



### **Hume on Practical Reason: Normativity and Psychology in *Treatise* 2.3.3**

David Phillips

*Hume Studies* Volume 31, Number 2, (2005) 299 - 316.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at

<http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use.

Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact

[humestudies-info@humesociety.org](mailto:humestudies-info@humesociety.org)

<http://www.humesociety.org/hs/>

# Hume on Practical Reason: Normativity and Psychology in *Treatise* 2.3.3

DAVID PHILLIPS

*Abstract:* I argue for an interpretation of Hume on practical reason different both from the traditional instrumentalist interpretation and the more recent nihilist interpretation. Both involve reading Hume as making normative claims. On the nihilist interpretation, Hume denies that either passions or actions can violate authoritative norms of reason; on the instrumentalist interpretation, Hume denies that passions can violate authoritative norms of reason, but holds that instrumentally irrational actions violate the one such authoritative norm. I argue instead for a purely psychological interpretation of T 2.3.3 and parallel passages in T 3.1.1. As I interpret him, Hume does not here even address the question whether passions or actions can violate authoritative norms. His conclusion is merely that a person's beliefs cannot conflict with her passions.

Hume has traditionally been taken to be the father of instrumentalism, the view “that reason's only practical role is working out and recommending action that best achieves the end of the agent.”<sup>1</sup> More recently, however, this traditional instrumentalist interpretation of Hume has been challenged. Christine Korsgaard,<sup>2</sup> Jean Hampton,<sup>3</sup> and Elijah Millgram<sup>4</sup> have argued instead for what I shall label the “nihilist interpretation,” according to which, in the crucial passages in *Treatise* 2.3.3 (SBN 413–8) and 3.1.1 (SBN 455–70), Hume denies that any actions can be contrary to reason.<sup>5</sup> These two conflicting interpretations differ in the normative

---

David Phillips is Associate Professor in the Department of Philosophy, University of Houston, Houston TX 77204-3004.  
Email: DPhillips@uh.edu.

claim they attribute to Hume, but share the idea that Hume does here make some normative claim, a claim about the extent to which there are authoritative norms of practical reason which actions can violate. My primary aim in this paper is to argue against this shared assumption. I articulate and defend a third interpretation according to which *all* that Hume does, in these passages, is to make a claim in faculty psychology: the claim that the products of the faculty of reason (beliefs) cannot conflict with the passions.

While my ultimate aim is to explicate and defend the psychological interpretation of these celebrated passages of Hume, I will begin by considering the relative merits of the two different normative interpretations. I will offer some criticisms of the case for the nihilist interpretation developed by Hampton, Korsgaard, and Millgram: an instrumentalist interpretation does not conflict with the text in as straightforward a way as Hampton supposes; Korsgaard is wrong to suggest that Hume commits himself to the view that no one ever violates the instrumental norm; and Millgram's reconstruction of Hume doesn't justify the attribution of a normative view to him. Nonetheless, I will argue, the nihilist interpretation is preferable to the instrumentalist interpretation.

I will go on to argue that the psychological interpretation is superior to both normative interpretations: when properly formulated, it allows us to make excellent sense of 2.3.3 and 3.1.1. And this has important implications. For, if we accept the psychological interpretation, Hume's view is a more attractive one. What he *does* say in 2.3.3 commits him only to a version (albeit a distinctively and problematically formulated and defended version) of a familiar and plausible idea: belief-desire psychology. He is *not* committed to any view about the possibility of authoritative evaluations of passion or action, neither holding (as the instrumentalist interpretation would have it) that such evaluations are limited to assessments of the instrumental adequacy of actions, nor holding (as the nihilist interpretation would have it) that no such evaluations are possible at all.

## 1. The Normative Interpretations

Begin with Hampton's presentation of the case for the nihilist interpretation and against the instrumentalist interpretation of Hume. As she characterizes instrumentalism, the instrumental theory of reason can be given the following tripartite definition:

- 1) An action is rational to the extent that an agent believes (reasonably) that it furthers the attainment of an end; *and*
- 2) Human reasoning involves the determination of means to achieve ends, in a way determined by the theory . . . *and*

- 3) These ends are in no way fixed by reason operating non-instrumentally; i.e., what makes them our ends is something other than reason. (57)

Hampton argues that Hume is not an instrumentalist because he rejects 1).

Hampton argues instead for the nihilist interpretation of Hume. As she reads him, Hume rejects instrumentalism because he denies the authority of the instrumental norm, which “directs us to pursue those objects and perform those actions that will be the most effective means to a desired end” (66). She writes,

Hume rejects completely the idea that [reason] has any normative authority over action. . . . For the . . . instrumentalist, to say that “reason has authority over us” is really to say that [*the*] *instrumental norm* has authority over us. But Hume rejects the idea that there is such an objective authoritative norm . . . [when we violate the instrumental norm] Hume insists that we do not violate any authoritative code of reason applying to action—because no such code exists. (66)

Hampton’s argument for attributing the nihilist position to Hume rests on a reconstruction of Hume’s underlying rationale, and a reading of some crucial passages. Both are problematic. The underlying rationale makes Hume a kind of more careful predecessor of John Mackie:<sup>6</sup> “Hume developed a conception of reason that is *not* a variant of the instrumental conception, because he appreciated that a true instrumental conception is problematic on naturalist grounds” (Hampton, 59).

The thought that Hampton here attributes to Hume is this: by holding that the instrumental norm is a valid norm of practical reason, we admit at least one categorical imperative, and thereby leave ourselves open to any general naturalistic objections to categorical imperatives. The problem is that Hampton cites no textual evidence that such concerns animate Hume. Indeed, as far as I can see, there is no textual evidence.

The second problem with Hampton’s case for the nihilist interpretation is a problem with her reading of specific passages of Hume. She fails to notice an important difference between T 2.3.3 and T 3.1.1, a difference which weakens her case for the nihilist as against the instrumentalist interpretation, but also ultimately helps us see how to make a better case. On 63–4 she comments on a passage from T 2.3.3, which she goes on to quote:

However, what is striking about the Humean conception is that it also says that instrumental reason has no authority over our actions! Consider Hume’s (famous) remarks in Book II:

’Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. ’Tis not contrary to reason for

me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledged lesser good to my greater, and have a more ardent affection for the former than the latter.

What is striking here is Hampton's suggestion that in this passage Hume is concerned with reason's authority *over actions*. Rather (allowing for current purposes what I am ultimately concerned to deny, that Hume is concerned with reason's *authority* at all), two of the three examples clearly concern not reason's authority *over actions* but its authority *over passions*. This matters because it means that most of this passage provides no evidence at all for the nihilist as against the instrumentalist reading; for the instrumentalist reading and the nihilist reading *agree* that Hume denies reason's authority over passions. And even the one case which may seem to concern actions arguably does not: the point about the person who chooses total ruin is plausibly a point just about the rational permissibility of this *preference*, not any further point about the rationality of action based on it. Indeed, the defender of the instrumentalist reading can plausibly claim that nothing in T 2.3.3 is directly concerned with the rationality of actions (the issue on which instrumentalist and nihilist readings disagree) as opposed to the rationality of passions (the issue on which they agree).

The trouble, however, with this limited defense of the instrumentalist interpretation is that, in T 3.1.1, Hume's focus explicitly shifts to the rationality of actions as well as that of passions. This shift is strikingly evident in the following passage:

Reason is the discovery of truth or falsehood. Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact. Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason. Now 'tis evident our passions, volitions, *and actions* [my emphasis] are not susceptible of any such agreement or disagreement. (T 3.1.1.9; SBN 458)

Moreover, Hume continues,

*Actions* do not derive their merit from a conformity to reason. . . . *Actions* may be laudable or blameable; but they cannot be reasonable or unreasonable. (T 3.1.1.10; SBN 458; my emphasis)

This suggests a better line of textual argument for the nihilist interpretation than Hampton gives. Just as, read normatively, in T 2.3.3 Hume denies that there are authoritative norms of reason applying to *passions*, so too, read normatively, in

comparable passages in T 3.1.1, Hume denies that there are authoritative norms of reason applying to *actions*. The proponent of the instrumentalist interpretation is forced into an unstable position, offering implausibly conflicting readings of these comparable passages; that is why the nihilist interpretation is superior to the instrumentalist interpretation.

A second prominent recent defender of the nihilist interpretation of Hume is Korsgaard. Her argument for the nihilist interpretation is also problematic. It depends in part on attributing to Hume the view that “no one is ever guilty of violating the instrumental principle” (Korsgaard, “Normativity,” 228). If she is right, then Hume does in these passages address the status of the instrumental principle. She develops her view, quoting Hume, thus:

Does Hume think that the instrumental principle, unlike the principle of prudence, is a rational requirement? If he does, then as the argument above shows, there should be cases in which Hume would be prepared to identify someone’s conduct as “instrumentally irrational,” that is, cases in which, without miscalculating or making a mistake, people fail or decline to take the means to their own “acknowledg’d” ends. Now Hume does not discuss this kind of case, but he does explicitly allow that actions can be irrational in two derivative ways. . . . Both of these are cases of mistake; the actions that result are not, strictly speaking, irrational. And after discussing them, Hume asserts:

The moment we perceive the falsehood of any supposition, or the insufficiency of any means our passions yield to our reason without any opposition.

This suggests that Hume thinks no one is ever guilty of violating the instrumental principle. Making a mistake, after all, is not a way of being irrational, and Hume thinks we do take the means to our ends as soon as mistakes are out of the way. But this is worrisome. How can there be rational action, in any sense, if there is no irrational action? How can there be an imperative which no one ever actually violates? (“Normativity,” 228)

But Korsgaard is wrong to suppose that Hume here expresses or suggests any view about the possibility of violating the instrumental principle. The sentence immediately before the one she cites reads as follows: “’Tis impossible, that reason and passion can ever oppose each other, or dispute for the government of the will and actions” (T 2.3.3.7; SBN 416). Taken in the proper context, Hume’s point is simply that cases of passion based on false belief do not involve conflict between

reason and passion. The correction of the false belief eliminates the passion caused by it: no residual conflict remains. This point does not entail that instrumentally irrational action is impossible. Instrumentally irrational action could be pictured as a matter of correct passions being overwhelmed by incorrect passions, rather than as passions conflicting with beliefs. Hume here neither takes nor suggests a view on the possibility of instrumentally irrational action.

An additional oddity of Korsgaard's treatment here is her suggestion that "making a mistake . . . is not a way of being irrational" ("Normativity," 228). Here her treatment contrasts sharply with Hampton's. In discussing the "curmudgeon refusing to do *y* even though he admitted to wanting *x*, and you issued a hypothetical imperative to him to the effect that he ought to do *y* to achieve *x*," Hampton makes the charge of mistake central:

If you would mean your hypothetical imperative to be a real *imperative* and not merely a statement of the causal facts, you would charge him with making a *mistake* if he didn't follow it, calling him "wrong" or "irrational." (69)

Here my sympathies are with Hampton. The idea of mistake seems to me central to the concept of irrationality, not, as Korsgaard suggests, irrelevant to it.

A third prominent recent defender of the nihilist reading is Millgram. Begin with Millgram's treatment of Hume's conclusion in T 2.3.3. After (sensitively and plausibly) rehearsing the two arguments Hume there offers, Millgram initially formulates the conclusion of Hume's second argument thus: "(4) A passion cannot be opposed (or, for that matter, endorsed) by reason; practical states of mind cannot be produced by reasoning." But Millgram then continues, "[the argument's] conclusion is tantamount to the claim that there is no such thing as practical reasoning, since if there were, reason would be able to endorse or oppose some motivational states" (78).

The argument Millgram gives for reformulating Hume's explicit conclusion is problematic, and the problem is a revealing one. I interpret the claim that there is no such thing as practical reasoning as the claim that there are no correct or incorrect ways to deliberate about passion or action. I grant that this claim entails the claim Hume explicitly makes, that reason and passion cannot oppose one another. But the converse entailment does not hold. It is perfectly possible that (objects of) the faculties of reason and passion be incapable of coming into conflict, even if there are correct or incorrect ways to deliberate about passion or action. This could be so if the correctness or incorrectness of these ways of deliberation was not the province of reason. Moreover, if Hume does think there are such correct ways of deliberating about passion or action, this *is* likely to be what he thinks about their correctness.<sup>7</sup> Hence Hume's conclusion that reason and passion cannot conflict

is *not* the conclusion that there are no such things as correct ways to deliberate about passion and action.

Millgram goes on to offer an interesting explanation *why* Hume should have accepted a position as counterintuitive as skepticism about practical reasoning. Millgram's explanation appeals to Hume's pictorial semantic theory. As Millgram sees it, Hume's semantic theory means that the only resource he has to distinguish different mental states including impressions, memories, beliefs, probability judgments, and imaginings, is vivacity. Hume assigns "bands on the vivacity spectrum to the different content-bearing mental states" (83). Millgram then plausibly argues that Hume will face insurmountable problems if he tries to fit desires or passions too into the same spectrum. Hence, Millgram claims, Hume's basic semantic views commit him to holding, as he does in the second of the arguments in T 2.3.3, that passions do not represent. Millgram takes it that this explanation also amounts to an explanation why "Hume *had* to be a skeptic about practical reasoning" (86). But, again, it does not. That passions cannot represent, and hence cannot be true or false, does not entail that they cannot be correct or incorrect. It entails only that their correctness or incorrectness can't be a matter of truth or falsity, of representational adequacy or inadequacy. It does not entail that there are no authoritative norms applying to passion or action. So in explaining why Hume had to think that passions cannot represent, Millgram has not shown that Hume had to be a skeptic about practical reasoning.

Despite these problems, I have argued, the nihilist interpretation is superior to the instrumentalist interpretation. But the problems help bolster the case for the superiority of the psychological interpretation over the nihilist interpretation, a case I will now develop further.

## 2. The Psychological Interpretation

To develop the psychological interpretation, and to contrast it with the two normative interpretations, it is helpful to distinguish two versions of an idea prominent in the relevant passages of Hume: the idea of a conflict (or lack thereof) between reason and passion. The idea's prominence is clear from fairly casual inspection. In the first sentence of T 2.3.3, Hume sets up the view he intends to argue against by noting that "nothing is more usual in philosophy . . . than to talk of the combat of passion and reason" (T 2.3.3.1; SBN 413). In the sentence immediately preceding the celebrated dictum that "reason is, and ought only to be the slave of the passions," he asserts that "we speak not strictly and philosophically when we talk of the combat of passion and of reason" (T 2.3.3.4; SBN 415). In summarizing, Hume claims that "'tis impossible, that reason and passion can ever oppose each other, or dispute for the government of the will and actions" (T 2.3.3.7; SBN 416).

We can, then, formulate the thesis Hume intends to reject thus:

(i) Reason and passion can conflict.

But we can then distinguish two different versions of this thesis:

(ia) *Psychological version*: A person's beliefs can conflict with her passions.

(ib) *Normative version*: Passions can violate authoritative norms.

According to the psychological interpretation, these are distinct and separable theses. Hume commits himself in these passages only to denying (ia), and, in so doing, he does not explicitly or implicitly commit himself to any view about (ib).

One way to see how sharply the psychological interpretation differs from either of the normative interpretations is to notice its implications for Hume's view on the rationality of passions. The instrumentalist and the nihilist interpretations share the idea that Hume denies that *passions* can violate authoritative norms of reason. They differ as to whether Hume denies that *actions* can violate authoritative norms of reason. But the psychological interpretation holds that Hume is committed to no view about whether even passions violate authoritative norms of reason; his only claim is that beliefs and passions cannot conflict.

While the psychological interpretation has not been much considered in the most recent literature on these parts of Hume, variants of it have certainly been articulated before. In *Hume's Moral Theory*, Mackie endorses Jonathan Harrison's reading:<sup>8</sup>

Harrison sums it up well: what Hume is getting at is that beliefs move us to action only if they are relevant to the satisfaction of a passion, and that reasoning, whether demonstrative or probable, affects our actions only in so far as it produces beliefs which are so relevant. (47)

Mackie goes on explicitly to characterize what Hume here advocates as a "psychological theory" (47). Michael Smith<sup>9</sup> distinguishes the "Humean theory of motivation" from the "Humean theory of normative reasons." Smith informally characterizes the Humean theory of motivation thus: "[M]otivation has its source in the presence of a relevant desire and means-end belief" (92). The Humean theory of normative reasons holds "that the rational thing for an agent to do is simply to act so as maximally to satisfy her desires, whatever the content of those desires" (130). The version of the psychological interpretation found in Mackie and Harrison can be expressed, employing Smith's terminology, as the claim that all that the passages from T 2.3.3 and T 3.3.1 do is to articulate the Humean theory of motivation: nothing in them amounts to an endorsement of the Humean theory of normative reasons. We will see below problems for this straightforward version of the psychological interpretation, which, I will argue, show that the psychological interpretation should be articulated differently, without focusing solely on motivational conflict between belief and passion. But even if the versions

of the psychological interpretation in the literature are not the best, versions of it are certainly extant.

Moreover, the psychological interpretation is in the spirit of some recent work on other aspects of Hume. Consider in particular Hume on induction, where a traditional skeptical (and hence normative) reading, according to which Hume's aim is to argue that judgments about the future are unjustified, has been challenged by a naturalistic reading, according to which Hume's aim is simply to show that inductive judgments are not the product of the faculty of reason.<sup>10</sup>

My argument for the psychological interpretation has two facets. In the last section I showed that, while the nihilist interpretation is the more plausible of the two normative interpretations, it is still problematic. In this section, I will argue that, by contrast, the psychological reading makes excellent sense of the text of T 2.3.3 and 3.1.1.

Begin with the start of 2.3.3. The text of T 2.3.3.1 to T 2.3.3.4 (SBN 413–5) provides significant and straightforward *prima facie* support for the psychological interpretation. Having introduced the idea of a combat between reason and passion, at the end of the first paragraph Hume announces his agenda: "In order to shew the fallacy of all this philosophy, I shall endeavour to prove *first*, that reason alone can never be a motive to any action of the will; and *secondly*, that it can never oppose passion in the direction of the will" (T 2.3.3.1; SBN 413). Here, having characterized the ideas he will oppose, Hume carefully indicates the grounds on which he will oppose them. And these grounds are simply, in effect, the denial that reason alone can ever motivate.

Moreover, the argument of the succeeding two pages fits similarly well with the psychological interpretation. Hume distinguishes the two ways the understanding "exerts itself," when making judgments about relations of ideas and about matters of fact, and argues that neither kind of judgment is, in itself, motivationally efficacious. Call this "the argument by enumeration." Having established the first of the two conclusions identified in the above quotation, Hume moves to show that it also implies the second conclusion: "But if reason has no original influence, 'tis impossible it can withstand any principle which has such an efficacy" (T 2.3.3.4; SBN 415). He concludes, "We speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (T 2.3.3.4; SBN 415). All this is fully consistent with the psychological interpretation. All Hume here does is to defend the claim that reason alone cannot motivate. As we have seen, doing so commits him to no position on (ib), on the question whether there are authoritative norms applying to passion (or action).

The central argument of T 3.1.1 also draws on the conclusion of T 2.3.3 in a way that is fully consistent with the psychological interpretation:

Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be deriv'd from reason; and that because reason alone, as we have already prov'd, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (T 3.1.1.6; SBN 457)

Again, the conclusion earlier derived in T 2.3.3 (expressed here twice in the two different variants of the second premise of the central argument of T 3.1.1, i.e., that “reason alone can never have [an influence on the actions and affections],” and that “reason of itself is utterly impotent [to excite passions, and produce or prevent actions]”) simply is, in effect, the denial that reason or its products alone can motivate.

The main apparent textual difficulties for the psychological interpretation come from the second line of argument Hume develops in T 2.3.3, a line of argument developed again in T 3.1.1. I will call this argument “the representative quality argument” and I will sometimes distinguish between the “first version” of the representative quality argument, found in 2.3.3, and the “second version,” found in 3.1.1. Hume first introduces the representative quality argument in 2.3.3 with the avowed intention simply of confirming the earlier conclusion:

As this opinion may appear somewhat extraordinary, it may not be improper to confirm it by some other considerations.

A passion is an original existence, or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification. When I am angry, I am actually possess with the passion, and in that emotion have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high. 'Tis impossible, therefore, that this passion can be oppos'd by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, considered as copies, with those objects which they represent. (T 2.3.3.4–5; SBN 415)

The representative quality argument presents an apparent problem for the psychological interpretation because it introduces the idea of a different type of conflict between reason and passion. Earlier in 2.3.3, in both the statement of Hume's agenda and the argument by enumeration, the kind of conflict between reason and passion whose existence Hume denies is a *motivational* conflict, a conflict consisting in reason and passions pushing us in different directions. In the representative quality argument, by contrast, the kind of conflict Hume has in mind is a conflict

consisting in passions being true or false. Hume argues that there can be no such conflict, because passions lack the representative quality that would render them capable of truth or falsity. In so doing, he may appear to commit himself to the normative thesis that there are no authoritative norms applying to passions.

The apparent problem is compounded by the celebrated examples he lists, after describing the two (improper) senses in which passions or affections can be unreasonable:

Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it. 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. (T 2.3.3.6; SBN 416)

Again, this passage appears to provide support for the normative interpretation. In saying that none of these passions are contrary to reason, Hume appears to be denying that there are authoritative norms applying to passions.

It is in responding to the apparent problem these passages raise that it becomes important to develop the more general version of the psychological interpretation, rather than the version (found in Mackie and Harrison) according to which Hume here claims only that there cannot be *motivational* conflict between reason and passion. The more general version adds a second idea: that there cannot be *cognitive* conflict between reason and passion either. But the guiding idea is still the same: reason and passion cannot conflict. Spelt out more fully, the more general version of the psychological interpretation says that in these passages, Hume really is, as advertised, offering a second line of argument designed to bolster the first. The two arguments fit together like this. First, in the enumeration argument, Hume argues that passions do not conflict with beliefs in the way passions characteristically conflict with each other, in terms of motivational impact. Second, in the representative quality argument, he argues that passions do not conflict with beliefs in the way in which beliefs characteristically conflict with each other, in terms of truth-conditions or cognitive content. He rules out both the possibility that passions can conflict with reason in the way that passions can conflict with each other, and the possibility that passions can conflict with reason in the way that beliefs can conflict with each other. In both cases, Hume is concerned only with the possibility of conflict between the faculties of reason and passion. Nothing he says addresses the possibility of authoritative standards for assessing passions.

Armed with this more general version of the psychological interpretation, return to the apparent problems presented by the later parts of 2.3.3. Consider first the apparent problem presented by the “representative quality” argument. To see how the problem can be addressed, it helps to look at the second version of the representative quality argument, the one Hume gives in T 3.1.1. He writes, “Whatever, therefore, is not susceptible of this agreement or disagreement, is incapable of being true or false, and can never be an object of our reason” (T 3.1.1.9; SBN 458). The key point here is that Hume does not emphasize the normative dimension of judgments of truth and falsity. His interest in these judgments is merely a matter of their making passions objects of, and thus potential competitors with other candidate objects of, the faculty of reason. Hume’s direct and immediate concern remains a concern with faculty psychology. That this is his concern comes out more in the next paragraph: “Actions may be laudable or blameable; but they cannot be reasonable or unreasonable: Laudable or blameable, therefore, are not the same with reasonable or unreasonable” (T 3.1.1.10; SBN 458). In giving the representative quality argument, Hume does not intend to rule out normative assessments of passions or actions; he intends only to argue that these assessments are not the product of the faculty of reason.

Now consider the problem presented by the famous examples of passions which, Hume claims, are not “contrary to reason.” The key here is to be clear that the phrase “contrary to reason” is ambiguous. One natural interpretation of this phrase is normative: to say of something that it is “contrary to reason” is to say that it is in some fundamental way mistaken, that it violates authoritative norms. But another interpretation is faculty psychological: to say of something that it is “contrary to reason” is to say that it conflicts with the deliverances of a specific psychological faculty, reason. When Hampton interprets these passages, she reads “contrary to reason” normatively. Hence she has Hume denying that passions violate authoritative norms of reason. But the better reading of the claims is the psychological reading. We are simply given further illustrations of the idea that passions cannot come into conflict with (the objects of) the faculty of reason.

### 3. Implications and Objections

The psychological interpretation has important implications for the understanding and assessment of Hume. It is helpful to consider these under two headings: what Hume *does* say in 2.3.3 (and parallel passages of 3.3.1) and what he *does not* say. Let me begin with what he does say. If I am right, what Hume develops in T 2.3.3 (and parallel passages of T 3.1.1) is a psychological theory. Smith gives a nice characterization of this psychological theory, which he calls “the standard picture of human psychology,” and clearly finds in 2.3.3:<sup>11</sup>

According to the standard picture of human psychology—a picture we owe to Hume (1888)—there are two main kinds of psychological state. On the one hand there are beliefs, states that purport to represent the way the world is. Since our beliefs purport to represent the world, they are assessable in terms of truth or falsehood, depending on whether or not they succeed in representing the world to be the way that it really is. And on the other hand there are desires, states that represent how the world is to be. Desires are unlike beliefs in that they do not even purport to represent the way the world is. They are therefore not assessable in terms of truth and falsehood. Hume concludes that belief and desire are therefore distinct existences. . . . The standard picture of human psychology is important because it provides us with a model for explaining human action. Crudely, our beliefs tell us how the world is, and thus how it has to be changed, so as to make it the way our desires tell us it is to be. An action is thus the product of these two distinct existences: a desire representing the way the world is to be and a belief telling us how the world has to be changed so as to make it that way. (7–9)

Of course, Hume does not express the standard picture of human psychology in exactly the way Smith would favor (significant charity is required to read direction of fit into Hume); and Hume's arguments for it are in a number of ways problematic. The representative quality argument is the biggest source of problems. Hume is wrong to think that passions cannot represent: anger, his own example, typically has cognitive content (one is angry with a person, institution, etc.). Now the obvious first line of defense of Hume here is to hold (as the distinction Smith makes between beliefs and desires in terms of direction of fit would have it) that passions do represent, but their content always has the direction of fit of desires, not beliefs. But this too seems wrong. Passions (including anger) often include elements with the direction of fit of beliefs, not desires: for example, the belief that someone has attacked you. The best thing to say, on Hume's behalf, is probably that what he calls passions always involve some element with the direction of fit of desire; that some such element is required for motivation; and that such elements cannot in themselves conflict with beliefs.<sup>12</sup>

There are further problems. Mackie notes two: the argument by enumeration does not really do enough to undermine the position of Hume's contemporary opponents, and there are possible counterexamples to the claim that beliefs and desires do not conflict (though not, Mackie thinks, devastating counterexamples) (*Hume's Moral Theory*, 47–50). More generally, as Smith suggests, Hume is too tempted to understand desires phenomenologically rather than dispositionally (104–11). But, despite these concerns, it does not seem excessively charitable to

see Hume in 2.3.3 as endorsing the standard picture of human psychology. And while that standard picture has its critics,<sup>13</sup> it is very widely accepted.

Equally important is what Hume does *not* say in 2.3.3 and 3.1.1. If the psychological interpretation is correct, Hume is not concerned, in 2.3.3 and parallel passages, to develop any view about whether actions or passions can be authoritatively assessed. In particular he does not, as the instrumentalist interpretation would have it, endorse a famous, very narrowly circumscribed view about the possibility of authoritative evaluation of passion and action. Nor does he, as the nihilist interpretation would have it, explicitly deny that any such authoritative evaluations can be made. We are then free to read other parts of Hume (perhaps other parts of the *Treatise*, perhaps the second *Enquiry*, perhaps “Of the Standard of Taste”) as containing an account of the possibility of authoritative evaluations of actions and passions.<sup>14</sup>

Now consider objections. First, it might be objected that the psychological reading makes Hume’s disagreement with his rationalist predecessors in effect a purely verbal one. If Hume can have (for example) character evaluations play the role that assessments of practical rationality play in other theories, then his view has no distinctive content.<sup>15</sup> Such an objection is, I think, mistaken. The psychological interpretation leaves Hume space for authoritative evaluations of passion and action in a way the normative interpretations do not. But the space it leaves has a distinctive shape. Any such evaluations cannot be a matter of passions or actions being true or false, of passions or actions conflicting with the deliverances of reason. Hume’s account of such evaluations will need instead to have a distinctively sentimentalist character; he is not merely a rationalist under another name.

Second, it might be thought that this paper just rehashes Elizabeth Radcliffe’s critique of the nihilist interpretation, so I need to say a little about how my critique of the nihilist reading relates to hers.<sup>16</sup> She focuses on practical *reasoning*, and argues that Hume does leave room for reasoning to influence action: “Hume’s discussion of motivation provides the material for attributing to agents not just a causal sequence that produces or explains their actions, but also a line of reasoning to a *conclusion* that precedes and can be causally connected to their actions” (254). To require that practical reasoning issue in norms is, she claims, to foist a Kantian understanding of practical reasoning onto Hume: “[T]he conclusion that Hume has no theory of practical reasoning is based on understanding instrumentalism solely as a system of Kantian hypothetical imperatives and asking Hume to meet misplaced expectations” (249). When instrumentalism is understood appropriately, Hume is an instrumentalist.

I have two connected objections to Radcliffe’s treatment. First, she doesn’t develop a detailed reading of the text of T 2.3.3. In the absence of such a detailed reading (of the kind I have developed above), the impression that Hume there does

explicitly take the nihilist view is likely to remain uneffaced. Second, Radcliffe seems to suppose that one could only read Hume as adopting an instrumentalist theory of normative reasons through illegitimately importing Kantian ideas into his empiricist framework. Qua empiricist, she thinks, Hume could not have addressed the question whether the instrumental norm is an authoritative norm of reason. Here I think she is mistaken. It is perfectly possible to be the sort of empiricist Hume is and to hold that there are no authoritative norms governing passions, but that there is one authoritative norm governing actions, the instrumental norm. This is one kind of instrumentalism easily read into Hume, given passages like this one: “Where a passion is neither founded on false suppositions, nor chuses means insufficient for the end, the understanding can neither justify nor condemn it” (T 2.3.3.6; SBN 416). I have argued above that it turns out to be a misreading. It is a misreading partly because, as Hampton and others point out, passages in T 2.3.3 that may look like endorsements of the authority of the instrumental norm are not, and partly because, as I argued above, to offer it is to read similar passages in T 2.3.3 and T 3.1.1 in implausibly different ways. But to show that it is a misreading, and to show that the nihilist interpretation too is a misreading, we need, in a way Radcliffe doesn’t, to work in detail through T 2.3.3 and T 3.1.1. The problem with the nihilist reading is not that Hume *could not* have taken a position on whether actions (or passions) violate authoritative norms of reason; it is that he *did not*. Having established that he did not, we can then offer, as Radcliffe does, a theory of practical *reasoning* that we can, if we like, call instrumentalist. But there is still an important sense in which Hume is not an instrumentalist: he does not accept the view that the instrumental norm is the only authoritative norm applying to actions, a view which he might have held, and which is often attributed to him. My conclusion, then, is similar to Radcliffe’s in that we both reject the nihilist interpretation of Hume, but I wouldn’t label the conclusion in the way she does, nor do I find her strategy of argument for it persuasive.

#### 4. Conclusions

In this paper I have argued for an interpretation of Hume on practical reason different both from the traditional instrumentalist interpretation and the more recent nihilist interpretation. Both involve reading Hume as making normative claims. On the nihilist interpretation, Hume denies that either passions or actions can violate authoritative norms of reason; on the instrumentalist interpretation, Hume denies that passions can violate authoritative norms of reason, but holds that instrumentally irrational actions violate the one such authoritative norm. I have argued instead for a purely psychological interpretation of T 2.3.3 and T 3.1.1. As I interpret him, Hume does not here even address the question whether passions or actions can violate authoritative norms. His conclusion, it turns out, is merely

that a person's beliefs cannot conflict with her passions. The view to which he is thus committed is both more attractive and less constraining than other recent interpreters have recognized. It is more attractive because what Hume does here commit himself to is (a distinctive version of) a familiar and plausible psychological theory. It is less constraining because he here develops no view at all about the possibility of authoritative assessments of passion or action.

## NOTES

Thank to Elijah Millgram, Bill Nelson, George Sher, an audience at the "Hume and His Critics" conference at Baylor University in 2005, and two anonymous referees for this journal, for most helpful comments on earlier versions of this paper.

1 I take this characterization of instrumentalism from Jean Hampton, *The Authority of Reason* (Cambridge: Cambridge University Press, 1998), 125. She calls the characterization "rough." It can be taken as an abbreviation of the more precise formulation I quote at the start of section 1, below.

2 Christine Korsgaard, "The Normativity of Instrumental Reason," (which I will refer to as "Normativity") in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (New York: Oxford University Press, 1997), 215–54, and Christine Korsgaard, "Skepticism about Practical Reason," *Journal of Philosophy* 83 (1986): 5–25.

3 Jean Hampton, "Does Hume Have an Instrumental Conception of Practical Reason," *Hume Studies* 21 (1995): 57–74.

4 Elijah Millgram, "Was Hume a Humean?" *Hume Studies* 21 (1995): 75–93.

5 References to the *Treatise* by numbered paragraph are to *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (New York: Oxford University Press, 2000). I then give corresponding page references to the older Selby-Bigge/Nidditch edition as "SBN." *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, revised by P. H. Nidditch, 2nd ed. (Oxford: Clarendon Press, 1978).

6 More careful because, though in chapter 1 of *Ethics* (Harmondsworth: Penguin, 1977), John Mackie argues against the existence of objective values, Hampton takes him there wrongly to suppose that hypothetical imperatives are naturalistically unproblematic. In the passage Hampton cites, Mackie says: "If you want X, do Y' (or 'You ought to do Y') will be a hypothetical imperative if it is based on the supposed fact that Y is, in the circumstances, the only (or the best) available means to X, that is, on a causal relation between Y and X. The reason for doing Y lies in its causal connection with the desired end, X: the oughtness is contingent upon the desire" (quoted in Hampton, 59).

7 I develop this thought more in Part 3.

8 Mackie, *Hume's Moral Theory* (London: Routledge, 1980), esp. chap. 3. Page references will be placed in the text. Jonathan Harrison, *Hume's Moral Epistemology* (Oxford: Oxford University Press, 1976), esp. 5.

9 See Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994), esp. chaps. 4 and 5. Smith himself does not suggest the psychological interpretation of Hume. Hume exegesis is not Smith's project, but he clearly takes 2.3.3 to contain endorsements both of the Humean theory of motivation and of the Humean theory of normative reasons.

10 For a nice treatment and introduction to the literature on these issues, see Don Garrett, *Cognition and Commitment in Hume's Philosophy* (New York: Oxford University Press, 1997), chap. 4.

11 Smith's reference in this passage to "Hume 1888" is a reference to the *Treatise*. It is clear from Smith's later discussion (104–5) that 2.3.3 is the main part of the *Treatise* Smith has in mind.

12 For similar criticisms of the representative quality argument, emphasizing also the ways in which Hume's claims about anger in 2.3.3 conflict with his treatment of anger elsewhere in Book 2, see Annette Baier, *A Progress of Sentiments* (Cambridge, Mass.: Harvard University Press, 1991), chap. 7, esp. 161–5. While Baier does not (of course) tackle the more recent interpretive literature with which I am mostly concerned, her discussion fits nicely with the psychological interpretation.

13 For interesting recent criticism, see Thomas Scanlon, *What We Owe to Each Other* (Cambridge: Harvard University Press, 1999), esp. chap. 1, section 8. For references to earlier criticisms, see Smith, *Moral Problem*, chap. 4.

14 Consider, as one example, Hume's treatment of the evaluation of sentiment in "Of the Standard of Taste," in *Four Dissertations* (1757; repr., Bristol: Thoemmes Press, 1995), 201–40. While there is controversy about how to interpret the view Hume there takes, it is clear that the evaluation of sentiments he countenances is not a matter of these sentiments straightforwardly conflicting with reason. As Hume puts it in characterizing the "species of philosophy" which "represents the impossibility of ever attaining any standard of taste," "All sentiment is right: because sentiment has a reference to nothing beyond itself. . . . But all determinations of the understanding are not right; because they have a reference to something beyond themselves, to wit, real matter of fact; and are not always conformable to that standard" ("Of the Standard of Taste," 208). The standard of taste is not a straightforward matter of sentiments being representationally inadequate. Instead, the standard of taste is the product of a consensus of qualified critics. As Mary Mothersill puts it, summarizing Hume's "official doctrine" (which she takes to be problematic), "There are certain properties—truth to nature in painting, verisimilitude in poetry, unity of action in drama—that are universally pleasing. . . . The essential properties are, however, very difficult to discern, and it requires native ability and exhaustive training and practice to be able to recognize them. Hence qualified critics (Hume's phrase is 'true judges') are rare. Where disputes about taste arise, it is the joint verdict of the true judges that settles the matter: that verdict is 'the standard of taste.'" Mothersill, "Hume and the Paradox of Taste," in *Aesthetics: A Critical Anthology*, ed. George Dickie, Richard Sclafari, and Ronald Roblin, 2nd ed. (New York: St. Martin's Press, 1989), 269–86, 271.

There is room for doubt about the consistency of Hume's position in "Of the Standard of Taste" with his position in the *Treatise*. Moreover, "Of the Standard of Taste" is concerned with standards for sentiments, not directly either for the sorts of passions most usually considered in T 2.3.3, or for actions. Still, the natural extension of the

position Hume takes in “Of the Standard of Taste” to passions and actions would not take authoritative norms governing passions and actions to involve any straightforward conflict between passions and beliefs.

15 Such an objection might be inspired by what Korsgaard says in “Normativity of Instrumental Reason,” 232–3, though she is there concerned, in a way I at this point am not, with the possibility of understanding Hume as an instrumentalist.

16 Elizabeth Radcliffe, “Kantian Tunes on a Humean Instrument: Why Hume is not *Really* a Skeptic About Practical Reason,” *Canadian Journal of Philosophy* 27 (1997): 247–70.