

WHAT KIND OF THEORY IS THE HUMEAN THEORY OF MOTIVATION?

Caroline T. Arruda

Abstract

I consider an underappreciated problem for proponents of the Humean theory of motivation (HTM). Namely, it is unclear whether it is to be understood as a largely psychological or largely metaphysical theory. I show that the psychological interpretation of HTM will need to be modified in order to be a tenable view and, as it will turn out, the modifications required render it virtually philosophically empty. I then argue that the largely metaphysical interpretation is the only plausible interpretation of HTM's central claim that desires are necessary conditions for motivation. This interpretation also fits better with the important roles that HTM plays in both moral psychological and metaethical debates.¹

The Humean theory of motivation (HTM) has received an incredible amount of philosophical attention.² One reason for this attention is that HTM has figured importantly in a wide variety of debates, most notably in debates about the nature of moral motivation, in theories of normative reasons, in anti-realist attempts to dispute the plausibility of moral realism and in theories of motivation in the philosophy of action. Given its importance, a great deal of ink has been spilled in the service of evaluating its truth, determining what precisely the main claims of the theory are and how strongly to interpret these claims. Nonetheless, in modern form, the most basic version of the theory states that desires are individually necessary conditions for agents to be motivated to intend to ϕ .

The purpose of this paper is not to rehash these debates; rather, it is to consider a more basic and overlooked aspect of HTM. Namely, it is unclear what *kind of theory it is*. It admits of two

¹ I presented an earlier version of this paper at 68th Annual Mountain-Plains Philosophy Conference, The University of Nevada-Reno. I would like to thank the audience and my commentator for their helpful questions. I would also like to thank an anonymous referee for this journal.

² See footnotes 4–16.

competing interpretations—one *largely metaphysical*, the other *largely psychological*. The upshot of clarifying which of these two interpretations is the right one is that we are in a better position to evaluate the truth of HTM and, thereby, the myriad debates in which it figures.

Let us begin by considering the basic meaning of the two possible interpretations of the kind of theory that HTM is. If HTM is a theory about the psychology of the agent, then its primary target is to explain how agents are psychologically unable or unlikely to form intentions without having various ends that they want to pursue and, thus, they need desires to provide this source of motivation. This type of interpretation understands HTM to specify the basic, defeasible natural laws about human psychology and thereby to explain the causal *likelihoods* that obtain for human beings. A nice exemplar of this type of view is Neil Sinhababu's (2009) defense of the Humean theory of motivation, to which I return in §3.

By contrast, a *largely metaphysical* interpretation of HTM identifies the causally efficacious and necessary antecedents for being motivated to act. These antecedents may be the same attitudinal states that the psychological accounts identify, but their role is different. On the metaphysical interpretation, the focus is these attitudes' or dispositions' *powers* or *capacities* to cause us to be motivated to act. If this enterprise sounds odd, consider that Nagel (1970, pp. 5–6) takes his account in *The Possibility of Altruism* to have a similar aim: "Human motivation possesses features which are susceptible to metaphysical investigation and which carry some kind of necessity."³

This is not to create a false dilemma between these two interpretations. Clearly, a theory can make both types of claims, particularly if the metaphysical claims put constraints on what an agent's psychology can be. But I argue that we ought interpret HTM in *largely metaphysical*, rather than *largely psychological*, terms. As I show, the psychological version of HTM will need to be modified in order to be a tenable view and, as it will turn out, it will need to be so significantly modified that it is rendered virtually philosophically empty. I then argue that the largely metaphysical interpretation is the only a plausible interpretation of HTM's central claim that desires are necessary conditions for motivation. This claim, it might seem, is even less plausible than the psychological interpretation, given worries that it makes motivation

³ See also Mele (2003, p. 112) for a similar gloss on the Humean view in the context of theories of moral motivation.

seem like the upshot of a mechanistic process. I conclude by sketching the beginnings of a coherent metaphysical interpretation of HTM that is decidedly not mechanistic and leaves room for agential authority.

With this said, my aim in this paper is not to defend HTM; rather, my aim is to produce a coherent account of it in order to set the stage for evaluating its truth. The plan for this paper is as follows. In §1, I undertake two tasks. I sketch the basic features of the psychological and metaphysical interpretations of HTM. I then identify what depends on distinguishing between these two competing interpretations. In §2, I discuss the various objections to the basic formulation of the psychological interpretation of HTM and propose a series of modifications in order to make it tenable. But although the version of the psychological interpretation of HTM with which we conclude is tenable, it turns out to be philosophically uninteresting, as I show in §3. I then argue that while the psychological interpretation may be an empirically interesting, testable theory about human motivation, it cannot do the work that HTM typically does for philosophical theories in metaethics and in theories of moral motivation. For this reason, I suggest that we should favor the metaphysical interpretation. I conclude, in §4, by sketching some possible ways of fleshing out the metaphysical interpretation.

1. The Psychological and Metaphysical Interpretations of HTM: What Is At Stake

Although HTM is used in myriad philosophical debates, there is some lack of clarity about what precisely a proponent of HTM defends. My suggestion is that the *most basic* version of HTM, represented by (1) below, is the one that admits of the two competing interpretations outlined above. To highlight the version of HTM under consideration here, let us consider the variety of ways HTM has been understood:

1. HTM specifies the motivational conditions under which agents form intentions to act. On this view, for an agent to intend to ϕ in light of some consideration a , she has an independent desire to ϕ or for the end that ϕ -ing brings about.⁴ This view can be taken to mean that desires are

⁴ See Smith (1994, pp. 92–3), who notes that this account is implicit in Davidson's (2001/1963) causal theory of action.

either individually necessary conditions for motivation or sufficient for motivation.

2. HTM identifies the non-cognitive attitudes that are necessary conditions for agents to take considerations to be reasons for action, where the reason in question is a motivating (or explanatory) reason.⁵
3. HTM identifies the conditions under which considerations constitute normative reasons, where those conditions comprise agents' non-cognitive attitudes.⁶
4. HTM identifies practical reason's scope and function, where only an agent's non-cognitive states (e.g., desires) can set the ends that she aims to pursue.^{7,8}

This list suggests two contexts within which HTM is used.⁹ First, HTM may be used in theories of explanatory and normative reasons, as exemplified in (2)–(4). Second, one might also take HTM to explain the necessary or sufficient conditions for being motivated to intend an action, as exemplified by (1). It is this version of HTM that, I will show, admits of two competing interpretations and that is the focus of this paper.¹⁰

Views about (1) do not speak with one voice about its meaning,¹¹ its aims¹² or the broader metaethical,¹³ moral psychological¹⁴ and action theoretic¹⁵ questions for which it is relevant. So let us take

⁵ See Smith (1994, p. 125) as an example. For a contrary view, see Miller (2008).

⁶ Williams (1981, pp. 102–05), Schroeder (2007, esp. chpt. 7), Sinhababu (2009), among others, are examples of this view. See Collins (1997), Dancy (1993; 2002, esp. pp. 85–104), Finlay (2009) and Parfit (1997) for challenges.

⁷ Some, such as Williams (1981), combine (3) and (4).

⁸ See Williams (1981) and Schroeder (2007) as examples of this view. See Nagel (1970, p. 30), Finlay (2006) and Korsgaard (1986, esp. pp. 15–17) for responses.

⁹ Compare with Sinhababu (2009, p. 465), Hubin (1999, pp. 32–3), Schroeder (2007, esp. pp. 2–3) and van Roojen (1995).

¹⁰ Notice that (1) provides a foundation for any of (2)–(4).

¹¹ For example, do desires independently, directly cause motivation or are they sometimes responsive to rational, prudential or moral considerations? Proponents of HTM (1) argue that HTM makes the former claim. Yet Nagel (1970), Smith (1994, chpt. 5), Scanlon (1998, pp. 95–100) and Darwall (1983, pp. 92–8) each shows that we can accept a version of HTM (1) in ways consistent with the latter. I discuss these claims at length in §3.

¹² See Pettit (1987), Schueler (2009), Setiya (2004), Shafer (2013) and Smith (1994) on this issue.

¹³ Blackburn (1984; 1995; 2001); Korsgaard (1986); Lillehammer (2002); Mackie (1977); McDowell (1995); Railton (1986); Shafer-Landau (2003); Smith (2004); Zangwill (2009).

¹⁴ Björklund et al. (2012); Brink (1986); Shafer-Landau (1998; 2000); Smith (2004).

¹⁵ As an example, see Mele's (2003, pp. 20–2) account of the "positive motivational base" to which desires contribute. Also see Schapiro (2001).

the weakest version of HTM—on which most who defend HTM would agree—as our focus. Namely, the claim that desires are individually necessary conditions or a sufficient condition for motivation.

The question now is how best to understand the nature of the conditions that desires provide. There are two plausible interpretations—one *largely psychological* and the other *largely metaphysical*. Let us turn first to the most basic version of the psychological interpretation:

The Psychological HTM: Desires are individually necessary or sufficient for motivation. They serve the function of exciting agents who possess them to intend to undertake a suitably related action in that they identify the action or its end as something toward which the agents in question have a pro attitude or pro disposition. These desires can be overridden or outweighed by other desires.

The above gloss captures the thought behind the Humean intuition that desires are, as Jonathan Dancy (1993, 13) puts it, “ert” while cognitive states, such as beliefs, are “inert.” That is, desires possess the capacity to motivate us to act, while beliefs and other cognitive states are motivationally neutral. Note that the gloss is ecumenical with regard to how desires exert their motivational force, given the neutral language of ‘pro attitude’/‘pro disposition.’ The reason for this is that I want to capture a variety of possible ways that desires can be understood to do so such that no objection to *The Psychological HTM* that I raise can be attributed to the understanding of desires at play rather than to the interpretation of HTM under consideration.¹⁶ Still, one might contend that what this view means depends on how we understand ‘desire’ here. I think that whether one understands desires as propositional attitudes or as dispositions results in the same *explanatory problems* for the psychological interpretation, as I discuss below.

¹⁶ Naturally, there are additional questions that one might ask about which view of desires’ abilities to do so is correct, as Schiffer (1976), Schueler (1995) and others have noted. Moreover, one might inquire further into how to understand desires—as propositional attitudes or dispositional states (Searle 1983; Schueler 1995; Smith 1994), as background or foreground states (Pettit and Smith 1990; Sinhababu 2009), among others—and whether ‘desire’ is one homogenous category of attitudinal states (Schueler 1995; Schapiro 2001). Although the answers to these questions may very well alter our understanding of HTM, they are secondary to the question of whether proponents of HTM are making a psychological or a metaphysical claim.

For the moment, let us set aside *The Psychological HTM* and turn to the most basic version of *The Metaphysical HTM*.

The Metaphysical HTM: If an agent is motivated to ϕ , then she has a desire to ϕ or a desire for the end that ϕ -ing brings about. This desire is causally efficacious, either as a necessary condition or a sufficient condition, in her coming to be motivated to ϕ .

The *Metaphysical HTM* understands motivation to be a state that is causally brought about by other states or attitudes. This is a largely metaphysical view insofar as it proposes an explanation of how certain attitudinal states or dispositions have the relevant *powers* or *capacities* to bring about the formation of other attitudinal states or dispositions—in this case, the intention to act.¹⁷ It follows, on this view, that if an agent is motivated to ϕ , then she must have had some desire that was causally relevant in bringing about this state.

But what is the nature of this causal efficacy? Put differently, *in virtue of which feature(s)* do desires cause us to form the intention to act? Is it, as Smith (1994) proposes, that they represent goals? Or perhaps it is simply their world-to-mind direction of fit that enables us to aim to make the world match their content. Answering this question is outside of the purview of this paper, given that I do not aim to show that any one particular version of *The Metaphysical HTM* is the right one; rather, I need only show that one of the two competing versions of HTM—the psychological or the metaphysical version—is the right type of interpretation of HTM's main claim. Once this issue is settled and it turns out that *The Metaphysical HTM* is the preferable version, we must then consider the specific features that are plausibly understood to be desires' causally efficacious characteristics.

2. How to Make the Psychological Interpretation of HTM Tenable

Since most defenses of HTM understand it in the psychological terms discussed above, let us turn first to evaluating whether this

¹⁷ My suggestion is in order for HTM to be plausible in any sense of the term, it must be understood as a metaphysical thesis about motivation. The specifics of such a theory are outside of the purview of this paper, and one might imagine that there are different, competing versions of the *largely metaphysical HTM*.

is the most plausible way to understand the kind of theory that it is. As it will turn out, the view will require significant modifications to be tenable, and these modifications will render it philosophically less important than it appears to be, particularly in central debates in metaethics.

Given that the view is intended to identify the basic psychological laws concerning motivation, it is clear that *The Psychological HTM* must specify its target population. The first modification required concerns the kinds of agents to whom this theory of motivation applies—namely, human (or non-human) agents that have the type of psychological make-up for which desires and other non-cognitive states are salient attitudes. Furthermore, we must add that these human agents are cognitively developed such that they can “put together,” as Schueler (2008: 106–108) describes it, the desire with the relevant beliefs about how best to satisfy it. The idea is that agents must have suitably developed cognitive abilities such that, in the face of a motivating desire, they are able to draw on the relevant beliefs that inform their pursuit of this desire. In this regard, HTM depends on the assumption that agents have the capacity for instrumental reasoning about the desires that motivate them. But this suggests that we ought not simply add “human agents,” for what we actually mean here is “cognitively developed human agents.”

This modification saves the view from being too widely applicable to any and all agents across very different possible worlds. In addition, it delimits the scope of HTM's account of the psychological laws about human motivation to human agents with developed cognitive capacities.

Yet additional problems would emerge if we were only to account for cognitively developed human agents in our modifications. *The Psychological HTM* only makes sense for those fully developed human agents with a *normal psychological make-up*, where ‘normal’ means that they tend to respond appropriately to their various attitudinal states. The psychopath does not, as Brink (1986) and Shafer-Landau (1998) each suggests, respond appropriately to her attitudinal states, most notably her moral judgments. Similarly, the agent who suffers from perennial *accidie*, or motivational apathy, will also fail to respond appropriately to what would normally be, on *The Psychological HTM*, motivating states. Since these cases and others like them should not be the testing ground for the truth of *The Psychological HTM*, we ought modify it further to exclude these cases as well:

The Modified Psychological HTM: Desires are individually necessary or sufficient for motivation. They serve the function of exciting **cognitively developed, psychologically normal human agents** who possess them to intend to undertake a suitably related action in that they identify the action or its end as something toward which the agents in question have a pro attitude or pro disposition. These desires can be overridden or outweighed by other desires.

These modifications fix the view so as to avoid the counterexamples above. But they point the way toward the need for a final modification. *The Modified Psychological HTM's* explanatory scope is limited to such a degree that it maps only motivational *regularities* among psychologically normal, cognitively developed human agents. After all, Jane may one day fall prey to *accidie*, while most other days she tends to be motivated by her desires. But if it maps only motivational regularities, it should be understood as an empirical view about human behavioral tendencies rather than a view about *the necessary conditions for human motivation*. Thus, *The Modified Psychological HTM* must be modified one last time:

The Empirically Responsive, Modified Psychological HTM: Desires **tend to be** sufficient for motivation. They **typically** serve the function of exciting cognitively developed, psychologically normal human agents who possess them to intend to undertake a suitably related action in that they identify the action or its end as something toward which the agents in question have a pro attitude or pro disposition. These desires can be overridden or outweighed by other desires.

Notice that, in order to make these final adjustments, the main claim associated with HTM—that desires are individually necessary or sufficient for motivation—had to be modified as well. Namely, if HTM is just a theory about psychological regularities, it can no longer be interpreted as specifying *necessary* conditions for motivations; rather, it must be revised such that its main claim is now a claim only about desires' sufficiency for motivation.

Here one might claim that I have gotten the relationship between desires and motivation, on the *psychological interpretation* of HTM, backwards. While my reading seems to entail that HTM must be able to show that desires *necessarily* motivate, what the proponent of the psychological interpretation of HTM actually argues is that if one is motivated to ϕ , one has a desire to ϕ .

The latter formulation is the main claim of the *psychological interpretation* of HTM, and nothing about what I have argued thus far denies this point. Nor have I argued that HTM must be understood as arguing that desires necessarily motivate. But what I have highlighted is that the *psychological interpretation, in explaining how desires play the role of exciting motivation*, ends up specifying a psychological connection between having a desire and being motivated to ϕ . This psychological connection is taken to be a claim about desires' ability (either alone or in concert with other necessary conditions for motivation) to excite agents to be motivated to ϕ . So for the *psychological interpretation* of HTM to be plausible, it must be able to specify the non-contingent, psychological laws (or psychological necessities) that obtain for human agents. Given what I have argued thus far, however, it looks like all the psychological HTM can specify are behavioral tendencies rather than psychological laws.

3. Why The Psychological Interpretation Fails, or Reasons to Favor the Metaphysical Interpretation

We have thus succeeded in making HTM, under a *psychological interpretation*, tenable. The question, however, is whether the version with which we concluded— *The Empirically Responsive, Modified Psychological HTM*—is philosophically useful. In short, does it answer to the needs and intuitions that have made HTM seem plausible?

3.1. HTM's Metaethical Promise and Why the Psychological Interpretation Does Not Deliver

On face value, one might think that *The Empirically Responsive, Modified Psychological HTM* is philosophically less important than traditionally thought insofar as it no longer provides an account of the necessary conditions for motivation. One might make this claim on the grounds that this version of the view accomplishes the following: (1) it maps psychological regularities rather than psychological requirements or laws; (2) it is radically defeasible; (3) it covers a set of contingent relationships between desires and the state of being motivated; and (4) its explanatory scope is limited to psychologically normal, cognitively developed human agents. Still, it is an interesting, testable empirical hypothesis, which, one might claim, has philosophical relevance for theories

of human motivation even if it is no longer an account of the necessary conditions for motivation.

But the fact that the *psychological interpretation* of HTM, in its most plausible rendering, is no longer able to account for the necessary conditions for motivation points to additional reasons to reject it and instead favor the *metaphysical interpretation*. To see why, let us consider the import that HTM is often taken to have in philosophical debates. If it turns out that the version of the *psychological interpretation* of HTM with which we concluded cannot do the work that HTM is often taken to do, then we have reason to favor the *metaphysical interpretation* (assuming that it can indeed do this work). Let us consider two roles that HTM often plays:

- (a) Its role in defenses of non-cognitivist anti-realist views in metaethics¹⁸ and
- (b) Its role in theories of moral motivation¹⁹

HTM figures in a variety of non-cognitivist, anti-realist views in metaethics. Although its role varies, it is generally marshaled as evidence that naturalist and non-naturalist varieties of moral realism are false. In broad brushstrokes, the argument goes as follows: moral judgments ought to be able to motivate us to act and the only states, on HTM, that can motivate are non-cognitive. Thus most varieties of moral realism are false, given that, if they were true, it would be impossible for moral judgments, which comprise beliefs about moral facts of the matter, to motivate (Shafer-Landau 2003; Smith 2004).

Accounts of moral motivation also draw on HTM. This is most common in judgment internalist views, proponents of which argue that there is a necessary (albeit defeasible) connection between moral judgment and moral motivation (Björklund et al. 2012; Bromwich 2010; Garrard and McNaughton 1998; Joyce 2002; Shafer-Landau 1998; Smith 2004; van Roojen 2002; Zangwill 2008). Typically, judgment internalist accounts are paired with HTM to refute moral realism in the ways described above.²⁰ They may also draw on HTM to challenge the judgment externalist,

¹⁸ For example, Blackburn (1984, 1995, 2001).

¹⁹ For a discussion, see Smith (1994, p. 61), Shafer-Landau (1998, p. 353; 2003, p. 121).

²⁰ Most judgment externalists, who argue that there is merely a contingent connection between moral judgment and moral motivation, challenge HTM, but they need not do so (Brink 1986; Shafer-Landau 2003).

who argues that there is a contingent rather than necessary connection between moral judgment and moral motivation.

While HTM's respective roles in these debates do not show which interpretation (*psychological* or *metaphysical*) of HTM is correct, they underscore the intuitions about motivation that HTM is supposed to capture. The relevant intuition here is that we are necessarily, rather than contingently, the kinds of agents for whom desires play a role distinct from that of cognitive states in the context of our exercise of our agential capacities.

If the test for whether to favor the *psychological* or the *metaphysical interpretations* of HTM is whether the *psychological interpretation*, as we modified it above, is useful in metaethical debates, we should consider whether the *psychological interpretation's* empirical import has any implications for the aforementioned metaethical debates. If the *psychological interpretation* can still serve the aforementioned or similar metaethical purposes, then we lack the ground to favor the *metaphysical* over the *psychological interpretation*.

As a psychological theory, HTM does not meet this burden. After the successive modifications required to make it tenable, it became a largely empirical claim about contingent behavioral regularities among cognitively developed, psychologically normal human agents. First, this type of view cannot play the role either in refutations of moral realism or in views about moral motivation discussed above. By contrast, a *metaphysical interpretation* of HTM can play such a role because it identifies the causally efficacious and, thereby, necessary role that some attitudinal states play in bringing about motivation. Second, if we think that the *psychological interpretation* entails substantive conclusions about human motivational regularities, then it is most likely to entail substantive conclusions about the plausibility of, say, a theory of blameworthiness. But it does not tell us anything directly substantive about whether moral realism or cognitivism is true—the very purposes to which, as discussed above in (a) and (b), HTM is often put. Third, the *psychological interpretation* of HTM sketches the boundaries of typical human agency. In this sense, one might convincingly argue that it identifies the upper-bounds for what we mean when we say that agents are *capable* of being morally motivated to act in light of their moral judgments. At best, then, it establishes that we are more easily morally motivated under some conditions rather than others. But this this does not establish that cognitivism is false nor does it independently establish that non-cognitivism is true.

Here a proponent of the *psychological interpretation* may respond by suggesting that one can grant that HTM is an empirical theory and yet retain (some of) its metaethical import. This response, if correct, would show that the argument that I have provided thus far does not definitively rule out the *psychological interpretation* of HTM in favor of the *metaphysical interpretation*. To determine whether this response is successful, let us consider a version of what I have been calling the *psychological interpretation* of HTM. As I noted earlier, Sinhababu (2009) defends such a view, although he does not use this terminology to describe his argument. He describes his view as follows: "Some Humeans regard the Humean theory as a conceptual truth about action. On their view, all possible agents have Humean psychologies. I regard the Humean theory merely as a truth about actual actions performed by human beings and about all the actions that humans are psychologically capable of performing" (467). He then goes on to show that, even given its largely empirical focus, this version of HTM is nonetheless relevant for metaethical debates in the following way: if normal human psychology is such that we are not the kind of creatures who are motivated by beliefs or by moral judgments that are truth-apt propositions, then it will be difficult to show how cognitivism or cognitivist theories of moral motivation are true (467–8).

As an aside, I think that aspects of Sinhababu's (2009: 471) view place it closer to the *metaphysical interpretation* than to the final version of the *psychological interpretation* (*The Empirically Responsive, Modified Psychological HTM*). One need only look to his (2009: 417–ff) discussion of the causal powers of desires' "hedonic" features. Although Sinhababu intends his account to apply to human agents in this world rather than all agents across all (or even just some) possible worlds, the view's target is to explain psychological *laws* concerning desires' causal force on human motivational mechanisms. More important, if what I have argued here is correct, *The Empirically Responsive, Modified Psychological HTM* is more contingent than a Humean such as Sinhababu would like. This is because the former specifies highly defeasible regularities in human behavior under typical conditions. Sinhababu's view seeks contingent psychological laws about human psychology, which identify a normal distribution and outlier cases. I think that the *metaphysical interpretation* is much better able to deliver on the latter aims, given the degree of variation for which *The Empirically Responsive, Modified Psychological HTM* allows.

Nonetheless, it is clear that Sinhababu aims to develop a version of HTM that specifies the basic psychological regularities of human motivation. Given this fact, we can treat his view as well as his discussion of his view's import for metaethics as a rough proxy for the *psychological interpretation* of HTM. Sinhababu (2009: 467–8) denies that the *psychological interpretation* of HTM has no metaethical import. In fact, he claims that it has *as much if not more* metaethical significance than the traditional interpretation of HTM:²¹

The puzzle still remains if one regards the Humean theory as a contingent truth concerning human psychology rather than a necessary truth applying to all possible agents. This is because cognitivism and internalism are not mere possibility claims—they are supposed to hold at least of the moral judgments of human beings in the actual world. Cognitivism is supposed to apply at least to the moral discourse of actual human beings. Internalism is usually regarded as a conceptual truth about the connection between moral judgment and action, and thus must be true in the actual world and beyond. So if we accept the Humean theory as a truth about actual human psychology, we cannot maintain the usual versions of both cognitivism and internalism, which apply to the actual world[...]. But if the Humean theory is true about actual human psychology, as I will argue, our moral judgments cannot satisfy both cognitivism and internalism. So even a modally modest version of the Humean theory will have great import for metaethics.

Sinhababu's suggestion is that what I am calling the *psychological interpretation* of HTM has important metaethical implications *even if* we grant that HTM establishes merely contingent human motivational regularities. His reason for making this claim is that a theory of human motivational regularities may present a picture of human psychology that is inconsistent with important metaethical theses about how and why we should be motivated to act morally. This inconsistency points in the direction, Sinhababu claims, of rejecting these metaethical theses.

One such inconsistency can be found between the *psychological interpretation* of HTM and the conjunction of cognitivism and judgment internalism. The cognitivist argues that we can be motivated

²¹ Note that he is responding to Smith's (1994) attempt to make cognitivism, judgment internalism and (a version of) HTM consistent.

by our cognitive states. Recall, from my discussion above, that the judgment internalist's thesis is that if we make a moral judgment, then, on pain of irrationality, we are necessarily motivated by it. Putting these two claims together, we get the claim that moral judgments are beliefs and that they necessarily (on pain of irrationality) motivate. Can we hold these two theses stable if the *psychological interpretation* of HTM is true? The answer, per Sinhababu, is 'no': If the *psychological interpretation* of HTM is true, then it is difficult to explain how moral judgments could plausibly be beliefs or how moral judgments, if they are beliefs, could motivate us to act. If they are beliefs, no *normal* human agent will ever be motivated by her moral judgments.²² Yet we want to show that normal human agents can be motivated by their moral judgments. So if the *psychological interpretation* of HTM is true and we want to show that normal human agents can be motivated by their moral judgments, we ought also reject the conjunction of cognitivism and judgment internalism.

It thus appears that if the *psychological interpretation* of HTM is true, then it still retains metaethical import. Though its import is, in one sense, less far reaching in that it does not purport to describe all possible worlds, it is no less weighty than originally thought.

To respond to Sinhababu's claim that the *psychological interpretation* of HTM has metaethical import, we must answer two questions. First, are motivational cognitivism and judgment internalism *fully undermined* if the *psychological interpretation* of HTM is true? Second, which interpretation of HTM is *better able to discharge* the aims of using HTM in these debates—which is to match up metaethical theories with a fair representation of how *relatively idealized* human agential capacities express themselves in this actual world?

Notice first that Sinhababu's claim sets a different bar for determining the metaethical significance of the *psychological interpretation* of HTM than the ones—claims (a) and (b) above—with which I began this section. Namely, he suggests that if the *psychological interpretation* of HTM is true, we create an empirical (though not a logical) barrier to defending some combination of cognitivism and judgment internalism. With this clarification in place, let us begin with first issue. It is true that cognitivism and judgment internalism are difficult to hold stable in their *traditional forms* if the *psychological interpretation* of HTM is true. But we *can grant* that the conjunction of cognitivism and judgment internalism is true

²² I owe this way of framing the point to an anonymous referee.

and grant the truth of the *psychological interpretation* of HTM by shifting slightly our understanding of what it means for moral judgments (understood as beliefs) to be capable of motivating us.

This slight shift requires us to revisit our understanding of what it means for an attitude to motivate us to act. Traditionally, it is thought—and here the Sinhababu-esque view would agree—that the conjunction of the three views does not hold because HTM requires that desires *directly* bring about motivation.²³ But nothing about the *psychological interpretation* of HTM prevents us from making the following claim: moral judgments both are beliefs and motivate us by virtue of the desires that we form *in light of them*.²⁴ This is the approach that Nagel (1970), Smith (1994: 160–ff), Scanlon (1998: 95–100) and Darwall (1983: 92–8) each takes, albeit for different purposes than I have here. Still, it can be used to marry the *psychological interpretation* of HTM with the conjunction of cognitivism and judgment internalism. What's more, it may be that this is a happy marriage between the three views. This is because the *psychological interpretation* can provide a clear picture of the route by which this desire formation occurs. If this is right, then the *psychological interpretation* of HTM does not have any particular implications in the direction of *undermining* the conjunction of cognitivism and judgment internalism. Rather, it identifies the psychological regularities that provide the context for interpreting these views' central theses.

Here one might worry that this “slight” revision involves more significant changes than it would initially appear. In particular, while these revisions are minor in the case of cognitivism, they are significant in the case of judgment internalism. And since the aim is to show that the *psychological interpretation* of HTM does not undermine the *conjunction* of motivational cognitivism and judgment internalism, we may indeed face a serious problem. The problem is as follows. I have suggested that it is clear how one's beliefs affect the set of desires one has, thereby neutralizing the supposed problem of making the *psychological interpretation* of

²³ This is a summary of the traditional account of HTM, which makes no distinction between what I have called the *metaphysical* and the *psychological interpretations* of HTM. To explain this traditional argument, we must then remain neutral between the two interpretations. So one should take ‘bring about motivation’ in ecumenical terms. It could mean ‘cause us to be motivated,’ ‘psychologically predispose us to be motivated,’ ‘dispose us to be motivated,’ etc.

²⁴ In light of these views, there is debate concerning Williams' (1981) view. This is because he calls desires “independent existences” but also claims that our “subjective motivational sets” can be changed by reason. See Finlay (2009, p. 1) for a discussion.

HTM consistent with cognitivism. But it is less clear how we might insert this picture of motivation into the combination of the *psychological interpretation* of HTM and judgment internalism. This is because it is not easy to explain how one forms desires in light of a specific type of belief (i.e., a moral judgment). More important, it is even more difficult to show how this happens in a reliable fashion that is governed by rationality (per the consequent in the judgment internalist's thesis that one is *necessarily* motivated on pain of irrationality).

I think that although this problem seems intractable, it is actually rather easy to solve. In light of what I described above as a slight shift in our understanding of motivation, we can grant that agents count as being motivated by beliefs when those beliefs shape the non-cognitive attitudes that the agent has. Moral judgments are one such example of where this influence *may* happen. But note that we need to show how it will, on pain of irrationality, happen in order to show that the *psychological interpretation* of HTM does not undermine judgment internalism. Given how we have framed the *psychological interpretation* of HTM, I think that it actually helps us to understand the psychological regularities under which this influence on non-cognitive states occurs. This is because psychological regularities are not immune to the influence of rationality, as Michael Smith (1994: 160–1) proposes. We can alter or develop nascent psychological regularities in response to what we believe to be rational. For example, I can recognize that I tend to favour friends who respond promptly to text messages and dislike those who take a long time to respond. I can recognize, in this instance, that I should work to change the set of nascent preferences or desires that I have on the grounds that they are irrational. Similarly, the judgment internalist can grant that *typically* moral judgments (as beliefs) have an effect on the sets of desires that we have. This is because the process of desire-formation is not wholly irrational nor is it immune to rational correction. In this way, the *psychological interpretation* of HTM is far from undermining the judgment internalist thesis; it may actually help to spell out precisely how moral judgments come to motivate us in reliable, regular ways.²⁵

Though this route to answering the first question sounds promising, it is worth noting that Sinhababu (2009: 465) blocks it by

²⁵ This is precisely the route that Smith (1994, pp. 160–1) takes to make a version of HTM, judgment internalism and cognitivism consistent with one another.

adding the following principle to the standard *psychological interpretation* of HTM: 'Desires can be changed as the conclusion of reasoning only if a desire is among the premises of the reasoning.' With this principle, one can no longer appeal to the idea that desires may be responsive to and changed in light of beliefs. And if this is the case, then we are unable to make the *psychological interpretation* of HTM consistent with central theses in metaethics such as judgment internalism or cognitivism (and, thereby, some forms of moral realism). What's more, if we cannot make these claims consistent, then it would seem that, *contra* my claim in this paper, the *psychological interpretation* of HTM does indeed have important consequences for metaethical debates.

While it is true that Sinhababu has technically blocked my answer to question one and thereby potentially derailed my thesis that the *psychological HTM* presents no serious metaethical problems, one must ask whether a standard version of the *psychological interpretation* of HTM, as I have described it above, must grant his added principle.²⁶ The answer to this question is clearly 'no.' Nothing about the standard *psychological interpretation* requires this principle nor do its central claims entail it. One might further note that Sinhababu's addition of such a claim actually vindicates, rather than challenges, my thesis that HTM, to deliver serious metaethical results, must appeal to something like the *metaphysical interpretation* or to a very different version of the *psychological interpretation* than is standardly supposed.

So in answer to the first question above, it is not the case that the conjunction of cognitivism and judgment internalism are *fully* undermined if a standard version of the *psychological interpretation* of HTM is true. It may even be the case that the *psychological interpretation* of HTM aids in explaining the conjunction of these two views, given that it appeals to a developed picture of normal human psychology. What, then, of the second question: which interpretation of HTM is better able to discharge the aims of using HTM in these debates—namely, to match up metaethical theories with a fair representation of how *relatively idealized* human agential capacities express themselves in this actual world?

I still think that the answer to this question is that the *metaphysical interpretation*, all things considered, does a better job than the *psychological HTM* on this front. Broadly speaking, the purpose of using HTM in metaethical debates is to have a relatively idealized picture

²⁶ I owe this point to an anonymous referee.

of human motivation set constraints on the picture that we paint of the relationship between the agent, moral judgments or beliefs, and actions.²⁷ HTM is typically used to roundly refute cognitivism or, by proxy through theories of moral motivation, moral realism. But as I have shown above, if one had this aim, then the *metaphysical interpretation* presents a more promising route. Even if we understand the *psychological interpretation* of HTM, as Sinhababu does, as presenting a picture of natural (defeasible) laws concerning human psychology, we do not have enough leverage from this picture to rule in or to rule out any particular ways of understanding this relationship. It may set some outer bounds on how we understand what human beings tend to prefer or what tends to activate their motivational capacities. But *any theory of human psychology*, Humean or not, will accomplish this task. So the *psychological interpretation* of HTM does not, *contra* Sinhababu, have any special metaethical significance. If it has any, it is no more significant than any other competing theory of human psychology. And if we want HTM to serve in this limited capacity, then we will need to determine how the *psychological interpretation* of HTM fares against competing, empirically informed theories of human psychology.

With that said, we can grant that the *psychological interpretation* may have more clear implications (albeit still limited) for the local debate between judgment internalists and externalists. This import is an artifact of this debate's reliance on human motivational variation to adjudicate the disagreement. This tendency is illustrated by Brink's (1986) case of the psychopath to defend judgment externalism and Shafer-Landau's (1998) variations on this case to further the case for this view of moral motivation.

3.2. *Does it Matter Whether the Psychological Interpretation Can Deliver on HTM's Metaethical Promise?*

Now one might say here—rightly—that determining the best interpretation of HTM does not depend on our interest in using it in other arguments. So why not simply take HTM as a largely empirical, contingent, radically defeasible account of our psychological tendencies, *regardless of whether it has any metaethical import*? Sure, this will take the rug out from under the metaethical aims (a)-(b) above, but this need not be the concern for someone who wants to defend HTM simply as a theory of motivation.

²⁷ Again, one might interpret Smith (1994) as presenting something akin to this type of view.

This route sounds promising. The problem, however, is that most defenses of HTM make the following *type of claim*:

Desires (and other non-cognitive states) have an essential, non-contingent feature that necessarily enables them to motivate us to act. Beliefs (and other cognitive states) lack this feature.

This suggests that any philosophically relevant version of HTM makes a claim about desires' *necessary* features in motivating us to act such that they serve as individually necessary conditions for or sufficient conditions for motivation. We can see this more clearly by looking at the often proposed features. They include: desires' direction of fit (Anscombe 1957; Searle 1983; Smith 1994), their ability to set ends or goals (Smith 1988; 1994), their "ert" quality (Dancy 2002: 81–2), the non-truth aptness of non-cognitive states, the dispositional characteristic that we retain them even when the world does not match up with their content (Smith 2004), among others. Of course, there are many problems with the above types of claims, most notably in the claim that desires' have a world-to-mind direction of fit (Copp and Sobel 2001; Clayton Coleman 2008; Platts 1997). For my purposes, these claims reveal what is often taken to be important about HTM. If taken on face value, HTM cannot mean *only* that beings with the normal psychology of a cognitively developed human being will be motivated by states of this type; rather, the only reasonable way to interpret those who use HTM in the context of these accounts is as defending a claim about those attitudinal states' *causal powers*. That is, they do not want to make a contingent claim about desires; they want to make a claim about the necessary features that desires have such that they are necessary conditions for motivation. The only way to produce such an account is to understand HTM as largely metaphysical rather than psychological in nature. While this conclusion does not entail that the *psychological interpretation* might not have empirical import for philosophical theories, it does suggest that insofar as we want to provide an account of the necessary conditions for motivation, the *metaphysical interpretation* is better suited to meet this need.

4. Conclusion

Nothing about what I have argued here shows that HTM is true or false; rather, I have shown only that, if HTM is true, it is a metaphysical theory rather than a psychological one. But considering

Pettit and Price's (1989) claim that adding desires as necessary causal antecedents for motivation adds a redundant explanatory element that amounts to saying "for x to cause y , the suitable conditions must obtain," it may be difficult to show how desires are causally efficacious in the way that the *metaphysical interpretation* proposes. One route is to argue that desires are causally efficacious in bringing about motivation when they are, as Schapiro (2009) suggests, not merely reflective of our "inner animal" but also reflective of our "outer person" and thus our agential authority. Defending this claim is outside the scope of this paper, but I have cleared the deck of the unpromising versions of HTM such that we can better evaluate those left behind.

The University of Texas at El Paso
El Paso, TX
United States
ctarruda@utep.edu

References

- Anscombe, G.E.M. (1957). *Intention* (Ithaca, NY: Cornell University Press).
- Björklund, F., G. Björnsson, et. al. (2012). 'Recent Work on Motivational Internalism.' *Analysis* 72 (1): 124–137.
- Blackburn, S. (1984). *Spreading the Word* (Oxford: Oxford University Press).
- (1995). 'Practical Tortoise Raising.' *Mind* 104 (416): 695–711.
- (2001). *Ruling Passions: A Theory of Practical Reasoning* (Oxford: Oxford University Press).
- Bratman, M. (2007). 'A Desire of One's Own,' In *Structures of Agency: Essays* (Oxford: Oxford University Press).
- Brink, D. O. (1986). 'Externalist Moral Realism.' *Southern Journal of Philosophy* 24: 23–41.
- Bromwich, D. (2010). 'Clearing Conceptual Space for Cognitivist Motivational Internalism.' *Philosophical Studies* 148 (3): 343–67.
- Clayton Coleman, M. (2008). 'Directions of Fit and the Humean Theory of Motivation.' *Australasian Journal of Philosophy* 86 (1): 127–39.
- Collins, A. W. (1997). 'The Psychological Reality of Reasons.' *Ratio* 2: 108–23.
- Copp, D. & D. Sobel. (2001). 'Against Direction of Fit Accounts of Belief and Desire.' *Analysis* 61 (1): 44–53.
- Dancy, J. (1993). *Moral Reasons* (Oxford: Blackwell Publishers).
- (2002). *Practical Reality* (Oxford: Oxford University Press).
- Darwall, S. (1983). *Impartial Reason*. (Ithaca, NY: Cornell University Press).
- Davidson, D. 2001a [1963]. 'Actions, Reasons, and Causes.' In *Essays on Actions and Events*, 3–19 (Oxford: Clarendon Press).
- Finlay, S. (2009). 'The Obscurity of Internal Reasons.' *Philosophers' Imprint* 9 (7): 1–22.
- (2006). 'The Reasons that Matter.' *Australasian Journal of Philosophy* 84 (1): 1–20.
- Hubin, D. (1999). 'What Is So Special About Humeanism?' *Noûs* 33 (1): 30–45.
- Hume, D. (2000 [1739/40]). *A Treatise of Human Nature*. Ed. D.F. Norton & M. J. Norton. (Oxford: Oxford University Press).
- Joyce, R. (2002). 'Expressivism and Motivation Internalism.' *Analysis* 62 (4): 336–44.

- Korsgaard, C. (1986). 'Skepticism about Practical Reason.' *The Journal of Philosophy* 83 (1): 5–25.
- Lillehammer, H. (2000). 'The Doctrine of Internal Reasons.' *The Journal of Value Inquiry* 34: 507–16.
- Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong* (London: Penguin Press).
- McDowell, J. (1995). 'Might there be External Reasons?' In *World, Mind and Ethics: Essays on the Philosophy of Bernard Williams*, eds. J.E.J. Altham & R. Harrison (Cambridge: Cambridge University Press).
- Mele, A. (2003). *Motivation and Agency* (Oxford: Oxford University Press).
- Miller, C. (2013). 'Identifying with Our Desires.' *Theoria* 79: 127–154.
- (2008). 'Motivation in Agents.' *Noûs* 42 (2): 222–266.
- Nagel, T. (1970). *The Possibility of Altruism* (Oxford: Clarendon Press).
- Parfit, D. (1997). 'Reasons and Motivation, Part I.' *Proceedings of the Aristotelian Society*, Supplementary Vol. 71: 99–130.
- Pettit, P. (1987). 'Humeans, Anti-Humeans, and Motivation.' *Mind* 96 (384): 530–533.
- Pettit, P. & H. Price. (1989). 'Bare Functional Desire.' *Analysis* 49 (4): 162–69.
- Pettit, P. & M. Smith. (1990). 'Backgrounding Desire.' *The Philosophical Review* 99 (4): 565–92.
- Platts, M. (1997 [1979]). *Ways of Meaning*, 2nd Ed (Cambridge, MA: MIT Press).
- Railton, P. (1986). 'Moral Realism.' *The Philosophical Review* 95 (2): 163–207.
- Scanlon, T.M. (1998). *What We Owe to Each Other* (Cambridge, MA: The Belknap Press).
- Schapiro, T. (2001). 'Three Conceptions of Action in Moral Theory.' *Noûs* 35 (1): 93–117.
- (2009). 'The Nature of Inclination.' *Ethics* 119 (2): 229–256.
- (2011). 'Foregrounding Desire: A Defense of Kant's Incorporation Thesis.' *Journal of Ethics* 15: 147–167.
- Schiffer, S. (1976). 'A Paradox of Desire.' *American Philosophical Quarterly* 13 (3): 195–203.
- Schroeder, M. (2007). *Slaves of the Passions* (Oxford: Oxford University Press).
- Schroeder, T. (2004). *Three Faces of Desire* (Oxford: Oxford University Press).
- Schueler, G.F. (1995). *Desire: Its Role in Practical Reasoning and the Explanation of Action* (Cambridge, MA: MIT Press).
- Setiya, K. (2004). 'Against Internalism.' *Noûs* 38 (2): 266–298.
- Shafer, K. (2013). 'Perception and the Rational Fore of Desire.' *The Journal of Philosophy* 110: 258–81.
- Shafer-Landau, R. (1998). 'Moral Judgment and Moral Motivation.' *The Philosophical Quarterly* 48 (192): 353–358.
- (2000). 'A Defense of Motivational Externalism.' *Philosophical Studies* 97(3): 267–291.
- (2003). *Moral Realism: A Defence* (Oxford: Oxford University Press).
- Sinhababu, N. (2009). 'The Humean Theory of Motivation Reformulated and Defended.' *Philosophical Review* 118 (4): 465–500.
- Smith, M. (1994). *The Moral Problem* (Malden: Blackwell Publishing).
- van Roojen, M. (1995). 'Humean Motivation and Humean Rationality.' *Philosophical Studies* 79: 37–57.
- (2002). 'Humean and Anti-Humean Internalism About Moral Judgments.' *Philosophy and Phenomenological Research* 65 (1): 26–49.
- Stampe, D. (1987). 'The Authority of Desire.' *The Philosophical Review* 94 (3): 335–381.
- Wallace, R. J. (2006). 'Normativity, Commitment, and Instrumental Reason,' In *Normativity and the Will* (New York: Oxford University Press).
- Tiberius, V. (2000). 'Humean Heroism: Value Commitments and the Source of Normativity.' *Pacific Philosophical Quarterly* 81: 426–446.
- Williams, B. (1981). 'Internal and External Reasons,' in *Moral Luck: Philosophical Papers 1973–1980* (New York: Cambridge University Press).
- Zangwill, N. (2009). 'Bires and the Motivation Debate.' *Theoria* 74: 50–59.

Copyright of Ratio is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.