HUME'S CORRECTION OF THE SENTIMENTS. INTERSUBJECTIVITY WITHOUT OBJECTIVITY

MAX BARKHAUSEN
New York University

This paper shows that, in order to understand Hume's sentimentalism and its meta-ethical implications properly, we must turn to his account of the correction of the sentiments required for moral judgment. I begin by discussing an interpretive question that has not yet received the attention it deserves: how should we understand the standard of impartiality to which his account of correction appeals? Once this question receives an answer, we come to see that Hume endorsed not only causal sentimentalism, the view that typical moral judgments are formed in response to moral sentiments, but also constitutive and epistemic brands of sentimentalism: moral sentiments constitute moral correctness and they can serve as an restricted guide to correct moral judgments. As I will argue, the resulting sentimentalist view entails a form of moral relativism.

1. Introduction

Hume's fame as a moral philosopher derives, in large part, from his sentimentalist view that morality is "more properly felt than judg'd of" (T 3.2.2.1; SBN 470). By contrast, his view that our moral sentiments should not serve as an unrestricted guide to moral truth, that, in fact, they stand in need of *correction*, has been appreciated by scholars but has not risen to the same degree of notoriety.

The present paper proceeds from the conviction that, in order to understand Hume's sentimentalism properly, we must turn to his account of the correction of the sentiments. I begin by discussing an interpretive question that has not yet received the attention it deserves: how should we understand the standard of impartiality to which his doctrine of correction appeals?

As we will see, the standard in play in the correction of the sentiments is rooted in the social function of morality, roughly, to accomplish co-ordination on mutually beneficial rules of conduct and to ensure that we act in accordance

Contact: Max Barkhausen <jmb714@nyu.edu>

with such rules even when they conflict with our passions, violent desires and preferences. To the extent that correction is required for judgments both about natural and about artificial virtues, it turns out that both are intimately tied to the co-ordinative function of morality, albeit in slightly different ways.

The understanding of the correction of the sentiments developed here yields an interpretation of Hume's sentimentalism. As we will see, Hume endorsed not only causal sentimentalism, the view that typical moral judgments are formed in response to moral sentiments, but also constitutive and epistemic brands of sentimentalism, according to which our moral sentiments constitute moral truth and moral sentiments can serve as a heuristics to, but not an unrestricted authority on, moral truth.

The present interpretation serves not only to shed light on Hume's sentimentalism but also on his anti-rationalist view that reason cannot discern moral truth: Hume can concede that reason can discern moral truth, but *only* relative to a shared set of emotional dispositions in response to which many of our moral convictions are formed.

Finally, Hume's sentimentalism has an important relativist consequence. On the interpretation developed here, correct moral judgments take moral sentiments as their basis, but they are subject to a standard of impartiality that is built into our moral concepts. Since there is more than one moral system compatible with this standard of impartiality, Hume is committed to a constrained relativism according to which each of these moral systems are equally correct.

These themes will be developed in linear order. Section 2 explains the nature of the impartial standard in play in Hume's account of correction and an interpretation of his sentimentalism; Section 3 contains an interpretation of Hume's anti-rationalism; Section 4 discusses the meta-ethical upshot of this interpretation; Section 5 concludes.

2. Hume's Correction of the Sentiments

2.1. Correction

Hume famously argued that morality is "more properly felt than judg'd of" (T 3.2.2.1; SBN 470). That is, he subscribed to a doctrine that is now called sentimentalism and in some ways goes back to the moral sense theorists, Hutcheson and Shaftesbury.¹ There are (at least) four distinct ways of understanding this doctrine:

Causal Sentimentalism: Typical moral judgments are formed *in response to*

1. The sentimentalist accounts of these moral sense theorists were importantly different from Hume's sentimentalism, of course; in particular, Hume was adamantly opposed to Hutcheson's view that our moral sense cannot be in terms of more familiar moral sentiments.

moral sentiments.

Constitutive Sentimentalism: True moral claims are true *in virtue of* the speaker's or her community's moral sentiments.

Epistemic Sentimentalism: The speaker's or her community's moral sentiments *justify* moral judgments.

Semantic Sentimentalism: Typical moral claims *are about or express* the speaker's or her community's moral sentiments.

For all I know, it is uncontroversial that Hume was a causal sentimentalist. He believed that typically our moral sentiments give rise to our moral beliefs. This is hardly a contentious view: even if, ultimately, morality is subject to rational discernment, it can be agreed that *typical* moral reasoners *typically* form moral beliefs in response to their sentiments.

Most would agree that Hume's sentimentalism amounted to something more than mere causal sentimentalism. There is, however, hardly any consensus about which of the above claims Hume would endorse in addition to causal sentimentalism. While I do not believe that Hume was a semantic sentimentalist,² I will refrain from discussing questions about Hume's view of moral semantics in the present paper. My focus are the constitutive and epistemic varieties of sentimentalism and, below, I will argue that Hume, in addition to being a causal sentimentalist, was in fact a constitutive and, in consequence, an epistemic sentimentalist, on a certain understanding of these views.

The key to understanding Hume's sentimentalism, as we shall see, is his account of the *correction of the sentiments*. Hume recognized that typical moral subjects hold one another to the expectation that their full-fledged, reflective moral judgments, unlike moral sentiments, be impartial, and often their judgments in fact are more impartial than the sentiments on which they are based.³ In what follows, I reserve the terms "judgment" and "belief" for full-fledged, reflective moral judgments, in contradistinction to whatever evaluative judgments may be part and parcel of our partial sentiments. This presupposes that, for Hume, there is a distinction between moral sentiments, on the one hand, and moral judgments proper, which rules out those non-cognitivist interpretations of Hume according to which our moral judgments do not amount to beliefs at all, and which have generally fallen out of favor.⁴

2. On this question, the reader might consult Cohon (2008).

3. Hume provides an extensive discussion of the partiality of our sentiments in T 3.1.1.

4. I agree with Garrett that our moral sentiments and our considered moral judgments should be understood as distinct kinds of moral evaluations. See Garrett (1997: 196). The view that Hume was a cognitivist in the sense assumed here is now mainstream. For a thorough defense of this view, see Cohon (2008).

For example, I resent the thief who stole my bicycle more than any other bike thief, but I believe that, all else being equal, all bike thieves deserve the same amount of resentment. Similarly, my sympathy concerns my family and peers—my ‘narrow circle’, as Hume would say—more than strangers. At the same time, I believe that certain moral privileges accrue to all moral agents equally, independently of our relationship and I approve of the same traits in my enemies as in members of my narrow circle, even if they are harmful to me.

This difference between the moral sentiments and our all-things considered moral judgments poses an explanatory demand on sentimentalists like Hume. Given that, on any understanding of sentimentalism, our partial sentiments are intimately related to our moral beliefs, the sentimentalist must explain how we succeed in reaching moral judgments more impartial than our sentiments.

It is for this reason that Hume introduces his doctrine of the *correction* of the moral sentiments.⁵ According to this doctrine, we learn to correct for the partiality of our sentiments in forming moral beliefs. The psychological mechanism at work in this correction, Hume says, is analogous to the the mechanism by which we form stable perceptual beliefs about the physical properties of the objects that we perceive.⁶ In the perceptual case, correction for perspectival and circumstantial distortions is required: objects may seem small or large to us depending on their distance, surfaces elliptical or circular depending on perspective, red or green depending on lighting. In the moral case, the moral sentiments vary with psychological contiguity; here, we learn to correct for the ways in which our sentiments are influenced by our psychological ties to our narrow circles.

An important question concerns the nature of the impartial standard in play in the correction of sentiments. Is this standard itself a matter of sentiment? Is it based on independent moral views? Conventions? To supply a correct understanding of this standard is a crucial and difficult task for those who believe that Hume espoused constitutive or epistemic sentimentalism. Recall that constitutive sentimentalism is the view that true moral claims are true *in virtue of* our moral sentiments or, in other words, that our moral sentiments are the truth-makers of moral claims. But if the truth-makers of moral claims themselves are sentiments, then how can it be that we need to correct for our sentiments in order to reach correct moral judgments? Epistemic sentimentalism, recall, is the view that our moral sentiments justify our moral judgments. If this is so, it ought to be explained, similarly, why we sometimes ought to correct for distortions in our evidence and what, precisely, renders some of the ‘data’ provided by our sentiments aberrant.⁷ While one might, for these reasons, perceive a tension be-

5. This doctrine is introduced in T 3.3.1.

6. See T 3.3.1.16; SBN 582.

7. Cohon (2008: 134 ff.) raises similar questions. Below, I will say more about how my interpretation relates to hers.

tween Hume's account of the correction of the sentiments on the one hand and constitutive and epistemic sentimentalism on the other, I believe that, in fact, the best interpretation of the correction of the sentiments favors a version of each doctrine. In the following, I will provide this interpretation.

2.2. *Moral Concepts*

The rules according to which we correct moral sentiments are intimately tied to our use of moral language and the rules that govern it, in turn. Hume emphasizes the importance of moral (and aesthetic) language whenever he turns to the correction of the sentiments and the possibility of consensus to which it gives rise. He stresses that correction for the biases of the senses is a common feature of all language, including moral language:⁸

Such corrections are common with regard to all the senses; and indeed it were impossible we could ever make use of language, or communicate our sentiments to one another, did we not correct the momentary appearances of things, and overlook our present situation. (T 3.3.1.16; SBN 583)

In "Of the Standard of Taste", Hume emphasizes, time and again, that, when our sentiments differ, consensus about matters of beauty and deformity is often owed to features of our discourse; in particular, he claims that the members of a linguistic community must agree in their applications of terms that import blame and praise:

But we must also allow that some part of the seeming harmony in morals may be accounted for from the very nature of language. The word virtue, with its equivalent in every tongue, implies praise; as that of vice does blame: And no one, without the most obvious and grossest impropriety, could affix reproach to a term, which in general acceptance is understood in a good sense; or bestow applause, where the idiom requires disapprobation.

There are certain terms in every language, which import blame, and others praise; and all men, who use the same tongue, must agree in their application of them. (EMPL 227–228)

A term that is understood in a good sense, for Hume, is one that commits us, by means of a linguistic requirement,⁹ to approve of the things to which it refers. Examples would be "good" or "beautiful"; to say that something is good or beautiful commits the speaker to certain forms of approbation. Were he to fail to

8. On the consensus-building role of moral language, see Jensen (1977), Garrett (1997) and Garrett (2015).

9. Garrett (2015: Chapter 5) argues that this is one of the distinctive features of the conceptual role of normative terms.

approve of things good or beautiful, she would have made a linguistic mistake. In the second half of the quote Hume claims that it is a precondition of this kind of language that ordinary speakers must agree in their applications of it.

In the second *Enquiry*, Hume adds that evaluative concepts that import praise and blame are grounded in the general interests of the community:

General language ... must affix the epithets of praise or blame, in conformity to sentiments, which arise from the general interests of the community. (EPM 5.42; SBN 228)

In order to proceed, we ought to briefly consider Hume's account of abstract ideas or, alternatively, notions or, as we might say, concepts. As Garrett has argued in detail,¹⁰ what it is to apply such concepts correctly and, we may assume, initial formation of these concepts, consists of four stages.

In the case of moral concepts (and other sense-based concepts),¹¹ the initial stage consists in the repeated activation of a response, such as feelings of approbation and disapprobation in the moral case. Then, in the next stage, the individual recognizes the external features that tend to give rise to these responses or, on Hume's account, the similarities in the ideas that give rise to a specific response. At this stage, a moral reasoner may be very much like a child who takes the meaning of "just" to be roughly synonymous with "disagreeable to me".¹²

The next and crucial stages require correction and abstraction. For example, in the case of perceptual judgments, we come to recognize that sometimes the response is present without the feature. Some objects look red even though they are not. Others are red even though they have never looked red to the perceiver. Correction and abstraction are required in order to generate what Don Garrett has called the "revival set" of a concept: the set of ideas, both possible and actual, to which it applies or, as we might say, its intension.

Our question, then, concerns the standard for correction in the moral case, specifically. What do we correct for, given that there is no standard independent of our moral sentiments that classifies some of our moral sentiments as aberrant and others as called for? In order to zone in on this question, we ought to better understand the ways in which our initial reactions are skewed and in need of correction. As Hume emphasized, a central *precondition* of our ability to reach moral consensus is that, in moral judgments, we overcome the partiality of the moral sentiments in response to which moral views are often formed.

We may distinguish between two ways in which our sentiments are partial, both of which Hume discusses under the label of partiality. The first concerns

10. See Garrett (1997: 197 ff.) and for a detailed and developed account, Garrett (2015: Chapters 4–5).

11. See Garrett (2015: Chapter 5) on causation and probability.

12. Cf. Piaget (1965).

the narrow scope of our moral sentiments: they tend to concern our narrow circle and not to generalize to those outside our narrow circle. Our emotional dispositions are limited. We cannot, for example, sympathize with every complete stranger or resent every wrongdoer. We are even more obviously incapable of sympathizing or resenting those who have never impinged on our consciousness at all.

The second kind of partiality corresponds most closely to the notion of partiality deployed in everyday discourse, such as when we speak of a partial referee in a sports match: our sentiments are likely to side with ourselves and our narrow circle just as a referee is likely to side with the team that has paid him off, or just as I am likely to resent the bicycle thief who stole my bicycle more fervently than any old bicycle thief. In general, I will resent others for actions that I, myself, do not feel guilty about when I subject others to their effects, and I am likely to sympathize with the plight of my narrow circle more than with that of strangers.

Hume does not explicitly formulate the general rules for correction of emotional 'biases' in reflective moral judgment. However, if he had wanted to give an explicit formulations of these, he could have told us not to assign different rights, permissions or obligations to different agents and not to evaluate their characters differently unless, at bottom,¹³ we can point to a relevant non-moral difference between these agents or their circumstances. He might have gone on to tell us that when a moral difference in rights, permissions, obligations, or character is explained in terms of a non-moral difference, the same non-moral difference ought to be correlated with the same moral difference throughout the reasoners' beliefs and assertions (all else being equal). Let us call these requirements *systematicity* and *impartiality*, respectively. Hume could easily have introduced these as general rules of moral thought and language explicitly, just as he explicitly formulated the rules that ought to govern causal reasoning. Each is constitutive of moral discourse and thought because it is a precondition of the co-ordinative function of moral language and thought.¹⁴ This is why Hume says that concepts that import praise and blame are rooted in the general interests of the community.

We may explain the role that these constraints play in moral conversation

13. The thought behind this qualification is that we often explain moral differences in terms of further moral differences, but that any non-moral difference needs to bottom out in a non-moral difference.

14. An aside: Hume here anticipates the treatment of moral supervenience advocated by contemporary non-cognitivists, such as Hare, Gibbard and Blackburn. See Gibbard (2003: Chapter 5), Blackburn (1988b; 1984; 1988a) and Hare (1952: 52). On these contemporary approaches, supervenience is simply a conceptual requirement on moral thought and language, and one of the functions of this conceptual requirements is to promote consensus and co-ordination.

in somewhat anachronistic terms while staying true to Hume's spirit. Each constraint ensures that, in conversation, we correct for one of our emotional biases. Each is a constraint on the conversational score,¹⁵ for each regulates the beliefs that a speaker may voice in conversation, taken together. For example, Al's belief that Bob oughtn't to be promoted is not partial taken individually. By contrast, his beliefs (or assertions) that he himself should be promoted and Bob shouldn't be, taken together, would be partial if he could not explain why his achievements, but not Bob's, justify the promotion.

Note that impartiality is not to be confused with a demand for equality, according to which everyone ought to give equal weight to everyone else's preferences or according to which everyone counts the same under all circumstances.¹⁶ For instance, Al does not fail to be impartial because he believes that everyone is permitted to act selfishly to some extent or that, if there were a fire at the White House, the president ought to be saved first. Though he would be partial if he thought that he, but nobody else, is permitted to act selfishly yet proved unable to explain why, or if he thought that, were the White House to burn, Barack Obama ought to be saved first without being able to point to some property of Obama's that explained this privilege, such as his being the president.

Instead of formulating the impartiality and the systematicity requirements explicitly, Hume chose a different emphasis when, in describing the general standpoint,¹⁷ he gave us a description of a judge whose moral judgments would conform to these rules precisely because her emotional dispositions would be favorable to their prescripts. Famously, Hume argued that moral approval is, and ought to be, determined by the sentiments that we experience from a detached, intersubjective perspective. He believed that, in assuming this standpoint, we take what he called "the general survey".¹⁸ This involves experiencing sympathy from various perspectives and not just our own. Relatedly, our moral judgments are supported by what Hume calls "extensive sympathy" in the *Treatise* and "the principle of humanity" in the second *Enquiry*, a capacity to empathize with and show concern for any human being, which is easily roused by the imagination.¹⁹

15. The notion of a conversational score was first introduced in Lewis (1979).

16. The term partiality is often used in this way in the literature on social choice and contractualist accounts for morality as, for instance, in Gauthier (1986).

17. The reader might turn to Sayre-McCord (1994) for an in-depth discussion of the psychology of the general standpoint and to Abramson (2001) and the citations therein for discussions of the role of sympathy in particular. Baier (1991) provides insightful discussions of nearly all aspects of the general standpoint.

18. See especially T 3.3.3.4

19. As, for instance, in this passage from the second *Enquiry*: "Virtue, placed at such a distance, is like a fixed star ... Bring this virtue nearer, by our acquaintance or connexion with the persons, or even by an eloquent recital of the case; our hearts are immediately caught, our sympathy enlivened, and our cool approbation converted into the warmest sentiments of friendship and regard" (EPM 5.43; SBN 230).

A reason for this difference in emphasis is a facet of Hume's moral psychology: he thought that typical moral reasoners are in fact capable of occupying the general standpoint and that this emotional exercise is our best means of reaching impartial and systematic moral judgments. But we can accept the view that systematicity and impartiality are constitutive of our moral concepts independently of Hume's moral psychology, and we may understand Hume's theory of the general standpoint as encompassing both an account of the rules constitutive of moral thought and language and of the psychology that explains our ability to conform to these rules. So we can accept Hume's account of moral concepts without agreeing with every facet of his moral psychology.

However, there is a further philosophical reason for Hume's emphasis on the general standpoint, which he himself does not highlight. The impartiality and the systematicity constraints do not go all the way to securing moral consensus, especially not towards consensus with those outside our narrow circles. This is because impartiality and systematicity impose no constraints on what counts as a *relevant* non-moral difference. We may assume that, within our narrow circles, moral sentiments will at least to some extent determine what counts as a morally relevant property: those traits that we approve of and disapprove of before correction sets in.

Since our sentiments vary with psychological contiguity, we may not assume that they will similarly determine the morally relevant properties when it comes to co-ordinating with those outside of our narrow circles. Systematicity and impartiality are, for instance, perfectly compatible with the view that those who are part of *my* narrow circle should enjoy certain privileges, that those who harm members of *my* narrow circle are evil but not those who harm members of other moral circles, and so on. Nonetheless, the reason why we accomplish coordination with those outside our narrow circles has to do with Hume's account of the general standpoint.²⁰ The sentiments that are distinctive of the general standpoint are required to render certain features, but not others, salient for generalization.

An important aspect of the general standpoint is our ability to empathize with others. This is what Hume calls sympathy. Sympathy endows us with a mutual understanding of the (non-moral) features to which our moral sentiments respond. It is this mutual understanding that renders certain non-moral properties salient when it comes to forming impartial and systematic moral views. For this reason, salience depends on our ability to experience sympathy.²¹ The

20. My understanding of the general standpoint is indebted to Mackie (2004: 129 ff.), Baier (1991: 179 ff.), Sayre-McCord (1994) and Sayre-McCord (1995). These scholars reject objectivist interpretations of the general standpoint and each provides an elucidating discussion of the speculative psychology that Hume's discussion of the general standpoint encompasses.

21. The sense of salience at work here is the informal one that Thomas Schelling first

morally salient properties are, in other words, those to which our moral sentiments respond in our respective narrow circles and they are rendered salient by our ability to experience sympathy and our mutual knowledge of this ability.

Cohon (2008: 139 ff.) has recently argued that impartial moral judgments are based on the feelings that we experience while assuming the general standpoint. This interpretation is, in part, motivated by her thesis that Hume subscribed to the moral sensing view, the view that “our basic awareness of vice and virtue is a direct apprehension by feeling” (Cohon 2008: 103). While I sympathize with many aspects of Cohon’s interpretation, we do not in fact have to assume that our impartial moral judgments are apprehensions of feelings that we experience only from the general standpoint. Rather, impartial moral judgments proceed from partial moral sentiments exclusively. Strictly speaking, the general standpoint is required only because sympathy renders certain properties salient for generalization between us and those outside our narrow circles. This is a plausible aspect of the resulting view because it allows Hume to say that impartial moral judgments transcend our capacity for pro-social sentiments in an important sense: our moral views concern strangers far removed from our narrow circles, future generations and generally many more people than we could plausibly be said to have feelings for.

Causal reasoning, on this account, does play a role in the formation of ordinary moral judgments, albeit a limited one. For determining whether one’s moral views are impartial and systematic might require inferential judgments, especially when, as is sometimes the case, our moral sentiments do not follow language in the process of correction. This is what Hume has in mind when he says in T 3.3.1 that “we have found [reason] to be nothing but a general calm determination of the passions, founded on some distant view or reflection” (T 3.3.1.18; SBN 583). But note that, even in these cases, reason alone is incapable of discerning the moral judgments in question. For they are still based on initial partial sentiments, just as our perceptual beliefs may be based on perceptual illusions that we have recognized as in some ways misleading.

2.3. *Intersubjectivity without Objectivity*

In my interpretation of Hume’s doctrine of the correction of the sentiments above, I have freely used words such as “bias” or “correct”. This language mirrors Hume’s claims that “experience soon teaches us this method of *correcting* our sentiments” (T 3.3.1.16; SBN 582) and that such “corrections are common with regard to all the senses” (T 3.3.1.16; SBN 583). It might be thought that this rhetoric calls for the notion of an *objective* standard for moral judgment. For

explored in his seminal Schelling (1960), albeit in the context of salient actions in co-ordination games.

our purposes, we can say that a standard for moral judgment is objective if it is counterfactually independent of our moral sentiments, attitudes, beliefs and conventions.²² Perhaps the most striking consequence of the view that there is an objective standard for moral judgment is that, if our moral sentiments were substantially different and our moral judgments differed substantially from our actual judgments, we would be mistaken.

Hume's definitions of virtue and his proto-utilitarian views give rise to at least some temptation to understand Hume in objectivist terms. As Garrett²³ points out, Hume defines virtue as "mental qualities, *useful* or *agreeable* to the *person himself* or *others*" (EPM 9.1; SBN 268) and, alternatively, as "whatever mental action or quality gives to a spectator the pleasing sentiment of approbation" (EPM Appendix 1.10; SBN 289). Hume also declares of morality that "her sole purpose is to make her votaries and all mankind, during every instance of their existence, if possible, cheerful and happy" and he argues that it "appears to be a matter of fact that the circumstances of *utility*, in all subjects is a source of praise and approbation" and that utility is "the *sole* source of that high regard paid to justice, fidelity, honour, allegiance, and chastity" (EPM 5.44; SBN 231). One way of understanding these 'definitions' of virtue is to interpret them as an objective standard for moral correctness.²⁴

On this interpretation, what makes an action right or wrong or a character good or bad is whether it conforms to a utilitarian standard, independently of our sentiments. If our sentiments were radically different, this would in no way threaten its existence or undermine its authority; rather, judgments based on radically different sentiments would be incorrect if they failed to track it. This objectivist understanding of Hume is compatible with causal sentimentalism and, in a sense, with epistemic sentimentalism, at least to the extent that our sentiments are in fact in conformity with the objective standard. At the same time, the justificatory force of the moral sentiments would be highly contingent. Clearly, objectivism entails the negation of constitutive sentimentalism, as the standard for correction is counterfactually independent of our moral sentiments.

Many will agree that objectivism is not a compelling interpretation of Hume.

22. This is often called mind-independence in contemporary meta-ethics. Many think of mind-independence as the very mark of objectivity. See, for instance, Driver (2012).

23. These definitions, Garrett notes, do allow us to infer whether an act is virtuous by means of causal reasoning, but they must not be understood as analyses of the meaning of moral terms or utterances; that, Garrett points out, would be anachronistic. See Garrett (1997: 201). Nor does Garrett think—and this becomes especially clear in his recent book (see Garrett 2015)—that the standard in question is independent of our sentiments.

24. As do utilitarians who believe that they follow in Hume's footsteps. A famous example is Bentham (1988: XIV), where Bentham alleges that Hume himself has demonstrated that "the foundations of all virtue are laid in utility." Glossop (1967) argues that "the essential doctrines of Utilitarianism are stated with a clearness and consistency not to be found in any other writer of the century" (87).

It is incompatible with the most interesting versions of sentimentalism with which Hume is often credited. Moreover, it is in tension with Hume's famous claim in T 3.1 that moral judgments cannot be derived from reason. This conclusion, as well as Hume's most (in)famous argument in its favor, are summarized concisely in the following passage:

Since morals, therefore, have an influence on the actions and affections, it follows, that they cannot be derived from reason; and that because reason alone, as we have already prov'd, can never have any such influence. Morals excite passions, and produce or prevent actions. Reason of itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason. (T 3.3.1.10; SBN 457)

Let us call the argument presented here Hume's argument from the *motivational inertia of reason* and its conclusion Hume's *anti-rationalism*. The argument turns on moral motivation: moral judgments come along both with emotional dispositions as well as motivations to act but neither can be the product of reason. Therefore, Hume argues, moral judgments cannot be based on reason. If objectivism is true, the epistemological import of Hume's anti-rationalism remains mysterious. If there is an objective standard for correct moral judgment, it seems that causal reasoning should have access to it. In fact, it should be our tool of choice in discerning and justifying moral truth.

These may not be decisive arguments against the objectivist interpretation, but note that we can avoid these difficulties by interpreting Hume as a non-objectivist. The interpretation of correction provided above provides the resources to do so. Impartiality and systematicity do not by themselves determine whether a moral view is correct; in fact, they only constrain permissible moral views given moral sentiments as inputs and given a set of morally salient non-moral properties that are determined by our partial sentiments and our capacity for sympathy. The standard for correction is therefore neither independent of our contingent emotional dispositions, nor could we determine, by means of causal inference, whether a moral claim conforms to the standard in question without taking our moral sentiments as premises.

It turns out, then, that the interpretation of Hume's correction of the sentiments given above entails a brand of constitutive (and, in consequence, epistemic) sentimentalism. For, on this view, our moral sentiments go into determining correct moral judgments, in conformity with the impartiality and the systematicity requirements. Accordingly, moral sentiments play an important justificatory role, even though the epistemic relation between the moral sentiments and moral judgments is not so straightforward as to license immediate inferences from, say, disapproval to negative moral judgments, because correct moral judgments are mediated by the requirements discussed above and our ca-

capacity for sympathy.

Hume's standard for correct moral judgment, on this account, is not objective, but, assuming that the moral sentiments of different moral subjects in a community are sufficiently similar, it provides an intersubjective basis for moral thought.²⁵ This is a very attractive feature of his view. As was noted above, the central function of moral language and thought, for Hume, is social co-ordination. What social co-ordination requires is intersubjectivity, but not objectivity. In other words, what Hume needs to supply is a standard of moral correctness that applies to each of us, given our shared psychological make-up, conventions and circumstances, but not a standard that applies independently of these features. This, on the interpretation given here, is precisely what he did.

Hume's standard of correction does not apply to the sentiments on which our moral judgments are based; that is, it does not require us to change these sentiments. Rather it imposes constraints on our moral views, *given the sentiments on which they are based*. Importantly, this is not to say that our moral views do not, in a sense, require us to have certain sentiments in turn. In fact, Hume explicitly argues that certain terms in our language commit us to sentiments like approbation and blame when he says that "no one, without the most obvious and grossest impropriety, could affix reproach to a term, which in general acceptation is understood in a good sense; or bestow applause, where the idiom requires disapprobation" (EMPL 227-228) and that "every tongue possesses one set of words which are taken in a good sense, and another in the opposite ..." (EPM 1.10; SBN 174). Recall that words that are taken in a good sense, for Hume, are those that, when used in declarative sentences, commit the speaker or hearer to approbation. But these social sentiments to which our moral judgments commit us, we may assume, are distinct from the partial and unsystematic sentiments on which they are based.

It will be helpful to consider a passage from "Of the Standard of Taste" and forestall a possible objection to the present view. In this passage, Hume distinguishes between healthy and deficient sentiments on which our aesthetic judgments are based and, at least if we plausibly assume that Hume's views on aesthetic judgments are broadly analogous to his views on moral judgments, this may seem to contradict my non-objectivist interpretation of correction and the claim that for Hume the sentiments on which our moral views are based do not themselves stand in need of correction:

In each creature, there is a sound and a defective state; and the former alone can be supposed to afford us a true standard of taste and sentiment. If, in the sound state of the organ, there be an entire or a considerable uniformity of sentiment among men, we may thence derive an idea of the

25. Cf. the discussion of intersubjectivity in Cohon (2008: Chapter 5).

perfect beauty; in like manner as the appearance of objects in daylight, to the eye of a man in health, is denominated their true and real colour, even while colour is allowed to be merely a phantasm of the senses. (EMPL 233–234)

Hume claims here that, with regard to taste, we may define a “true” standard of taste in terms of the sentiments of a sound, as opposed to a deficient, subject. It is tempting to interpret “true” as objective in the sense defined above. Moreover, it is easy to imagine Hume claiming that, similarly, we can define a “true” standard of virtue in terms of the reactions of a sound subject. But would this not contradict the view that it is not the sentiments themselves that must be corrected and, worse, doesn’t it imply that there is a standard for the correction of sentiments independent of what we feel and think?

I think not. In Hume, the distinction between sound and deficient emotional dispositions should be understood in terms of the function of our moral practice: while sound emotional dispositions lead to views that promote the co-ordinative function of morality, deficient ones do not. If emotions are deficient or sound, they are so relative to the function of morality; but this does not mean that, unless we already desire to sustain this function, we are rationally compelled to develop the emotional dispositions that will help us do so or pass the judgments that they support. This is, of course, in keeping with Hume’s famous remark in T 2.3.3 that it is “not contrary to reason to prefer the destruction of the whole world to the scratching of my finger” (T 2.3.3.6; SBN 416).

The standard of correction here explained has its basis in the social function of our moral practice. For this reason, the correction of the sentiments is in some ways similar to the artificial virtues. Let me explain. It is well-known that Hume conceived of the artificial virtues in terms of their function of social co-ordination: roughly, to accomplish co-ordination on mutually beneficial rules of conduct and to ensure that we act in accordance with such rules even when they conflict with our passions, violent desires and preferences. This understanding of the artificial virtues permeates, for example, Hume’s discussion of property norms and gives rise to his famous metaphor of the two men who are each pulling the oars of a boat for mutual benefit (T 3.2.2.10; SBN 490).

As we have seen, impartiality is crucial to securing intersubjective consensus on both natural and artificial virtues. So, like the artificial virtues, the rules of moral thought and language discussed above are vital to the co-ordinative function of our moral practice and conformity is in our mutual interest. Hence, we may understand the correction of the sentiments and the rules that govern it in terms of their co-ordinative function. Yet, unlike the artificial virtues, the rules of correction do not govern social behavior directly but, rather, correct application of moral concepts in accordance with the social function of morality in general. Insofar as correction applies both to judgments concerning the natural and the

artificial virtues, it follows that, to some extent, both are ultimately rooted in the same social function.

An interesting question is whether failing to correct for partiality in our applications of moral concepts would be irrational.²⁶ For example, would a person be irrational if he took his moral sentiments at face value and passed moral judgments partial to his narrow circle in ways ruled out by the standard of correction explained above?²⁷ Since rationality, for Hume, is a function of the agent's desires, this of course depends precisely on the motivational make-up of the agent. Now, as we just saw, correction is required in order to reach moral consensus and hence in order for us, as a collective, to settle on mutually beneficial rules of conduct. Being a party to the mutually beneficial co-operative schemes to which correct usage of moral language leads is arguably in most of our interests and, precisely for this reason, for most of us, our desires render correct use of moral language rational.

3. Anti-Rationalism

The interpretation of Hume's account of correction advocated here can shed light on his anti-rationalism and his argument from the inertia of reason. Consider, first, his claim that moral distinctions are not derived from reason. Just how this claim should be understood is a matter of dispute. On what Rachel Cohon calls the traditional understanding of Hume, he argues from judgment internalism, the view that moral judgments necessarily come along with some degree of motivation to act in accordance with them, and the motivational inertia of reason against the view that moral judgments are beliefs or have truth-conditions. As a consequence, Hume, on the traditional interpretation, is made out to be an old-fashioned expressivist who believes that moral judgments are neither true nor false but are expressions of sentiment. One major difficulty this view faces is that it appears to be at odds with the correction of the sentiments because, if moral judgments are mere expressions of emotion, it is unclear why, in forming moral beliefs, we ought to correct for the partiality of our sentiments. Other (to my mind, convincing) reasons have been given in the literature.²⁸ Whether or not these arguments are decisive, they provide sufficient motivations to interpret Hume's anti-rationalism in a way that does not foist the traditional view on us.²⁹

An alternative understanding of Hume's anti-rationalism that avoids the traditional interpretation is at hand once we take into account his functional ap-

26. Thanks to an anonymous referee for raising this question.

27. Friends tell me that "unfair!", for their young children, is roughly synonymous with "I don't like it!"

28. See especially Cohon (2008: Chapter 1).

29. Cf. Cohon (2008).

proach to morality. It is intimated in the following quote from the second *Enquiry*:

Extinguish all the warm feelings and prepossessions in favour of virtue, and all disgust or aversion to vice: Render men totally indifferent towards these distinctions; and morality is no longer a practical study, nor has any tendency to regulate our lives and actions. (EPM 1.8; SBN 172)

There are two ways of understanding Hume's claim that, in the absence of feeling, we would be indifferent to moral distinctions. One is that our recognition of moral distinctions would lose its motivational force and the other that we would be incapable of recognizing moral distinctions at all. We must appreciate this in order to understand Hume's argument from the inertia of reason. For it is only meant to establish the former claim. In fact, Hume denies the latter claim, at least on a certain reading. There is in fact a sense in which reason alone—more specifically, causal reasoning—can discern moral distinctions:

The only object of reasoning is to discover the circumstances on both sides, which are common to these qualities; to observe that particular in which the estimable qualities agree on the one hand, and the blamable on the other; and thence to reach the foundation of ethics, and find those universal principles, from which all censure or approbation is ultimately derived. As this is a question of fact, not of abstract science, we can only expect success, by following the experimental method, and deducing general maxims from a comparison of particular instances. (EPM 1.10; SBN 174)

Taken out of context, Hume's denial that reason can discern moral truth and his ambition to "find those universal principles, from which all censure or approbation are ultimately derived" may appear puzzling. It may also seem puzzling that he denies that moral questions are questions of fact yet finds it obvious that universal moral principles are empirical facts that we can discern following the experimental method. But what Hume acknowledges in the above quote is the possibility of true inductive generalizations about the moral judgments that moral reasoners will condone given their emotional dispositions and their implicit knowledge of the rules of moral discourse and thought, and hence he allows that reason can discern the moral claims that would be acceptable to a moral reasoner or a group of moral reasoners.³⁰

These judgments are correct relative to the underlying emotional dispositions of the group, but not independently of these dispositions.³¹ We can discern the intension of our moral concepts by means of causal reasoning that proceeds from

30. Thanks to Rachel Cohon for helpful discussion.

31. In Garrett (2015: Chapter 4), Garrett draws a distinction between productive and

our emotional dispositions and our implicit knowledge of the general rules that regulate moral language and thought. For example, given our emotional make-up and the rules constitutive of moral thought and language we will by and large condemn wanton cruelty and judge that wanton cruelty is wrong. Anyone familiar with our psychology, socialization and linguistic practice is in a position to believe this generalization about us and may therefore come to recognize inductively that wanton cruelty is in the common ground of 'wrongness'. However, what gives a moral concept its intension are these emotional dispositions and general rules.

To inquire into the 'universal principles' of moral judgment is, in fact, one of Hume's important concerns; it is the project he is engaged in when he tries to discern what all virtues have in common and that ultimately leads him to the 'definitions' of virtue and the proto-utilitarian statements discussed in Section 2. These, therefore, should be understood as claims that describe, at a very high level of generality and from a perspective that takes into account the coordinative function of our moral practice, the extension of our moral concepts relative to our actual sentiments, and not as providing an objective standard of moral correctness independent of these sentiments.

It is because of Hume's functional approach that he dismisses the "monkish virtues"³²—celibacy, fasting, penance, mortification, self-denial, humility, silence, solitude—as virtues. For while some of us may be disposed to call these virtues, they are not good for anyone and therefore detached from the beneficial function of our moral practice. Finding universal principles of moral judgment, then, is an inquiry into our dispositions to judge, but dispositions that are not suitably related to the function of our moral practice may be discarded as insignificant when it comes to determining the intension of moral concepts. So the enterprise is not solely descriptive but is best understood in light of Hume's functional understanding of morality and moral concepts. It is, as it were, a normatively charged project in the social sciences: to discern the character traits that we regard as virtuous and to explain how they sustain co-operative behavior; if a character trait that is regarded as a virtue fails to have a function, it is dismissed as deficient. Moral concepts, for Hume, are essentially functional concepts.

Anyone can discern the intension of moral vocabulary without thus incurring a rational obligation to apply the vocabulary to instances in its common ground. Let me use an analogy in order to illustrate this view: enlightened subjects were quite aware of how the term "witch" was used centuries ago but they were under

reactive definitions of moral concepts: we can either define moral concepts in terms of our emotional reactions (reactive) or, alternatively, in terms of the features that give rise to these reactions (productive). Making use of Garrett's helpful distinction, my point is that, if it weren't for the underlying shared reactions, the possibility of giving a productive definition by means of causal reasoning would be forestalled.

32. See (EPM 9.3; SBN 270) for Hume's discussion of the monkish virtues.

no rational pressure to call anyone a witch (prudential reasons resulting from social pressure notwithstanding). Nor would their knowledge of the common ground of “witch” motivate them to believe in sorcery or to kill innocent women. Similarly, knowledge of the common ground of moral concepts does not provide a justification for our moral beliefs or for our moral practice.

In summary, there are two distinct arguments in the much-debated passages on the motivational inertia of reason. First, while Hume acknowledges that reason can yield correct beliefs about the origins of moral evaluations and the common ground of our moral concepts, these beliefs are not moral beliefs proper because they are not grounded in sentiments with motivational force. The tie between moral judgment and motivation is best explained in terms of the hypothesis that moral beliefs are by and large formed in response to the moral sentiments. That moral beliefs are not by and large products of inductive generalization but are, rather, formed in response to moral sentiments, is an essential feature of our moral practice. If our practice lacked this feature, it would not be a moral practice, understood in functional terms, because we would fail to be motivated to act in accordance with our moral beliefs. As mentioned above, it should be noted that abductive reasoning and induction play some, if a limited, role in the formation of moral judgments. For these tools of reason may be required in order to ensure that the moral beliefs accord with such rules constitutive of moral language and thought as impartiality and systematicity.

Second, causal reasoning can only discern what moral judgments are correct relative to a shared core of moral sentiments because Hume endorses constitutive sentimentalism. In other words, if it were not for the moral sentiments, there would be nothing that reason can discern. This, I believe, is what Hume means when he says that “the rules of morality ... are not conclusions of reason” (T 3.3.1.10; SBN 457), for the sentiments go into Hume’s explanation of our dispositions to apply moral judgments and even though reason can discern what corrections our sentiments require, the standard of correctness in question only applies to moral reasoners who come to the table of moral debate with a set of emotional, intuitive and partial moral evaluations.

4. Constrained Relativism

Moral concepts, for Hume, are functional concepts that serve the purpose of mutually beneficial co-ordination on rules of conduct. How does co-ordination work on Hume’s picture? Roughly, our moral sentiments lead us to adopt some moral rules that prescribe behavior beneficial to those in our narrow circles. So a degree of co-ordination with those closest to us is accomplished without much reflective effort, simply in virtue of the emotions we experience towards them. Co-ordination with those outside our narrow circles is accomplished by means

of our following further rules that are constitutive of moral language, rules that require us to correct for the partiality and lack of systematicity in our moral sentiments. The standard of correct moral judgment, we have seen, is constituted in part by our moral sentiments and in part by general rules: the systematicity and the impartiality requirements.

An important entailment of the resulting meta-ethical view is a denial of moral objectivity: what is right or wrong, good or evil, and so on, depends on our emotional dispositions and linguistic conventions. If either were different, different moral judgments would be correct. But even if we hold our emotional dispositions fixed, it is very plausible that there are numerous ways of co-ordinating on mutually beneficial sets of rules in accordance with the systematicity and impartiality requirements. The moral rules on which we have actually converged single out only one among many ways in which we might have co-ordinated on mutually beneficial rules of conduct, given shared emotional dispositions on which our reflected moral judgments are based. Another way of saying this is that there are numerous determinate sets of moral views that we could have adopted and which would have served the co-ordinative function of our moral practice equally well, given these dispositions.

It will help to illustrate this with an example. The hawk-dove game is often understood as an illustration of the need for social co-ordination with regards to property. In this game, two parties compete for a resource and each has two strategies available: *hawk* and *dove*. Hawks are willing to fight, while doves are concessive. When a hawk meets a dove, the dove will concede and the hawk gets all. When two doves meet, they will share the resource equally. When two hawks meet, they will fight, each incur a cost and do worse than two doves willing to share. Imagine now that the game represents competition for territory, and imagine there that the players can occupy either the role of owner or intruder. There are now more complex strategies, such as the 'Bourgeois strategy', *play hawk if owner, dove if intruder*, and the 'Robin Hood-strategy', *play dove if owner, hawk as intruder*.

It can be shown that each of these strategies is socially optimal—a population all of whose members play one of these strategies does better than any other population. We can think of a population playing the Bourgeois strategy as a population that has implemented property norms fairly close to our own. The important point is that the same population could have co-ordinated on the Robin-Hood strategy without loss. We may imagine that each alternative is compatible with the respective populations' moral sentiments—the Bourgeois community despises Robin Hoods, and the Robin Hoods despise the Bourgeois. Moreover, each norm is obviously compatible with systematicity and impartiality: differences in expected behavior are explained in terms of the morally relevant properties of either being owner or intruder.

5. Conclusion

I began this paper with two promises: first, that understanding Hume's account of the correction of moral sentiments in moral judgments is key to understanding his sentimentalism and his anti-rationalism; second, that this understanding would yield an important meta-ethical consequence. These promises have now been met.

On my understanding of the standard for impartiality involved in the correction of the sentiments, impartiality is best understood in terms of general rules that concern the correct application of moral concepts: a requirement for impartiality and a requirement for systematicity. These requirements, recall, concern our moral beliefs taken jointly, not individually. It follows that moral judgments are still true or false in virtue of our moral sentiments, even though they are subject to further constraints. Moreover, it follows that our moral sentiments can serve as an imperfect guide to moral truth. This is to affirm brands of constitutive and epistemic sentimentalism, respectively.

Hume's argument from the inertia of reason to anti-rationalism consists, as we have seen, of two separate strands of thought. First, Hume argues that, even though causal reasoning can in principle discern correct moral judgments given a shared core of moral sentiments, causal sentimentalism is essentially true of moral judgment because it ensures that we are by and large motivated to act in accordance with our moral judgments. Second, he argues that causal reasoning can discern moral truth *only* relative to a shared core of moral convictions, precisely because moral truth is in part constituted by this core.

An important meta-ethical upshot of Hume's constitutive sentimentalism and his standard of correction is a form of moral relativism: we cannot legislate between different systems of moral belief that are based on the sentiments shared within a community and conform to the standard of impartiality

Acknowledgments

I am grateful to Hsue Qu and Dan Waxman as well as two anonymous referees for this journal for providing detailed and very helpful feedback on earlier drafts of this paper. I have presented previous versions of this paper twice, once at a conference on Sentiment and Reason at the University of Buffalo and once in Dominik Perler's 'Kolloquium für theoretische Philosophie' at Humboldt Universität zu Berlin, and I am grateful for the invaluable feedback I received on both occasions. I have greatly benefitted from discussions with Christian Barth, Sebastian Bender, Rachel Cohon, Don Garrett, Max Hayward, Jennifer Marusic, Dominik Perler, Stephan Schmid, Geoff Sayre-McCord and Kelley Schiffman.

Abbreviations

T = David Hume (1978). *A Treatise of Human Nature* (2nd ed.). L. A. Selby-Bigge and P. H. Nidditch (Eds.). Clarendon. Page numbers indicated by "SBN".

EHU = David Hume (1975). *An Enquiry concerning Human Understanding*. In *Enquiries concerning Human Understanding and concerning the Principles of Morals* (3rd ed.). L. A. Selby-Bigge and P. H. Nidditch (Eds.). Clarendon. Page numbers indicated by "SBN".

EPM = David Hume (1975). *An Enquiry concerning the Principles of Morals*. In *Enquiries concerning Human Understanding and concerning the Principles of Morals* (3rd ed.). L. A. Selby-Bigge and P. H. Nidditch (Eds.). Clarendon. Page numbers indicated by "SBN".

EMPL = David, Hume (1874–5). Of a Standard of Taste. In T. H. Green and T. H. Grose (Eds.), *The Philosophical Works of David Hume* (Vols. 1–4). Longman, Green.

References

- Abramson, Kate (2001). Sympathy and the Project of Hume's Second Enquiry. *Archiv für Geschichte der Philosophie*, 83(1), 45–80.
- Baier, Annette (1991). *A Progress of Sentiments: Reflections on Hume's Treatise*. Harvard University Press.
- Bentham, Jeremy (1988). *A Fragment on Government*. Cambridge University Press.
- Blackburn, Simon (1984). *Spreading the Word: Groundings in The Philosophy of Language*. Oxford University Press.
- Blackburn, Simon (1988a). How to Be an Ethical Antirealist. *Midwest Studies in Philosophy*, 12(1), 361–375.
- Blackburn, Simon (1988b). Supervenience Revisited. In Geoffrey Sayre-McCord (Ed.), *Essays on Moral Realism* (59–75). Cornell University Press.
- Cohon, Rachel (2008). *Hume's Morality: Feeling and Fabrication*. Oxford University Press.
- Driver, Julia (2012). What the Objective Standard is Good for. In Mark Timmons (Ed.), *Oxford Studies in Normative Ethics* (Vol. 2, 28–44). Oxford University Press.
- Garrett, Don (1997). *Cognition and Commitment in Hume's Philosophy*. Oxford University Press.
- Garrett, Don (2015). *Hume*. Routledge.
- Gauthier, David (1986). *Morals by Agreement*. Oxford University Press.
- Gibbard, Allan (2003). *Thinking How to Live*. Harvard University Press.

- Glossop, Ronald J. (1967). The Nature of Hume's Ethics. *Philosophy and Phenomenological Research*, 27(4), 527–536.
- Hare, Richard M. (1952). *The Language of Morals*. Clarendon Press.
- Jensen, Henning (1977). Hume on Moral Agreement. *Mind*, 86(344), 497–513.
- Lewis, David (1979). Scorekeeping in a Language Game. *Journal of Philosophical Logic*, 8(1), 339–359.
- Mackie, John Leslie (2004). *Hume's Moral Theory*. Routledge.
- Piaget, Jean (1965). *The Moral Judgement of the Child*. Free Press.
- Sayre-McCord, Geoffrey (1994). On Why Hume's General Point of View Isn't Ideal—and Shouldn't Be. *Social Philosophy and Policy*, 11(1), 202–28.
- Sayre-McCord, Geoffrey (1995). Hume and the Bauhaus Theory of Ethics. *Midwest Studies in Philosophy*, 20(1), 280–298.
- Schelling, Thomas C. (1960). *The Strategy of Conflict*. Harvard University Press.