How Simple is the Humean Theory of Motivation?

ABSTRACT: In recent discussions of the Humean Theory of Motivation (*HTM)*, several authors – not to mention other philosophers around the proverbial water cooler – have appealed to the simplicity of the theory to defend it. But the argument from simplicity has rarely been explicated or received much critical attention – until now. I begin by reconstructing the argument and then argue that it suffers from a number of problems. Most importantly, first, I argue that *HTM* is unlikely to be simpler than even close competing theories, and second, it is unlikely that a plausible version of the theory will be very simple. Moreover, I argue that a convincing case for *HTM* is likely to have to show that it is more virtuous than defenders have done so far.

In recent discussions of the Humean Theory of Motivation (*HTM*), several authors – not to mention other philosophers around the proverbial water cooler – have appealed to the simplicity of the theory to defend it (Heath 2008, ch. 1; Persson 2019; Smith 2013; Sinhababu 2009, 2017). This argument from simplicity has received surprisingly little critical attention. The argument sometimes appears for itself, and sometimes in discussions where *HTM* is defended *vis-à-vis* other theories in other dimensions as well, such as regarding how explanatorily powerful they are with respect to various features of action. In both cases, simplicity tends to be used as a tiebreaker in virtue of which we are supposed to prefer *HTM* to other contending theories about what action is. The point is that *HTM* uses fewer entities in its explanation of action than the other theories, so it is preferable to them.

I want to push back against the idea that simplicity favours *HTM*. Even assuming that simplicity is a virtue, I shall argue that it does not lend support to the view in comparison with relevant competing theories. To show that, I aim to explicate the argument from simplicity and show that it suffers from two main problems. It is not at all clear whether *HTM* is simpler than even close competing theories, and it is not at all clear whether a *plausible* version of *HTM* will be very simple. Less centrally, I shall also argue that it is doubtful whether simplicity alone could play the key role in the justification of a theory of action that it sometimes seems taken to be – other theories might be more virtuous on aggregate, or there might be virtue stand-offs. These considerations indicate that simplicity does not lend any significant kind of support to *HTM*.[1]

To be clear, I see the argument from simplicity for *HTM* as a rather general argument in the philosophy of action which seems to be tempting to many philosophers and attributable to many in print (e.g. Heath 2008, ch. 1; Persson 2019; Smith 2013; Sinhababu 2009, 2017). Some preemptive remarks should however be made about Neil Sinhababu's (2009, 2017) version: as he arguably has developed the argument from simplicity the most, I shall use his discussion to develop a generalised version of the argument from simplicity, but this paper is not primarily a critical

discussion of Sinhababu's views *qua* the views Sinhababu actually holds.[2] Rather, my aim is to argue that it is inadmissible to argue that *HTM* gains any significant kind of support from simplicity when it is compared with other relevant theories. This is a worry for all philosophers who might be tempted to think that simplicity supports *HTM*. It is of general interest as it tells us something about how *HTM* may or may not be defended, and it matters in particular for philosophers who might be tempted to rely on the argument to defend *HTM*.

Here is the paper plan. I introduce *HTM* in section (1). In section (2), I reconstruct the argument from simplicity for *HTM*, expanding on and generalising from Sinhababu's (2017) version. In section (3), I explicate the account of simplicity in the argument. In sections (4) and (5), I then criticise the key premise that *HTM* is simple in the relevant sense. There are other relevant comparable theories that are simpler, and the best forms of *HTM* need not be very simple themselves. Less centrally, in section (6), I note that it is not obvious whether simplicity (even in combination with explanatory power) makes the theory sufficiently virtuous to be acceptable. I conclude in section (7).

(1) *HTM*

In its standard form, *HTM* is a causal theory of action. According to causal theories of actions, actions are actions in virtue of being events caused by the right cause, in the right way.[3] On *HTM*, the right cause is a *belief-desire pair*, suitably linked up, which causes events – usually with an 'in the right way' clause added to avoid the problem of deviant causation.[4] Hence, an event is an action if (and, ordinarily, only if) and because it stems from a belief-desire pair in the right way.

For Humeans, a belief is a mental state that represents the world, and a desire a mental state that motivates an agent to interact with the world. These states are modally distinct from each other; it is always possible to have one without the other.[5] But if one is motivated to act – i.e. has a desire – and combines it with a representation of how one might achieve its aim – i.e. a belief – and this pair of states then cause an event in the right way, the event is an action.

Most defenders of *HTM* also add that this belief-desire pair constitutes a motivating reason for the action. A motivating reason is a reason that explains and rationalises why the action was done, making it intelligible. This is explicit in Donald Davidson's famous (1963) formulation of *HTM*. It looks like this:

> (*C1*) *R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property.

(*C2*) A primary reason for an action is its cause. (Davidson, 1963, p. 687; 693)

If we add the assumptions from above – actions are events, and an anti-deviant causation-clause ought to be added to the theory – Davidson's formulation is still the paradigmatic version of *HTM*. When a particular version of *HTM* is needed to make sense of the discussion below, I will use this expanded version of Davidson's view plus the standard terminological conventions that Davidson's 'pro attitude' is equivalent to 'desire' and that 'primary reason' is equivalent to 'motivating reason' as my default interpretation of what *HTM* is. However, the arguments below are arguments against all versions of *HTM* that are supposed to gain support from simplicity.

(2) The Argument from Simplicity

Several authors have defended *HTM* by appealing to its simplicity, parsimony, or economy (I use these terms interchangeably). The defences have taken slightly different forms. Some authors only invoke the point implicitly (Persson, 2019), others mention simplicity or economy (Heath, 2008; Smith, 2013), but in the most elaborate defence, Sinhababu takes the simplicity of *HTM* to interplay with its explanatory power.[6] Doing so, he argues that *HTM* lies at the heart of explanations of – as the title of Sinhababu (2017) goes – 'action, thought, and feeling.' But simplicity plays the same role for the authors whether or not it gets explicitly linked up with explanatory power: it is a tiebreaker between competing theories with the implicit assumption that the competing theories are worth comparing to *HTM* in the dimension of simplicity.[7]

Generalising from the different ways of defending *HTM* by appealing to simplicity, we get the following argument:

(1) The best form of *HTM* gives us an explanation of action which is (powerful and) simpler than competing theories.
(2) If the best form of *HTM* gives us an explanation of action which is (powerful and) simpler than competing theories, then we should prefer it to less simple competing theories.

(C) We should prefer the best form of *HTM* to less simple competing theories.

With the parentheses around explanatory power in the formulation of the argument, I mean to indicate that the potential explanatory power of *HTM* can be bracketed for now. How can I do that when explanatory power very clearly features in Sinhababu's discussion? It bears repeating that the argument from simplicity just presented is a generalised reconstruction of a consideration appealed to by many authors, not just Sinhababu. Various authors' use of the notion of simplicity are implicit in the argument. As Sinhbababu's discussion of it is the most extensive and developed

one in the literature, I draw on it to a significant extent in my reconstruction of the argument here and below, but it is not the sole contributor to the argument as I reconstruct it.

Instead, abstracting from Sinhababu – who does not say very much about what explanatory power involves – I take explanatory power to be a theory's ability to explain some set of data well for some relevant interpretation of 'well'. For example, the theory might increase our understanding, avoid being sensitive to changes in background conditions, have a precise *explanandum*, increase the factuality of our beliefs, integrate new information with old, be cognitively salient – or have some sort of conjunction of several of these properties (cf. Ylikoski and Kuorikoski 2010). 'Well' could also, arguably, be interpreted as having additional properties still – but it does not matter which interpretation one ultimately would like to opt for here.

This is because we can bracket explanatory power, construed in this broad way, by holding it fixed between relevant competing theories for now. Independently of which theories we ultimately think are the most powerful, it is plausible to take many theories to be at least reasonably explanatorily powerful on this broad interpretation of the notion. And then it is interesting to compare *HTM* with other reasonably powerful theories to see whether simplicity supports it over the others – whereas it is less interesting to compare *HTM* with views that can be easily dismissed.[8] Hence, the argument focuses on whether *HTM* gains support from simplicity when contrasted with competing theories which are strong competitors – and we can bracket the question of explanatory power when comparing it with *such* theories.

The theories that I take to be particularly interesting to compare to *HTM* in the dimension of simplicity are theories that are commonly endorsed and are, like *HTM*, versions of the otherwise plausible causal theory of action. Commonly endorsed and otherwise plausible theories are strong 'test cases' – if *HTM* does not do better than them when it comes to simplicity, it does not gain any significant kind of support from simplicity in general. More concretely, what I shall do in my discussion below is then to compare *HTM* with some belief-based and besire-based causal theories of action, where the idea is that actions are events caused by beliefs or besires (i.e. mental states that both represent the world and motivate), respectively.[9] And I shall argue that *HTM* is not simpler than belief- and besire-based theories.

Of course, various authors will have various arguments in favour of or against various views – Humeans are not likely to think belief- or besire-based causal theories of action are very explanatorily powerful, and non-Humeans will worry about *HTM*. Therefore, one might again worry that I am saying too little about *which* theories are explanatorily powerful. However, philosophical disagreement is no objection to using such theories for the sake of comparing their respective simplicity. There is almost always good-faith disagreement about philosophical theories,

so no theory can be straightforwardly ruled out just because of that. But commonly endorsed causal theories of action all seem to me to be sufficiently powerful to deserve a place at the table of simplicity comparisons, regardless of whether one might have arguments speaking in favour of or against them.

The last point can be illustrated with the various dimensions of explanatory power introduced above. Non-Humean causal theories are likely to increase our understanding of action because they take action to consist of events caused by the right cause in the right way, they have the same *explanandum* as *HTM* (action), integrate a pre-theoretical folk-psychological notion of action with new information based on something in the world (events caused in certain ways), and they are cognitively salient (as they are on the table in the literature). The only dimension listed where the non-Humean causal theories are not straightforwardly powerful concerns whether they increase the factuality of our beliefs – but that is because it would be question-begging to assume that *any* theory of action, Humean or not, is factive prior to evaluating it fully. Hence, abstracting from particular arguments for or against *HTM*, these competing theories seem powerful, and therefore like strong test cases for *HTM* in the dimension of simplicity.[10]

Regardless of which theory ultimately is the most powerful one, then, we can try to determine whether *HTM* is simpler than the others: because they seem pre-theoretically roughly equally powerful, and disagreement does not straightforwardly rule out one theory or another, we can bracket their explanatory power here. Instead, we can try to see whether simplicity could function as a tiebreaker between *HTM* and other theories.

We can even run a similar argument about them whether or not we care about explanatory power as a theoretical virtue. If one cares about explanatory power and holds the theories to be equally powerful, it will be a tiebreaker between equally powerful theories. If one does not care about explanatory power on its own, it will be a tiebreaker by itself. What I am after here, then, is to discuss whether simplicity serves as a tiebreaker which supports *HTM*, independently of whether one wants to link it up with explanatory power.

(3) Qualitative simplicity

However, before I can say more about the (lack of) simplicity of *HTM*, we need an interpretation of what 'simplicity' is supposed to mean here. Unlike other authors, Sinhababu does say some things about how he understands it, and a plausible interpretation of the virtue can be distilled from his discussion. Hence, I shall use his discussion to develop an appropriate interpretation of simplicity – but here, too, I want to emphasize that the view I shall distil out of his discussion will

go beyond exegesis, for I will also give a brief independent defence of its appeal in the argument from simplicity.

With that *caveat* in mind, Sinhababu's core idea is that:

> Including motivational states that aren't desire [in explanations of action, thought, and feeling] commits [dialectical opponents] to additional types of causal powers, so a psychology built around the Humean Theory will be simpler than one built around their views. If the Humean Theory accounts for all the phenomena, its simplicity will make it superior. Usually I'll agree that things my opponents invoke are real (such as intention, willpower, and agency), and show that the motivation they produce simply is that of desire. Sometimes I'll reject the entities or processes they invoke (such as motivationally potent beliefs and libertarian free will). Both the former reductive move and the latter eliminative move preserve the Humean Theory's simplicity, invoking no fundamental motivational states other than desire. (Sinhababu 2017, p. 13)

HTM, he thinks, is simple in the sense that it does not introduce non-Humean entities that could help making explanations powerful, but instead explains all actions in terms of belief-desire pairs (as well as several other phenomena in similar terms). Sinhababu thinks that even non-Humeans are committed to entities like beliefs and desires, even (at times) in action explanations (Sinhababu, 2017, ch. 1). Nothing more is needed here than what everyone already is committed to; no additional entities or powers or processes are required. That is why *HTM* fits within a simpler total account of our psychologies.

This explication gives us an interpretation of his take on simplicity. The relevant overarching notion is *ontological simplicity*, which is concerned with the number and complexity of entities invoked by a theory.[11] Inspired by David Lewis (1973), we can distinguish between *qualitative* and *quantitative simplicity* as different kinds of ontological simplicity. Qualitative simplicity means that one invokes as few *types* of entities as possible in an explanation, whereas quantitative simplicity means that one invokes as few *token instantiations* as possible of the phenomenon one is explaining, regardless of how many types one is committed to.

The types in question should be understood liberally, however, as generalisations over some (sets of) similar features of our ontology, not necessarily entities in the sense of objects proper or kinds (whether natural, social, or other). Qualitative simplicity matters whether or not we are talking about objects *simpliciter* or other properties of objects. It is, for example, qualitatively simpler *both* to explain some alleged sightings of the Loch Ness monster as hallucinations (so they are not sightings of *any* entity) and to explain them as sightings of logs in the water (so they are sightings of an object, but one that lacks the property of being a cryptid) than it is to explain the

sightings as sightings of an actual existing cryptid. In the former case, we do without an object. In the latter case, we do without a property of an object. To the extent that we care about qualitative simplicity, we should do with as few entities as possible for our explanatory purposes in at least one of these ways.

Because Sinhababu attempts to avoid extravagant entities, he seems to endorse the qualitative take on simplicity. In fact, he explicitly invokes Lewis – who endorses qualitative simplicity – for his purposes on p. 15 of Sinhababu (2017). Hence, the idea is that extravagant entities in action explanations should be eliminated from our ontology altogether, or at least reduced to beliefs or desires (Sinhababu 2017, ch. 1). Accordingly, the theory with the least number of entities in the explanation of action will be counted as the better one – and I will focus the rest of my discussion on qualitative simplicity, leaving quantitative simplicity on the side.

There is a potential worry in the background here, however. It is one thing to establish that Sinhababu invokes qualitative simplicity, and quite another to think that qualitative simplicity is what genuinely matters. So why is it a theoretically helpful interpretation of the notion of simplicity in the argument from simplicity for *HTM*?

There are two reasons. First, even assuming that ontological simplicity is a virtue (for any reasons we may prefer), it simultaneously seems like quantitative simplicity does not plausibly matter when we talk about the extension of the set of actions. This is because it would be unhelpful to think that as few events as possible should count as actions. Philosophers differ widely when it comes to what they take the extension of agency to be – on one extreme, philosophers like John Hyman think that any purposive movement involves agency (Hyman 2015, ch. 2). On that view, any purposive movement could plausibly count as an action. On the other, Christine Korsgaard thinks that (human) agency involves constitutive commitments to the categorical imperative as a form of self-integrating principle, so (human) movement that does not does not count as action (Korsgaard 2009, ch. 4-5). As there are such extremely divergent views of action and agency, it seems hard to find much pre-theoretical reason to prefer one theory to another, and hence also much pre-theoretical reason for why we should treat action or agency as being as quantitatively simple as possible.

Second, qualitative simplicity should be distinguished from the virtue of theoretical unity. At times, Sinhababu writes as if simplicity does the same work as unity (Sinhababu 2017, p. 49). But the notions are different. It is possible to have a unified explanation – in the sense that one explains all tokens of some phenomenon – which involves many entities without simplicity. For example, a constitutive explanation of water as H_2O is as unified as one that introduces a new, distinct, element to explain all water tokens.[12] Focusing on qualitative parsimony means that we

are narrowing down our conception of simplicity to one virtue rather than one that confuses several virtues.

(4) Is *HTM* simpler than the alternatives?

With *HTM*, the argument from simplicity, and qualitative simplicity characterised, it is now time to evaluate the argument from simplicity. There are two problems with premise (1) in the argument (i.e. 'The best form of *HTM* gives us an explanation of action which is (powerful and) simpler than competing theories'). They indicate that qualitative simplicity does not support *HTM*, for *HTM* need not in fact be simpler than other competing theories. I shall discuss the first problem in this section and then discuss the second problem in the next.

The first problem is that it is unclear why desires would feature in qualitatively simple explanations of action. Assume, like everyone else in the debate, that we should not be eliminativists but rather want and can provide and explanation of what makes action 'action'. With this assumption granted, desires may well be redundant when we contrast *HTM* with close theoretical competitors, for qualitatively simpler views that invoke fewer entities in the explanation of action are available. To make that point, I shall contrast *HTM* with two other widely embraced forms of the causal theory of action: belief-based theories, according to which beliefs suffice to motivate actions, and besire-based theories, according to which an additional kind of mental states – besires – does that.

As per the methodological discussion in section 2, I want to emphasise two methodological points about this comparison before starting it. First, there may be independent arguments in favour of either these theories or *HTM*, for example based on their respective explanatory power. But I have bracketed explanatory power, so I shall not pay them much attention here. Second, these theories are just two *illustrative examples* of alternative possible theories about action. They have been chosen for that role because they are commonly defended causal theories of action, but one may of course present more complicated versions of them than the cursory versions I present here, or other accounts of action still. Hence, comparisons between *HTM* and other theories of action, causal or not, may reveal even greater problems still for *HTM*. But, again, belief- and besire-based views are strong test cases: if *HTM* is not qualitatively simpler than widely endorsed and similar theories such as belief- or besire-based causal theories, the argument from simplicity seems remarkably flawed.

Belief-based theories, then, are views according to which beliefs suffice to motivate action, and desires, if they exist at all, are to be understood in another way, but do not take part in action

explanations except possibly insofar as they are treated as a kind of beliefs.[13] Otherwise, these theories work exactly like *HTM*. Furthermore, on besire-based theories, besires, which are mental states that both represent and motivate, can do the explanatory work behind action instead of beliefs or desires – but they are otherwise identical to the others.[14] On both belief- and besire-based views, however, there is only one type of mental state involved in explaining actions: beliefs and besires, respectively. It follows that we get qualitatively simpler explanations of actions than what *HTM* can provide.

One potential worry about this point is that belief- and besire-based theories might not be very simple themselves. After all, the beliefs in question might have to be fairly complex, for at least some of them would have to explain what we take desires to be, making them play unexpected psychologies roles. Moreover, many philosophers are sceptical about besires. Besires seem to have two directions of fit, yet accepting mental states with two directions of fit is not very parsimonious. Accordingly, one might worry that appealing to beliefs or besires involves appealing to extra entities too. In other words, what we gain on simplicity in action explanations by appealing to such states might be something that we lose on the simplicity of the mental states involved.

However, *any* mental states we appeal to in the explanation of action are going to have to be fairly qualitatively complex. This is because they have all kinds of properties. In particular, exactly the same worry shows up for desires, which Humeans of course must appeal to. For example, on Smith's conception of desires as motivating states, they have many different properties – they are all and only all motivating states, but they can also be either occurrent or non-occurent, be more or less easy to know, have different etiologies by being caused by normative beliefs or appear independently of these, and so on (Smith 1994, ch. 4). And Sinhababu's conception of desires is arguably even more complex. He thinks that desires are not just dispositions to be motivated; they also have a hedonic aspect (so one usually tends to gain pleasure when they are satisfied), direct attention, come in two flavours (by being either positive desires or aversions), and intensify by being vividly imagined (Sinhababu 2009; 2017, ch. 2).

In response, perhaps one might think that the theories could become simpler by being based on some unified account of why they have these properties deeper down.[15] Perhaps, as *inter alia* Smith's famous account suggests, this is so because they are motivational dispositions. But treating them as motivational dispositions would not make them qualitatively simpler in any relevant sense. Qualitative simplicity, as characterised above, is not just a virtue about which kinds or entities we appeal to in our explanations, but also of how many properties we ascribe to entities – remember how both interpreting sightings of the Loch Ness monster as hallucinations and as sightings of logs gives us more parsimonious interpretations of the sightings than veridical

interpretations give us. So whether the properties of desires are made simpler by treating desires as motivational dispositions does not speak to the issue.[16]

Hence, it seems to me like the minutiae of any account of mental states involved in action explanations on causal theories of action are going to have to be rather complex, in the sense that they all feature lots of different properties. For present purposes, this means that no theory clearly seems to do much better than any other insofar as the complexity of the mental states involved is concerned. Hence, I stick to discussing which theory of action is simplest *in action explanations*. And then other views formulated without desires, such as views formulated solely in terms of beliefs or besires, remain simpler than *HTM*.

One may wonder whether simplicity in the explanation of action is what matters here, however. Sinhababu, at least, does not seem to think that action is the only *explanandum* for *HTM*. He defends *HTM* as part of a package deal which is supposed to explain action, thought, and feeling, and he even explicitly aims to defend his view as one that provides an explanation of a broad range of phenomena associated with motivation – and which does not interfere with explanations elsewhere in psychology – rather than a view which explains action alone (Sinhababu 2017, p. 15). This, he thinks, is for good reason:

> 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. (Sinhababu 2017, p. 15)

There are, however, several problems with this assumption about our *explananda*, so I shall conclude this section by arguing, *contra* Sinhababu, that explaining action is just enough for a theory. If that is right, the worry about the simplicity of *HTM* remains an issue.

Assume, with Sinhababu, that what we most fundamentally are after is a simple theory of everything in our psychology. If that is right, his own argument appears redundant – developing a theory that grand is such a massive undertaking that we have no idea about whether Sinhababu's own view approximates it. It goes some way towards explaining some phenomena, but just explaining some phenomena still seems to leave it rather local, and its dialectical force therefore limited. It says nothing about qualia, the nature of mental states like emotions or intuitions, complex psychological phenomena ranging from depression to self-deception to personalities, how beliefs are related to knowledge, etc. Additional assumptions are necessary for saying anything

about these things. Of course, Sinhababu's view could be correct if he makes the right additional assumptions and these turn out to be consistent with or even support the rest of his framework, but his theory is not extensive enough to guarantee answers to all these questions at present.

However, if we instead make the plausible assumption that Sinhababu's *explanandum* is less than an entire psychology, but still remains ambitious as he attempts to explain action, thought, and feeling at once, the *explanandum* seems arbitrarily combined. Why should we attempt to explain action, thought and feeling rather than one or two of these phenomena, or some other set of features of a complete psychology?

Often, philosophers do not discuss explicitly how *explananda* should be picked out. But one way to think about that – which also seems to me in line with standard practice – is to attempt to explain some phenomenon that we tend to think hangs together and which will be philosophically rewarding if we manage to explain it, either for its own sake because we reach truths about, or at least some understanding of, it, or because understanding it matters because we can put that understanding to use to explain other phenomena. 'Action' is just one concept, so it seems to hang together, and understanding it seems both likely to be philosophically rewarding for its own sake and for the sake of other debates, for example in the philosophy of mind or in moral philosophy. This indicates that it is a better *explanandum* than Sinhababu's conjunction of action, thought, and feeling, as they do not straightforwardly appear to hang together.

In the last block quote above, Sinhababu does however seem to give an argument against discussing action (or anything else) on its own. He seems to suggest that, at least insofar as we are concerned with parsimony, we should not opt for explanations of singular phenomena such as action. This is because explanations of singular phenomena would rule out (good) explanations based on multiple explanatory factors. It would always be simpler to provide an explanation for each phenomenon with some additional independent force which only explains that phenomenon. If that is right, there does indeed seem to be something problematic about taking action to be an *explanandum* on its own. But I do not think the point that we can add a primitive entity to explain any phenomenon to maintain simplicity helps him. Other theoretical virtues than simplicity come into play here. This indicates that action indeed may be discussed on its own.

How so? Information integration was one of the properties of explanatory power introduced in section 2 above: a theory which integrates old and new information seems to be a more explanatorily powerful than one that does not. Now, we might think of information integration as conducive to explanatory power or to something else – maybe to theoretical conservativism, or maybe it is a *sui generis* virtue – but regardless, together with Sinhababu's example here, it shows that other considerations than simplicity matter. This is because introducing

a new entity to explain some phenomenon does not plausibly integrate our explanation of this new phenomenon with our older information – and this lack of theoretical integration makes it unreasonable to suggest separate fundamental forces to explain each separate phenomenon. Analogy: if we have a theory of mammal evolution in biology and also want a theory of the evolution of birds, then it is intuitively more virtuous to also explain why birds have evolved using the same evolutionary theory, for that integrates our explanations. But if so, we should try to make sure that our theories are virtuous when it comes to different virtues, not just separate simple theories.

With these points in mind, it seems just fine to aim to discuss the nature of action rather than our psychologies in general for the purpose of providing an explanation of some phenomenon. And we can discuss the nature of action by making use of otherwise virtuous theories – such as theories that are explanatorily powerful, conservative, fruitful, etc. – rather than theories that just posit some random force behind what makes something an action (where 'action' is the phenomenon we have singled out). We can then try to see which one of these is the most virtuous, including which one is the qualitatively simplest.

So, to summarise: holding explanatory power fixed, there is no reason to think that we should not explain action in terms of beliefs or besires, using suitable notions of these states. This would still generate qualitatively simpler explanations of action than *HTM*.

(5) Is a good version of *HTM* simple?

The second main problem for *HTM* when it comes to premise (1) is that it is unclear whether the *best* form of *HTM* is very qualitatively simple, both in contrast with other Humean theories and with non-Humean ones. Let us grant, for now, that some form of *HTM* indeed is fairly qualitatively simple. Then Humeans may still need to invoke more entities than those that feature in Davidson's paradigmatic version of the view. If so, it seems very plausible that one or many of the more complicated versions of *HTM* may be theoretically better than Davidson's view with respect to other virtues than simplicity. This entails that the virtue of simplicity is not likely to lend support to a plausible version of *HTM*, so premise (1) suffers from another problem as well.

We can exemplify this problem by using Smith's (2009) version of *HTM*. According to this version of the view, a third entity – a capacity for being instrumentally rational – must be added to beliefs and desires to explain actions.[17] If we need such a capacity, it might be that a properly formulated Humean view is not so simple, after all; it features a third entity alongside beliefs and desires. If this is right, it means that the best version of *HTM* might involve more than the

Davidsonian one I have used to illustrate *HTM*. But the virtue of simplicity would still have us supporting the Davidsonian theory – or maybe one that is simpler still.

In response, it might be claimed that the extra assumption makes *HTM* more explanatorily powerful. If so, a version of *HTM* which is explanatorily powerful and simple might still need it. This thought does not improve it for present purposes, however. I argued at some length in section (2) that we should bracket explanatory power for present purposes, and we should not deviate from that standard here.

Furthermore, it may be other things than explanatory power that could improve *HTM* here. For example, Smith (2009) uses his appeal to instrumental rationality primarily as a way to solve the problem of deviant causation. That may be good for the theory because it makes it more coherent with our background beliefs (if we believe that deviantly caused events are not actions) and because it is more fruitful (as an argument for instrumental rationality invites many new research questions), rather than because it is more powerful than other forms of *HTM*.

In fact, it need not matter how Smith or others defend their extended forms of *HTM*; the point is more general still. It may be that *any* more complicated form of *HTM* should be preferred for reasons that do not have anything to do with explanatory power – such views may be better at solving various problems or have various other theoretical virtues.

Another possible reply to the second worry for *HTM* is to insist that even though it is not the simplest version of *HTM* that we ought to accept, any version of *HTM* is still *fairly* simple. That might be true. But it is unclear if a view with extra assumptions – such as a capacity for rationality – is qualitatively simple *enough* to be supported by an appeal to simplicity in contrast with a theory, Humean or not, that does without them. Again, why could not all the explanatory work be done by beliefs or besires? If so, the most plausible form of *HTM* does not gain much support from simplicity.

(6) Premise (2) and beyond

I have now presented two problems for the alleged qualitative simplicity of *HTM*. They are both problems for premise (1). However, there are also two distinct theoretical risks for the second premise in the argument (i.e. 'If the best form of *HTM* gives us an explanation of action which is (powerful and) simpler than competing theories, then we should prefer it to less simple competing theories').[18] These risks indicate that premise (2) is not strong enough to be conclusive for accepting *HTM*, independently of whether *HTM* is simple and powerful or not.

I do want to emphasise that these are theoretical *risks* rather than conclusive arguments against the premise, however. If *HTM* is stronger than other views when it comes to explanatory power and simplicity, that is still an important theoretical result. It shows that we might prefer it to other theories all else being equal. But all else need not be equal, and then these risks appear. This is because, as I have been suggesting, qualitative simplicity and explanatory power are not likely to exhaust the list of theoretical virtues. There are several other virtues, such as conservativeness, elegance, and theoretical fruitfulness – feel free to plug in any that might matter. The other virtues are not plausibly reducible to qualitative simplicity and explanatory power, and while it is possible that simplicity and power are the most important virtues in contrast with the others, it is not at all clear why they would be. That generates two risks that undermine the strength of premise (2).

The first risk is that if *HTM* is not motivated by all the virtues, it might not be the most virtuous theory on aggregate. It might well be the case that another theory is more virtuous because it is virtuous in many other ways than being simple and powerful, whereas *HTM* only is virtuous in two ways in premise (2). Assume, for example, that we are metaethical cognitivists who think that moral judgements are cognitive rather than purely non-cognitive states, realists who think that these judgements sometimes are true, and motivational internalists who think that moral judgements motivate. Then a besire theory seems to explain moral motivation very straightforwardly: some moral judgments involve grasping moral truths, and doing so motivates action. Then the besire theory is conservative because it fits our background beliefs, it is elegant because it unifies our metaethics and moral psychology, and it is fruitful because it generates research questions about how to differentiate besires from other judgements about the world. So explanatory power and simplicity do not do enough to support *HTM* by themselves in contrast with the otherwise virtuous pet besire theory just presented.

The second risk is that focusing on the two virtues in the argument without considering other possible virtues can lead to virtue stand-offs. This is because several theories may be equally virtuous because they all have different virtues, and this is so even if *HTM* is the simplest and most explanatorily powerful view. *HTM* might have general explanatory power and simplicity going for it, but the besire theory from the last paragraph still seems to be conservative, elegant, and fruitful. That makes it unclear whether we should accept *HTM* rather than some other theory.

With these two risks in mind, premise (2) seems rather weak. Perhaps we have some reason to prefer *HTM* to other theories in virtue of its alleged explanatory power and qualitative simplicity, but it remains unclear if that is *enough* reason to prefer it when it is compared with other theories.

In response, one could perhaps argue that explanatory power and qualitative simplicity are especially important in contrast with the other virtues. This makes particular sense in the case of explanatory power: we presumably want powerful explanations, regardless of how that should be understood. But that is hardly the case for qualitative simplicity. Famously, it is not obvious whether the world is qualitatively simple – and it may in fact very well be the case that increased theoretical sophistication may lead us away from simplicity rather than towards it, since some argue that that is how psychology has tended to progress over time (e.g. May 2018, p. 196). Of course, it might be the case that the world is simple for some non-obvious reason, but then one would have to show what reason that is to defend simplicity with the non-obvious reason. Accordingly, it is not obvious why simplicity is especially important.

Another possible way to avoid the risks might be to appeal to the joint-carvingness of the theory. The idea here would be that power and simplicity are related to joint-carvingness, and insofar as a theory carves reality at its joints, it is for that reason better than others, even though it might clash with other theories about the virtues. But then joint-carvingness, too, seems like a theoretical virtue. Taking joint-carvingness to matter this much seems based on a substantive judgement about which theoretical virtue we should prefer as an adjudicator, but that judgement needs defence. So more work remains to show that simplicity is important enough to be able to play a deep theoretical role in a defence of *HTM*.

(7) Conclusion

Again, much more work must be done to defend *HTM* by appeal to its theoretical virtues. Such a defence must show that a plausible version of *HTM* is simple in comparison with competitors, *contra* the arguments in sections (4) and (5). In particular, it must be shown that *HTM* is the simplest explanation in contrast with widely endorsed and similar ones, and it must be shown that the best version of *HTM* is simple. Moreover, it would also be good to show that *HTM* does well enough with respect to the other virtues to avoid the risks that other theories might be more virtuous on aggregate or that there are virtue stand-offs. Fortunately, I do not think it is impossible to defend *HTM* by appealing to other, or maybe even all, of the virtues. But attempting to do so is a task for another occasion.

Acknowledgements

This paper has been long in the writing and versions of it have received comments from far too many philosophers to list. But I want to thank audiences at Leeds (in particular, members of the postgraduate community when I did my PhD between 2015-2019), ECAP9 at LMU Munich, and

the Humeanisms conference hosted by the Hungarian Academy of Sciences in Budapest in 2018 for comments on early presentations of some of the material in here. Special thanks go to Adina Covaci, Alexios Stamatiadis-Bréhier, Alison Toop, and Will Gamester – as well as Neil Sinhababu for letting me have an early look at Humean Nature and Jessica Isserow for encouraging me to have that look. Finally, I want to thank Chris Cowie and reviewers at this and other journals for helping me improve the mature draft.

References

Alvarez, Maria. 2010. *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford: Oxford University Press.

Davidson, Donald. 1973. "Freedom to Act." In *Essays on Freedom of Action*, edited by Ted Honderich, 137—156. London: Routledge.

Davidson, Donald. 1963. "Actions, Reasons, and Causes." *Journal of Philosophy* 60 (23). 685—700. doi:10.2307/2023177

Gregory, Alex. 2021. *Desire as Belief: A Study of Desire, Motivation, and Rationality*. Oxford: Oxford University Press.

Heath, Joseph. 2008. *Following the Rules. Practical Reasoning and Deontic Constraint*. Oxford: Oxford University Press.

Hempel, Carl. 1961. "Rational Action." *Proceedings and Addresses of the American Philosophical Association* 35: 5—23.

Hyman, John. 2015. *Action, Knowledge, & Will*. Oxford: Oxford University Press.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Lewis, David K. 1973. *Counterfactuals*. Oxford: Blackwell.

May, Joshua. 2018. *Regard for Reason in the Moral Mind*. Oxford: Oxford University Press.

Persson, Ingmar. 2019. *Reasons in Action: A Reductionist Account of Intentional Action*. Oxford: Oxford University Press.

Schroeder, Timothy. 2004. *Three Faces of Desire*. Oxford: Oxford University Press.

Setiya, Kieran. 2007. *Reasons without Rationalism*. Princeton, NJ: Princeton University Press

Sinhababu, Neil. 2017. *Humean Nature: How Desire Explains Action, Thought and Feeling*. Oxford: Oxford University Press.

Sinhababu, Neil. 2009. "The Humean Theory of Motivation Reformulated and Defended." *Philosophical Review* 118 (4): 465—500. doi:10.1215/00318108-2009-015

Smith, Michael. 2013. "A Constitutivist Theory of Reasons: Its Promise and Parts." *LEAP: Law, Ethics, and Philosophy* 1: 9—30. https://raco.cat/index.php/LEAP/article/view/294565

Smith, Michael. 2009. "The Explanatory Role of Being Rational." In *Reasons for Action*, edited by David Sobel and Steven Wall, 58—80. New York: Cambridge University Press.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.

Velleman, J. David. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Ylikoski, Petri and Kuorikoski, Jaakko. 2010. "Dissecting Explanatory Power." *Philosophical Studies* 148 (2): 201—219. doi:10.1007/s11098-008-9324-z

---

[1] This is, however, not to say that simplicity does not support *HTM* at all. *HTM* clearly involves fewer entities than a theory of action that takes all human actions to consist of human bodily movements caused by supernatural divine alien lizardmen from beyond space and time. That theory is, among many other things, too obviously unparsimonious to take seriously.

[2] It is, in fact, not entirely clear to what extent Sinhababu might endorse the version of the argument I shall discuss. See section 2 in general and endnote 10 in particular below for more discussion.

[3] Strictly speaking, nothing forbids Humeans or other causal theorists from taking what is caused to be something else than an event, such as a process. But most Humeans take causation to be a relation between events, so I shall stick with this assumption.

[4] Deviant causal chains appear when the world obstructs an intended way to cause an action but then accidentally makes one reach one's intended aim anyway. See Davidson (1973) for the *locus classicus* of the discussion of the issue.

[5] It is unclear how these states should be characterised in more detail, however. A common view is that beliefs should be characterised in terms of their aiming to fit the world, whereas desires aim to make the world fit them (cf. Smith 1994, ch. 4). But whether that characterisation is correct or not, the full story about the distinction remains controversial.

[6] Sinhababu: 'Many think that causal explanations of human motivation require a richer ontology of psychological states and processes than the Humean Theory allows. I'll use the properties of desire to explain the phenomena they say Humeans can't explain, and then use Occam's razor to cut away their additional entities' (Sinhababu, 2017, p. 12). The idea is that *HTM* and the desires involved in it are highly explanatorily power, making it able to explain the same things as less parsimonious theories, but then simplicity – or Occam's razor – supports going with *HTM* rather than with other explanations. This means that it serves like a tiebreaker between them.

[7] This is true even if one were to treat simplicity as a feature of explanatory power, as one might want to do if one were to take explanatory power to be a master theoretical virtue which incorporates all others. That would generate a terminological difference from me, but not a substantive dispute, for one could treat all other dimensions of explanatory power as fixed but break out simplicity and compare *HTM* with other views in that dimension. I want to thank an anonymous reviewer for bringing up this potential worry.

[8] We can therefore, for example, safely rule out discussing views like the one with the lizardmen from endnote 1.

[9] For some belief-based theories, see Alvarez (2010) and Gregory (2021). For some besire-based ones, see Velleman (2000) – on my reading – and Setiya (2007). I shall, however, not defend the plausibility of causal theories of action here. That would be a book-length endeavour.

[10] Individual philosophers could still think that various theories fail to be explanatorily powerful in the final analysis – Sinhababu might, for example, think belief- or besire-based theories lack explanatory power to the extent that they are not interesting to compare with *HTM* when it comes to their simplicity. It is certainly logically possible to think that these theories are simpler than *HTM* but also uninteresting: perhaps he thinks they do not give desires enough of an explanatory role to explain important parts of our psychology. If so, the jury is out on what implications my criticism of the argument from simplicity as I have formulated it will have for Sinhababu, for if other theories are too far off the mark when it comes to explanatory power, he might not want to treat them as relevant competing theories to *HTM*. Whether he does – or whether they are – will depend on things like how explanatorily powerful they are and how steadfast or conciliatory we should be in the face of philosophical disagreement.

However, none of this is an issue for the main argument of the paper. As stated in the introduction, its aim is to argue that it is inadmissible to argue that *HTM* gains any significant kind of support from simplicity when it is compared with other relevant theories. This is a point about the general dialectical move of defending *HTM* by appealing to simplicity, regardless of what one might think about which theories count as relevant. My discussion focuses on the general argument from simplicity in the dialectic, not on discussing individual philosophers' sets of views.

[11] There are also other, less significant, forms of simplicity in the literature. *Syntactic simplicity* is concerned with the number and complexity of the hypotheses in a theory, and *ideological simplicity* is concerned with the ideas expressible by the terminology of the theory. The argument for simplicity for *HTM* does not plausibly involve them: it is concerned with entities in the world.

[12] Admittedly, a simple theory which explains some *explananda* might be said to impose a kind of explanatory unity in virtue of explaining them by using few entities. One gives the same explanation of the *explananda*, after all, by appealing to the same entities. But whether or not there is such a correlation between simplicity and unity, the virtues are not the same, as per the water example.

[13] For some theories featuring beliefs in this way, see Alvarez (2010) and Gregory (2021).

[14] For some theories featuring besires, see Velleman (2000) – on my reading – and Setiya (2007).

[15] I want to thank an anonymous reviewer for this journal for raising this possibility.

[16] To be clear, the line of reasoning here generalises to other accounts that attempt to unify desires than Smith's, such as Schroeder (2004)'s attempt to identify desires with structures in the reward system in the brain – in Schroeder's case, representations conducive to reinforcement learning. It is, admittedly, unclear whether an account like Schroeder's fits *HTM* in the first place as he does not hold that desires motivate intrinsically – but it can nevertheless be used to illustrate the generalisation point.

The reason it can do that is that what is at issue is the fact that the account takes desires to have lots of properties. This question does not turn on what kind of entities they are: structures in the brain, motivational dispositions, natural kinds, psychological kinds which are not natural kinds, or something else. For example, Schroeder-style desires have the properties of 'tending to cause pleasure (in normal circumstances)', 'tending to cause motivation (in normal circumstances)', 'constituting things as rewards', and similar. These properties ramp up the complexity of the desires quickly, and that worry generalises – any approach to desire is likely to become very complex in the sense that it involves many properties, regardless of how one would to answer questions about the ontology of the desires.

[17] See also Hempel (1961) for an earlier version of the view – but note that Smith presents other versions of *HTM* elsewhere, such as in his (1994).

[18] There is also a general risk: several theories could be equally virtuous in general. But this could be true for any virtue.