states other than desires can motivate action:

Nietzsche nonetheless shares with Kant a basic conception of agency as action that is not determined by immediate experiences and desires. For, just as Kant defines agency as motivated action that is 'free' in the sense that it 'can [. . .] be affected but not determined by impulses', Nietzsche presents the agency of the sovereign individual as consisting of an ability to 'forget' and thus 'digest' experiences and desires which allows for 'a little tabula rasa of consciousness.'

(Bailey 2013)

Bailey and I agree that Nietzsche doesn't accept anything like Kant's metaphysically extravagant account of the will. But to argue that Nietzsche 'considers the 'sovereign individual' to be 'free' simply to will actions without being determined to them by immediate experiences and desires' is to lose what Nietzsche finds most interesting about this individual.

What Nietzsche wants to tell us about the sovereign individual's 'independent, protracted will and the *right to make promises*' is that it's composed of immediate desires (GM II: 2). This is how Nietzsche explicitly describes the sovereign individual: as having 'an active *desire* not to rid oneself, a desire for the continuance of something desired once' (GM II: 1). If forced to contemplate breaking promises, we might imagine the sovereign individual reacting with intense displeasure, as is typical of people forced to vividly imagine outcomes that they strongly desire not to happen.

Nietzsche's history of how the sovereign individual was created in GM II: 3 invokes procedures suitable for brutally conditioning animals into fearing new things. Writing that 'If something is to stay in the memory it must be burned in: only that which never ceases to *hurt* stays in the memory', he describes horrific punishments for criminals, including boiling, flaying and being torn apart by horses. He writes that 'With the aid of such images and procedures one finally remembers five or six 'I will not's', in regard to which one had given one's *promise* so as to participate in the advantages of society – and it was indeed with the aid of this kind of memory that one at last came "to reason!"' These gruesome procedures aren't ways to create a new motivational faculty constitutive of reason – hence Nietzsche's scare quotes. But they are ways to condition creatures into having new desires. Nietzsche explains the existence of conscience not by positing a new motivational state other than desire, but by invoking familiar processes we already have in our psychological theories. Simplicity is preserved.

GM II: 1 begins: 'To breed an animal *with the right to make promises* – is not this the paradoxical task that nature has set itself in the case of man?' The story Nietzsche tells throughout the essay is about how to forcefully instill new desires in an animal through images of pain so that it'll behave as its masters require. To understand the sovereign individual as keeping promises because he somehow acquired a motivational state irreducible to desire is to lose the animal, the paradox, and Nietzsche's brutal story of how the task was achieved.

Continuities between human and animal motivation

As I've already begun to describe, Nietzsche and Hume are happy to emphasize that the motivational theory that they ascribe to humans is simple enough for animals to share.

166 *Neil Sinhababu*

Throughout the *Treatise*, Hume stresses the continuities between human and animal motivation. Books I and II of the *Treatise* extend his view of human psychology into the animal kingdom. After presenting his account of human probabilistic reasoning in Book I, he announces that ‘no truth appears to me more evident, than that beasts are endowed with thought and reason as well as men’ (1.3.12). The rest of the section, titled ‘Of the Reason of Animals’, describes animals as reasoning in the same way humans do. Book II contains two sections, ‘Of the Pride and Humility of Animals’ and ‘Of the Love and Hatred of Animals’, which follow Hume’s descriptions of how these mental states operate in humans and argue that they operate similarly in animals (2.1.12, 2.2.12). A century before Darwin, Hume suggested a unified picture of human and animal psychology:

But to pass from the passions of love and hatred, and from their mixtures and compositions, as they appear in man, to the same affections, as they display themselves in brutes; we may observe, not only that love and hatred are common to the whole sensitive creation, but likewise that their causes, as above-explained, are of so simple a nature, that they may easily be supposed to operate on mere animals. There is no force of reflection or penetration required. Every thing is conducted by springs and principles, which are not peculiar to man, or any one species of animals.

(2.2.12)

Similarly, Nietzsche explicitly accounts for human moral judgement in terms of a basically animal motivational psychology. Consider D 26, titled ‘Animals and morality:’

The beginnings of justice, as of prudence, moderation, bravery – in short, of all we designate as the *Socratic virtues*, are *animal*: a consequence of that drive which teaches us to seek food and elude enemies. Now if we consider that even the highest human being has only become more elevated and subtle in the nature of his food and in his conception of what is inimical to him, it is not improper to describe the entire phenomenon of morality as animal.

This animalistic view of human moral psychology continues through the Second Essay, where human psychology is described as a type of animal psychology well over a dozen times. As previously mentioned, GM II begins by describing the sovereign individual as ‘an animal *with the right to make promises*’. This is an ‘animal which needs to be forgetful’, but which has ‘bred in itself an opposing faculty, a memory’. GM II: 16 treats the formation of the bad conscience as a dramatic step in animal development, describing the result as ‘an animal soul turned against itself’. Discussing guilt before God, GM II: 23 exclaims: ‘Oh this insane, pathetic beast – man! What ideas he has, what unnaturalness, what paroxysms of nonsense, what *bestiality of thought* erupts as soon as he is prevented just a little from being a *beast in deed!*’

While most of the other animals don’t feel the bite of conscience, as they don’t have souls turned against themselves, conscience motivates action and causes displeasure about past action just like desire usually does. Just as desires to eat or have sex move humans and animals to pursue food or mates, and cause displeasure upon seeing that they’re unattainable, conscience moves humans to do particular actions that one regards as right, and causes displeasure if one has failed to do so. Perhaps pigs and squirrels don’t feel the bite of conscience because they lack concepts of the self, of agency or of

1st Proofs: Not for Distribution

morality, which are necessary for recognition that one has acted wrongly. We can still share their desire-belief motivational structure, while differing in being able to desire and believe additional things. A creature capable of grasping these and other concepts can point the same animal passions in new directions – towards doing the right thing, or towards not displeasing God.

The Humean Theory is built to encompass both human and animal motivation with maximal simplicity. The variable E (for ends, the objects of initial desires) can represent an object of animal desire, just as it can represent an object of human desire. A (for action) and M (for means) can represent whatever actions and means the animal can understand. In some individual cases, Humean explanations may fill in these variables to posit more attitude-contents than Kantian explanations do. But this doesn't make psychology as a whole any more complex, because plausible psychological theories will already allow that humans can desire and believe a vast range of things. Since commitment to a vast range of desire-contents and belief-contents is unavoidable, pursuing simplicity at that level is folly. Simplicity is available, however, at the level of attitude-types and systematic ways that attitude-types interact. The Humean Theory exhibits this kind of simplicity in encompassing all of mammalian and possibly avian motivational psychology. Human beings appear as a complicated case that the theory accounts for with the same simple set of attitude-types and interactions, befitting our nature as complicated animals.

Hume and Nietzsche in the history of philosophy

This section considers how Hume and Nietzsche saw the history of moral psychology, and explores the complicated historical relationship between them.

Both Hume and Nietzsche recognize that their views of motivation are unpopular, as the philosophical tradition has glorified the role of reason in driving motivation and reasoning. Consider how Hume introduces his theory of motivation in the *Treatise*:

Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason, to give the preference to reason, and assert that men are only so far virtuous as they conform themselves to its dictates. Every rational creature, it is said, is obliged to regulate his actions by reason; and if any other motive or principle challenge the direction of his conduct, he ought to oppose it, till it be entirely subdued, or at least brought to a conformity with that superior principle. On this method of thinking the greatest part of moral philosophy, ancient and modern, seems to be founded; nor is there an ampler field, as well for metaphysical arguments, as popular declamations, than this supposed pre-eminence of reason above passion. The eternity, invariableness, and divine origin of the former have been displayed to the best advantage: The blindness, unconstancy, and deceitfulness of the latter have been as strongly insisted on. In order to shew the fallacy of all this philosophy, I shall endeavour to prove first, that reason alone can never be a motive to any action of the will; and secondly, that it can never oppose passion in the direction of the will.

(2.3.3)

Hume describes the rationalist consensus he sees in both ancient and modern philosophy of action, and sets himself directly against it.

168 Neil Sinhababu

Nietzsche also criticizes philosophers' high opinion of reason at length. In *Twilight of the Idols*, he disparages the psychological health of ancient philosophers who esteemed reason highly: 'The moralism of the Greek philosophers from Plato on is pathologically conditioned; so is their esteem of dialectics. Reason-virtue-happiness, that means merely that one must imitate Socrates and counter the dark appetites with a permanent daylight – the daylight of reason' (TI 'The Problem of Socrates' 10). He alludes to the evolutionary kinship between humans and animals as he follows Hume in noting the divine origin his opponents ascribe to reason: 'And in India, as in Greece, the same mistake was made: 'We must once have been at home in a higher world (instead of a very much lower one, which would have been the truth); we must have been divine, for we have reason!'' (TI 'Reason' in Philosophy: 5). The modern philosopher whom Nietzsche criticizes most ferociously along these lines is Kant: 'An action demanded by the instinct of life is proved to be *right* by the pleasure that accompanies it; yet this nihilist with his Christian dogmatic entrails considered pleasure an *objection*' (A 11). In fighting back against Kantian moral psychology, Nietzsche defends much of the same conceptual territory that Hume had over a century before.

The interesting similarities between Hume and Nietzsche go far beyond motivation. Craig Beam discusses their criticisms of Christian monkish virtues and ascetic ideals (Beam 1996). Peter Kail notes their interest in naturalistic explanations and deflationary views of causation (Kail 2009). Katsafanas' own defence of Nietzschean constitutivism ties the position more closely to Hume than to Kant. Nietzsche seems to have arrived at these Humean views without any awareness that Hume shared them. He knew of Hume mainly as the sceptic about causation and synthetic *a priori* judgements who roused Kant from his dogmatic slumber. Beam notes that 'his knowledge of Hume was rather sketchy and did not go much beyond the conventional image of Hume as an epistemologist and empiricist' (301).

I find nine explicit references to Hume in Nietzsche's writing. The most favourable is a note from 1885–1886 in which he remarks, 'We have no 'sense for the *causa efficiens*': here Hume was right' (WP 550). In the *Untimely Meditations*, he attributes to Hume a quip about people who expect their old age to be better than their youth, though Hume got it from Dryden. The other references treat Hume mostly as someone whose theoretical philosophy Kant responded to (GS 357; WP 92; WP 101; WP 530) or as less pessimistic than Schopenhauer (GS 370, NCW 5), often making a sociological point about Germans. The first paragraph of BGE 252 is the harshest:

They are no philosophical race, these Englishmen: Bacon signifies an attack on the philosophical spirit; Hobbes, Hume, and Locke a debasement and lowering of the value of the concept of 'Philosophy' for more than a century. It was against Hume that Kant arose, and rose; it was Locke of whom Schelling said, *understandably*, 'je meprise Locke', in their fight against the English-mechanistic doltification of the world . . .

Nietzsche's oft-expressed distaste for English philosophy here exceeds his similarly oft-expressed distaste for his fellow Germans. And while he would soon call Kant a 'catastrophic spider' and write, 'The instinct which errs without fail, *anti-nature* as instinct, German decadence as philosophy – *that is Kant!*' (A 12), he sides here with his hated foe against Hume. One wonders how his prejudices would've been affected if he had known that Hume was Scottish, not English. Nietzsche seems to have been as unaware

1st Proofs: Not for Distribution

of Hume's practical philosophy as of his ethnic background, so looking into his explicit references to Hume isn't a very helpful way of figuring out whether he shares the Humean theory of motivation. A better way is to explore what he and Hume wrote about motivation, and that's what I've done.

One reason why Nietzsche and Hume found themselves on the same sides of so many debates is that a mutual antagonist stood between them. This antagonist is Kant, famously roused from his dogmatic slumber by Hume and criticized fiercely by Nietzsche. While it's unclear how much of Hume's practical philosophy Kant read, Henry Allison and Manfred Kuehn suggest that he was aware of Hume's scepticism about practical reason (Kuhn 2001). The battle against Kantian moral psychology may have led Nietzsche to take Humean positions, not knowing that Hume had occupied those positions more than a century before.

Korsgaard and reflective endorsement

Now I'll discuss Korsgaard's view that rational endorsement is essential to human action, and that the ability to gain distance from our desires in reflection explains human moral agency. The phenomenon of reflective endorsement itself raises a problem for Humeans: how can they explain how we evaluate our desires and decide whether or not to act on them?

In *The Sources of Normativity*, Korsgaard claims that action requires reflectively endorsing the motives on which we act (Korsgaard 1996). She ascribes this view to both herself and Kant: 'being human we must endorse our impulses before we act on them' (122). Here's how she characterizes this kind of reflective endorsement:

I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I act? Is this desire really a reason to act? The reflective mind cannot settle for perception and desire, not just as such. It needs a reason. Otherwise, at least as long as it reflects, it cannot commit itself or go forward.

(93)

While Nietzsche and Hume offer a unified account of human and animal motivation, Korsgaard sees the capacity for rational endorsement as an essential difference in the 'structure of our minds' (92). While 'the human mind is self-conscious in the sense that it is essentially reflective' (92), 'a lower animal's attention is fixed on the world. Its perceptions are its beliefs and its desires are its will. It is engaged in conscious activities, but it is not conscious of them' (92–93).

Korsgaard maintains this view of action in her recent work, approvingly discussing Plato and moving to Kant:

The soul that drinks in response to thirst does so not merely because it has an appetite to drink, but because it 'nods assent to [the appetite] as if in answer to a question'. Having an appetite for something and giving that appetite the nod are not the same thing. The soul does not act from appetite, but from something that endorses the appetite and says yes to it. Even when conflict is absent, then, we can see that there are two parts of the soul. To put it Kant's way: in the human

170 *Neil Sinhababu*

soul, the experience of choosing to act on an incentive – the experience of adopting a principle – is distinct from the experience of the incentive itself.

(Korsgaard 2009: 139)

The internal quotation is from Socrates. Here again Korsgaard distinguishes having a desire and endorsing it as two separate events that precede action.² In agreeing with Kant about the experience of incentives and acting on them, she also claims that these events are distinguished in the phenomenology of deliberation. She takes the mental states responsible for these events as two separate parts of the human soul – inclination (or appetite or desire) and reason. As she writes, ‘inclination presents the proposal; reason decides whether to act on it or not, and the decision takes the form of a *legislative act*’ (154). There certainly are cases in which action feels this way.

In *Sources*, Korsgaard explains the theoretical significance of reflective endorsement for her account of agency. She sees it as necessary for giving behaviour the status of action, because it allows agents to be involved in actions while desire doesn’t allow this:

if I am to constitute myself as the cause of an action, then I must be able to distinguish between my causing the action and some desire or impulse that is in me causing my body to act. I must be able to see myself as something that is distinct from any of my particular, first-order, impulses and motives, as the reflective standpoint in any case requires. Minimally, then, I am not the mere location of a causally effective desire but rather am the agent who acts on the desire.

(227–8)

When Thales went stargazing and accidentally fell into the pit, his falling wasn’t an action because he didn’t rationally choose to do it – it just happened. The force of gravity didn’t come from his self, but from the outside, so falling wasn’t his action. And according to Korsgaard, if desires and means-end beliefs were able to suddenly take control of me and cause my limbs to move about in a certain way without any input from reason, that wouldn’t be my action either. That wouldn’t be the process that constitutes action, in which desires merely propose actions and reason independently decides to act on them. Korsgaard quotes Kant himself in support of this view: ‘we cannot consciously conceive of a reason which consciously responds to a bidding from the outside with respect to its judgments’ (101). She sees reason and agency entering the process of motivation through reflective endorsement, so that behaviour that didn’t involve reflective endorsement wouldn’t be decided by reason and thus wouldn’t involve action. I am distinct from my desires (Korsgaard uses ‘desire’ and ‘impulse’ as falling into the same class from the point of view of her theory.) For an action to be mine, it has to involve me.

In ‘Self-Constitution in the Ethics of Plato and Kant’, Korsgaard develops this objection to the Humean Theory, comparing two different models of how reason and desire interact to produce action (Korsgaard 1999). The first is the Combat Model, according to which various motivational forces struggle against each other to produce action, and the strongest ones win. She sees Hume as basically subscribing to this model of action, with the caveat that the only forces entering into combat are passions – ‘He simply argues that reason is not a force’ (100). She says that ‘there are a few questions Hume should have asked first, for the Combat Model makes very little sense’ (100). Her

central objection seems to be that 'The Combat Model gives us no clear picture of the person who chooses between reason and passion' (101).

Humeans can present a clear picture of persons as constituted by their underlying mental states, including desires. If what I'm made of includes my desires, what they cause is what I cause, and the actions they cause are mine. (If the statue is made of clay, the downward pressure on the scale caused by the clay is caused by the statue.) This follows the principle of Humean Self Constitution which I defend in Chapter 10 of *Humean Nature*: 'Agents are constituted in part by all of their desires, and aren't constituted by any other motivational states' (167). They can also respond more aggressively by considering cases that her theory doesn't handle very well. Nomy Arpaly suggests some of these – fast actions that proceed impulsively and without reflection, and cases of rational akrasia where we make the rational choice against our reflective judgement instead of being guided by it (Arpaly 2003). But I'll focus on reflective endorsement and its role in constituting the self here.

Korsgaard might argue that desires aren't the sorts of things that can play such a role in constituting a person, so Humean Self-Constitution must be wrong. Let's return to cases that fit Korsgaard's picture of reflective endorsement, in which we consider desires and act on them only after reason endorses them. She sees such cases as showing that desire (even combined with means-end belief) is insufficient for action. We only act when we endorse the desire. Desire here seems like something separate from us, which we can reflect on just as we can reflect on objects outside ourselves in deliberation. Then it looks like we shouldn't be identified with desire. So a Humean reduction of the self to desire has to explain some difficult phenomenological data. Even if spontaneous action favours Nietzsche and Hume's view, reflective action seems to favour Plato and Kant's view.

Whether Kantian ethical theories apply to humans depends on the nature of our motivational structure. Essential to Kantian ethics is the idea that reason can set its own ends, and that it isn't merely a slave of the passions. Kantians hold that we're moral agents while the animals aren't because our motivational structure isn't just a bunch of desires pushing against each other. Kant himself understood ethics as a system of synthetic a priori judgements, which must be derived independently of desire if we are to be autonomous and rational moral agents.

As a contemporary Kantian constructivist, Korsgaard sees morality as arising from reflective endorsement itself. She writes in *Sources* that the need for reflective endorsement of impulses 'sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question' (93). On Korsgaard's view, it's from our capacity to ask about whether we're justified, from a reflective standpoint, that normativity comes into the world. As she puts it later, 'Reflection gives us a kind of distance from our impulses which both forces us, and enables us, to make laws for ourselves, and it makes those laws normative' (129). If reflection turns out to merely be a way that desires push against each other, it wouldn't give practical agents any distance from their desires. So we wouldn't be able to make laws for ourselves, and normativity wouldn't apply to our actions – at least on the Kantian conception of law and normativity.

If Humeans can show that reflection just consists in desires pushing on each other, there will be no room for Kantian conceptions of law and normativity to get a grip on us. This will require an account of what's going on in the cases of reflective

172 Neil Sinhababu

endorsement that Korsgaard discusses. How can Humeans explain how we can think and feel as we reflectively endorse or reject our desires? Nietzsche provides the answer.

The Nietzschean account of reflective endorsement in *Daybreak* 109

This section develops a Humean account of reflective endorsement from D 109. First I'll lay out what Nietzsche takes to be happening in cases where we reflect on a desire, either endorsing or rejecting it. Then I'll explain why this account is simpler than Korsgaard's, and how it maintains the combat model of desire, preventing us from getting the distance from our desires that Kantian moral agency requires.

Nietzsche begins the section, titled 'Self-mastery and moderation and their ultimate motive' by announcing that he finds 'no more than six essentially different methods of combating the vehemence of a drive'. True to Humean views of motivation, these methods don't involve reason stamping out the vehement drive simply by the force of willpower or a normative judgement against it. First, one can avoid opportunities to gratify it, perhaps allowing it to wither away in the absence of positive reinforcement that would come from gratification. Second, one can impose upon oneself a schedule as to when one gratifies it, so that the drive naturally arises only at those times and one is free from it at other times. Third, one can wildly over satisfy it in the hopes that disgust at over satisfaction will give rise to a motivation against it. Fourth, one can try to build negative associations (as Christians do with the idea of the Devil or of Hell) with the object of the drive. Fifth, one can focus all one's mental energies on something else and pursue that with such intensity that one has no energy left to serve the disfavoured drive. Sixth and finally, one can do things that weaken one's entire constitution so that all of one's drives will be dragged down.

There is no seventh method like 'Judging the vehement drive to be bad, reason eliminates its motivational force or generates a new motivation to counteract it'. Creating or eliminating mental states simply by rational inference doesn't require feelings of disgust or thoughts of the Devil. Nietzsche suggests these roundabout methods because he agrees with Hume that reason is merely the slave of the passions. Since reason lacks the power to create or eliminate passions by itself, all it can do is understand the nature of drives and try to set up situations that weaken them. That's how Nietzsche's six methods work.

After listing these methods, Nietzsche tells us what happens when we reflect on such vehement drives:

that one desires to combat the vehemence of a drive at all, however, does not stand within our own power; nor does the choice of any particular method; nor does the success or failure of this method. What is clearly the case is that in this entire procedure our intellect is only the blind instrument of another drive which is a rival of the drive whose vehemence is tormenting us: whether it be the drive to restfulness, or the fear of disgrace and other evil consequences, or love. While 'we' believe we are complaining about the vehemence of a drive, at bottom it is one drive which is complaining about another; that is to say: for us to become aware that we are suffering from the vehemence of a drive presupposes the existence of another equally vehement or even more vehement drive, and that a struggle is in prospect in which our intellect is going to have to take sides.

This is how Nietzsche understands the phenomenology of reflecting negatively on one's desires. One's attention in such cases is directed by another desire of equal or greater strength, which causes negative feelings towards the desire reflected on.

Nietzsche's account is grounded in desire's typical psychological effects. Desire directs my attention at things I associate with its object, at ways to attain the object, and at obstacles to attaining it. Hume tells us how desire, 'making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and effect'. (2.3.3). It's pleasant to discover ways to attain the object of our desire, and unpleasant to discover obstacles to attaining it. Hungry people attend to food while lustful people attend to those they sexually desire. Both also attend happily to opportunities and unhappily to obstacles to attaining what they desire.

People don't typically attend to the desire that is directing their attention. As Philip Pettit and Michael Smith point out, this desire stands in the background of deliberation, not the foreground (Pettit and Smith 1990). The hungry think more about food and how to get it than about hunger, while the lustful think more about sexual interaction and how to achieve it than about lust. If there ever were creatures whose lust directed their attention on lust itself rather than on how to enter into the desired sexual interactions, there is a good reason why they didn't become our ancestors.

This is how a desire to attain peace or avoid disgrace will make me cast my view on peace and the avoidance of disgrace, and also on whatever unruly desire stands in the way of achieving it. It won't direct attention to itself. So it may lead us to think that the only desire we have is the unruly one we're unhappily reflecting on. The first-personal phenomenology of reflecting unhappily on a vehement and unruly desire is elegantly explained by taking a rival desire to direct unhappy attention onto it. Korsgaard writes that 'From a third-person point of view, outside of the deliberative standpoint, it may look as if what happens when someone makes a choice is that the strongest of his conflicting desires wins. But that isn't the way it is *for you* when you deliberate' (100). But in fact, that's exactly how it is for you when you have two conflicting desires, and you reflect on one from the point of view of the other. Korsgaard is right about what the phenomenology of reflection is like. But she doesn't recognize that the phenomenology of considering one desire from the viewpoint of another is precisely that way.

With the Nietzschean explanation of reflective endorsement, a broadly Humean account of motivation can explain why some actions are spontaneous and others are reflective. In spontaneous cases, I'm typically affected by only one desire, and it gets me to act straightaway (I may act thoughtlessly and then regret it, if I only realize afterward that the action prevents the satisfaction of another of my desires). In reflective cases, the course of action that one desire makes me attend to is brought into critical focus by another desire before I can act. If one desire is much stronger than the other, its attention-directing and motivational effects may cause it to dominate my reflection and motivate action accordingly. But if the desires are of roughly equal strength, I may be indecisive about what to do, or torn between them. In situations involving even more desires and other psychological states that interact with them, increasingly complex psychological phenomena are possible.

Nietzsche's account of reflective endorsement is simpler than Korsgaard's in terms of psychological state types. Explaining reflective endorsement in terms of two desires having their typical effects leaves no need for a separate, motivationally effective part of the soul that is practical reason acting independently of desire. The motivational

174 Neil Sinhababu

effects of desire explain our behaviour, and the attention-directing and hedonic-emotion-causing effects of desire explain how we focus on things and how we feel about them. The Nietzschean account explains all the behavioural and phenomenological data of reflective endorsement using only the Humean Theory's simple ontology of psychological state types. Kantians can't plausibly argue that we don't have desires, or that our desires lack the attention-directing and hedonic features essential to the Nietzschean account of reflective endorsement. This leaves Humeans with a simpler theory that fits the data. Einstein (1934) writes that 'It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience' (165). In psychology, the Nietzschean account of reflective endorsement moves Humeans closer to this supreme goal.

If Nietzsche is right, what becomes of Korsgaard's thesis that moral agency arises from a kind of reflection that involves more than desire? If a system of desires pushing against other desires leaves no room for normativity, the answer is that the Nietzschean account reveals humans to be so animalistic that morality doesn't apply to us. Hume didn't think that the psychological similarities between humans and animals had this consequence, and neither do I. But Kant, Plato, and Korsgaard set the bar for moral agency much higher than Hume did. If human reflective endorsement is just another instance of desires combating other desires, it won't help humans clear this high bar. The distinctive motivational structures required for Kantian moral agency don't appear in the human mind. To avoid the bizarre conclusion that human beings can't be moral agents, we'll have to endorse a simpler account of moral agency, like the one that Hume offers. This blocks the path from Kantian conceptions of moral agency to a Kantian normative ethics that applies to human beings.

Perhaps one can imagine creatures whose motivational and deliberative systems are as Korsgaard imagines them. And perhaps a Kantian moral theory would apply to them. But the Nietzschean account of reflective endorsement reveals that you and I are not such creatures. Our desire-driven moral psychology is Humean, all too Humean.

Notes

- 1 This erroneous view is shared by influential defenders of directions-of-fit analyses like Smith (1994).
- 2 Risse (2007) offers useful Nietzschean criticisms of this Kantian view.

References

- Arpaly, N. (2003) *Unprincipled Virtue*, Oxford: Oxford University Press.
- Bailey, T. (2013) 'Nietzsche the Kantian?' in J. Richardson and K. Gemes (eds.), *Oxford Handbook of Nietzsche*, Oxford: Oxford University Press.
- Beam, C. (1996) 'Hume and Nietzsche', *Hume Studies*, 22.2: 299–324.
- Einstein, A (1934) 'On the Method of Theoretical Physics', *Philosophy of Science*, 1.2: 163–9.
- Hume, D. ([1739] 2000) *A Treatise of Human Nature*. Oxford: Oxford University Press.
- Kail, P. J. E. (2009) 'Nietzsche and Hume: Naturalism and Explanation', *Journal of Nietzsche Studies* 37: 5–22.
- Katsafanas, P. (2008) *Practical Reason and the Structure of Reflective Agency* (dissertation).
- (2013) 'Nietzsche's Philosophical Psychology' in J. Richardson and K. Gemes (eds.), *Oxford Handbook of Nietzsche*, Oxford: Oxford University Press.
- (2016) *The Nietzschean Self*. Oxford: Oxford University Press.

1st Proofs: Not for Distribution

Nietzsche's Humean theory of motivation 175

- Korsgaard, C. M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.
- (1999) 'Self-Constitution in the Ethics of Plato and Kant', *Journal of Ethics*, 3.1: 1–29.
- (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Kuehn, M. (2001) *Kant: A Biography*, Cambridge: Cambridge University Press.
- Pettit, P. and Smith, M. (1990) 'Backgrounding Desire', *Philosophical Review*, 99.4, 565–92.
- Risse, M. (2007) 'Nietzschean 'Animal Psychology' versus Kantian Ethics' in B. Leiter and N. Sinhababu, *Nietzsche and Morality*, Oxford: Oxford University Press.
- Sinhababu, N. (2009) 'The Humean Theory of Motivation Reformulated and Defended', *Philosophical Review*, 118.4: 465–500.
- (2013) 'The Desire-Belief Account of Intention Explains Everything', *Noûs*, 47.4: 680–96.
- (2017) *Humean Nature*. Oxford: Oxford University Press.
- Smith, M. (1994) *The Moral Problem*. Oxford: Oxford University Press.