

Neil Sinhababu: *Humean Nature*

Budapest, 03/07/2017

(References are to section numbers. Since I worked from the uncorrected proof of the manuscript, I didn't want to use page numbers.)

I. Framework

The meta-ethical setup is akin to Smith's moral problem (1.2; 11.5):

1. Cognitivism: Moral judgments are beliefs
2. Internalism: Moral judgments can produce their own motivational force
3. The Humean Theory of Motivation (HTM)

ISSUES:

(i) *Desire-Belief Theory of Reasoning as part of HTM*

HTM has two parts (1.1): DBTA (Desire-Belief Theory of Action) and DBTR (Desire-Belief Theory of Reasoning). Question is: why is DBTR included? (cf. Smith's treachery). I discuss DBTA later, now I would like to ask questions about the role of DBTR.

DBTR (1.1): "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."

- (a) "Someone without the relevant desires, then, can't be motivated to act by any chain of reasoning that begins only in beliefs – "reason alone can never be a motive to any action of the will." But is this obviously something the Humean needs to deny? Strictly speaking, you are not motivated by a chain of reasoning but, if DBTA is true, desires. Beliefs, if DBTR is wrong, might produce some of your desires, but it will still be desires that motivate you. Cf. Mele's Antecedent Motivation Thesis (AMT) and Nagel's motivated desires.
- (b) What other reasons are given for including DBTR in HTM? 1.2: historically genuine (Hume quote above, e.g.) and meta-ethically relevant. How much do we have to care about the first (this is an analytical piece, after all) and does the contrast between rationalism and Humean theories really disappear if DBR is taken out? Sure, reason is then not entirely powerless in facing the passions, but still, all it can do is to create *mental states*, it cannot create *action*. And it is the latter, I think, that really matters for a psychological theory like HTM.
- (c) The discussion in 2.5: It appears to hold for certain desires (sex, appetites), but not necessarily for others. However, the scope of DBTR, as presented, is all-encompassing. (Again, cf. Nagel's motivated desires or think of R Brandt's cognitive psychotherapy approach.) Hence it is not enough to show that *some* desires come under DBTR but that *all* do: has this been shown, however? I don't think so.

(ii) The methodological principle

“Of the theories best fitting the data, we should choose the one with the simplest ontology of fundamental explanatory entities.” (1.3) But it is not that simple, is it? We have: 1. Simplicity of ontology; 2. Simplicity of explanation (this sort of distinction NS also has: he talks about the simplicity *and* explanatory power of the Humean theory). Is this not a balancing act then? Take all the complex explanations provided later (e.g. 3.1 – feeling of obligation; 3.2 – the Darwall case; 8.3 – Stripada; Chapter 9 – the different silencing phenomena). It would be natural, to me, to hold that there is a point at which an ontologically simpler theory becomes problematic due to its overly complex, convoluted (etc etc) explanation of psychological phenomena.

NS says (1.3): “We seek the simplest total psychological theory, not the simplest explanation of any individual phenomenon. Often an individually simple account of one phenomenon should be rejected because it adds new fundamental entities to our total psychological theory. Otherwise no phenomenon could be explained by multiple factors – it’s always simpler individually to invoke a single new fundamental force that provides a full explanation, cluttering our overall theory with a fundamental force for each phenomenon.”

This suggests that 1 trumps 2, but to me this sounds implausible. Anyone?

(iii) The understanding of cognitivism

This connects to my queries re DBTA – to put it simply: is the only relevant cognitive state belief? The contrast looks overly simplifying. It is strange that more complex or ‘unorthodox’ cognitive states are rarely considered in the book (e.g. 1.2 – desires; 2.5 – desire-as-belief), while such proposals abound in the literature. More to follow below.

(iv) The larger meta-ethical picture

As noted, NS reproduces a version of Smith’s moral problem: one cannot accept 1-3 theses above because the result is what NS dubs ‘incapabilism’: humans are unable to make moral judgments (because: judgment would have to be a belief that by itself motivates flying straight in the face of HTM, which we take to be granted given the arguments in many chapters of the book).

As NS admits (11.1), there is a tension here since 1 and 2 are typically read as conceptual (or metaphysical) theses, while he construes HTM (unlike e.g., Smith?) as an empirical theory. So, why couldn’t one say that empirics do not refute conceptual analyses and therefore HTM is wrong? NS argues that this is not so and I agree. However:

- (a) There is still another way out: why not embrace incapabilism? NS thinks this is absurd; it’s a price not to be paid (1.2). Is this obviously so? Cf. Smith’s analysis of witches and then of moral reasons: we have a conceptual claim and then check if the world conforms – are there witches as defined? Are there moral reasons as defined? Perhaps there aren’t. Analogously, we could say: this is what moral judgment is like conceptually (cognitive and motivates by itself), but well, there is nothing in the world that conforms to it. Our definition stands, but the world doesn’t oblige (as it were). NS: This is absurd because humans would then be *psychologically* incapable of making moral judgments (1.2). I find this too strong. Surely, some kind of *error theory* could then still be true:

people would be making moral judgments conceived (not necessarily thought of by the people themselves) along cognitivist and internalist lines (e.g. the way people make the judgments and the way they react to them bears all the hallmarks of belief or something like this), but they would be wrong/in error. Then you would have the usual error theory options: 1. Eliminativism; 2. Fictionalism; 3. Conservationism (as in J Olson's book). 1 might mean that they cannot be making moral judgments but 2 and 3? Why can we not have an error theory about moral judgment?

(b) The noncognitivist option (11.4): it is rejected using a version of the *moral attitude problem* – what is the mental state the noncognitivist needs and does it exist in the actual world? Query: Why do we need a *single* mental state? Why not work with a complex of mental states (perhaps building on each other)? Cf. Blackburn's emotional staircase idea. Sure, we have to start somewhere and that's a single mental state, but then all other complex things are built out of this one building block. Must this basic state be queer (non-actual)? I am not convinced.

II. DBTA (Desire-Belief Theory of Action)

It reads (1.1): "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."

ISSUES:

I think NS would agree (but I can also put this as a query) that there are really three separate theses involved in DBTA:

1. Motivation: motivation is produced by desire plus belief.
2. Independence: desire and belief are independent existences.
3. Existence: both desire and belief are real existences.

1 is typically taken to be *the* defining thesis of DBA, but as 11.2 shows, 3 is important too (Nagel's logical ghosts), and as the discussion of desire-as-belief shows (2.5), 2 has to be there, too. Or?

Should this setup be accepted, my questions concern 1 (i-iv) and 2 (v); I take 3 to be refuted:

(i) *What kind of a belief?*

In 1.1, NS makes it clear that the belief involved can be of the usual instrumental (causal – 3.4) or of the constitutive kind (cf. B Williams on practical deliberation). This is fine as far as it goes, but is this enough, or, taken from a different angle, is this not too restrictive? Take the example of moving hands. Some Humeans would not force us to say that there is an instrumental or constitutive belief present. E.g. Smith (yes, for this purpose, I read him as a Humean...) says that the belief "I can do A" suffices; and Mele goes even farther talking about intrinsically motivated actions as needing only desires (that involves what he calls the representative element of an action-plan) and no beliefs. Any principled reason for the restriction? It doesn't seem to me that this would compromise DBTA. Or would it? If so, why?

(ii) *Dancy's argument for pure cognitivism*

Dancy's views are not much considered in the book, except for a short detour on his rejection of what he calls psychologism (9.1) and to which I will come back later. But before

he argues against psychologism, Dancy first devotes an entire chapter of his *Practical Reality* to refute DBTA (the core of his argument also appears in earlier writings). His argument for what he calls pure cognitivism is however not considered at all in the book – is this for a reason? (And yes, I published on this, yes...) Since it is a direct, albeit non-empirical challenge on HTM, its absence is curious. (I know that NS wouldn't find Dancy's argument convincing since he makes it clear – which is the central premise of D's argument – that not all desire involves/produces/provides/encompasses (which? Cf Mele) motivation. (And yes, I published on this, too.)

(iii) Desire strength

Considering how important a role it plays in NS's account of the properties of desire (it appears in the definition of the hedonic, motivational, as well as the attention property/aspect), it is strangely undertheorized – almost in fact forgotten about – in the book.

In 7.1, we are told that desire strength is understood dispositionally and that we can use different measures for different theoretical purposes. Here and elsewhere (8.3), it appears suggested though that the two typical measures have to do with desire's motivational strength and its phenomenological (hedonic) property. Besides wanting to hear more about this, I find this unclarity strange considering that the properties of desire in Chapter 2 are all defined by reference to desire strength (using proportionality). These definitions suggest that e.g. motivation is proportionate to the strength of the desire – but then how could it be that the desire's strength *just consists in* its motivational force? These definitions, in other words, appear to suggest that desire strength is something primitive in our account of the properties of desire. (In fact, this is clearly put in 9.3: "Strength of desire explains this phenomenological difference, just as it explains motivation.")

(iv) The combination problem (not that one)

It's a kind of magic, isn't it: how do beliefs and desire combine? NS's solution (5.1, 5.2, 6.3) is attractively simple: the attention aspect of desire makes sure the combination comes about. This property of desire consists in directing attention to whatever the agent associates with the object of desire – including, of course, means of satisfying the desire. End of story.

- (a) I did wonder, though, if this is enough to get 1 (Motivation) on the table whenever we need it. My primary worry is that all this is contingent. Yes, I know we are talking about psychology, but even within that admitted contingency, there is this simple question: There are plenty of things an agent can associate with the object of desire, what guarantees – which appears to be the case given the ubiquity of human action based on desire-belief pairs (if DBTA is true) – that the agent almost always (NS says at one point: 'often') picks out the means-end (or constitutive) connection? Isn't *this* magic?
- (b) Furthermore, consider intrinsically motivated actions (Mele) and/or immediate bodily movements or expressive actions (Hursthouse, Döring) – in short, all those cases where even attributing a suitable belief to the agent is not easy. How would combination happen in these cases? Surely, it has to be sub-conscious, an automatic process of some sort. Is that covered by the attention aspect? How would it work?

(v) *Alternative cognitivist accounts*

Another curiously missing sustained discussion in the book is that concerning cognitivist accounts of motivation. This isn't to say that there is nothing: the desire-as-belief model is discussed (2.5), the desire idea is mentioned (1.3, although not discussed), Döring's affective perception theory of emotion is discussed (3.4, although not as a cognitivist account...), Tenenbaum's desire theory is discussed (7.4, although, in his view, as NS points out, it is belief that motivates), Kauppinen's name is mentioned (4.2). However, most of these discussions do not concern the idea that cognitivism is not exhausted by moral judgments as beliefs. Moreover, there are plenty more theories in the literature that are not represented at all: e.g., Kauppinen's view of moral intuition; Zagzebski on the thinning of moral judgment; Little on virtuous perception (interpreting McDowell), and so on and so forth.

(a) I suspect that this omission derives from the methodology endorsed by NS (my queries are above re this method): since HTM/DBTA can explain all the relevant phenomena, there is no point 'inventing' new mental entities to do the job. However, and I suppose this can be another issue with NS's method, it is not clear that these cognitive states are in this way *ad hoc* or at least superfluous. Their proponents – Little, Zagzebski, Kauppinen, many of the emotion theorists – do use empirical data and theories to support their claims. Could they just be side-lined using Occam's razor? This is far from clear to me.

(b) Take NS's own emotional perception model: representation – emotional disposition – moral feeling – moral judgment (belief) [on the lower branch we have: emotional disposition – motivation]. I am somewhat puzzled here. Where exactly is the *emotion* in this picture? These are supposed to be all *causal* connections implying that the cognitive, conative, and affective aspects of emotions do not form a single, integrated whole but at least some of them are mere add-ons? This certainly contrasts with the perceptual theory of emotion that NS claims to be happy with (4.1), as well as with Kauppinen who is also approvingly cited (4.2). Not to mention Döring who is criticised elsewhere, but is herself an advocate of a particular strand of the perceptual theory ('affective perception'). So what exactly is going on here? For me a natural reading of NS's emotional perception theory of moral judgment would be along – well, what the 'perceptual' qualifier actually suggests – cognitivist lines: it would not deny that moral judgments are beliefs, but it would deny that what motivates is a desire. Instead, as almost (?) all these people claim, what motivates us – emotions, moral intuitions, virtuous perceptions – are mixed states that include, *constitutively*, both a cognitive and conative (as well as affective) element. They are more like desires or desire-entailing-beliefs, with two directions of fit. Most of these people (or all) are then internalists about *these* states, not about belief.

III. Reasons (Humean Psychology of Reasons; Chapter 9)

This is intended to be a psychological theory operating with mental states, not a theory of *normative* reasons. There is also a clear parallel to the emotional perception theory of moral judgment (although there are also differences – see below). We have: representation – desire – experience of salience – normative belief [on the lower branch: desire – motivation].

ISSUES:

(i) *The model itself (as in 9.1)*

As with the emotional perception model of moral judgment (as belief), one could query why desire (in the role emotion/emotional disposition plays in the moral version) cannot be read along perceptualist lines as being constituted by the perceptual experience of salience (and 'shape' – a concept NS doesn't use but is typically mentioned elsewhere alongside 'salience' in the literature). Take e.g. Tenenbaum's theory of desire (and there are others). The idea, again, that normative judgment expresses a belief (hence no internalism concerning *belief* follows) could still be kept and also that desire motivates. But desire would now again be a cognitive-cum-conative state (denying Independence in my setting out of DBTA) and internalism might well be true of *this* state.

(ii) *Morality is not about reasons (9.7)*

The above reasoning could be strengthened if it was the case that moral judgment is about moral reasons. Then what I said above about the emotional perception model could be directly involved. However, NS denies that morality would be about reasons.

- (a) It is not clear to me what kind of 'rationalism' (strong – conceptual or metaphysical – connection between morality and reasons) is denied here. I suppose, given the emphasis on this being a psychological theory of reasons (and morality), the relevant rationalism is of the judgment kind (this is how I formulated it above). But then most of 9.7 appears to be about the existence version: the one about facts and normative reasons, rightness and wrongness etc. Now, we can give it to NS that he is right about existence rationalism (or call it as you wish), but why would this imply judgment rationalism? I.e. that it is possible that while facts about what is right and wrong are not facts about reasons, moral judgments are still normative judgments?
- (b) The examples of indifferent and evil agents. I suppose many would read these in the opposite way: as a *reductio* of the Humean position. This is what, to pose himself a challenge that he then claims to answer, Schroeder does: he calls it the too few reasons problem. To put it another way, these examples can also be taken to beg the question as far as the plausibility of the Humean position is concerned. More than needs to be said in either direction.
- (c) Perhaps reasons are not needed because they help explain the psychological significance of morality. But if reasons fundamentalism is true, then they are needed to explain the *normative* significance of morality. To psychologize it: if morality is not about reasons and moral judgments are not reasons-judgments, then, Scanlon and others would say, they are not *normative* judgments either. Surely, this would be a massive loss!
- (d) "The significance of morality isn't that it gets you what you want." But, of course, the emotional perception model of moral judgment does insist that moral emotions involve desires (to produce motivation), so in a way, while morality might be *more* about moral feelings/experiences, it is also about getting what we want...
- (e) The Foot example (and then Buffy). I am not sure I am convinced. Foot's original point was that categoricity cannot be the hallmark of morality because etiquette has the same

feature. From this it doesn't have to follow that neither provides reasons: all that follows (if at all, since of course Kantians would deny that Foot is right) is that moral reasons cannot be determined by categoricity (only) because then etiquette would also provide us with moral reasons, which we wouldn't want to admit. This is not the same as saying that neither morality, nor etiquette provides reasons – the challenge is rather to explain how morality can give us moral reasons that etiquette cannot.

(iii) Silencing (a cluster of phenomena in 9.3-9.5)

First, it is good to have two distinctions on the table: one between subjective and objective silencing (whether silencing occurs on the level of judgment or on the level of facts/properties); the other between rational and motivational silencing (whether silencing saps away the reason-giving force or the motivation – on the Humean model, though, the two might well go together).

- (a) The phenomenon B Williams calls practical necessity (also: moral incapacity) in 9.3 (if it is not that, then I would say this is another challenge to the Humean model): many deny that this is a merely psychological phenomenon. Some argue that it is a *sui generis* experience, i.e., irreducible (certainly to desires).
- (b) The phenomenon of Scanlon's bracketing in 9.4: Why invoke second-order desires? I would just go for 'swamping' (and yes, I've published on this...) That is, I would go with Schroeder (and Grice and Hubin) and say that the reasons involved are so decisively overridden that they are simply not on the agent's horizon.
- (c) The phenomenon of choosing among reasons for the same course of action in 9.5: "The Humean Theory allows reason-choosing by allowing a desire with content referring to the other desires to block the strongest desire supporting some action from motivating it, so that a weaker desire is the sole motivator." First, I would like to know how this blocking works; if anything, this looks an invention/*ad hoc*. Secondly, we are talking about blocking the desire from motivating the action – but this phenomenon surely is not about *motivational* silencing? All that we have is that the agent chooses one reason instead of another – it is unlikely that even rational silencing is involved. But it seems the Humean is not only forced to read this into the case but also motivational silencing. I find this implausible.

(iv) Against Dancy's rejection of psychologism (9.2)

"Schroeder's view avoids psychologism by allowing reasons to be outside the mind, where things promoting desire-satisfaction often are. While desires are the background conditions of reasons, they aren't the reasons themselves."

I don't like this solution:

- (a) First, Dancy doesn't have a problem with putting mental states in the background: he takes beliefs to be enablers in action-explanation, after all (he is hard pushed to find a role for them...) He would still say, though, that desire has absolutely no role to play (because of his argument for pure cognitivism) in action-explanation, neither in the foreground, nor in the background. In short, it is not entirely correct to consider, in the

present Humean context, his argument against psychologism without his argument for pure cognitivism.

- (b) Second, if action-explanation is causal and desire-belief pairs have the relevant causal powers only, but they are not reasons, then it follows that reasons don't explain action. This is certainly something Dancy would not accept: his whole argumentation is premised on the idea that we have to give an account of what it is for an agent to act for a reason. Nor is he alone with this idea, most of the literature, as far as I can tell, would agree with him on this. The only alternative to this is to give up the idea that action-explanation is causal. This is in fact what Dancy does, but would the Humeans want to do it? NS says that "Davidson's view that desires are reasons was an advance in 1963, but philosophy has made progress." Would he be also willing to see the rest of Davidson's position go? If they do, they would give up almost everything ('almost' since there is the vexed matter of pure cognitivism pro and con) that at least Dancy thinks makes their position in the philosophy of action interesting?
- (c) For these reasons, I think the Humean would be better off endorsing an alternative position, which, contrary to NS's remark ("all the self-described Humeans agree with Dancy that reasons don't cause action, while maintaining that desire-belief pairs do"), is well-represented in the literature (Alvarez, Mantel etc): there are two kinds of reasons when it comes action-explanation, the motivating and the explanatory. The former are non-psychological (facts, propositions, take your pick – or read my papers on it...), the latter are mental states. (In fact, in the 2003 PPR symposium on Dancy's book, M Smith and others proposed just this idea.)