

Hume as an Error Theorist

Rafael Graebin Vogelmann

Universidade Federal do Rio Grande do Sul

1. Introduction

The broad outline of Hume's account of our capacity to draw moral distinctions is fairly clear: reason alone cannot draw these distinctions, rather we must resort to feelings or sentiments of a particular kind to distinguish between virtue and vice. That much can be inferred from the very titles of the first two sections of book 3 of Hume's *Treatise of Human Nature*.¹ There is no doubt that Hume is, in some sense, a *moral sentimentalist*. Hume's description of the psychological process by means of which we come to approve or condemn an action is also reasonably clear: we take the action as the external sign of a character trait;² that lead us to consider the idea of this trait and whether it produces pleasure or pain on people close to an agent endowed with it; we then sympathize with this pleasure or pain, experiencing an equivalent (if weaker) sympathetic

¹ In what follows Hume's *Treatise of Human Nature* is cited with notations of the form *T b.p.s.p* / *SBN pqr*, the lower-case letters here standing for arabic numerals. Numerals immediately following *T* indicate book, part, section, and paragraph numbers in David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (Oxford: Clarendon Press, 2007); numerals following 'SBN' indicate page number in David Hume, *A Treatise of Human Nature*, 2nd ed., ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Oxford University Press, 1978). *An Enquiry concerning the Principles of Morals* is cited with a notation of the form *2E s.p.p*. Numerals immediately following *2E* indicate section, part (when the case) and paragraph numbers in David Hume, *An Enquiry concerning the Principles of Morals*, ed. J. B. Schneewind (Cambridge: Hackett Publishing, 1983). *EMPL* followed by a number indicates a page in David Hume, *Essays Moral, Political, and Literary*, ed. Eugene F. Miller (Indianapolis: Liberty Fund, 1987).

² "If any action be either virtuous or vicious, 'tis only as a sign of some quality or character." (T 3.3.1.4 / SBN 575).

feeling,³ which, on its turn, leads to one of four indirect passions (love or hatred, pride or humility) – these passions are the moral feelings, the perceptions of approval and disapproval, which guide our moral judgments.⁴ Hume's view about the nature of moral convictions or opinions is much harder to pin point. He has been interpreted as a noncognitivist, according to whom to hold a particular moral opinion is to have certain feelings or passions;⁵ as a dispositional cognitivist, according to whom to believe that an action or character trait is virtuous or vicious is to believe that it is such as to cause feelings of approval or disapproval in a spectator that satisfies certain conditions;⁶ as a naturalist cognitivist, according to whom moral beliefs mobilize ideas copied from the impressions of approval and disapproval but ascribe a dispositional quality to the action or person evaluated;⁷ and, finally, as a projectivist according to whom we come to think of actions or character traits as virtuous or vicious because we project our moral feelings onto these objects of evaluation. According to this view, moral convictions incorporate an error: we think of actions and character traits themselves as virtuous or vicious, but they are not, in themselves, virtuous or vicious. It, therefore, represents Hume as an error theorist.⁸ My goal in this paper is to defend a projectivist reading of Hume's account of moral convictions.

This interpretation faces some challenges. First, its emphasis on error seems to fit poorly with Hume's friendly attitude towards moral distinctions. Right at the beginning of the of the *Enquiry Concerning the*

³ That the pain or pleasure which arises from the contemplation of a character trait is produced by sympathy is the main thesis of T 3.3. Exactly how these feelings are connected to moral appraisal is a harder to ascertain. According to Hume, in order to arrive at a stable judgment and correct distortions in our sympathetic feelings we have to adopt a *general point of view* (T 3.3.1.15 / SBN 581-2). How to characterize this point of view and how it corrects for distortions in sympathy is a matter of controversy. For an illuminating discussion of these issues see Sayre-McCord (1994) and Cohon (2008, ch.5).

⁴ "The pain or pleasure, which arises from the general survey or view of any action or quality of the mind, constitutes its vice or virtue, and gives rise to our approbation or blame, which is nothing but a fainter and more imperceptible love or hatred." (T 3.3.5.1 / SBN 614).

⁵ This view is defended by Flew (1963), Blackburn (1984, ch. 6) and Smith (2009).

⁶ This view is defended by Lo (2009, p. 60-66). Mackie (1980, p.66-69) considers but ultimately rejects this view.

⁷ This view is defended by Cohon (2008, p.111-3).

⁸ See Mackie (1980, p.71-2), Stroud (1993) and Kail (2007, ch. 7-9).

Principles of Moral, for instance, Hume claims that those “who have denied the reality of moral distinctions, may be ranked among the disingenuous disputants” and recommends that an “antagonist of this kind” be left alone for eventually “mere weariness” will bring him to the “side of common sense and reason” (2E 1.2). It may seem that an error theorist would be exactly this kind of “antagonist”. Second, the passages that provide the strongest textual evidence for a projectivist reading seem to have such unpalatable consequences that interpretative charity recommends one to seek an alternative interpretation. Hume first introduces the idea of projection in his discussion in the *Treatise* of the origins of the idea of the necessity involved in causal relations. He claims that this idea is derived from the impression of the mind’s propensity to move from the impression of an object to an idea of its usual attendant and that, therefore, the necessity of a causal relation is something that “exists in the mind, not in objects” (T 1.3.14.22 / SBN 165). To explain how we come to think of objects themselves as connected by causal relations, Hume claims that “the mind has a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects discover themselves to the senses” (T 1.3.14.25 / SBN 167). The same propensity is supposed to explain how we come to think of objects as having a particular smell, for instance. The claim that virtue and vice “may be compared to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in objects, but perceptions *in the mind*” (T 3.1.1.26 / SBN 469, my emphasis) strongly suggests, therefore, that moral thought should receive the same projectivist treatment dispensed to our thought about causality. However, it seems hard to take seriously the claim that virtue and vice are perceptions in the mind. As Stroud has pointed out “if vice were a perception or feeling in the mind, then in saying that I get a certain feeling from contemplating X I would be saying that I get vice from contemplating X. And that is incoherent” (STROUD, 2003, p.181).⁹

In light of these challenges, the other interpretations of Hume’s view about the nature of moral convictions mentioned above may seem more attractive for, one could argue, they can satisfactorily account for his sentimentalism while avoiding the problems faced by a projectivist reading.

⁹ A very similar complaint is made by Cohon (2008, p.114-5).

But this is a mistake: in the next two sections I argue that the noncognitivist and the dispositional reading fail to account for a number of claims Hume puts forward in stating his moral sentimentalism. The shortcomings of these views will reveal that the best way in which to conceive of a moral judgment in light of Hume's text is as a belief that mobilizes an idea that is a copy of a moral feeling. In section 4 I argue that once we understand moral judgments in this way and once we have clearly in view the nature and origin of the idea these judgments mobilize we can see that moral judgments, as Hume conceives of them, involve a peculiar kind of category error: they involve predicating of an action or character trait an idea that cannot be so predicated because it is not the idea of a property or quality. The idea that we project our moral feelings onto the world is introduced, I will argue, as an attempt at explaining why we systematically incur in this category mistake.

If this interpretation is correct, then it sets Hume apart from other error theorists. Error theories concerning moral judgment usually consist in the conjunction of two independent theses: (a) that moral judgments ascribe a particular quality to the action, person or character trait they evaluate and (b) that as a matter of fact the quality in question does not exist. This is the form Mackie's error theory takes (see MACKIE, 1991, ch. 1) and this is how he reads Hume (see MACKIE, 1980, p.71-2). Kail reads Hume in the same way: moral judgments ascribe *essential value* (i.e., non-instrumental value) to their objects but we have independent reason to believe that there is no such thing as essential value (KAIL, 2007, p.169). On the reading I am proposing, Hume's view does not take this form. It does not depend on a negative metaphysical thesis to the effect that virtue and vice are not actual features of reality. Instead, it rests solely upon considerations about the nature of the ideas that are mobilized in moral judgments and moral thought in general. It claims not that moral judgments make metaphysical assumptions about reality that turn out to be false but rather that our moral ideas are not at all ideas of qualities that could be truthfully predicated of actions or character traits, if not in our world at least in a different possible world.

2. *The noncognitivist reading*

Hume has often been read as a noncognitivist about moral convictions because it has seemed to many that he was arguing for noncognitivism at T 3.1.1. In this section, after claiming that “those judgments, by which we distinguish moral good and evil” are perceptions (T 3.1.1.2 / SBN 456) he puts forward the *Argument from Motivation*: “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” (T 3.1.1.6 / SBN 457). It may seem plausible to reconstruct this argument thus:

- (i) The perceptions in which moral convictions consist can, by themselves, influence actions;
- (ii) Beliefs cannot, by themselves, influence actions;
- (iii) Therefore, the perceptions in which moral convictions consist are not mere beliefs.

A belief, Hume takes himself to have already established, is a vivid idea (T 1.3.5.7 / SBN 86 and T 1.3.10.3 / SBN 119-120). If vivid, strong ideas cannot influence our actions, neither can weaker ideas. Given that there are only two types of perceptions (ideas and impressions), the perceptions in which moral convictions consist must then be impressions of some sort or, at the very least, a compound of ideas and impressions. And given that, according to Hume, the only impressions that are capable of moving us to action are direct passions, it follows that to hold a moral opinion is to have a certain direct passion or some mental state capable of producing, on its own, a direct passion (perhaps accompanied by some idea or belief). Therefore, if this is a correct reconstruction of Hume’s argument, he is committed to a noncognitivist view of moral convictions.¹⁰

¹⁰ Contemporary noncognitivism is usually formulated as a thesis about the linguistic function of moral statements or about the nature of the speech act of making a moral claim. Noncognitivists usually hold that moral claims express attitudes or prescribe rules instead of describing the world. It is implausible to read Hume as putting forward a view of this kind. Clearly, the question with which he is concerned is about the nature and origins of a particular kind of mental state – a particular perception. It is still plausible, nevertheless, to describe him as a psychological noncognitivist if he does defend the view that to hold a particular moral opinion is to have a particular passion or an impression and not simply

That, however, is *not* a plausible reconstruction of the *Argument from Motivation* for premise (ii) is not a plausible reconstruction of the claim that reason alone has no influence on actions. This thesis is supposed to have been established in T 2.3.3. In this section Hume argues for the notorious view that reason is, and ought only to be, the slave of the passions. His argument is the following: “The understanding exerts itself after two different ways, as it judges from demonstration or probability”; “the first species of reasoning alone is ever the cause of any action” (T 2.3.3.2 / SBN 423-4); the second species of reasoning alone can only direct impulses that are produced by prospects of pain or pleasure; therefore, “reason alone can never produce any action, or give rise to volition”; it could only prevent action “by giving an impulse in a contrary direction to our passion”, but if it could do that, then it would be able to produce actions as well (T 2.3.3.4 / SBN 414-5); therefore, one cannot be deterred from some actions nor impelled to others by reason *alone*. Let us refer to the conclusion of this argument as the thesis of the *Inertia of Reason*. This argument has frequently been interpreted as an argument for the *Inertia of Belief*, i.e., the view that beliefs cannot produce the kind of motivational impulse that could deter or impel one to action. And that interpretation is indeed natural: reasoning alone produces beliefs; that being the case, if beliefs alone could produce motivational impulses, it would seem to follow that reason alone has an influence in our actions. No wonder this argument has frequently been reconstructed thus:

- (I) Reasoning is either demonstrative or probable;
- (II) The beliefs that are produced by demonstrative reasoning do not produce, by themselves, motivational impulses;

to have a belief. Furthermore, if we take the argument above as an argument for a conclusion about the linguistic function of moral statements it is invalid. We can admit (iii) that when we hold a moral opinion we are in a mental state that involves more than a mere belief and consistently hold that moral statements describe the world, instead of expressing the speaker’s attitudes or prescribing rules. That one has certain beliefs may be entailed by the fact that one has successfully made a promise without it being the case that in saying “I promise to...” one is thereby expressing those beliefs or asserting their content (that is not the linguistic function of this phrase). In the same way, that one has certain attitudes may be entailed by the fact that one holds a particular moral view without it being the case that when one states one’s view using moral terms one expresses these attitudes (see JOYCE, 2009, p.33-35).

- (III) The beliefs that are produced by probable reasoning do not produce, by themselves, motivational impulses;
- (IV) Therefore, beliefs produced by reasoning do not produce, by themselves, motivational impulses.¹¹

The first thing to notice about this argument is that its conclusion is not the same as premise (ii) above. Premise (ii) is the claim that beliefs *in general* cannot influence action or, which is the same from Hume's perspective, cannot produce motivational impulses. Conclusion (IV) is restricted to beliefs that are the outcome of reasoning. We could move from (IV) to (ii) only by assuming that all beliefs are produced by reasoning. That, however, is a false supposition and one Hume is not willing to make. A belief according to Hume is vivid idea and impressions can produce vivid ideas without any contribution from reason. Impressions, Hume tell us, carry conviction with them (T 1.3.5.7 / SBN 86). When I see a red book on the table in front of me I come to believe that there is a red book in the table simply because impressions are such as to give rise to vivid copies of themselves. There are, therefore, beliefs that are not the product of reasoning but rather the product of impressions, and these beliefs are beyond the scope of the argument for the *Inertia of Reason*.

If we substitute (IV) for (ii) in the *Argument from Motivation*, we end up with an invalid argument:

- (i) The perceptions in which moral convictions consist can, by themselves, influence actions;
- (IV) Beliefs *derived from reason* cannot, by themselves, influence actions;
- (iii) Therefore, the perceptions in which moral convictions consist are not mere beliefs.

Clearly, the conclusion that could be drawn from (i) and (IV) is not (iii) but rather "the perceptions in which moral convictions consist are not mere beliefs *derived from reason*" (an anti-rationalist thesis rather than a noncognitivist one). The possibility that moral judgments or convictions consist in beliefs that are derived from impressions rather than some form

¹¹ See, for instance, Mackie (1980, p.47), Stroud (2003, p.157), Pidgen (2009, p.97) and Sandis (2009, p.146-7).

of reasoning remains open, given that for all we know these beliefs could influence our actions.¹²

One could insist that even though Hume is not arguing for the unqualified version of the *Inertia of Belief* thesis (according to which no belief can produce motivational impulses) in T 2.3.3, he is committed to this thesis and appeals to it in the *Argument from Motivation*. That is implausible from an exegetic point of view, for Hume formulated the second premise of that argument as the claim that *reason* has no influence on actions and clearly refers the reader to the “proof” provided at T 2.3.3. But, worse still, there is textual evidence that Hume rejected not only the unqualified version of the *Inertia of Belief* thesis but also the weaker claim (IV), which restricts the thesis to beliefs derived from reason.

In presenting the argument for the *Inertia of Reason* Hume claims that “’Tis obvious, that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity” (T 2.3.3.3 / SBN 414). A prospect of pain or pleasure, however, is a causal belief about the pain or pleasure that I would experience were certain circumstances to obtain. Hume seems to be claiming then that this particular kind of causal belief is capable of producing a motivational impulse. And this interpretation is reinforced by the claims Hume puts forward in the section entitled “Of the Influence of Belief”. In this section Hume claims that impressions of pleasure and pain (call these *hedonic impressions*) influence the will (T 1.3.10.2 / SBN 118) and that the ideas derived from these impressions have a similar effect when they become vivid enough, i.e., when they become beliefs (T 1.3.10.3 / SBN 119). That happens, because impressions and ideas differ only in strength and vivacity: when an idea becomes vivid enough, therefore, its effects become similar to the effects of the corresponding impressions. Belief causes “an idea to imitate the effects of the impressions” (T 1.3.10.3 / SBN 119). Presumably, the beliefs that mobilize the ideas derived from hedonic impressions are beliefs concerning the pleasure and pain that is produced by particular objects, i.e., prospects

¹² Formulating the *Argument from Motivation* as an inference to the best explanation, as Smith (2009, p.105) does, makes no difference here. If beliefs that are not produced by reasoning can influence our behavior, then the hypothesis that moral convictions are beliefs that are directly derived from impressions and the hypothesis that moral convictions are impressions of some kind provide equally good explanations of the fact that moral convictions are capable of influencing our behavior.

of pleasure and pain. These beliefs, therefore, have an influence on the passions that mimics the influence of impressions of pleasure and pain. Given Hume's commitment to the idea that to influence the will is to produce an impulse towards a particular action (T 2.3.3.4 / SBN 415), we should understand the influence of hedonic impressions in this way: they produce motivational impulses in the form of passions of propensity or aversion. Prospects of pain and pleasure should, then, have the same effect. These prospects are causal beliefs, produced by causal reasoning. It follows that Hume is committed to the view that some beliefs produced by causal reasoning do produce, by themselves, motivational impulses – a claim that directly contradicts the thesis of the *Inertia of Belief*, even in its qualified version.

The only way to avoid this conclusion is to deny that hedonic impressions produce emotions of propensity and aversion by themselves, i.e., to hold that they can only produce these motivational states by interacting with a preexisting desire. The only desire that could play this role would be a general desire to experience pleasure and to avoid pain. There are two ways in which to explain how the interaction between a hedonic impression and this desire could work, but neither can account for the claim that prospects of pain and pleasurable mimic the motivational effects of hedonic impressions. (A) A natural way in which to conceive of this interaction is this: suppose I eat a slice of pineapple; that produces a pleasurable sensation; I then come to think of pineapples as pleasurable and this new belief interact with the standing general desire for pleasure to produce a derived desire for pineapples. But if that is the case, it is not the pleasurable impression that is producing the new desire, but rather the interaction between the previous existing desire for pleasure and the belief that pineapples are a source of pleasure, i.e., a prospect of pleasure. And that cannot be right: prospects of pain and pleasure mimic the motivational effects of hedonic impressions; therefore, the capacity of these impressions to produce those effects cannot depend on prospects of pain and pleasure. (B) It must be simply a matter of fact that the interaction between the pleasurable impression and the general desire for pleasure directly produces a derived desire or aversion for the source of the impression. But if this is how hedonic impressions produce motivational impulses then prospects of pain and pleasure do not mimic the motivational effects of hedonic impressions. Prospects of pain and pleasure are causal beliefs and

Hume has a story to tell about how causal beliefs interact with preexisting emotions of propensity or aversion: these beliefs extend the propensity or aversion for a particular object *O* to other objects that are known to bring about *O* (T 2.3.3.3 / SBN 414). Prospects of pain or pleasure would interact with a preexisting general desire to experience pleasure and avoid pain in exactly this way. But it is not only prospects of pain and pleasure that can interact with motivational impulses to produce this kind of effect – all causal beliefs can do so. And most causal beliefs are totally unrelated to impressions of pleasure and pain – their capacity to produce derived desires by interacting with preexisting desires is not a reflection of a power that belongs to hedonic impressions but is rather an original power of the human mind. Therefore, if prospects of pain and pleasure produced new desires in this way, they do not mimic the effects of hedonic impressions but simply behave as ordinary causal beliefs.

We do better, then, to take at face value Hume's claim that hedonic impressions influence the will: they do so *by themselves*. Prospects of pain and pleasure, a kind of causal belief produced by causal reasoning, mimic the effects of these impressions, therefore influencing, *by themselves*, the will and our actions. I conclude that Hume is not committed to the *Inertia of Belief* thesis, not even if it is restricted to beliefs derived from reason.¹³ We should, then, reject the noncognitivist reconstruction of the *Argument from Motivation*. That argument is best understood as an argument for a conclusion about the psychological origins of moral convictions (these perceptions are not the product of reasoning) rather than an argument for

¹³ This entails that (I)-(IV) is not a correct reconstruction of the argument for the *Inertia of Reason*. How to properly reconstruct this argument is not the topic of this paper. I just want to call attention to the fact that reason why the argument was reconstructed as an argument for the *Inertia of Belief* was the supposition that "if reason alone produces a belief and this belief alone produces a motivational impulse, then reason alone is capable of producing motivational impulses". The only way to reconcile the claim that prospects of pleasure and pain (which are causal beliefs produced by reasoning) produce, by themselves, motivational impulses with the thesis of the *Inertia of Reason* is to deny this assumption. Cohon makes a compelling case for the view that this assumption can be reasonably denied if we take reason to consist in a process (or several processes). We can then understand the argument for the *Inertia of Reason* as an argument for the claim that the processes in which reason consists produce only beliefs, not motivational impulses, even though some of these beliefs produce, by a different process, motivational impulses (see COHON, 2008, p. 74-77).

a conclusion about the psychological nature of moral convictions (that these perceptions are not beliefs).¹⁴

That is enough to remove the main reason to adopt a noncognitivist reading of Hume's moral sentimentalism. But one could hold that even though he is not arguing for this view in T 3.1.1 that is, nevertheless, his view. Indeed, that is suggested by the way in which he formulates his view at the beginning of T 3.1.2:

... the course of the argument leads us to conclude, that since vice and virtue are not discoverable merely by reason, or the comparison of ideas, it must be by means of some impression or sentiment they occasion, that we are able to mark the difference betwixt them. Our decisions concerning moral rectitude and depravity are evidently perceptions; and as all perceptions are either impressions or ideas, the exclusion of the one is a convincing argument for the other. (T 3.1.2.1 / SBN 470)

This passage strongly suggest that moral convictions are impressions rather than beliefs for it is natural to take "decisions concerning moral rectitude and depravity" to be moral convictions. And this suggestion is reinforced by the subsequent claims that the impressions in questions are pleasurable or painful feelings we get from contemplating a character and that, in the case of virtue, these pleasurable feeling itself "constitutes our praise or admiration" (T 3.1.2.3 / SBN 471). That suggestion, however, is misleading. Hume himself, in the same section of the *Treatise*, acknowledges

¹⁴ This raises the question of how exactly to understand the *Argument from Motivation*. I have suggested, in the previous note, that the best way to reconstruct the claim that reason has no influence on the will is as the claim that the processes in which reason consist do not yield motivational impulses but only beliefs. To correctly reconstruct the argument we should build around this premise. One possible reconstruction is this: (i) moral convictions can, by themselves, influence the will; (ii*) reasoning processes cannot, by themselves, influence the will; (iii*) therefore, moral convictions are not the product of reasoning processes. However, this argument has the same form as (a) hedonic beliefs can, by themselves, influence the will; (b) reasoning processes cannot, by themselves, influence the will, (c) therefore, hedonic beliefs are not the product of reasoning processes. Given that Hume accepts (a) and (b) and denies (c), he has to declare arguments with this form invalid. Here again, I find Cohon's suggestion to be illuminating. She reconstruct the argument thus: (A) the processes of moral discrimination can, by itself, influence the will; (B) reasoning processes cannot, by themselves, influence the will; (C) therefore, the process of moral discrimination is not a reasoning process (COHON, 2008, p.82). This is a valid argument and it fits well with Hume's formulation of his conclusion: "'tis in vain to pretend, that morality is discover'd only by a deduction of reason" (T 3.1.1.7 / SBN 457).

the possibility that one can hold a moral conviction without having an impression of the kind that constitutes our praise or condemnation:

Nor is every sentiment of pleasure or pain, which arises from characters and actions, of that peculiar kind, which makes us praise or condemn. The good qualities of an enemy are hurtful to us; but may still command our esteem and respect. 'Tis only when a character is consider'd in general, without reference to our particular interest, that it causes such a feeling or sentiment, as denominates it morally good or evil. 'Tis true, those sentiments, from interest and morals, are apt to be confounded, and naturally run into one another. It seldom happens, that we do not think an enemy vicious, and can distinguish betwixt his opposition to our interest and real villainy or baseness. But this hinders not, but that the sentiments are, in themselves, distinct; and a man of temper and judgment may preserve himself from these illusions. (T 3.1.2.4 / SBN 472)

Only sentiments of pleasure and pain derived from the disinterested contemplation of a character can constitute or lead to the perceptions of praise and condemnation. We are, however, likely to mistake interested sentiments for moral sentiments and that can lead us, for instance, to "think [a virtuous] enemy vicious". Clearly, to think an enemy vicious is to hold a moral conviction. Given that in this case the agent does not have a genuine moral sentiment, the moral conviction cannot be identified with one such sentiment. Nor can it be identified with the interested sentiment which the agent mistakes for a moral sentiment, for "a man of temper and judgment" could have the interested sentiment in question and lack the moral conviction that his enemy is vicious. No other impression is a plausible candidate to play the role of moral conviction in this example. Therefore, given that "as all perceptions are either impressions or ideas, the exclusion of the one is a convincing argument for the other", it follows that the moral conviction in this case is a vivid idea, i.e., a belief. This is a strong counterexample to the noncognitivist interpretation of Hume's sentimentalism.

Nevertheless, Hume clearly identifies the perceptions of praise and blame, approval and disapproval, with impressions of a particular kind. A feeling of satisfaction of a particular kind "constitutes our praise or admiration". Latter we are told that "our approbation or blame [...] is nothing but a fainter and more imperceptible love or hatred." (T 3.3.5.1 / SBN 614) and love and hatred are, according to Hume, simple impressions

of reflection (T 2.2.1.1 / SBN 329). We should, therefore, distinguish the feelings of moral approval and disapproval, which are impressions, from moral convictions, which, at least in some cases, are ideas. In particular, the claim that our “decisions concerning moral rectitude and depravity” are impressions should be read in light of this distinction: these “decisions” are not moral convictions but rather the moral feelings of approval or disapproval by means of which distinguish virtue and vice.

3. *Dispositionalism*

With the distinction between moral feelings and moral convictions in view, we can ask what the content of these convictions is. A number of passages suggest that Hume takes moral convictions to be causal beliefs. Consider these examples:

...when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment of blame from the contemplation of it. (T 3.1.1.26 / SBN 469)

The hypothesis which we embrace is plain. It maintains, that morality is determined by sentiment. It defines virtue to be *whatever mental action or quality gives to a spectator the pleasing sentiment of approbation*; and vice the contrary. (2E Appendix I.10)

Both passages suggest that to claim or to believe that an action or character trait is virtuous is to claim or to believe that the action or character trait is such as to produce a moral feeling of approval in a suitable spectator. And it is not hard to find in Hume’s text indications of how to characterize this suitable spectator. First, Hume holds that as a matter of fact moral feelings arise only from the disinterested contemplation of a character trait and are grounded on a sympathetic response to the feelings of the people affected by the character trait in question. Second, the relevant moral feelings are those experienced not from the particular point of view each of us occupies but from a “steady and general point of view” which corrects for distortions in our sympathetic responses (T 3.3.1.14-5 / SBN 580-2). It is not at first view implausible, then, to read Hume as claiming that to believe an action or character trait to be virtuous or vicious is, roughly, to believe

that it would cause a moral feeling of approval or disapproval in any sympathetic and disinterested spectator that occupies the general point of view, or, which is the same, that any such spectator is disposed to experience the appropriate moral feeling upon contemplating the action or character trait in question.¹⁵

I will refer to this interpretation as the “dispositional reading” of Hume’s account of moral convictions. According to this reading, Hume is a cognitivist regarding moral convictions: he takes these convictions to represent moral facts. Furthermore, he is a naturalist cognitivist, for moral facts turn out to be ordinary natural facts, namely, facts about the psychological dispositions of human beings to experience moral feelings under certain (very specific) hypothetical conditions. Moral convictions, then, turn out to be causal beliefs about the psychological effects of the contemplation of character traits.

Despite its appeal, this reading faces some very serious exegetic problems. First, it is reason that “discovers the connection of causes and effects” (T 3.1.1.12 / SBN 459). Causal beliefs are the product of probable reasoning. But Hume takes pains to argue for the view that moral convictions are not the product of reasoning. The dispositional reading, therefore, fits poorly with the arguments of T 3.1.1. Second, if moral convictions were causal beliefs, we would have to infer a character to be virtuous after observing that it regularly produces feelings of approbation in suitably positioned spectators, for all causal beliefs are grounded in the observation of a constant conjunction between the objects we take to be connected by a causal relation (T 1.3.6.4 / 88). But we “do not infer a character to be virtuous, because it pleases: But in feeling that it pleases after such a particular manner, we in effect feel that it is virtuous. The case is the same in our judgments concerning all kinds of [...] sensations” (T 3.1.2.3 / SBN 471). Third, the claim that we do not infer a character to be virtuous but rather feel it to be virtuous suggests (once the noncognitivist reading is off the table) not only that we form moral convictions without inference but that moral convictions are directly produced by moral feeling. For if our “decisions concerning moral rectitude and depravity” are feelings of some sort (T 3.1.2.1-2 / SBN 470) and these are not the starting point of an inference to a moral conclusion, how else could they lead us to “pronounce

¹⁵ For a defense of one such view see Lo (2009, p. 60-66).

an action blame-able or praise-worthy” (T 3.1.1.3 / SBN 456)? That suggestion is strengthened by the claim, at the end of the passage just quoted, that the same is true of moral judgments and judgments concerning sensations. As observed above, for Hume there are at least two ways in which to arrive at a belief: we can do so by reasoning or an impression can directly produce a belief. Beliefs concerning the color of objects, for instance, are usually arrived at in the latter way: to see, say, a red tomato is to have a visual impression; in normal circumstances, this impression is copied, giving rise to a vivid idea of a red tomato; if that idea is strong enough, then it will be a belief. This is a process of belief formation that involves no inference at all. What Hume seems to be claiming in the passage just quoted, then, is that moral convictions are arrived at in a similar way: just as I come to believe that a particular object is red by seeing its color, I come to believe that a character trait is virtuous by feeling its virtue, i.e., by experiencing a feeling of approval upon contemplating it.¹⁶ Now, beliefs that are the direct product of impressions are simply vivid copies of these impressions. This entails that the entirety of their content is derived from those impressions. So, for instance, if I come to believe that “this tomato is red” upon seeing the tomato, the belief in question is nothing but a vivid copy of the impression I had when I saw the tomato and the ideas it mobilizes are ones that could be derived from the impression itself. In particular, the idea of redness that constitutes this belief is a simple idea – a copy of the simple impression of redness. To the extent that the belief that the tomato is red is directly produced by the visual impression, therefore, its content cannot be properly represented as “this tomato is such as to cause an impression of redness in a human observer under such and such lighting conditions”. A belief with this content would be much more than a copy of the visual impression I have when I look at the tomato, for it mobilizes ideas that cannot be derived from that impression. In particular, the idea of cause is not a copy of any sensible impression. In a similar way, to the extent that moral convictions are directly produced by moral feelings, their content is not adequately represented by the dispositional reading.

The dispositional reading may be on to something nevertheless. Rachel Cohon has offered an interpretation of Hume’s account of moral convictions that does justice to the passages that motivate the dispositional

¹⁶ The analogy between sense perception and moral evaluation is drawn by Cohon (2008, p.103-4).

reading while at the same time avoiding the objections I just presented. According to Cohon, moral ideas (i.e., the ideas of virtue and vice) are copies of the moral feelings of approval and disapproval. Like any other idea, moral ideas can “attain great forcefulness and liveliness, which renders them beliefs” (COHON, 2008, p.105). The main error of the noncognitivist reading, Cohon argues, is to construe Hume as recognizing only one type of moral perceptions, namely, moral feelings, and ignoring the possibility that these feelings can be copied and give rise to ideas. The dispositional reading, while recognizing moral ideas, errs in taking the beliefs that mobilize those ideas to be causal beliefs and, thus, in ascribing causal content to moral ideas (COHON, 2008, p.106). But what, then, are moral beliefs about? The crux of Cohon’s interpretation is the claim that while moral ideas are copies of moral feelings, the moral convictions or opinions that are constituted by those ideas represent “a person, trait, or action as having a moral property”, namely, the property we *feel* when we have a moral feeling (COHON, 2008, p.105). What property is that? Hume is clear in stating that virtue and viciousness are not properties of the objects taken in themselves:

Take any action allow’d to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions, and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action.” (T 3.1.1.16 / SBN 468-9).

Take any action allow’d to be vicious: Wilful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call *vice*. In which-ever way you take it, you find only certain passions, motives, volitions, and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You never can find it, till you turn your reflection into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. (T 3.1.1.16 / SBN 468-9).”

If we can depend upon any principle, which we learn from philosophy, this, I think, may be considered as certain and undoubted, that there is nothing, *in itself*, valuable or despicable, desirable or hateful, beautiful or deformed;

but that these attributes arise from the particular constitution and fabric of human sentiment and affection. (EMPL 162, my emphasis)

The claim that there is no vice to be found in the object can be understood in two ways: one may take it to mean that there are no moral properties (and then, if one holds that moral convictions represent actions or character traits as having moral properties one will be lead to an error theory) or it may mean that the objects of moral evaluation have no moral properties when considered in themselves, i.e., when considered abstraction made of their relation to us and, particularly, to our feelings. Cohon defends the latter reading. An action or character trait *in itself* has no moral properties because these are relational properties and one term of the relation is a human psychological reaction (COHON, 2008, p.115). Moral properties are response-dependent – to have a moral property is actually to figure in a relation with our subjective responses. Indeed, they may very well be dispositional properties, like “the tendency a trait has to cause disapproval [or approval] in the well-informed and imaginative human observer” (COHON. 2008, p.112). That does not mean, however, that to have a moral conviction is to have a belief whose content is properly captured by the dispositional reading. When we think of an action or character trait as virtuous or vicious we mobilize a simple moral idea rather than the complex idea of a tendency to cause certain feelings. Nevertheless, it may be that, unbeknownst to us, the properties we *feel* when we have moral feelings of approval or disapproval are dispositional properties. In exactly the same way, in thinking the tomato before me to be red, I am not thinking that the surface properties of the tomato are such as to cause a sensation of redness in human observers under such and such lighting conditions, even if as a matter of fact that is what it is for something to be red (COHON, 2008, p.111).

In sum, according to Cohon, although moral convictions are usually directly produced by moral feelings (as beliefs about the colors of objects are directly produced by visual impressions) and although the ideas mobilized in these convictions are simple moral ideas (copies of moral feelings), to ascribe viciousness to an action or character is to ascribe “that property, whatever it is, that I sense when I disapprove” (COHON, 2008, p.112). Given that the property we feel when we experience moral feelings is a relational property whose nature can be spelled out in naturalistic

terms, it follows that moral claims have naturalistic truth-conditions (which as a matter of fact are sometimes satisfied – see COHON, 2008, p.99-100), even if these cannot be read off moral ideas.¹⁷

Cohon's interpretation has the virtue of providing a unified account of both the skeptical-sounding and the realist-sounding passages in Hume's text, while holding on to the claim, central to Hume's sentimentalism, that moral convictions are directly guided by our moral feelings. I will now argue, however, that the centerpiece of Cohon's interpretation – the claim that, even though moral ideas are simple ideas derived from moral feelings, moral beliefs represent their object of evaluation as having a particular property and are sometimes true – is unattainable from the perspective of Hume's theory of ideas.

For Hume, all that is ever present to the mind are its perceptions (T 3.1.1.2 / SBN 456). To think of something (for instance, to think of a unicorn) is to have a perception of a particular kind, namely, an idea. To believe is likewise to have a perception, namely, a vivid idea. Ideas are the stuff of which beliefs and thoughts are made. The quality, the content and the effects of a belief are all inherited from the ideas that constituted it. What the discussion of the problems faced by the dispositional reading shown is that the ideas of virtue and vice that constitute moral convictions are simple ideas, copies of feelings of disinterested approval or disapproval (which are phenomenologically distinct from other similar feelings – T 3.1.2.4 / SBN 471-2). And Cohon's reading incorporated this conclusion. But if moral ideas of virtue and vice are copies of feelings of approval and disapproval, then these are the only ideas that can constitute thoughts and beliefs about those feelings. That is because these feelings are indirect passions and passions are simple impressions (T 2.1.2.1 / SBN 277) – meaning that we cannot form an idea of them by putting together ideas copied from other impressions. In the same way, the idea that is a copy of the impression of redness is the only idea that can constitute thoughts about the color red. Now, there is no doubt that we do think about (in contrast with experiencing) the feeling of disinterested approval we have when

¹⁷ Cohon's reading is clearly committed to an externalist view of the reference of moral terms, such as 'virtue' and 'vice'. The reference of these terms is not fixed by the content of our moral ideas, considered in themselves. Rather, moral terms refer to that property, whatever it may be, that causes moral impressions and, consequently, produces in us moral ideas and beliefs. That commitment is explicit at Cohon (2008, p.112).

contemplating certain traits of character. We may think of this feeling, for instance, in order to compare it to feelings of love that arrive from the contemplation of the good someone made us and determine what they have in common and in what they differ. To have one such thought, to think of this feeling, is simply to have the idea of virtue before one's mind. So much so that one who has never experienced this kind of feeling will be unable to entertain a thought about it (particularly, this person will be unable to conceive of its peculiar phenomenological feel), just as a blind person cannot conceive of the colors she never saw (T 1.1.1.9 / SBN 5). This means that the moral ideas of virtue and vice are not ideas of properties, but ideas of a particular kind of mental state, a particular kind of feeling. Given that the content of a thought is determined by the ideas that constitute it, and that to think of the feeling of approval is merely to entertain the simple idea of virtue, it follows that we cannot even conceive of moral properties. We cannot, for instance, conceive of the property of being virtuous (as something that belongs to a character trait) in order to compare it to the property of being, say, socially encouraged. For when we bring to mind the idea of virtue, the thought in question will be constituted by exactly the same idea as the thought concerning the feeling of disinterested approval, and so will be the very thought. That is, we will be thinking about a feeling, not a property of character traits.

So, the ideas of virtue and vice are ideas of certain feelings. Given that assumption, how could the belief that a particular character trait is virtuous or vicious represent that trait as having a particular property (as Cohon would have it)? That would only be possible if the contribution the ideas of virtue and vice make to the content of the thought they constitute changes as the context changes. Compare these two beliefs: (a) that Nero's cruelty produces in us an unpleasant feeling of disapproval and (b) that Nero's cruelty is vicious. The first belief (which is *not* a moral belief) is constituted by the idea of Nero's cruelty, the idea of vice (since this is the idea we mobilize in thinking about the feeling of disapproval) and the idea of the particular relation in which they stand, namely, a causal relation. It does not represent Nero's cruelty as having a particular property, but rather represents it as the cause of a particular feeling. If we lose the idea of the causal relation, we get belief (b). If Cohon is correct, this belief represents Nero's cruelty as having a particular property. Given that the idea of Nero's cruelty continues to represent the character trait it represented at (a), Cohon

would have to claim that the subtraction of the idea of causal relation changed the representational content of the idea of vice. It stood for a feeling in (a), now it stands for a property. It is very implausible, however, to ascribe this view to Hume. He consistently holds that simple ideas (such as the moral ideas of virtue and vice) represent that of which they are a copy¹⁸ and never hints at the possibility that the same simple idea could itself represent different things when combined with different sets of ideas.

Furthermore, even if we admitted that when incorporated to moral beliefs moral ideas represent a property, (a) and (b) cannot be simultaneously true given Hume's understanding of truth. Contradiction to truth, Hume tell us, consist in the "disagreement of ideas, consider'd as copies, with those objects, which they represent" (T 2.3.3.5 / SBN 415). Presumably, truth consists in the agreement of ideas, *considered as copies*, with the objects they represent. Claim (a) is true. That means that the simple ideas that constitute this belief agree with that which they represent, when considered as copies of those things. In particular, the idea of vice in this case represents the feeling of disapproval and is an accurate copy of this feeling. But then (b) cannot be true, even if we admit that it represents Nero's cruelty as having a particular property. For if that is so, the idea of vice in this case represents a property. In order to assess the truth of this belief we would have to consider whether the idea of vice agrees with the property it represents, i.e., whether it is an accurate copy of it. But it cannot be, for, given that (a) is true, that idea is an accurate copy of a feeling and, therefore, cannot be an accurate copy of a property (a thing of a completely different ontological kind). Therefore, belief (b) *as Cohon would construct it* cannot be true.

Cohon, then, is right in claiming that moral ideas are copies of moral feelings of approval and disapproval. But that view is incompatible, at least in the framework of Hume's philosophy, with the view that moral beliefs represent actions, person or character traits as having a particular property *and* that some moral claims are true. These conclusions pave the way to the interpretation that presents Hume as an error theorist.

¹⁸ "Ideas always represent the objects or impressions, from which they are deriv'd, and can never without a fiction represent or be apply'd to any other." (T 1.2.3.11 / SBN 37). Consider also "[...] all ideas are deriv'd from impressions, and are nothing but copies and representations of them [...]" (T 1.1.7.5 / SBN 19).

4. *Projectivism*

In the last section I argued that, according to Hume, moral ideas of virtue and vice are copies of moral feelings of approval and disapproval and that, consequently, these ideas are ideas of feelings and not ideas of properties. I also argued that this entails that if moral claims like “Nero’s cruelty is vicious” ascribe a property to the character trait they evaluate, then they cannot be true. But this is a problematic conclusion: if a moral claim ascribes a property and we have no reason to suppose that the property in question is never instantiated by character traits, then it should be possible for one such claim to be true. This reveals that, given the commitments of Hume’s sentimentalism and his account of truth, there is some incoherence involved in taking moral beliefs to represent their object of evaluation as having a particular property. The best way in which to make sense of this, I will now suggest, is to read Hume as a projectivist and, therefore, as an error theorist of a peculiar kind.

The view that the ideas of virtue and vice are ideas of certain feelings rather than ideas of properties allow us to make sense of Hume’s baffling claim that virtue and vice “may be compared to sounds, colours, heat and cold, which, according to modern philosophy, are not qualities in objects, but *perceptions in the mind*” (T 3.1.1.26 / SBN 469, my emphasis). We can now understand this as the claim that the ideas of virtue and vice are ideas of certain feelings and that when we mobilize these ideas in thought we think about these feelings (these perceptions) and not about any quality or property that could be ascribed to objects or, more properly, to character traits. As noted above, this view has what seem to be some very unpalatable consequences: if the idea of vice is the idea of a feeling of disapproval, to think that I get a feeling of disapproval from contemplating a trait of character is to think that I get *vice* from contemplating that trait. And that, Stroud holds, is an incoherent thought (STROUD, 2003, p.181). But Hume does not take that to be an incoherent thought: the idea of vice is the idea of a feeling of disapproval and, therefore, it is perfectly intelligible to think that I get *vice* from contemplating a certain character trait. Not only that, but this is the kind of thought in which the idea of vice is properly employed. This also allows us to understand why Hume saw no contradiction between the claim that virtue and vice are perceptions and the mind and the immediately preceding claim that “when you pronounce any action or character to be vicious you mean nothing, but that from the constitution of

your nature you have a feeling or sentiment of blame from the contemplation of it" (T 3.1.1.26 / SBN 469). In this passage he is not defending a view about what kind of property virtue and vice are (for he immediately goes on to say that they are not qualities in objects) nor is he defending a dispositionalist conception of the content of moral statements, rather he is indicating the kind of thought in which the ideas of an action or character trait and moral ideas of vice or virtue are properly conjoined – namely, thoughts about the capacity of actions or traits of character to produce the kind of feeling moral ideas stand for.¹⁹

But if virtue and vice are indeed perceptions in the mind rather than properties it follows that when we think of a character trait or action as vicious, as we certainly do, we are predicating a feeling of a character trait or action. And that is a nonsensical thought: feelings are the kind of thing we can have, not the kind of thing that can be instantiated by a character trait or an action. Predicating virtue or vice of a character trait or action involves, therefore, a category error – particularly, the error of taking the idea of a mental state that can be ascribed to thinking beings as the idea of a property or quality that could be predicated of something. Surely this leads to the view that moral claims and beliefs systematically incorporate an error, but Hume would not flinch before this consequence, for he holds a very similar view regarding our causal beliefs.

According to Hume, we have no idea of the kind of necessity involved in a causal relation as something that belongs to objects themselves. The only idea we have of this necessity is the idea of a mental thing: the "determination of the thought to pass from causes to effects and from effects to causes" (T 1.3.14.22 / SBN 166). On that account, this necessity "is something, that exists in the mind, not in objects" (T 1.3.14.22 / SBN 165). The idea of this necessity, therefore, is the idea that we have at our disposal to think about this peculiar determination of the mind (as when, for instance, we entertain the thought that this determination is produced by

¹⁹ The same explanation does not apply to another passage that is usually evoked in favor of the dispositional reading: "The hypothesis which we embrace is plain. It maintains that morality is determined by sentiment. It defines virtue to be whatever mental action or quality gives to a spectator the pleasing sentiment of approbation; and vice the contrary." (2E Appendix I.10). This passage can be made compatible with the reading I am proposing if we suppose that in it Hume is not presenting a claim about the meaning of moral terms but rather a view about what is the feature that is shared by all those character traits that we, as a matter of fact, classify as virtues.

the experience of a constant conjunction between the cause and the effect). This idea does not represent anything “that does or *can* belong to the objects” (T 1.3.14.19 / SBN 164, my emphasis). Nevertheless, we “suppose necessity and power to lie in the objects we consider, not in our mind” (T 1.3.14.25 / SBN 167). That is, we ascribe the idea that is a copy of a particular determination of the mind to objects themselves – and we commit this mistake systematically. The situation is exactly the same as with moral ideas.

The proneness to this mistake requires an explanation, and that is why Hume introduces the idea of projection. We make this mistake systematically because the mind has “a great propensity to spread itself on external objects, and to conjoin with them any internal impressions, which they occasion, and which always make their appearance at the same time that these objects discover themselves to the senses” (T 1.3.14.25 / SBN 167). The explanation seems to be that because certain internal impressions always accompany the impression of the object, the ideas derived from these impressions become associated. This would usually lead to a causal belief (to the effect that the object produces the feeling) but because we fail to distinguish the internal impression from the perception of the object we simply ascribe the idea of the internal impression to the object as one of its properties: “The agreeable quality is thought to lie in the object, not in the sentiment; and that merely because the sentiment is not so turbulent and violent as to distinguish itself, in an evident manner, from the perception of the object” (EMPL 165).²⁰

In the case of moral ideas, because the contemplation of a character trait is usually accompanied by a moral feeling of approval or disapproval, we end up ascribing the idea of this feeling as properties to the character trait (and consequently to persons who instantiate this trait and to actions that follow from it). A projectivist explanation concerning moral beliefs is offered in this notorious passage from the *Enquiry Concerning the Principles of Morals*:

²⁰ Hume does not offer a detailed account of this psychological process. The way in which he introduces it (“‘Tis a common observation, that the mind has a great propensity to spread itself on external objects” - T 1.3.14.15 / SBN 167) suggest that he takes this propensity to be an original principle of the human mind that we observe and cannot explain further.

Thus the distinct boundaries and offices of *reason* and of *taste* are easily ascertained. The former conveys the knowledge of truth and falsehood: The latter gives the sentiment of beauty and deformity, vice and virtue. The one discovers objects, as they really stand in nature, without addition or diminution: The other has a productive faculty, and gilding or staining all natural objects with the colours, borrowed from internal sentiment, raises, in a manner, a new creation. (2E Appendix I.21).²¹

We can now understand talk of “gilding or staining all natural objects with the colours, borrowed from internal sentiment” as the claim that moral beliefs involve predicating of objects like actions and character traits an idea that is actually a copy of a sentiment. They involve, in this manner, a projective error. Stroud argues against this view. According to him, we “do not think that an act of willful murder itself has a feeling of disgust or disapprobation, any more than we think that a painting on a wall has a sentiment of pleasure or awe. That is nonsense in each case.” (Stroud, 1993, p.261). My point is that this is exactly Hume’s view. His investigation of the nature and origin of moral ideas reveal them to be ideas of certain feelings and that renders moral judgments nonsense. The task of an investigator of human nature is then to explain why we systematically make this mistake and how it eluded us for so long. The notion of the propensity of the mind to spread itself on objects is an attempt at providing one such explanation.²²

²¹ This passage, by itself, does not provide very strong evidence for the projectivist reading. Cohon’s reading can accommodate it. Cohon holds that Hume’s point in this passage is that virtue and vice are relational properties that are produced by interaction between the contemplated character trait and the spectator’s mind – in this sense it is a new creation (Cohon 2008, p. 122-3). Obviously, however, the projectivist reading can take this passage, as well as the anti-realist passages considered in the previous section, at face value, i.e., as claiming that there are no moral properties.

²² One may, of course, wonder whether Hume is successful in offering this explanation. Stroud (1993) thinks he is not. But the reason why Stroud finds Hume’s projective explanation unsatisfactory is that it does not explain how we can *intelligibly* think about the virtuousness or viciousness of actions or character traits (1993, p.267-8). That, I am suggesting, is not something Hume sets out to explain. He actually holds that we cannot think intelligibly about such things – moral thoughts involve a category error that render them nonsensical. Stroud does suggest that this may be Hume’s view (1993, p.262-3) and, as far as I can see, provides no exegetic reason to reject this interpretation. It may be that, on account of the theory of ideas which Hume assumes, the same considerations that lead to the thesis that moral beliefs are nonsensical attributions of feelings to actions or character traits would lead to the conclusion that not even attributions of ordinary empirical properties (such as the property of being round) are intelligible (STROUD, 1993, p.268) and

One might object that this leads us right back to the noncognitivist reading. For the beliefs that result from the projective error are nonsensical and traditional noncognitivism held precisely that because moral statements are nonsensical they have no truth-value and, therefore, express an attitude or emotion instead of describing reality.²³ But it would be a mistake to adopt a noncognitivist reading of Hume's sentimentalism for this reason. Hume's topic is not the linguistic function of moral statements or terms. Rather, when he raises questions about moral judgments he is concerned with the psychological nature of a particular kind of perception (T 3.1.1.2-3 / SBN 456). The question he is concerned with in T 3.1.1 is not whether moral statements describe the world or merely express an attitude, but whether our "decisions concerning moral rectitude and depravity" are ideas or impressions, and he argues for the latter (T 3.1.2.1 / SBN 470). The only noncognitivist thesis that can be plausibly ascribed to Hume, therefore, is the claim that moral convictions themselves are not beliefs but rather feelings or sentiments of some sort. I have argued in section 2 that while Hume clearly holds that the *decisions* about morality are impressions (namely, moral feelings), he takes moral convictions to be beliefs – vivid ideas produced by moral feelings. The fact that the beliefs in question are nonsensical (that they consist in the predication of a feeling to an action, person or character trait) does not affect this claim. One may insist that the beliefs in question cannot be strictly speaking false (although Hume himself does not hint at this possibility), but that does not change the fact that they are beliefs (vivid ideas rather than impressions), that they are never true and that they systematically incorporate an error, a misunderstanding of the idea they mobilize.

The projectivist reading of Hume's sentimentalism, therefore, fits remarkably well with the text. The claims at the beginning of the *Enquiry Concerning the Principles of Morals* seem to provide a recalcitrant passage, however. There Hume claims that those "who have denied the reality of moral distinctions, may be ranked among the disingenuous disputants" who are to be brought to the "side of common sense and reason" (2E 1.2).

if that is the case we would have a *reductio ad absurdum* of Hume's argument for the projective view. If that is a real difficulty, it is not one of which Hume was aware. The fact that Hume's projective view is vulnerable to this difficulty does not change the fact that he held this view.

²³ See Ayer (1946, ch. 6).

And it seems that to claim that no moral belief is true is the same as to deny the reality of moral distinctions. In order to accommodate this remark within the projectivist reading we have to hold that the skeptic Hume is denying to engage with in this passage is not the one who denies the truth of moral beliefs but rather one that refuses to draw moral distinctions. Hume identifies the perception "by which we distinguish moral good and evil" with moral impressions, i.e., feelings of disinterested approval or disapproval (T 3.1.1.2 / SBN 456). These feelings, I have argued, directly produce moral beliefs. Obviously, these feelings are not under our control. It is a fact that the contemplation of certain character traits produces approbation and the contemplation of others produces disapprobation. Moral beliefs, as beliefs in general, are, therefore, "more properly an act of the sensitive, than of the cogitative part of our natures" (T 1.4.1.8 / SBN 183). We simply find ourselves drawing moral distinctions, approving and disapproving of certain actions and character traits and having certain moral beliefs. The skeptic Hume is urging us to ignore in the passage quoted above is the one who claims not to make moral distinctions, who claims that when faced with the noblest virtue and the most despicable vice he does not come to believe that the first is commendable and the other vile. That, Hume is claiming, is mere pretend and we need not bother with it.

5. Conclusion

I have argued that the noncognitivist reading of Hume's moral sentimentalism, according to which to have a moral conviction is to have a particular passion instead of a belief, fails because Hume himself acknowledges the possibility of holding a moral conviction without having an impression of the kind that constitutes our praise or condemnation. That is the case of the agent that declares a virtuous enemy vicious while experiencing no feeling of moral disapproval. Given that for Hume perceptions are either impressions or ideas, the fact that in this case there is no impression with which the moral conviction could be identified entails that it is actually an idea, more particularly, a belief.

The dispositionalist reading incorporates the conclusion that moral convictions are beliefs but conceives of them as causal beliefs about the power of certain character traits to produce moral feelings in suitable spectators. That reading also fails, I have argued, because it cannot account

for Hume's claim that we feel rather than infer virtue and vice. The best way in which to understand this claim is as the claim that moral convictions are directly produced by certain impressions, namely, moral feelings, and, therefore, that moral ideas of virtue and vice are copies of those feelings (rather than complex ideas of the power of certain character traits to produce moral feelings).

Rachel Cohon's cognitivist interpretation of Hume's sentimentalism aimed at making (a) the view that moral ideas are copies of feelings compatible with the claims that (b) moral beliefs ascribe certain properties to the actions or character traits they evaluate and (c) that, at least on occasion, they are true. I have argued, however, that these claims are actually incompatible. According to (a) the ideas of virtue and vice are ideas of feelings and then, even if we were willing to admit that the contribution these ideas make to the content of the thoughts they constitute is different in different context, so that according to (b) they could stand for moral properties instead of feelings when incorporated into a moral belief, these belief would all be false for the moral ideas, being ideas of feelings, would not be faithful copies of the properties they stood for.

The shortcomings of these readings reveal that Hume takes moral ideas of virtue and vice to be ideas of feelings of approval and disapproval and that moral beliefs (which conjoin these ideas with ideas of actions or character traits) involve a kind of category error – we take the idea of a feeling to be the idea of a property. The idea of the projection of feelings onto the world is introduced as an attempted explanation of why and how we systematically make this mistake.

This projectivist reading fits rather well with Hume's text. If it is correct, then Hume is an *error theorist*, according to whom moral convictions are beliefs but can never be true. At the same time, however, it sets Hume apart from other error theorist. Error theories, such as Mackie's (1991, ch. 1), usually consist in the conjunction of two independent theses: (a) that moral judgments ascribe a particular quality to the action, person or character trait they evaluate and (b) that as a matter of fact the quality in question does not exist. Clearly that is not Hume's approach. He does not provide an account of our ideas of moral properties and then argues that properties thus conceived do not exist and, therefore, that moral beliefs incorporate false assumptions about reality. His point is rather that once we understand the origin and nature of our moral ideas we can see that they are not ideas of

properties but of feelings and that, therefore, these ideas, as the idea of the necessity involved in causal relations, “represent not any thing, that does or *can* belong to the objects” (T 1.3.14.19 / SBN 164, my emphasis). Moral beliefs are not only false as a matter of fact – they involve a category error that renders them incapable of being true. Hume’s moral error theory, therefore, does not rest upon a metaphysical claim about the furniture of reality. It rests solely upon considerations about the origin and nature of our moral ideas – the proper object of Hume’s science of human nature.

References

- AYER, A. 1946. *Language, Truth, and Logic*. Dover Publications.
- BLACKBURN, S. 1984. *Spreading the Word*. Oxford: Oxford University Press.
- COHON, R. 2008. *Hume’s Morality*. New York: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199268443.001.0001>
- FLEW, A. 1963. On the Interpretation of Hume. *Philosophy*, 38: 178–181.
- HUME, D. 1978. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge e P.H. Niddich. Oxford: Clarendon Press.
- HUME, D. 1983. *An Enquiry concerning the Principles of Morals*. Ed. J. B. Schneewind. Cambridge: Hackett Publishing.
- HUME, D. 1985. *Essays, Moral, Political, and Literary*. Ed. E. F. Miller. Indianapolis: Liberty Fund.
- HUME, D. 2007. *A Treatise of Human Nature*. David Fate Norton and Mary J. Norton (eds). Oxford: Clarendon Press.
- JOYCE, R. 2009. Expressivism, Motivation Internalism, and Hume. In Pigden, C. (ed). *Hume on Motivation and Virtue*. 30-56. Nova York: Palgrave Macmillan.
https://doi.org/10.1057/9780230281158_2
- KAIL, P. J. E. 2007. *Projection and Realism in Hume’s Philosophy*. Nova York: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199229505.001.0001>
- LO, N. Is Hume Inconsistent? – Motivation and Morals. In C. Pigden, (ed). *Hume on Motivation and Virtue*. 57-79. Nova York: Palgrave Macmillan.
https://doi.org/10.1057/9780230281158_3
- MACKIE, J. L. 1980. *Hume’s Moral Theory*. Londres: Routledge.
<https://doi.org/10.4324/9780203359891>
- MACKIE, J. L. 1991. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.

PIGDEN, C. 2009. If Not Non- Cognitivism, Then What? In: C. Pigden (ed). *Hume on Motivation and Virtue*. 80-104. Nova York: Palgrave Macmillan.

https://doi.org/10.1057/9780230281158_4

SANDIS, C. 2009. Hume and the Debate on ‘Motivating Reasons’. In: Pigden, C. (ed). *Hume on Motivation and Virtue*. Nova York: Palgrave Macmillan.

https://doi.org/10.1057/9780230281158_7

SAYRE-MCCORD, G. 1994. On Why Hume’s “General Point of View” Isn’t Ideal – and Shouldn’t Be. *Social Philosophy and Policy* 11(1): 202–228.

<https://doi.org/10.1017/s0265052500004350>

SMITH, M. 2009. The Motivation Argument for Non-Cognitivism. In: C. Pigden, (ed). *Hume on Motivation and Virtue*. 105-120. Nova York: Palgrave Macmillan.

https://doi.org/10.1057/9780230281158_5

STROUD, B. 1993 “Gliding” or “staining” the world with “sentiments” and “phantasms”. *Hume Studies* 19: 253–72.

STROUD, B. 2003. *Hume*. Londres: Routledge.

RESUMO

Neste artigo considero e rejeito uma leitura não-cognitivista do sentimentalismo moral de Hume (segundo a qual ele identifica convicções morais com impressões de um tipo particular) bem como uma leitura disposicionalista (segundo a qual Hume concebe convicções morais como crenças causais a respeito do poder de traços de caráter de produzir certos sentimentos em espectadores apropriados). Sustento que as falhas dessas leituras mostram que Hume é mais bem compreendido como um teórico do erro, de acordo com quem embora convicções morais sejam crenças elas jamais são verdadeiras. Em contraste com teorias do erro contemporâneas, contudo, a tese de Hume não se baseia em uma alegação metafísica para efeito de que não há propriedades morais. Antes, ele sustenta que ideias morais não são ideias de qualidades que possam ser corretamente predicadas de ações ou traços de caráter, mas ideias de sentimentos e que, portanto, crenças morais incorporam sistematicamente um erro categorial.

Palavras-chave: *Hume, Sentimentalismo Moral, Convicções Morais, Projetivismo, Teoria do Erro.*

ABSTRACT

In this paper I consider and reject a noncognitivist reading of Hume’s moral sentimentalism (according to which he identifies moral convictions with impressions of particular kind) as well as a dispositional reading (according to which Hume takes moral convictions to be causal beliefs about the power of character traits to produce certain feelings in suitable spectators). I argue that the shortcomings of these views show that Hume is best understood as an error theorist, according to whom although moral

convictions are beliefs they are never true. In contrast with contemporary error theories, however, Hume's view is not grounded on a metaphysical claim to the effect that there are no moral properties. He holds instead that moral ideas are not at all ideas of qualities that could be truthfully predicated of actions or character traits but rather ideas of feelings and, therefore, that moral beliefs systematically incorporate a category error.

Keywords: *Hume, Moral Sentimentalism, Moral Convictions, Projectivism, Error Theory.*

Recebido em maio de 2019.
Aprovado em janeiro de 2020.