

Hume's Representation Argument Against Rationalism¹

by
Geoffrey Sayre-McCord
University of North Carolina/Chapel Hill

Hume offers a barrage of arguments for thinking that morality is not based on reason. Two arguments, in particular, are discussed in detail by Rachel Cohon and David Owen in "Hume on Representation, Reason and Motivation." They call the first the Representation Argument and the second the Motivation Argument.

The Representation Argument maintains that neither passions, nor volitions, nor actions can be contrary to reason on the grounds that (i) being contrary to reason involves things "consider'd as copies" disagreeing with "those objects they represent" and (ii) neither passions, nor volitions, nor actions are copies, or representations, of "any other existence or modification" at all. Indeed, Hume claims that they "have no more reference to any other object, than when I am thirsty, or sick, or more than five foot high" (T. 415). Thus the Representation Argument "proves directly," he says, "that actions do not derive their merit from a conformity to reason..." (T. 458).

At the same time, Hume thinks, the Representation Argument establishes that "reason can never immediately prevent or produce any action by contradicting or approving it" (T. 458). And this claim (which Cohon and Owen call the Inertia Thesis) is then mobilized in the Motivation Argument, in conjunction with the observation the morality does have the sort of immediate impact on action that reason lacks, to show that reason "cannot be the source of the distinction betwixt moral good and evil."² Hume puts the argument this way:

The merit and demerit of actions frequently contradict, and sometimes controul our natural propensities. But reason has no such influence. Moral distinctions, therefore, are not the offspring of reason. Reason is wholly inactive, and can never

¹. This paper was originally delivered as a comment on "Hume on Representation, Reason and Motivation" by Rachel Cohon and David Owen at the University of Nottingham at the 1996 meeting of the International Hume Society.

². The version of the argument Cohon and Owen focus on is:
(i) "...morals...have an influence on actions and affections..."
(ii) "...reason alone... can never have any such influence"
(iii) "...it follows, that [morals] cannot be deriv'd from reason."

Of course it follows, only with an additional premise that Hume makes explicit: that an active principle cannot be derived from a passive principle.

As Cohon and Owen see things, this line of argument -- from the Representation Argument, through the Inertia Thesis, to the conclusion of the Motivation Argument -- faces three serious problems, which they try to deal with on Hume's behalf.

The first problem is that passions (and maybe volitions, too) quite obviously do make "reference" to other objects. As Hume himself recognizes in Book II of the Treatise, our passions are characteristically directed. Anger, for instance, is unlike free-floating anxiety precisely in that in being angry we are usually angry *at someone* (either ourselves or others) *for having done* (or failed to do) *something*. Contrary to what Hume asserts in presenting the Representation Argument, it is just not true that anger has "no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high" (T. 415). So, the argument looks to have an obviously false premise.

The second problem is that the conclusion of the Representation Argument "does not look at all like a claim about what reason alone can cause" (p. 12), whereas the Inertia Thesis the argument supposedly establishes is quite clearly a (negative) causal claim. So, Cohon and Owen suggest, the Representation Argument apparently cannot establish the Inertia Thesis in the way that Hume thinks it does.

The third problem is that Hume himself, in Book I, rightly recognizes that certain beliefs -- those concerning the prospect of pleasure and pain -- do motivate even as they are the product of (causal) reasoning. So, the Motivation Argument, which the Inertia Thesis supports, seemingly flies in the face of what, elsewhere, Hume recognizes to be true about the motivational effects of reason.

All told, Hume's arguments look to be in sad shape.

Are things really this bad? Cohon and Owen do not think so, and I believe they are right. But we disagree quite substantially on why these "problems" are not nearly so serious as they seem. In what follows I will do my best to characterize their response to these three problems as well as my own, in each case trying to emphasize the ways in which my response differs from theirs.

The First Problem

In connection with the first problem, Cohon and Owen spend a fair bit of time describing Hume's grounds for thinking that we can form no idea of the causes of our impressions of sense. I am a little unclear as to why. In places they suggest that our inability to form such an idea establishes that such impressions fail to represent. The thought seems to be that we would have reason to think impressions represent only if we had reason to think they

resembled what causes them.³ But even if this is reasonable, our inability to form an idea of the causes of our impressions would show only that we have no reason to think they represent, not that we have reason to think they do not. And the Representation Argument turns on the latter, stronger, claim. In any case, Cohon and Owen recognize that Hume does not explicitly argue that impressions do not represent.⁴ And I believe they recognize as well that if he were to argue that impressions of sense do not represent *on those grounds*, the argument would not extend to impressions of reflexion (e.g., the passions), the causes of which we can conceive (since they are our own ideas and impressions). Thus, while it is clear that Hume thinks passions do not represent, I doubt he did or would defend that view with the argument Cohon and Owen appear to be advancing.

Nonetheless, I think Hume is committed, and rightly so, to the first premise of the Representation Argument. And that introduces a problem insofar as that premise flies in the face of what we know about the intentionality of the passions, about their being, characteristically, directed at someone or something.

Towards the very end of their discussion of the first problem, Cohon and Owen acknowledge that they have so far said nothing about the directedness of the passions. Yet they claim that "Hume has no problem in accounting for this" (p. 9). All they add by way of explanation, though, is that Hume thought passions such as hatred are "associated with an idea of another, and Hume has no problem with ideas representing" (p. 9). They do not spell out how the associated idea serves to make the passion itself directed, nor do they explore how the account they mention might be reconciled with the Representation Argument.

I suspect that Hume's appeal to (merely) associated ideas will not account adequately for the intentionality of the passions. But I also think that a broadly Humean account can be developed. In any case, I will leave the issue to one side. More relevant here is the question of how Hume might resolve the apparent tension between the first premise of Representation Argument and recognizing that the passions have intentional content.

³. Along they way they also appear to endorse the claim that if something resembles its cause it represents its cause, as when they note that impressions and ideas "come in resembling pairs, and *as the former cause the latter*, the latter represent the form, and what they represent is impressions" [p. 6, my italics]. Of course, that two things resemble one another, even when one causes the other, does not establish that the second is a copy of the first. My son Thomas noted this and pointed out that he might say "yes" to an offer my other son accepted with the same word, and say it because his brother did (since he likes to do things with his brother), even though he was not copying his brother.

⁴. In places Hume seems to assume that some impressions of sense may well represent. See, for instance, T. 233.

As it happens, I³ think the resolution is pretty obvious, at least in broad outline.⁵ Hume did, I think, make a mistake in saying that our passions "have no more a reference to any other object, than when I am thirsty, or sick, or more than five foot high" for just the reasons highlighted by the first "problem". But having a reference to another object is not the same as representing it.

The difference is important because the key premise of Hume's Representation Argument is that a passion contains no "representative quality, which renders it a copy of any other existence or modification," and it may lack such a quality even if it does, in some sense, make reference to some other existence or modification. The argument's conclusion -- that passion cannot be "oppos'd by, or be contradictory to truth and reason" -- depends only on granting (1) that such an opposition or contradiction "consists in the disagreement of ideas, consider'd as copies, with those objects which they represent" and (2) that passions do not represent (or copy) any objects in the required way. Both might be granted even if one holds that our passions do make reference to other objects.

Clearly, a fully satisfying defense of Hume along these lines would require spelling out how it is Hume might mark the difference between merely making reference to something and representing it. And Hume is on notoriously shaky ground when it comes to accounting for representation. Moreover, there are some nice complications to Hume's account of the passions (with his appeal there to the subject, object, and cause of our passions) that promise to give ideas, "consider'd as copies", a prominent place. Even so, I expect that the intentional content Hume rightly sees as bound up with our passions will not serve to render the passions themselves representational in a way that allows them to count as true or false, as accurate or inaccurate representations.⁶

Thus the first "problem" with Hume's Representation Argument is really, at most, a problem that arises if one thinks the only argument for seeing passions

⁵. In fact, the resolution I favor is basically the same, I think, as one Rachel Cohon offers in "On an Unorthodox Account of Hume's Moral Psychology," *Hume Studies*, vol. XX, number 2, November 1994, pp. 179-194. (See, especially, page 189.)

⁶. Hume maintains that the ideas that are bound up with our passions are conceptually distinct from them. A particular feeling of love, for instance, might itself be present even absent any thought about others or their qualities. Yet one could instead hold that having an idea of another person as having certain qualities is a proper part, rather than merely a natural concomitant, of the feeling of love. In that case, the idea which forms a part of the passion might be contrary to reason thanks to the idea "disagreeing" with the object that it represents. Even then it is unclear that the passion, which includes the false idea, is itself either false or contrary to reason, for "even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment." (T. 415). When Hume draws this last distinction, between the passion being unreasonable and the judgment upon which it is founded being unreasonable, he seems to have in mind that the passion is causally founded on the judgment, not that it contains the judgment as a part. But I think the causal reading is not essential to the point. As Hume notes, as long as the judgments upon which the passions are founded are not false "the understanding can neither justify nor condemn it" -- and that seems to be true regardless of whether the judgment is merely a cause of, or is instead a part of, the passion in question.

as non-representational is that they lack intentional content altogether, and I think there is no good reason to think that.

The Second Problem

The second problem arises when it comes to accounting for the support the Representation Argument is supposed to provide for thinking that moral distinctions are not founded on reason alone. The dubious step is from the Representation Argument's conclusion -- that passions and actions cannot conflict with reason because they lack representational content -- to the Motivation Argument's premise -- that reason by itself is motivationally inert. Hume obviously thinks the one establishes the other. Yet, as Cohon and Owen point out, "It is hard to see straight off how an argument proving that passions and actions cannot represent something else by resembling it, and in this sense cannot conform to reason, is supposed to show that reason alone cannot motivate actions" (p. 12). Might not reason have a causal impact on passions and actions even though passions and actions can neither conform to nor contradict reason? How does the non-causal conclusion of the Representation Argument support ("in a single step") the negative causal claim advanced by the Inertia Thesis?

Cohon and Owen consider and reject the proposal that the Inertia Thesis is not, after all, a causal claim. As this proposal would have it, "the claim that reason alone cannot produce action is an ontological -- almost a logical -- thesis rather than a causal one: passions, volitions, and actions cannot be the conclusions of bits of reasoning, because they are of the wrong ontological category; they are 'realities' rather than representations that can be true or false" (p. 14). Read in this way, the Inertia Thesis follows immediately from the Representation Argument. Yet so interpreted, the Inertia Thesis is useless when it comes to the Motivation Argument. As Cohon and Owen note, the Motivation Argument will go through only if the claim that "reason alone cannot produce passions or actions" is a causal claim, since the second premise of the argument -- that "morals excite passions, and produce or prevent actions" -- is clearly a causal claim. To interpret the Inertia Thesis as anything other than a causal claim is to introduce an equivocation that renders the Motivation Argument invalid.

So the problem is to explain how it is that the Representation Argument's conclusion supports the Inertia Thesis, when the latter is construed as denying the *causal* efficacy of reason alone. To be honest, I do not really understand the solution Cohon and Owen propose -- or rather, I think I understand what they propose, but I do not see why they think it solves the problem they identify. However, because their solution to this problem is simultaneously advanced by them as a solution to the third problem, I shall put off discussing it.

In the meantime, I want to make a point about the second problem, before turning to the third problem and the solution Cohon and Owen offer to both.

⁵
The second problem gets its bite because of the apparent mismatch between the evidently non-causal conclusion of the Representation Argument and the patently causal nature of the Inertia Thesis. As Cohon and Owen emphasize, we learn from the Representation Argument that neither passions nor actions "can be oppos'd by, or be contradictory to truth and reason," but we are to infer from that the Inertia Thesis' claim that reason alone can cause neither passions nor action. They find the connection mysterious. To see why they should not, it is important to recognize that Hume's claim concerning the inertia of reason is that "reason can never immediately prevent or produce any action *by contradicting or approving of it*" (T. 458) -- important because this (negative) causal claim *does* follow directly from the Representation Argument's non-causal conclusion that reason can neither contradict nor approve of actions (or passions or indeed anything else that is non-representational in nature). Just as I cannot cause you to do something by legitimately ordering you to do it, if I cannot legitimately give you orders at all, reason cannot cause an action by contradicting or approving of it, if reason cannot contradict or approve of actions at all. The non-causal claim concerning reason's inability to contradict or approve of things that are non-representational in nature underwrites directly the relevant negative causal claim.

Of course, there might still be a problem here. Thinking back to the equivocation that threatened the proposal that the Inertia Thesis is non-causal, one might worry that my version of the Thesis (causal though it is) would likewise not be adequate to the Motivation Argument. After all, what I am suggesting is that Hume's claim that "reason alone cannot motivate action" should be interpreted as "reason cannot cause an action *by contradicting or approving of it*." And it is natural to wonder whether the Inertia Thesis, so interpreted, can play the relevant role in Hume's Motivation Argument. Will the Inertia Thesis be enough to establish the relevant contrast with morals? It will, if Hume's claim concerning morals is that morals can, unlike reason, cause an action by contradicting or approving of it. And I think this is just what Hume does claim when he contrasts morality with reason by saying that while "reason can never immediately prevent or produce any action by contradicting or approving of it," good and evil "are found to have that influence" -- that is, morality does (sometimes) immediately prevent or produce action by contradicting or approving of it (T. 458).

On this view, the role of reason, when it is operating alone, is to discover truth or falsehood, where "Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact" (T. 458). In this capacity, reason will approve of or contradict only what is "susceptible of any such agreement or disagreement". Since neither passions nor actions are so susceptible, neither passions nor actions will be approved of or contradicted by reason operating alone. Significantly, this is compatible with recognizing that reason, when operating in concert with sentiments and passions, may well lead to the approval or contradiction of some action. Yet in these cases the approval or contradiction at stake is not a matter of the actions

exhibiting the truth or falsehood discovered by reason, but a matter of harmony or conflict between the actions and a person's sentiments or passions.

I do not mean here to offer a full-scale defense of this reading of the Inertia Thesis. But I do hope to have said enough, first, to raise worries about pressing the second problem, in the way Cohon and Owen do, in terms of the mismatch between a non-causal claim and the causal claim it is supposed to support and, second, to suggest that Hume might have a fairly straightforward and elegant way of tying together the conclusion of the Representation Argument with the Inertia Thesis.

Keeping in mind that I have not yet talked at all about Cohon and Owen's solution to the second problem, let me turn to the third problem -- since the second and third are supposed to be solved in one fell swoop.

The Third Problem

Whatever one thinks of the first two problems, the third threatens to serve as (at least an *ad hominem*) *reductio* of Hume's whole line argument. Hume himself grants that certain beliefs -- concerning prospective pleasures and pains -- motivate. And these beliefs seem obviously to be the product of (causal) reasoning alone. Thus, such beliefs look to be a powerful counterexample to Hume's claim that reason by itself fails to motivate.

One might try to avoid the difficulty by insisting that such beliefs do not motivate. But Hume pretty clearly will not take that route, at least in so far as he sticks to what he says in the section of Book I called "Of the influences of belief." Alternatively, though, and this is the response recommended by Cohon and Owen, one might maintain that such beliefs are not the products of reasoning alone. This seems implausible, given that a belief to the effect that some course of action will cause one pain seems pretty clearly to be an unproblematic judgment concerning a matter of fact. However, Cohon and Owen urge, it is not so implausible as it seems, once one reconsiders "what 'reason alone' means" (p. 17). They suggest that "reason alone...is reasoning considered apart from any passions and any feelings of pleasure or pain" (p. 18). Thus, to say that reason, alone, cannot produce passions or actions is to say that reasoning apart from any passions and apart from any feelings of pleasure or pain will fail to cause either passions or actions. If we take this suggestion to heart, we will see that our beliefs concerning future pleasure or pain are not, after all, the products of reason alone since, absent all feelings of pleasure and pain, we could form no ideas (and so no beliefs) concerning either present or future pleasures or pains. As Cohon and Owen point out:

A being who possessed only reason alone, in the sense we are suggesting Hume intended, would be one who had Humean reason but no passions and no feelings of pleasure or pain. Such a being... could not have the ideas of the passions, nor those of pleasure or pain, since he could not experience the originals. Consequently he could not form any beliefs about

them, even ⁷ if his causal reasoning were perfect. Understood in this way, reason alone does not produce any beliefs about the prospects of pain or pleasure either.

Thus the third problem is (in a way) neatly solved: the beliefs that cause passions and actions -- those concerning prospective pleasure and pain -- are not available to those endowed only with reason, so Hume can consistently acknowledge that such beliefs do motivate while denying that reason by itself motivates.

The same proposal, concerning how we should interpret Hume's claims concerning 'reason alone', is offered (I think) as a solution to the second problem, as well. But here I am less clear on how the solution is supposed to go. If we embrace Cohon and Owen's interpretation of 'reason alone' then we can see that the Inertia Thesis will be given the required causal reading -- reason alone, it turns out, really cannot cause action, because a person possessing only reason will fail to have the sorts of perceptions that cause actions. So, their proposal avoids the pitfalls facing the suggestion that the Inertia Thesis should be given a non-causal reading. But how is the Thesis (so interpreted) supported by the Representation Argument's conclusion? How is that argument even relevant? What Cohon and Owen say is that the Representation Argument establishes that actions and passions "can never be the outcome of reason" (p. 19).

Given Cohon and Owen's view that the second problem arises because the conclusion of the Representation Argument looks to be non-causal, while the Inertia Thesis is obviously causal, I am supposing that they are here making a causal claim -- that actions and passions cannot be, themselves, the *causal* output of any reasoning process. But if that is what they mean, then I do not see how the Representation Argument establishes the point. What grounds does the Representation Argument offer for thinking that the process of reasoning will not somehow cause impressions that in turn motivate action even in those as yet unfavored by such impressions? I do not see any. In fact, the *process* of reasoning might well be expected to give rise to all sorts of impressions -- headaches, exhaustion, exhilaration -- some of which might be motivational in their upshot.

Cohon and Owen's claim might, of course, be given a non-causal reading. They may be thinking that passions and actions cannot be, themselves, the output of any reasoning process, because they assume, for instance, that something counts as an output of reasoning only if it is a conclusion of some argument -- which it can be only if it has representational content.⁷ However, if

⁷ This interpretation is suggested by Cohon and Owen's otherwise implausible claim that "Reason produces only ideas or representations" (p. 18). The passages they cite in support of this claim do recommend thinking that reason is such that only representations are contrary or conformable to it, but not -- so far as I can see -- that reason is such that only representations might be caused by it.

the claim is given some such non-causal reading, then they will face exactly the problem they saw plaguing the non-causal proposal they explicitly consider.

Either way, then, Cohon and Owen's proposed reading of "reason alone" in the Inertia Thesis fails to bridge the gap they find between the Representation and Motivation Arguments. As I have said, I am not myself bothered by the problem in the first place. But that is because I see the Representation Argument's non-causal conclusion as establishing just what is necessary to underwrite the particular (negative) causal claim Hume uses in the Motivation Argument -- that reason can neither prevent nor cause an action *by contradicting or approving of it*. In other words, I do not see the gap in the first place.

What about Cohon and Owen's solution to the third problem? Here again I find myself less bothered by the problem, in the first place, than they are. I have already mentioned two responses one might offer to the difficulty -- one might deny that beliefs concerning future pleasure and pain motivate or one might deny, as Cohon and Owen do, that such beliefs are the product of reason alone. Yet there is a third reply available: one might grant that such beliefs do motivate, and grant too that they are the product of reason alone, but insist that their ability to motivate as they do is not similarly a product of reason but is, instead, the result of our affective constitution. On this view, our beliefs concerning pleasure and pain are indeed the product of reason, but their effect on us depends on something reason alone cannot provide -- our being disposed to pursue pleasure and avoid pain.

This last reply seems to me to fit both the spirit and the letter of what Hume has to say about the influence of belief. The perception of pleasure and pain have been, Hume notes, "implanted in the human mind... as the chief spring and moving principle of all its actions" (T. 118). We are so constituted that these perceptions have a regular and predictable effect. Still, things might have been otherwise. Another being, differently disposed, might perfectly well have impressions of pleasure and pain (along with the corresponding ideas) and yet remain unmoved. Like us, such a being would be able, through reason alone, to reach conclusions concerning future pleasure and pain; but unlike us, such a being would find such prospects a matter of indifference.⁸

Assuming, as I think we should, that this reply is available to Hume, the third "problem" raises no more difficulty for Hume than do the first two.

Given this reading, though, one may well wonder whether Hume is right in claiming, as he does in the Motivation Argument, that morality, unlike reason, can by itself influence actions by contradicting or approving of them. Will not Hume have to say that our moral judgments, no less than our judgments about the prospect of pleasure and pain, have their influence only thanks to our being disposed, say, to pursue what is right and avoid what is wrong? And will not

⁸. Incidentally, I know a number of real people who are apparently indifferent to future pleasure and pain.

⁹
this need for the affective disposition show that morality alone, like reason alone, is inert?

Hume is, I believe, committed to thinking that the effect of our moral attitudes depends on our affective dispositions. Yet that is not the same as saying that their effect depends on something other than morality. The affective dispositions that lead us to act (or not) on our moral attitudes are not themselves independent of morality. Indeed, a crucial element of being moral is being such that one's moral opposition to, or approval of, an action actually leads one to act appropriately. While people are, as Hume notes "*often* goven'd by their duties," they are not always [T. 457, my italics]. Whether and to what extent people are "deter'd from some actions by the opinion of injustice, and impell'd to others by that of obligation" turns, in the end, on whether and to what extent they are moral (not on whether they are rational) [T. 457].⁹

Conclusion

In the end, I think none of the three problems pose any threat at all to Hume's argument. I have tried here to explain why each problem arises only on a misreading of Hume and have, along the way, sketched, without defending in detail, the line of argument that carries Hume from the Representation Argument, through the Inertia Thesis, to the conclusion of the Motivation Argument. The view that emerges poses, I believe, a strong challenge to a long-standing and venerable strain of rationalism in ethics. In particular, it challenges those versions of rationalism that hold that a fully rational being could not fail to be moral. Needless to say, I have not here explored the force of the challenge or the variety of ways a rationalist might respond. My aim has only been to show that the challenge is more internally coherent, and indeed more plausible, than Cohon and Owen make it out to be.

⁹. On this reading, the Motivation Argument presupposes neither non-cognitivism nor emotivism. In fact, Hume's apparent recognition that people sometimes fail to be moved by their moral duties -- by their opinions concerning injustice and obligation -- suggests that he rejects the sort of motivational internalism non-cognitivism and emotivism are, standardly, taken to entail.